The modified dividend–price ratio

Vassilis Polimenis [a],[*],[1], Ioannis M. Neokosmidis [b]

[a] *Business Division, Department of Economics, Aristotle University, 54124 Thessaloniki, Greece*
[b] *Department of Economics, Aristotle University, 54124 Thessaloniki, Greece*

ABSTRACT

We show that log-dividends (d) and log-prices (p) are cointegrated, but, instead of de facto assuming the stationarity of the classical log dividend–price ratio, we allow the data to reveal the cointegration vector between d and p. We define the modified dividend–price ratio (mdp), as the long run trend deviation between d and p. Using S&P 500 data for the period 1926 to 2012, we show that mdp provides substantially improved forecasting results over the classical dp ratio. Out of sample, while the dp ratio cannot outperform the "simplistic forecast" benchmark for any useful horizon, an investor who employs the mdp ratio will do significantly better in forecasting 3-, 5- and 7-year returns with an R^2_{OS} of 7%, 26% and 31% respectively. In some sense mdp can be considered as a de-noising of the classical ratio as it addresses the major weakness in dp, its presumed inability in revealing business cycle variation in expected returns. Unlike dp, mdp exhibits positive correlation with the risk free return component, and can discern if a low dividend state coincides with a low yield state.

© 2016 Elsevier Inc. All rights reserved.

1. Introduction

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 predictive slope coefficients for longer horizons.

Miller and Modigliani (1961) argued that dividend policy is irrelevant, and that stock prices should be driven by the "real" variable which is the earnings power of corporate assets. Yet, from early on dividend yields attained special importance as a forecasting variable due to the straightforward participation of the dividend yield in return formation, and its highly persistent dynamics which could provide predictability in long forecasting horizons via the mechanism outlined above.

Cochrane (1992, 2011) argues that for long horizons, long-run return and/or dividend growth predictability have to coincide with the variability of the log dividend–price ratio (dp)[2] Actually Cochrane (2011) goes one step further in arguing that (surprisingly) dp has no information about future dividend growth, and that almost all variation in dividend yields is driven by variation in discount rates. Powerful as it may be, this finding is based on two main assumptions, a) the *stationarity* of dividend yields and b) the assumed ability to recursively extend the Campbell and Shiller (1988) approximation to *infinity*.[3] Furthermore, there are some major problems with the predicting performance

---

* Corresponding author.
E-mail addresses: polimenis@econ.auth.gr (V. Polimenis), ineokosm@econ.auth.gr (I.M. Neokosmidis).
[1] Tel.: +30 2310 99 1317.

[2] In this paper lowercase letters always denote logs: $d_t = \log D_t$, $p_t = \log P_t$, and $r_t = \log R_t$.
[3] Engsted, Pedersen, and Tanggaard (2012) study the error of the Campbell–Shiller approximation in the presence of a non-stationary dp.

Table 1
Summary Statistics.
We present the summary statistics for annual returns, equity premia, risk free rates, classic dividend–price ratio (dp) and our modified dividend–price ratio (mdp). The table shows the correlation matrix between 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 2012.

	r_t	re_t	rf_t	dp_t	mdp_t	Mean	Std	AR(1)
r_t	1					0.09	0.20	0.06
re_t	0.99	1				0.06	0.20	0.05
rf_t	0.03	−0.12	1			0.04	0.03	0.93
dp_t	−0.25	−0.24	−0.05	1		−3.35	0.45	0.87
mdp_t	−0.34	−0.39	0.35	0.69	1	−2.05	0.26	0.70

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 & Welch, 2003). 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 Thompson (2008).

1.1. The non-stationarity of the dividend yield

Econometrically, most researchers argue that 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. At the same time, the majority of empirical studies on return predictability, cannot reject statistically (if not economically) the hypothesis of the presence of a unit root in the dividend–price ratio (Goyal & Welch, 2003; Lettau & Van Nieuwerburgh, 2008; Lettau & Ludvigson, 2005 among others).

We can see from summary statistics presented in Table 1, that the dividend–price ratio dp has an autocorrelation $\varphi = .87$. Clearly, this is a local alternative that unit root tests have not enough power to detect. Furthermore, it is known, as early as Kendall (1954) that typical estimation methods will tend to highly underestimate true persistence in finite samples.[4] In the following sections, we present robust econometric evidence against the stationarity of the classical dp. Not only is stationarity rejected via a straightforward ADF testing for dp, but using the more powerful test of a restriction on the cointegration vector for d and p we reject the hypothesis that log-dividends and log-prices are linked with a long run relationship of the form (d-p).[5] Econometrically, dp is at best a near non-stationary process.

Economically, the unquestionable requirement that stock prices cannot be far from corporate fundamentals for too long has often been interpreted in a strict sense that the log dividend–price ratio is stationary either in the full sample or at least in specific subsamples. The classical thinking about the behavior of dividend yield ratios is that dividends should represent a more or less "fixed" fraction of earnings, and earnings should represent a more or less "fixed" fraction of prices. Thus, most contemporary literature de facto assumes that the classical dp is a *stationary process* and should not include any trends. Generally speaking though, this is an economic requirement, which depends on a particular sample, rather than a hard fact. Actually, corporate officers 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 non-stationary 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 non-stationarity 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 is the next logical step still satisfying a "fundamentals" based asset pricing philosophy. In this paper we modify the dividend–price ratio by relaxing the stationarity assumption for the classical dp_t, and assuming 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 allow the data to reveal the "true" cointegration vector $[1, -\beta]$.

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−βp. We may then think of β as the unique population parameter that "fine tunes" dp by revealing the 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 (possibly) small I(1) noise term.

$$dp_t = mdp_t + (\beta - 1)p_t. \tag{1}$$

By not de facto assuming an *unreliable rejection* of the non-stationarity null for dp, the modified ratio presents a more reliable alternative, which allows for a richer representation of the d.g.p. Also, at $\varphi = .70$, 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.

In our setup, β provides the drift ratio between d and p. Roughly speaking, a β < 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 non-stationarity 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.

We show that an investor who employs the modified ratio (mdp) will improve his Out-of-Sample forecasting of 3-, 5- and 7-year returns with an R^2_{OS} of 7%, 26% and 31% respectively. Furthermore, an investor who has seen enough of the small (due to super-consistency) required early sub-sample to reliably infer population values for the cointegration coefficient between d and p, will actually improve his forecasts of the 5- and 7-year returns by an astonishing R^2_{OS} of 49%, and even attain a 3-year R^2_{OS} of 34%.

---

[4] Actually, even the Kendall bias correction for autocorrelation $-(1 + 3\varphi)/T$ is low.
[5] That the $[1, -1]$ vector spans the cointegration space.

Finally, a competing approach that may also 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 & 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. Yet, allowing for breaks that are impossible to predict ex ante has little value for return predictability and forecasting, and will produce a weak out of sample R^2 (see Lettau & Van Nieuwerburgh, 2008). This is clearly not the case with our parsimonious approach that produces significant out-of-sample forecasting gains.

2. Econometric methodology

In this paper, we employ the multivariate Johansen (1991, 1995) methodology in order to formulate, estimate and test the modified ratio. In order to provide comprehensive testing in the presence of multiple cointegrating relations,[6] the Johansen test estimates a Vector Error Correction (VEC) model. In the general setup, one may have n-dimensional time series and there may be multiple cointegrating relations among the variables. At its core, the Johansen method uses the size of the eigenvalues of an impact matrix $C = AB'$ to infer its rank. Specifically, the method infers the cointegration rank by testing the number of eigenvalues that are statistically different from 0. Although the method may appear to be very different from the Engle and Granger (1987) approach, Johansen's maximum likelihood approach is essentially a generalization of the augmented Dickey and Fuller (1979) test for unit roots in many dimensions.

If w is a two-dimensional vector $w_t = [d_t \ p_t]'$, and there exists a cointegrating vector b, then $b'w_{t-1}$ is the "error" in the data that quantifies a deviation from the stationary mean at time $t - 1$. Error correction in our context manifests itself as the tendency of a cointegrated dividend and price series to revert to a common stochastic trend. 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

$$mdp_t = d_t - \beta p_t. \tag{2}$$

The dividend and price series correct from the "disequilibrium" that mdp represents at rates captured by a vector of their specific *adjustment speeds* a, thus forming a multiplicative error-correction term $ab'w_{t-1}$ that needs to be added to a simple VAR model explaining jointly price change (Δp) and growth (Δd) dynamics and thus produce the so-called vector error-correction VEC(q) model

$$\Delta w(t) = \sum_{i=1}^{q} B_i \Delta w(t-i) + a(b'w_{t-1} + c_0) + c_1 + u(t). \tag{3}$$

The Johansen test for *deterministic* cointegration above[7] addresses many of the limitations of the workhorse of cointegration estimation, the Engle–Granger method. Since we are only using a two-dimensional vector $w = [d \ p]'$, for us the main benefits of the Johansen method is that it avoids the two-step procedure,[8] and thus provides a framework

for testing restrictions on the cointegrating relations b (and the adjustment speeds a) in the VEC model.

While it is true that the trace and maximum eigenvalue cointegrating rank tests in Johansen are derived under the assumption of Gaussian iid innovations, it has been shown that the standard rank tests based on asymptotic critical values remain asymptotically valid even in the presence of conditionally heteroskedastic shocks,[9] and in particular the trace statistic is more robust to both skewness and excess kurtosis.

3. Results

High quality return data for the S&P 500 index, with and without dividends, are available from CRSP since 1926. Below 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[10] spans the most recent 87 year period that ranges from January of 1926 to December of 2012. We only use nominal data throughout the paper.[11]

When constructing the classical log dividend–price ratio

$$dp_t = d_t - p_t = log(D_t/P_t)$$

we need to employ an annual horizon in order to cancel 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. The most common annual dividend–price ratio is based on the following computation

$$DP_t = \frac{D_t}{P_t} = \frac{\sum_{i=0}^{11} D(t-i)}{P_t}. \tag{4}$$

When we are endowed with monthly gross returns, $R(t) = \frac{P_t + D(t)}{P_{t-1}}$, and the monthly returns due to price gain alone (without dividends) $X(t) = P_t/P_{t-1}$ respectively, the monthly dividend for month (t) is given by[12]

$$D(t) = \left(\frac{R(t)}{X(t)} - 1\right) P_t.$$

A secondary method is to form a dividend–price ratio by re-investing interim dividends. This 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.[13] Thus the literature mainly employs the simpler and less volatile annual sum construction above for DP, as it presumably represents more purely dividend policy decisions of firms.

---

[6] This is of no consequence here as we are dealing with a 2-dimensional vector.

[7] Eq. (2) may be slightly different depending on whether we assume that are no intercepts or trends in the cointegrating relations and there are no trends in the data. Here we assume that the log series have linear trends but the cointegration relationship contains only a constant. This specification is a model of deterministic cointegration, where the relations eliminate both stochastic and deterministic trends in the data.

[8] A concern related to the Engle–Granger method is that it is a two-step procedure, with the 1st regression employed to estimate the residual series, and the 2nd regression to test for a unit root. Errors in the 1st estimation are automatically carried into the 2nd. Moreover, the estimated, and not observed, residuals require different tables of critical values for standard unit root tests.

[9] Seo (1999); Boswijk (2001); Kim and Schmidt (1993) among others suggest that the standard procedures are asymptotically valid both for unit root and cointegration. Nevertheless, the unit root tests have size distortions in small samples. Rahbek, Hansen, and Dennis (2002) find that the usual procedures in order to test for cointegration based on the multivariate settings are asymptotically valid in the presence of multivariate conditional heteroscedasticity (for further analysis on this concept see Harris and Sollis, 2003).

[10] The data are from the Goyal and Welch database, available at http://www.hec.unil.ch/agoyal.

[11] See also the discussion in Campbell and Shiller (1988). Engsted and Pedersen (2010) find that long-horizon predictability depends on whether returns and dividend growth are measured in nominal or real terms.

[12] We should not confuse $D(t)$ the monthly dividend for month t, with D_t the ending at month t annual dividend.

[13] Chen (2009) finds that the annual (from monthly) dividend construction can have significant implications on estimated dividend growth predictability, as the reinvestment assumption makes dividends inherit a lot of the intra-year realized return volatility.

3.1. Integration and co-integration of the series

In our notation, $\boldsymbol{w}_t = [d_t \quad p_t]'$ represents the vector of underlying log dividend and price series, and stationarity of the classical dp is robustly tested in a straightforward manner as a restriction $\boldsymbol{b} = [1 \ {-}1]'$. In order to test for cointegration we must specify, as a first step, the deterministic components which are involved in both the short and long run dynamics and the optimal lag length (q). As shown in Eq. (3), we proceed by considering that the log series have linear trends but the cointegration relationship contains only a constant. This specification is characterized as testing for *deterministic* cointegration among trending series.

For the optimal lag selection, we first estimate the above model in an unrestricted form like a VAR model in levels with a high initial number of auto-regressive lags and then we test for significance in the higher order autoregressive coefficients. Estimating an initial VAR in levels is crucial for the convergence properties of the usual test statistics. An extended reference on this subject can be found in Hamilton (1994, ch.18) and Toda and Yamamoto (1995). We start by assuming a maximum order of 12 lags and after we condition down to a more parsimonious representation based on the Hannan–Quinn criterion. We conclude on using 7 optimal lags for VAR and thus 6 lags for VECM.

Table AI presents the results. Trace tests[14] show that the series are cointegrated with a cointegration relationship of the form,

$$mdp_t = d_t - 0.8017 p_t. \tag{5}$$

The second panel of Table AI presents results for testing the restriction that the vector $[1 \ {-}1]$ spans the cointegration space based on the Johansen procedure on [d p], and it is shown that $[1 \ {-}1]$ does not span the cointegration space. As the Johansen procedure is essentially a multivariate generalization of the augmented Dickey–Fuller test for unit roots, this is more powerful empirical proof of the nonstationary behavior of dp that deals with the low power of unit root tests against highly persistent alternatives.

Finally, in order to measure the exact error correction feedback mechanism we estimate the entire VECM in Eq. (3) and present the results in Table AII. We see that both dividends and prices exhibit valid speed of adjustment dynamics (correction sign) and significant coefficients.

3.2. In-sample predictability

In this section, we present the main univariate forecasting regressions based on the classical dividend–price ratio (dp) and the modified ratio (mdp) respectively. We formulate continuously compounded returns, equity premia, and dividend growth for 1, 3, 5 and 7-year horizons ($h = 1, 3, 5, 7$) using monthly S&P 500 data.

Table 2a presents the results of return, equity premium and dividend growth predictability for S&P 500 based on the following forecasting regression. For return predictability, the left hand variable is the time-t *future* log return (r), for one, three, five and seven years ahead[15]

$$r_t(h) = a + cx_t + u_t(h).$$

In the above regression, the predictor variable represents either the classical dp ratio or the modified ratio mdp. Standard errors are GMM corrected based on the Hansen–Hodrick formula.[16]

As we can see, for both 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

Table AI

Cointegration Test and The null Hypothesis of $[1 \ {-}1]$.

We apply the Johansen testing procedure assuming trending series and no trend on cointegration relationship. The pair [d p] tests for a cointegration relationship between the 12 month summed-up dividends (d) and prices. The 2nd panel presents results for the restriction test that $[1 \ {-}1]$ spans the cointegration space between (d,p). A (*,**) denotes the 5% and 1% rejection level respectively. Data are overlapping annual spanning the period, 1926–2012.

Panel A	#Coint.vec	Trace test stat
	0	19.35*
	≤1	0.24
		5% critical value
	0	15.49
	≤1	3.84
Panel B	H0: $[1 \ {-}1]$	χ^2-stat
		10.42**

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 $dp_t = mdp_t + (\beta - 1)p_t$. 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 (c), significance (t-stats), and log-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[17] "theoretical limit" of 1. When extending the forecast horizon to seven years, classical slopes have gone from roughly 0.40 to 0.50, 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.

3.3. A first attempt at an economic explanation

While the Miller–Modigliani irrelevance theorem implies no reason for dividends to play a role in determining equity price levels or equity returns, in reality dividends have always been at the center focus of many investors. At the same time there is a substantial amount of recent evidence that suggests that share repurchases have substituted for dividend payments over the last 20 to 25 years. If dividend and repurchase policies are not independent and in fact *substitute* each other, the classical dividend yield will not correctly account for such substitution and *misinterpret gradual long-run changes in payout policy as business cycle variation*. It seems plausible, that the mdp ratio explains long run returns better by removing a stochastic trend in the classical dp and capturing the business cycle variation in expected returns.[18]

In order to better understand how mdp works, it is instructive to plot both ratios against future realized returns. Fig. 1 plots the 5-year future

---

[14] We also calculated Maximum-Eigenvalue test statistics and had similar findings.

[15] Similarly for the h-year realized equity premium ($re_t(h)$) or realized dividend growth ($\Delta d_t(h)$).

[16] In order to correct heteroskedasticity and correlation effects Newey–West estimates of the standard errors have also been tried with no change on the significance of the findings.

[17] 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."

[18] Another methodology suggested by scholars is to use a repurchase-adjusted dp. We report in the last subsection some preliminary findings.

Table AII
VECM results between dividends and prices.
This table presents the results from the VECM estimation between dividends (d) and prices (p) using the multivariate Johansen methodology. Data are overlapping annual spanning the period, 1926-2012. (*) and (**) denote significance at the 5% and 1% rejection level respectively.

	a	$\Delta w(t-1)$	$\Delta w(t-2)$	$\Delta w(t-3)$	$\Delta w(t-4)$	$\Delta w(t-5)$	$\Delta w(t-6)$	c_1
$\Delta d(t)$	-0.009770	0.001761	0.026325	0.284771	0.061383	0.059586	0.216636	0.001098
	(0.00251)**	(0.03050)	(0.03039)	(0.03031)**	(0.03029)*	(0.03035)*	(0.03045)**	(0.00065)
		0.014078	0.013924	-0.000125	-0.011817	0.013378	0.020326	R^2
		(0.01126)	(0.01127)	(0.01127)	(0.01125)	(0.01126)	(0.01128)	0.226167
$\Delta p(t)$	0.013824	0.008963	0.089282	0.050096	0.097928	-0.193692	-0.148287	0.004256
	(0.00704)*	(0.08533)	(0.08504)	(0.08481)	(0.08475)	(0.08491)*	(0.08519)	(0.00182)*
		0.098233	0.004153	-0.086519	0.046227	0.083947	-0.039416	R^2
		(0.03151)**	(0.03154)	(0.03154)**	(0.03147)	(0.03151)**	(0.03155)	0.038372

realized long run returns against current dp and mdp levels. 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.

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 realized equity premium, we can intuitively deduce that the enhanced performance of mdp in predicting future returns comes from its robust capacity in predicting the return component 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.

It is important to economically discuss the positive correlation of mdp with future risk free returns. 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. If companies that consistently pay dividends attract a certain type of investor (clientele) 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*. This reasoning could be at the heart of the strong forecasting power of our modified dp. As mdp is driven by a strong positive correlation with risk free returns (0.35), a low mdp print can discern that a low dividend state coincides with a low yield state. This should be compared with the near orthogonality of the classical dp (-0.05). One way to understand why the

classical ratios don't share such forecasting 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 to enhance return forecasts.

3.4. Out-of-sample performance

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 against the forecasting ability of a simple benchmark for a real time investor. Campbell and Thompson (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_{OS}^2 statistic,

$$R_{OS}^2 = 1 - \left[\sum_{k=1}^{\tau} \left(r_{t+k} - \hat{r}_{t+k} \right)^2 / \sum_{k=1}^{\tau} \left(r_{t+k} - \bar{r}_{t+k} \right)^2 \right].$$

This measure compares the OS performance of a predictor variable that predicts r against the "simplistic forecast" benchmark that utilizes the simple average of past returns $\bar{r}$ as forecast. The OS coefficient of determination R_{OS}^2 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_{OS}^2 directly measures the welfare benefits (for a mean–variance investor with a given risk aversion coefficient) of the increased portfolio returns achieved by using the predictor variables.

Table 2a
Predictability of returns, equity premia and dividend growth.
Standard errors are GMM corrected. (Data are annualized constructed from monthly observations from 1926 to 2012.)

|		c	t(c)	R^2		c	t(c)	R^2
$r_t(1)$	dp_t	0.09	1.57	0.03	$r_t(3)$	0.27	3.10	0.11
	mdp_t	0.21	2.84	0.07		0.65	5.03	0.23
$r_t(5)$	dp_t	0.42	4.19	0.19	$r_t(7)$	0.51	3.48	0.25
	mdp_t	1.04	9.23	0.41		1.16	12.51	0.49
$re_t(1)$	dp_t	0.09	1.67	0.04	$re_t(3)$	0.28	3.19	0.12
	mdp_t	0.17	2.20	0.04		0.53	3.17	0.16
$re_t(5)$	dp_t	0.44	3.67	0.21	$re_t(7)$	0.55	3.23	0.30
	mdp_t	0.83	4.19	0.26		0.89	4.52	0.29
$\Delta d_t(1)$	dp_t	0.02	0.63	0.00	$\Delta d_t(3)$	0.01	0.13	0.00
	mdp_t	0.08	1.28	0.02		0.08	0.71	0.01
$\Delta d_t(5)$	dp_t	0.04	0.50	0.00	$\Delta d_t(7)$	0.00	0.04	0.00
	mdp_t	0.18	1.23	0.02		0.12	0.93	0.01
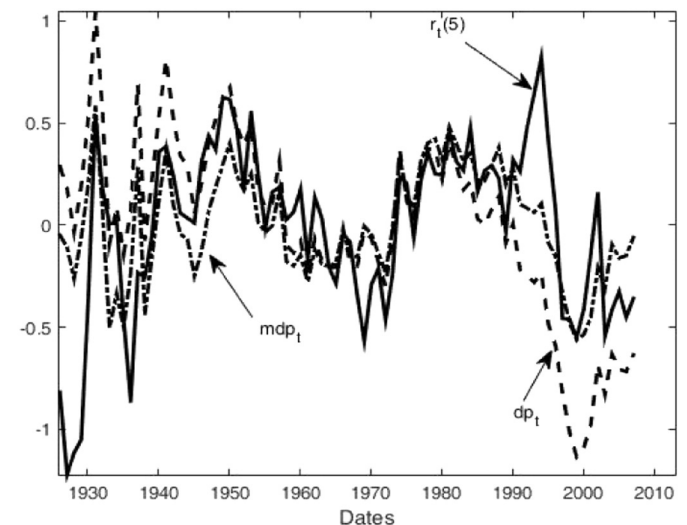
Fig. 1. Evolution of dp and mdp against forward looking 5-year returns.

Table 2b

Univariate forecasting of long run risk free rates.

We run univariate regressions between long run risk free rates, $rf_t(h) = \sum_{j=1}^{h} rf_{t+12j}$, with the competing dividend–price ratios (dp_t, mdp_t) as regressors. Data are annualized constructed from monthly observations with an overlapping rolling window from 1926 to 2012. Standard errors are GMM corrected.

		c	t(c)	R^2
$rf_t(1)$	dp_t	− 0.00	− 0.69	0.00
	mdp_t	0.04	2.71	0.11
$rf_t(3)$	dp_t	− 0.01	− 0.48	0.01
	mdp_t	0.12	2.02	0.13
$rf_t(5)$	dp_t	− 0.02	− 0.39	0.00
	mdp_t	0.21	2.08	0.15
$rf_t(7)$	dp_t	− 0.04	− 0.49	0.01
	mdp_t	0.27	2.02	0.14

We present the Campbell–Thomson OS coefficient of determination, for predicting returns and equity premia for 3-, 5- and 7-year horizons. We divide the data sample in two periods. Initially, we utilize a 15-year minimal *estimation* period (1926–1941). The remaining sample, extending beyond the estimation period (until 2012), constitutes the *evaluation* period. We choose 15 years for the initial estimation period as 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 (see the discussion in Welch & Goyal, 2008).

While calculating dp from current data is straightforward, in order for the econometrician to construct mdp, the true long-run coefficient (β) between d and p needs to be estimated first. On a first approach, the straightforward method is to re-estimate the cointegration coefficient on a *recursive* (R) basis, each time using only data up to a certain point t. This means a two-step procedure for every time t: a) a cointegration coefficient b_t with d and p data only up to time t is estimated, and b) $mdp(R) = d − b_t p$ for this period is formed, so that finally a forecasting regression of returns against that mdp can be run.

As we can see, the classical dp ratio cannot provide positive R^2_{OS} values, meaning that it fails to outperform the simplistic forecast benchmark for all short- to medium-term horizons. To get the dp ratio to (marginally) outperform the "simplistic forecast" we need to utilize a long 7-year horizon. On the other hand, the modified ratio provides forecasting benefits as fast as in predicting the 3-year forward return. An investor who employs the modified ratio (mdp) will improve his Out-of-Sample forecasting of 3-, 5- and 7-year returns with an R^2_{OS} of 7%, 26% and 31% respectively.

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. As the investing horizon gets longer, the modified ratio R^2_{OS} is increasing and reaches a 31% strong gain for seven years ahead. The statistical significance of R^2_{OS} does not matter here because we believe that the sign is clear. Even a small R^2_{OS} can provide great investment benefits for investors who would otherwise

Table 3

Out Of Sample (OS) evaluation.

We present OS results for classical and the two modified dp ratios: one with a recursive procedure mdp(R) and one where the entire sample is used to estimate the cointegrating coefficient mdp(P). Data are overlapping annual spanning the period, 1926–2012.

Realized future returns	rt(3)	rt(5)	rt(7)
dp	− 0.03	− 0.02	0.14
mdp (R)	0.07	0.26	0.31
mdp (P)	0.34	0.49	0.49

Realized future premia	ret (3)	ret (5)	ret (7)
dp	− 0.01	0.00	0.16
mdp (R)	− 0.11	− 0.01	0.03
mdp (P)	0.20	0.28	0.26

think that "…returns will be as they always have…" (see Campbell & Thompson, 2008; Rapach, Strauss, & Zhou, 2010).

3.5. Performance of the population mdp

Actually even the strong performance of mdp(R) depicted in Table 3 is conservative, and the true forecasting benefit of mdp for a large sample is probably even higher. Although it is more agreeable from an informational point of view, the recursive procedure carries great sampling errors, and puts the predicting power of mdp(R) at a disadvantage. This happens because, when estimating b_t recursively, we run forecasting regressions of future returns against an mdp proxy, and not against the true mdp that would be produced by the use of the *population* coefficient β.

Furthermore, *due to the super-consistency* of the cointegration estimator only a small early sub-sample is required to reliably infer *population* values for the cointegration coefficient between d and p. Thus, in Table 3 we not only present the recursive performance of mdp(R), but also estimate a *population* (or long-run) mdp(P) where the cointegration coefficient is estimated using the full sample. Effectively, the difference between mdp(P) and mdp(R) measures the forecasting gain for an investor who has seen enough data to recover the population coefficient β. As shown in Table 3, an investor who has seen enough of this early subsample, will actually improve his forecasts for the 5- and 7-year returns by an astonishing R^2_{OS} of 49%, and even attain a surprising 34% Out-of-Sample 3-year R^2 statistic.

A concern with evaluating the performance gain of the population mdp(P) is whether a practitioner operating in the early part of our sample, and estimating cointegration coefficients without access to enough historical data, could have exploited the full forecasting power of mdp(P) to his advantage. This "look ahead" concern, when we try to examine the out-of-sample power of our modified ratios, is well documented by Lettau and Ludvigson (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 also in Lettau and Ludvigson (2001), the appropriate estimation strategy for measuring the full forecasting power of the modified ratios, could be to use the full sample, because sufficiently large samples of data are necessary to recover the true cointegration coefficients. Assuming that the investor knows the population coefficient is not a heavy requirement because cointegration coefficients are
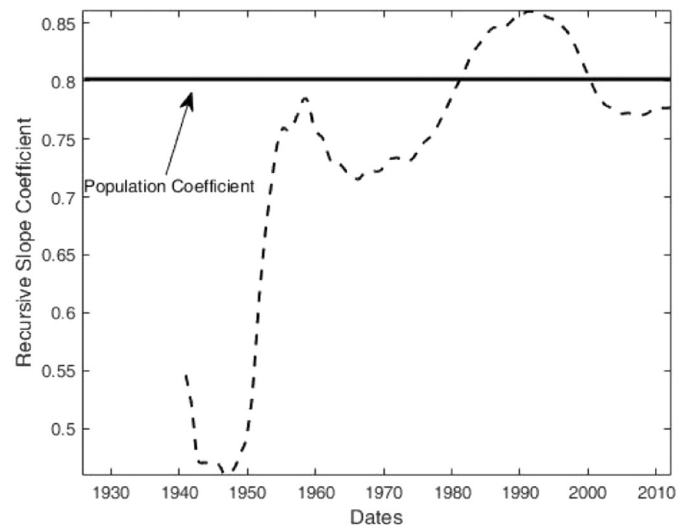

Fig. 2. Convergence of a recursively estimated cointegration b coefficient to its population value β.

super-consistent, converging to their true values at a rate proportional to the sample size T.

Fig. 2 shows a graph of recursively estimated b coefficients over the sample. As shown in Fig. 2, an investor who has seen as little as 30 years of data may treat estimated coefficients as long-run b values during the second-stage forecasting regressions.

Besides the econometrics, on a practical level we feel that the true performance gain should be thought of as somewhere between the performance of mdp(R) and that of mdp(P); i.e. we view them more as a low and high limit on the forecasting gain of modifying dp. In any case, and regardless of whether one uses the recursive or population methodology to measure the performance of our modified ratio, comparable levels of OS performance to the tune of 30% (or 50% for the population statistic) have not been achieved by any other forecasting indicator that we know of.

3.6. Further robustness checks

a) One robustness check,[19] is to consider single-equation, multivariate regressions of the form $r_t(h) = a + b_1 d_t + b_2 p_t + u_t(h)$ instead of using the trend deviation mdp as the single right-hand side variable. Under the null hypothesis that the left-hand-side variable is stationary, while the right-hand side variables are I(1) with a single cointegrating 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 7%, 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, even when using data from the whole sample period (as in our long-run mdp).

b) A second important robustness check is to test the performance of mdp and dp in economically meaningful subsamples. One such subsample is the period after 1965. Even though we leave this exercise along with other robustness checks for a follow-up study, a first analysis for the 3-year horizon reveals that the performance gain of mdp becomes stronger in the 1965–2012 subperiod. Specifically, in the 1965–2012 multivariate subsample regression $r_t(3) = a + c_0 dp_t + c_1 mdp_t$, the coefficient of mdp is significant while for dp we cannot reject the null hypothesis.

c) As a third robustness check, we investigate the possible role of stock repurchases. Some scholars argue that as there were persistent changes in firm payout policies in the 1990s, we should adjust dividend–price ratios for repurchases (Fama & French, 2001; Grullon & Michaely, 2002; Boudoukh et al., 2007). We run multivariate regressions for 3-, 5- and 7-year horizons with both a repurchase adjusted dividend–price ratio and mdp on the right hand side competing to explain returns. In all three horizons, mdp "wins" as it comes out strongly significant while the repurchase adjusted dp is not significant.

4. 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 β as the unique parameter that "fine tunes" dp, and reveals the true long-run equilibrium between d and p, by removing a possibly small I(1) "noise" component. Indeed, using S&P 500 data for the period 1926 to 2012, we show that mdp is more informative than classical dp, and provides substantially improved forecasting results over the classical dp ratio for medium and long horizons from 3 to 7 years. As we show, one source for the gain of the modified ratio in forecasting returns is due to its enhanced ability to forecast their risk free component. Depending on whether one uses the recursive or population methodology to form mdp, the performance gain of modifying dp lies between a low of 30% and a high limit to the tune of 50% for the population method.

References

Blough, S.R. (1992). The relationship between power and level for generic unit root tests in finite samples. *Journal of Applied Econometrics, 7*, 295–308.

Boswijk, H.P. (2001). "Testing for a unit root with near-integrated volatility" (No. 01–077/4). *Tinbergen institute discussion paper.*

Boudoukh, J., Michaely, R., Richardson, M., & Roberts, M.R. (2007). On the importance of measuring payout yield: Implications for empirical asset pricing. *The Journal of Finance, 62*(2), 877–915.

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

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

Chen, L. (2009). On the reversal of return and dividend growth predictability: A tale of two periods. *Journal of Financial Economics, 92*(1), 128–151.

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

Cochrane, J.H. (1992). Explaining the variance of price–dividend ratios. *Review of Financial Studies, 5*, 243–280.

Dickey, D.A., & Fuller, W. A. (1979). "Distribution of the estimators for autoregressive time series with a unit root", Journal of the American Statistical Association, v.74, 427–431.

Engle and Granger, (1987), "Cointegration and error correction: Representation, estimation and testing," Econometrica, v.55, 251–276

Engsted, T., & Pedersen, T.Q. (2010). The dividend–price ratio does predict dividend growth: International evidence. *Journal of Empirical Finance, 17*(4), 585–605.

Engsted, T., Pedersen, T.Q., & Tanggaard, C. (2012). The log-linear return approximation, bubbles, and predictability. *Journal of Financial and Quantitative Analysis, 47*(03), 643–665.

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

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

Fama, E.F., & French, K.R. (2002). The equity premium. *The Journal of Finance, 57*, 637–659.

Goyal, A., & Welch, I. (2003). Predicting the equity premium with dividend ratios. *Management science: journal of the Institute for operations research and the management sciences, 49*, 639.

Graham, B., & Dodd, D. (1934). *Security analysis.* New York: McGraw-Hill.

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

Hamilton, J.D. (1994). *Time series analysis. vol. 2,* Cambridge Univ Press.

Harris, R., & Sollis, R. (2003). *Applied time series modelling and forecasting.* Wiley.

Johansen, S. (1991). Estimation and hypothesis testing of cointegration vectors in Gaussian vector autoregressive models. *Econometrica, 59*, 1551–1580.

Johansen, S. (1995). *Likelihood-based inference in cointegrated vector autoregressive models.* Oxford: Oxford University Press.

Kendall, M.G. (1954). Note on bias in the estimation of autocorrelation. *Biometrika, 41*(3–4), 403–404.

Kim, K., & Schmidt, P. (1993). Unit root tests with conditional heteroscedasticity. *Journal of Econometrics, 59*(3), 287–300.

Lettau, M., & Ludvigson, S. (2001). Consumption, aggregate wealth, and expected stock returns. *The Journal of Finance, 56*, 815–849.

Lettau, M., & Ludvigson, S.C. (2005). Expected returns and expected dividend growth. *Journal of Financial Economics, 76*, 583–626.

Lettau, M., and Van Nieuwerburgh S. (2008). "Reconciling the return predictability evidence", Review of Financial Studies 21, 1607_1652.

Miller, M.H., and Modigliani F. (1961). "Dividend policy, growth, and the valuation of shares", The Journal of Business 34, 411_433.

Perron, P. (1989). The great crash, the oil price shock, and the unit root hypothesis. *Econometrica: Journal of the Econometric Society, 1361–1401.*

Rahbek, A., Hansen, E., & Dennis, J.G. (2002). *ARCH innovations and their impact on cointegration rank testing. Centre for Analytical Finance working paper, 22.* (pp. 15), 15.

---

[19] This approach was initially proposed in Lettau and Ludvigson (2005) for the case of *cay.*

Rapach, D.E., Strauss, J.K., & Zhou, G. (2010). Out-of-sample equity premium prediction: Combination forecasts and links to the real economy. *Review of Financial Studies*, *23*, 821–862.

Seo, B. (1999). Distribution theory for unit root tests with conditional heteroscedasticity. *Journal of Econometrics*, *91*(1), 113–144.

Toda, H.Y., & Yamamoto, T. (1995). Statistical inference in vector autoregressions with possibly integrated processes. *Journal of Econometrics*, *66*(1), 225–250.

Welch, I., & Goyal, A. (2008). A comprehensive look at the empirical performance of equity premium prediction. *Review of Financial Studies*, *21*, 1455–1508.