Essex Finance Centre
Working Paper Series

Working Paper No 75: 01-2022

"Stock returns predictability with unstable predictors"

"Fabio Calonaci, George Kapetanios, Simon Price"

Essex Business School, University of Essex, Wivenhoe Park, Colchester, CO4 3SQ
Web site: http://www.essex.ac.uk/ebs/

Stock returns predictability with unstable predictors

Fabio Calonaci
Queen Mary University of London

George Kapetanios
Kings College, University of London

Simon Price*
Essex Business School, MMF and CAMA

January 6, 2022

Abstract

We re-examine predictability of US stock returns. Theoretically well-founded models predict that stationary combinations of I(1) variables such as the dividend or earnings to price ratios or the consumption/asset/income relationship often known as CAY may predict returns. However, there is evidence that these relationships are unstable, and that allowing for discrete shifts in the unconditional mean (location shifts) can lead to greater predictability. It is unclear why there should be a small number of discrete shifts and we allow for more general instability in the predictors, characterised by smooth variation variation, using a method introduced by Giraitis, Kapetanios and Yates. This can remove persistent components from observed time series, that may otherwise account for the presence of near unit root type behaviour. Our methodology may therefore be seen as an alternative to the widely used IVX methods where there is strong persistence in the predictor. We apply this to the three predictors mentioned above in a sample from 1952 to 2019 (including the financial crisis but excluding the Covid pandemic) and find that modelling smooth instability improves predictability and forecasting performance and tends to outperform discrete location shifts, whether identified by in-sample Bai-Perron tests or Markov-switching models.

JEL codes: G17, C53

Keywords: returns predictability, long horizons, instability.

*Corresponding author: sgprice@essex.ac.uk

1 Introduction

In this paper we re-examine predictability of US stock returns using a new methodology to model structural change in predictors. Theoretically well-founded models predict that stationary combinations of I(1) variables such as the dividend or earnings to price ratios or the consumption/asset/income relationship often known as CAY may predict returns. However, there is evidence that these relationships are unstable (Lettau and Nieuwerburgh (2008)), and that allowing for discrete shifts in the unconditional mean (location shifts) can lead to greater predictability. Lettau and Nieuwerburgh (2008) identify two (or possibly three) potential break points. But it is unclear why there should be a small number of discrete shifts, and we allow for more general instability in the predictors that is characterised by smooth variation and provide a test of the null hypothesis of cointegration (a stationary predictor). We apply this methodology to the three predictors specified above and find that modelling smooth instability improves predictability and tends to outperform discrete mean shifts.

There is a long empirical literature on the predictability of stock returns that initially uncovered evidence that returns were predictable at short horizons by various variables, notably financial ratios such as the dividend or earnings price ratios. This was initially interpreted as rejecting the efficient markets hypothesis, or EMH, and met with scepticism. But Campbell and Shiller (1988) and John Y. Campbell (1988a) pointed out that predictability from financial ratios was perfectly consistent with present value (PV) asset pricing and time varying returns, which are very plausible, so predictability was not a rejection of the EMH.

Early work was marred by low significance and explanatory power, and researchers began to use returns cumulated over long horizons to increase the signal to noise ratio. Consequently the (cumulated) returns became long-memory with complicated error structures. The econometrics of this were controversial, and in practice various HAC corrections were applied. The title of Boudoukh et al. (2005) ('The myth of long-horizon predictability') neatly captures the scepticism. However, Cochrane (2008) argues forcefully that as the dividend to price ratio does deviate from its mean and as there is a clear theoretical relationship between it and future returns and dividends, it is ultimately hard to avoid the conclusion that predictability must exist.

In this literature predictors are often justified by an appeal to the SDF depending on the business cycle, which opens the door to a very large number of candidate predictors.[1] However, dividend and earnings financial ratios are rooted in the PV asset pricing theory, which makes them attractive from a theoretical perspective. Moreover, there are other economic or financial constructs derivable from the same theory.

[1] Witness the long list of predictors examined in Welch and Goyal (2008).

One such is rooted in consumer behaviour. Lettau and Ludvigson (2001) and Lettau and Ludvigson (2004) argue that deviations from households' long-run consumption/asset/income relationship (CAY) should have predictability power for similar reasons, and demonstrate this for US data.[2] Wright (2004) and Price and Schleicher (2005) look at Tobin's Q and returns predictability using a similar argument.

Yet the evidence remains mixed. Many were sceptical, and there is controversy about appropriate tests. Ang and Bekaert (2007) ask "Stock Return Predictability: Is it There?" and conclude that it is not. The title of Welch and Goyal (2008)'s reworking of the analysis includes the word 'comprehensive', and correspondingly performs exhaustive permutations. They conclude there is predictability but there are many instabilities. Rapach and Wohar (2006) examine the structural stability of predictive regression models of U.S. returns. Paye and Timmermann (2006) likewise look for instability in predictive regressions, and find plenty of evidence for it. This is a recurrent theme in the literature.[3] But the source of the instability may not necessarily lie in the parameters of the model, but in the behaviour of the predictors, and this is the focus of our paper.

As an example of this, CAY itself met with a great deal of scepticism (eg in Brennan and Xia (2005)), and Bianchi et al. (2021) (WP version Bianchi et al. (2016)) subsequently revisited the data and allowed CAY to change using a two-state Markov-switching cointegrating framework, which turned out to be very important for predictability. They set out a theoretical model that justifies this, but it raises a question of why there should exist discrete regimes (rather than continuous variation). In earlier and frequently cited work Lettau and Nieuwerburgh (2008) re-examine predictability and financial ratios, mainly focussed on the dividend-price ratio, allowing for mean (location) shifts in the predictors. They suggest that there were one or two mean shifts (in 1991, or in 1954 and 1994) which need to be taken into account, as they often are in subsequent work by other authors. But again, it is unclear why there should be discrete shifts.[4]

In the wider literature there is plenty of evidence that structural change[5] is endemic in empirical macroeconomic models.[6] Ways of modelling this now constitute an enormous literature with a diverse range of approaches. Much of the focus has been on forecasting (comprehensively surveyed in Rossi (2013))

[2]Fernandez-Corugedo et al. (2007) show the same for the UK.

[3]Recent work includes Rodrigues et al. (2019).

[4] Favero et al. (2011) offer a theoretical explanation driven by long-run demographic trends, which is more consistent with smooth movements than mean shifts.

[5]The terms 'structural change', 'time variation' and 'instabilities' are commonly used in the literature: we consider each to be equivalent.

[6]In a forecasting context, this was prominently brought to attention by Stock and Watson (1996). Later examples that are more rooted in macroeconomic models include Cogley et al. (2010), Cogley and Sargent (2001) and Cogley and Sargent (2005) using TV-VARs on US inflation dynamics, Benati (2008) on UK macroeconomic dynamics, and Sims and Zha (2006) using a regime-switching VAR and Barnett et al. (2014) examining a range of models using UK data.

but structural models now frequently incorporate relevant methods. As yet there is no consensus on how to handle instabilities. Unforecastable permanent abrupt exogenous parameter shifts particularly in location are what is often described as 'structural change'; there may be regime shifts triggered by endogenous processes as in smooth transition models or probabilistic shifts between discrete regimes as in Markov-switching models; or smooth parameterised stochastic time series processes, often random walks. And structural change and long memory or unit root processes blur together: stochastic trends may be thought of as a succession of structural breaks. If we have a clear idea of what these break processes are, then we can apply the appropriate model. But often we are ignorant about the process and it is unclear why a particular model should be chosen. In practice models often incorporate parameter variation from random TV coefficient (RC) models, driven by persistent – often random walk – processes, bounded to avoid explosive outcomes. An alternative approach is to use kernel methods to model time variation using deterministic processes, but Giraitis et al. (2014a) develop an estimation method for RC models using a kernel-based nonparametric technique as an alternative to state-space methods. They demonstrate that with only mild conditions the method has good properties such as consistency and asymptotic normality in a range of macro-relevant contexts. This is the approach we adopt here. Moreover, the technique can be used to estimate cointegrating relationships, where the residuals from the filtered relationship can be tested for cointegration using a KPSS test, which Monte Carlo evidence Kapetanios et al. (2020) shows has good properties. Even in the absence of non-stationarity in the relationships there may be strong persistence in predictors. Dealing with this has been a major problem in the predictability literature, and our method offers a new approach to resolving this.

In this paper we therefore apply this methodology to model time varying long-run potentially cointegrating relationships in the US CAY and financial ratios, and explore whether they improve predictability. The answer is that they do, and improve both predictability and forecasting performance against not only the unadjusted fixed-parameter relationships but also alternative methods of allowing for location shifts (pre-testing for discrete breaks and applying a Markov switching approach).

In the next section we briefly discuss present value conditions, before outlining our methodology in Section 3. In the next section we present the results of our break-adjustment and smoooth location shift procedures. In Sections 5 and 6 we present the results of predictability and forecast tests, while the final section concludes.

2 Present Value conditions

2.1 Financial ratios

Campbell and Shiller (1988) introduced the log-linear approximation to the present-value condition for asset prices known as the Campbell-Shiller decomposition. As this is well-known, we will not give more detail here, except to state the critical relationship

$$d_t - p_t = -\frac{k}{1-\rho} + E_t[\sum_{j=0}^{\infty} \rho^j (-\Delta d_{t+1+j} + r_{t+1+j}]$$

where d is log dividends, r is returns, k is a constant and $0 < \rho < 1$ is another constant that is likely to be close to unity. So current $d_t - p_t$ is high when either future dividends are expected to grow slowly, or when future returns are expected to be high. Consequently current $d_t - p_t$ is informative about (predicts) future dividends or returns. If dividend growth is roughly constant, then future returns should be predictable. This argument is expressed in terms of dividends but applies equally to earnings e_t.

2.2 CAY

Consumption, wealth and non-wealth ('labour') income are related *via* the intertemporal budget constraint, which holds for individuals and in aggregate, and the Campbell-Shiller decomposition can be extended to this case. A simple derivation starts from the accumulation equation for wealth.

$$W_{t+1} = (1 + R_{w,t+1})(W_t - C_t)$$

where W_t is all wealth including human capital, $R_{w,t}$ is the return on wealth and C_t is total consumption (including durables). Define $r \equiv \log(1 + r)$ and let lower case letters denote log variables. Campbell and Mankiw (1989) linearise around the long-run consumption to wealth ratio, assuming that is stationary, obtaining

$$\Delta w_{t+1} \approxeq k + r_{w,t+1} + (1 - 1/\rho_w)(c_t - w_t)$$

where ρ_w is the steady-state ratio of saving to wealth, $(W - C)/W$, and k is a constant. Solving forward, we obtain

$$c_t - w_t = \sum_{i=1}^{\infty} \rho_w^i (r_{w,t+1} - \Delta c_{t+i})$$

and it follows that

$$c_t - w_t = E_t \sum_{i=1}^{\infty} \rho_w^i (r_{w,t+i} - \Delta c_{t+i})$$

This can be extended to incorporate labour income Y_t by assuming a return on human capital so that if A_t refers to non-human wealth then

$$cay_t = c_t - \omega a_t - (1-\omega)y_t = E_t \sum_{i=1}^{\infty} \rho_w^i \{(\omega r_{a,t+i} + (1-\omega) r_{h,t+i}] - \Delta c_{t+i}\} + (1-\omega) z_t$$

and z_t is a stationary random process driving income or human capital. The constant in this expression is suppressed. At one level this is trivial relationship - simply an (intertemporal) accounting identity. Its usefulness is that as the budget constraint is forward looking it incorporates agents' expectations about drivers of wealth and returns (the discount rate). In particular, the current ratio of consumption to (weighted) wealth and labour income is determined by future returns to non-human and human wealth (human capital) and consumption growth. The terms on the right hand side are all stationary and as $\rho_w < 1$ the equality ensures that the left hand side is also stationary. As consumption and wealth are non-stationary, the implication is that $\{c, a, y\}$ form a cointegrating relationship with associated vector $\{1, -\alpha, -\beta)\}$. All this is predicated on the validity of a first-order approximation and the assumption of a constant return on human capital.

Campbell and his collaborators introduced this approach to help understand the dynamics of consumption. Later Lettau and Ludvigson (2001) used it to shed light on the predictability of returns debate, and the 'CAY' relationship became a popular albeit controversial addition to the arsenal of returns predictors. However, the estimated relationship subsequently showed signs of instability in US data. Bianchi et al. (2016) adopt a novel Markov-switching approach to estimating the parameters in *cay* allowing for infrequent two-state mean-shifts, which helps predictability. An alternative view might be that the CAY relation is evolving continuously and that the parameter ω is not constant. Inspection of the data reveal that the wealth and labour income ratios to consumption do not appear to be stationary. 'Labour' income here also includes net transfers which in this context are driven by changing shares of taxation on labour and profits. The ratio of labour income to has also been far from stationary, falling by 44% from its peak in 1974 to the end of the sample. So it seems possible that a time-varying approach to CAY would be illuminating, and might shed light on the predictability debate.

3 Econometric methodology

3.1 Cointegration with TVPs

In this section, we briefly present the econometric methodology explored in Kapetanios et al. (2020)[7] for inference in a simple cointegrating regression model in the presence of time varying parameters. The literature on time varying cointegration includes classic papers such as Bierens and Martins (2010), who offers a test for time variation against the null of invariant cointegration, but here we offer a robust estimation method and a test for cointegration. The analysis is conducted by, first, extending the kernel estimators of Giraitis et al. (2018) to a cointegrating regression setup and proving consistency; and, second, proposing a cointegration test which can detect cointegration when the parameters are not constant.

Let y_t and the k elements of x_t be unit root processes. The standard definition of cointegration is that there exists some vector of constant parameters β in the linear regression

$$y_t = x_t'\beta + u_t \tag{1}$$

where u_t is an $I(0)$ process and β is a $k \times 1$ vector.

We generalise this to allow for a persistent and bounded β_t such that in the regression

$$y_t = x_t'\beta_t + u_t \tag{2}$$

β_t is a $k \times 1$ vector of time varying parameters. We assume that β_t satisfies

$$\sup_{|s|<s_0} \|\beta_t - \beta_{t-s}\| = O_p\left(\left(\frac{s_0}{t}\right)^{\gamma}\right) \text{ for some } 0 < \gamma \leq 1. \tag{3}$$

Condition (3) implies that the sequence of parameters drifts slowly with time, a property that is sufficient for consistent estimation of β_t. This covers deterministic piecewise differentiable processes assumed in the work of Dahlhaus on locally stationary processes (e.g. Dahlhaus (2000) or Dahlhaus and Polonik (2006)). Condition (3) also includes stochastic parameter processes exhibiting a degree of persistence necessary for consistent estimation of stochastically driven time variation. These include bounded random walk processes, as well as some fractionally integrated processes. In addition, parameters satisfying (3) can feature a combination of deterministic trends and (structural) breaks.

Under the parameter time variation framework of (3), an extremum estimator for β_t is derived by minimising an objective function $\widehat{\beta}_t = \arg\min_\beta \sum_{j=1}^{T} k_{tj} u_j^2$:

$$\widehat{\beta}_t = \left(\sum_{j=1}^{T} k_{tj} x_j x_j'\right)^{-1} \left(\sum_{j=1}^{T} k_{tj} x_j y_j\right)$$

[7] A longer version of the paper is Kapetanios et al. (2019).

where the weights k_{tj} are generated by a kernel, $k_{tj} := K\big((t-j)/H\big)$, where $K(x) \geq 0$, $x \in \mathbb{R}$ is a bounded function and H is a bandwidth parameter such that $H \to \infty$, $H = o(T/\log T)$. The kernel estimator $\widehat{\beta}_t$ is a simple generalisation of a rolling window estimator of the form

$$\widehat{\beta}_t = \left(\sum_{j=t-H}^{t+H} x_j x_j'\right)^{-1} \left(\sum_{j=t-H}^{t+H} x_j y_j\right).$$

We assume that K is a non-negative bounded function with a piecewise bounded derivative $\dot{K}(x)$ such that $\int K(x)dx = 1$. For example,

$$K(x) = (1/2)I(|x| \leq 1), \quad \text{flat kernel,}$$
$$K(x) = (3/4)(1-x^2)I(|x| \leq 1), \quad \text{Epanechnikov kernel,}$$
$$K(x) = (1/\sqrt{2\pi})e^{-x^2/2}, \quad \text{Gaussian kernel.}$$

If K has unbounded support, we assume in addition that

$$K(x) \leq C\exp(-cx^2), \quad |\dot{K}(x)| \leq C(1+x^2)^{-1}, \quad x \geq 0, \qquad \text{for some } C > 0,\ c > 0. \tag{4}$$

When x_t is stationary, β_t is bounded away from zero, and for simplicity if $\gamma = 1/2$, Giraitis et al. (2018) show that assuming a martingale difference error process $\hat{\beta}_t - \beta_t = O_p\left(\left(\frac{1}{H}\right)^{1/2}\right) + O_p\left(\left(\frac{H}{T}\right)^{1/2}\right)$. Further, if x_t is a unit root process and β_t is deterministic, then Phillips et al. (2017) have shown consistency and derived rates for $\widehat{\beta}_t$.

We wish to test the hypothesis that u_t is an $I(0)$ process. In Kapetanios et al. (2020) the cointegrating KPSS test is extended with a statistic based on the kernel estimate $\widehat{\beta}_t$. We define the model's residuals by

$$\hat{u}_t = y_t - x_t'\widehat{\beta}_t$$

and the KPSS test statistic by

$$CI = \frac{T^{-2}h\sum_{j=1}^{T} S_j^2}{\hat{s}^2}$$

where $h = H/T$, $\hat{s}^2$ is an estimate of the long run variance of $\hat{u}_t$ and $S_{[Tr]} = \sum_{j=1}^{[Tr]} \hat{u}_j$. The asymptotic distribution of the test statistic CI is given by the following expression

$$T^{-2}h\sum_{j=1}^{T} S_j^2 = T^{-1}\sum_{j=1}^{T}\left(T^{-1/2}h^{1/2}S_j\right)^2 \Longrightarrow Q^2$$

where $Q = \sqrt{2}\int_{-1}^{1} K(s)dB^*_{y,(s+1)/2}$ and B^* denotes Brownian motion. In Kapetanios et al. (2020) the finite sample properties of the test are explored in Monte Carlo experiments.

3.2 Long-horizon prediction

Returns are close to white noise although often with detectable structure in their time-dependence, while predictors are frequently much more persistent. This feature has generated a large literature on how best to model these data. The most frequent response in the empirical finance literature has been to cumulate returns to increase the signal to noise ratio and amplify the persistence in the data. This comes at a cost, as the data to be predicted are overlapping and the implied error structure is complex. The pragmatic response is to use a HAC correction to the standard errors to enable inference to proceed. HAC corrections have been controversial. Partly as a response to this, Kostakis et al. (2015) develop a method based on IVX estimation which has been widely adopted. We, however, offer an alternative, as the methodology developed in Giraitis et al. (2014b) and Giraitis et al. (2018) which we use removes persistent components from observed time series, that may otherwise account for the presence of near unit root type behaviour. Our approach may therefore be seen as to IVX methods where there is strong persistence in the predictor. An example of this property of the GKY methodology is provided in Kapetanios et al. (2020) where use of a time varying cointegration correction reveals the presence of stationarity in presumed cointegrating relations that would be missed if standard cointegration analysis were undertaken. This methodology may, Nevertheless, we do use IVX methods as a robustness check. Specifically, Harvey et al. (2021) have recently developed a new test with attractive features which we apply.

3.3 Forecasting

Predictability is examined within sample and effectively tests the predictions of the PV asset pricing model. But it is also of interest to see if predictors have out-of-sample forecasting power. While this is interesting, it should be clear that it is not a test of the hypothesis of interest. In sample-predictability may exist but without significant forecasting improvements resulting.

4 Results: time variation in predictors

4.1 CAY

The theoretical decomposition that makes CAY an interesting object is based on a first order approximation around mean values and assumes a stationary return to human capital. Figures 1 and 2 show the log ratios of consumption and income[8] to wealth. Until the mid-1990s it was arguable that these ratios were roughly constant. Nothing in the theory of consumer behaviour or indeed growth models in general requires these ratios to be stationary, but the accuracy of the first order linearisation employed above depends upon deviations from the mean being in some sense bounded, and this is less persuasive over the full sample. The consumption and income ratios fall by 32 and 44% respectively from their peaks in 1974 to the end of the sample. Part of this may be due to a decline in the return on human capital, reflected in the fall in the US labour share (Elsby et al. (2013)). These mean changes and the unobserved return to human capital suggest that flexible parametrisation in the location parameter may offer a better approximation.

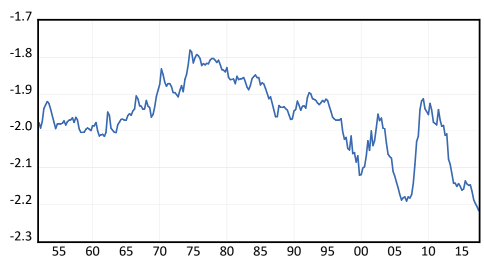

Figure 1: Income to wealth ratio (y-a)

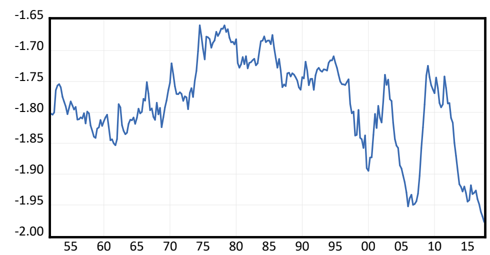

Figure 2: Consumption to wealth ratio (c-a)

4.1.1 Candidate predictors

Bianchi et al. (2016) have estimated time varying versions of CAY which they used to predict returns and as inputs into other analysis. Lettau generously provides updates of the data[9] which we employ in our analysis with the vintage from 1952Q1 to 2019Q3. We truncate the sample at 2019 to avoid the Covid pandemic period. Recall that CAY is interpretable as a potentially cointegrating residual.[10] Our nomenclature is that c_t is log consumption, a_t log assets and y_t log income. Data definitions are in Bianchi et al. (2016). In contrast to the data used in Lettau and Ludvigson (2001) where consumption was restricted to non-durables and services, it is defined as Personal Consumer Expenditure, a comprehensive measure. A subscript t on parameters indicate time variation. We estimate three relationships which we use to construct measures of CAY.

The first is fixed parameters in the long-run relationship

$$c_t = \beta_0 + \beta_1 a_t + \beta_2 y_t + \varepsilon_t$$

[8] Meaning labour income and government transfer payments.

[9] URL `https://sites.google.com/view/martinlettau/data`.

[10] Lettau and his co-authors estimate this using DOLS (Stock and Watson (1993)).

and $\widehat{\varepsilon}_t$ is interpreted as a cointegrating residual, an estimate of CAY, which we denote by CAYF (for fixed).

CAYF: *fixed parameters*

$$cayf_t = c_t - \beta_1 a_t - \beta_2 y_t - \beta_0$$

We estimate this by simple OLS, an Engle-Granger regression.

We estimate two versions of the TV CAY.

The first is a Markov-switching model with locations shifts, where the CAY is again treated as a potentially cointegrating relationship.

CAYMS: *time varying mean: Markov-switching*

$$cayms_t = c_t - \beta_1 a_t - \beta_2 y_t - \beta_{0,t}$$

Here $\beta_{0,t}$ is driven by a two-state Markov-switching process. We place no restrictions on the fixed parameters β_1 and β_2. The estimate of β_0 is the expected value using the estimated probabilities and state coefficients. As in Bianchi et al. (2016) we examine the smoothed ($CAYMSsm$) and filtered ($CAYMSfil$) estimates.

For the second we estimate the cointegrating relationship using our non-parametric method allowing for time variation in the mean (location).

CAYTVM: *time varying mean*

$$caytvm_t = c_t - \beta_1 a_t - \beta_2 y_t - \beta_{0,t}$$

where again as in Bianchi et al. (2016) both β_1 and β_2 are fixed.

4.1.2 Estimated relationships

Table 1 reports estimates for the models described above for the full sample (1952Q1 to 2019Q3) For time varying coefficients (indicated by asterisks)averages are reported. Figure 3 reports the time varying parameters, with the fixed OLS coefficients for reference.

Table 2 reports tests for stationarity (cointegration). With the KPSS test the null (stationarity) is strongly rejected for the Bianchi et al. (2016) Markov-switching and fixed parameter CAYs. It also

Table 1: Estimates

	CAYF	CAYTVM	$CAYMSsm$	$CAYMSfil$
β_0	0.441	2.733*	2.492*	2.488*
β_a	0.218	0.322	0.502	0.574
β_y	0.801	0.785	0.828	0.795

The table reports the estimates for the models described in the text. Where asterisks indicate TV coefficients the averages are reported. Sample 1952:Q1 to 2019:Q3.

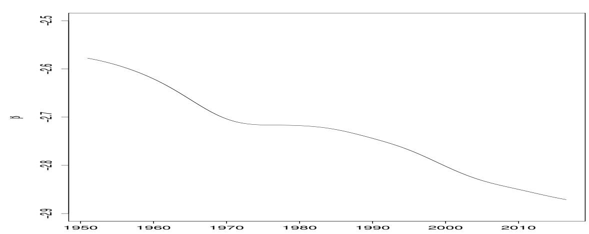

Figure 3: Time varying coefficient β_0 for $CAYTVM$

rejects for our smooth variation method, but only at 7%. As expected, the fixed parameter measure cannot reject a unit root on any test. For the Markov-switching models the evidence is mixed, with only the Phillips=Perron test clearly rejecting. By contrast there is clear rejection of a unit root for the smooth time varying estimate. So the weight of the evidence supports stationarity for our measure.

Finally, Figures 4 to 7 report the various versions of CAY.

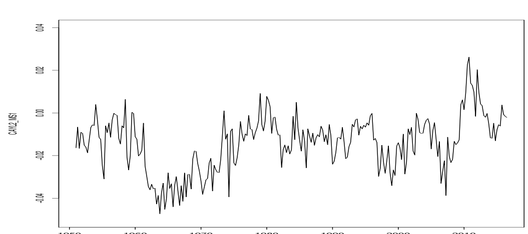

Figure 4: $CAYMSfil$

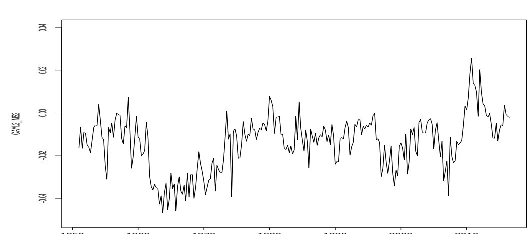

Figure 5: $CAYMSsm$

Table 2: Tests for stationarity

	$CAYTVM$	CAYF	$CAYMSsm$	$CAYMSfil$
KPSS	0.07	0.02	0.01	0.01
ADF	0.01	0.45	0.10	0.13
PP	0.01	0.45	0.01	0.01
EG1	0.02	0.17	0.18	0.11
EG2	0.02	0.18	0.19	0.10

The table provides the p-values of the KPSS, Dickey-Fuller (ADF), the Phillips-Perron (PP) and Engle-Granger tau-statistic (t-statistic) (EG1) and normalized autocorrelation coefficient (EG2) (which we term the z-statistic). The null hypothesis of the KPSS test is stationarity and the alternative a unit root with deterministic trend. For ADF, EG and PP the null hypothesis is a unit-root and the alternative is stationarity.

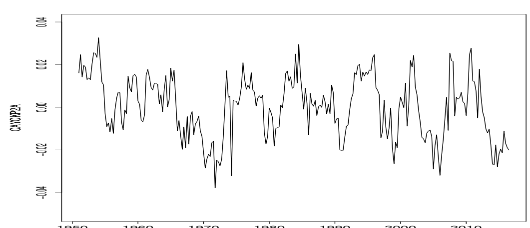

Figure 6: CAYTVM

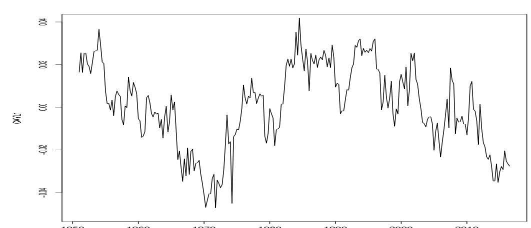
Figure 7: CAYF

4.2 Time variation in financial ratios

Well before the analysis was extended to CAY, John Y. Campbell (1988b) had shown that the dividend price ratio should be informative about future returns or dividends, and the same applies equally to earnings.[11] However, Lettau and Nieuwerburgh (2008) demonstrated that these relationships are unstable and that allowing for discrete shifts in the unconditional mean (location shifts) can lead to greater predictability. They identified at most two candidate shifts, either one in 1991 or two in 1954 and 1994. As already observed, it is unclear why there should be a small number of discrete shifts and we apply our smoothly varying methodology to this case as well. There are many potential predictors (see Welch and

[11] As this is well known, we do not spell it out.

Goyal (2008)) but we restrict attention to the dividend and earnings price ratios as these have the clearest theoretical interpretations. Specifically, we examined $d_t - p_t$ and $e_t - p_t$, where d_t is log dividends, p_t log price and e_t log earnings. The models in that case would be

$$DP_t = d_t - p_t + \alpha_t^d + \epsilon_t^d$$

$$EP_t = e_t - p_t + \alpha_t^e + \epsilon_t^e$$

That is, estimating a time-varying mean. But following Lettau and Nieuwerburgh (2008) and with 13 more years of data than they used, we also look for discrete location shifts using the Bai-Perron[12] procedure and use these to demean the two series. Moreover, as with CAY we estimate two-state Markov-switching models (Hamilton (1996)).

Looking first at the Markov-switching estimates, there is a high-mean, low-asset valuation regime that prevails from 1976:Q2 to 2001:Q2, and a low-mean, high-asset valuation regime in two subperiods at the beginning and end of our sample, namely 1952Q1-1976Q1 and the post-millennial period 2001Q2-2013Q3. Were the regimes known with certainty, we could we could immediately compute the adjusted CAYs. However, as the regimes are unobservable and associated only with an estimated probability, we weight the intercept estimates by their probabilities at each point in time.[13] We consider two estimates, the filtered and smoothed probabilities.[14]

Using the Bai-Perron method, in each case we identify four (different) break dates, reported in Table 3. The identified breaks around 1974 and 1985 are close for both variables.[15]

Table 3: Break dates for DP and EP means

DP	EP
1974Q2	1961Q4
1985Q3	1973Q4
1994Q1	1985Q4
2007Q4	1997Q1

Table 4 reports tests for stationarity for each variant. For the unadjusted DP we comprehensively reject stationarity or cannot reject non-stationarity. Once again, the KPSS test offers only weak evidence for

[12] Bai and Perron (1998), Bai and Perron (2003).

[13] Bianchi et al. (2016) apply a related technique to CAY and similarly identify three clear subperiods characterized by two regimes for the mean.

[14] As do Bianchi et al. (2016).

[15] For DP multiple breaks include one in 1994, a date identified by Lettau and Nieuwerburgh (2008). The break they identified in 1954 could in principle not be in our sample as the sample trimming leaves insufficient observations.

stationarity (p value 0.10) for the discretely mean adjusted and smooth adjusting measures, although in each case the ADF and PP test reject a unit root. A unit root is rejected in all the adjusted cases, but for the KPSS test stationarity is rejected for the MS versions for EP.

Table 4: Tests for stationarity

	DP	Adjust DP	EP	Adjust EP	$\hat{\varepsilon}_{\bar{DP}_{TV}}$	$\hat{\varepsilon}_{\bar{EP}_{TV}}$	MS^{sm}_{DP}	MS^{sm}_{EP}	MS^{fil}_{DP}	MS^{fil}_{EP}
KPP	0.01	0.10	0.01	0.10	0.10	0.10	0.10	0.01	0.10	0.01
ADF	0.48	0.01	0.09	0.01	0.01	0.01	0.04	0.03	0.06	0.03
PPP	0.40	0.01	0.04	0.01	0.01	0.01	0.03	0.02	0.03	0.02
EG1	0.55	0.03	0.07	0.01	0.01	0.01	0.05	0.02	0.04	0.04
EG2	0.56	0.03	0.06	0.01	0.02	0.01	0.02	0.02	0.05	0.05

DP and EP are the raw ratios: Adjust indicates that the ratios have been mean-adjusted. $\hat{\varepsilon}_{\bar{X}_{TV}}$ is the residual from the time varying regression for $X = DP,\ \ DE$ while MS^{sm}_X and MS^{fil}_X refer to smoothed and filtered MS residuals. The table provides the p-values of the KPSS, Dickey-Fuller (ADF), the Phillips-Perron (PP) and Engle-Granger tau-statistic (t-statistic) (EG1) and normalized autocorrelation coefficient (EG2) (which we term the z-statistic). The null hypothesis of the KPSS test is stationarity and the alternative a unit root with deterministic trend. For ADF and PP the null hypothesis is a unit-root and the alternative is stationarity.

Figures 8 to 12 show the various variations on the adjustments - mean adjusted, smooth trends and the Markov-swithching residuals. As is apparent from the data the financial crisis in 2008 had a large albeit temporary effect on the financial ratios. However, while the dividend ratio spikes upwards, largely reflecting falling prices, for earnings the opposite is the case. This turns out to have a material impact on predictability from earnings after 2007.

5 Results: predictability and forecast performance

In this section we present results for predictability and forecasting performance for various predictors, including our time varying models of the dividend and earnings to price ratios and CAY and for forecasting two reference models (a constant and the unconditional sample mean of the excess return, $r - r_f$). Predictability is measured by the significance of the coefficient on a single lagged predictor in a regression of cumulated returns over various periods, and forecast performance is measured by a fully recursive procedure where the relationships are estimated in an expanding window and evaluated out-of-sample, and also using full-sample estimates of the modified predictors and predictive regressions estimated over a rolling window. We examine two samples, one (1952 to 2007) pre-financial crisis and another to 2019 (thus

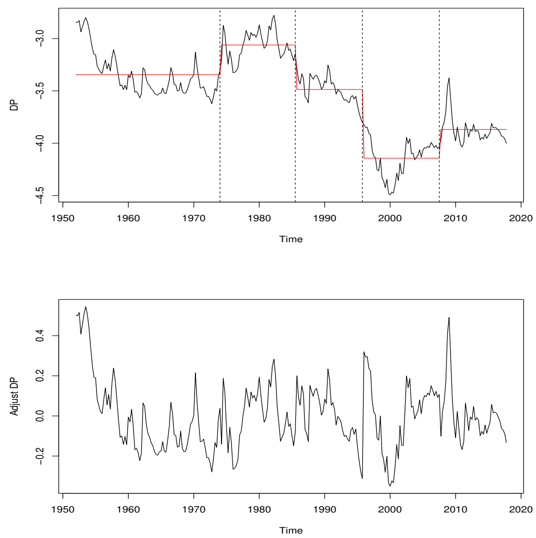

Figure 8: Dividend Price ratio breaks in mean and adjusted series

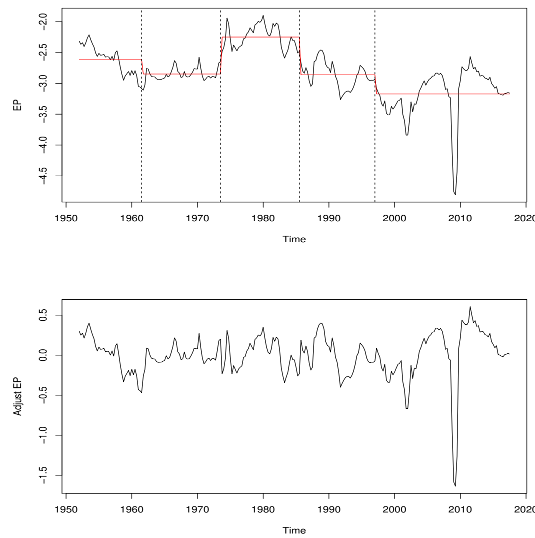

Figure 9: Earnings Price ratio breaks in mean and adjusted series

excluding the Covid period). The results show that our method generally outperforms the alternatives on both predictability and forecast performance, with the exception of the deeranings to price ratio, where as we observed above there is a large spike in 2010.

5.1 CAY

Table 5 reports evidence on long horizon predictability. Given the profound impact of the financial crisis for robustness we examine both the period to 2007 (pre-financial crisis) and the full sample to 2019. As is usual and as the econometric theory predicts, the coefficients' magnitude generally increases with horizon, as does the size of the HAC-corrected t-statistics and R^2. Criteria for 'good' predictors include the comparative magnitude of the t-statistics and R^2. Looking at the period to 2007, all the time varying methods perform much better than the fixed-coefficient CAY. However, the $CAYTVM$ dominates at all horizons. In the longer sample, the performance of the MS models deteriorates, but the smooth time variation model continues to dominate.

Another criterion is forecast performance. Table 6 presents the RMSE from rolling forecasts where the predictor is generated non-recursively, thus using information from the entire sample.[16] The length of

[16]Thus this is a quasi out-of-sample approach as estimation uses the entire sample (except for the benchmark using the

Figure 10: Time varying mean DP and EP with original data

the estimation rolling window is that of the initial period 1952Q1 - 1980Q4. The benchmarks are a constant and the mean value of excess returns in the rolling estimation sample, $r - r_f$ wher r_f is the risk-free rate (the T bill rate). We find that over the shorter sample (1952-2007) for horizons $h = 1$ and 2 the $CAYTVM$ does best but thereafter the filtered CAY is preferred. By contrast, in the full sample (1952-2019) $CAYTVM$ is uniformly preferred.

Table 9 reports Diebold-Mariano (DM) tests for all possible one-sided forecast comparisons. The most interesting results are those between the TV measure and the best performing MS measure, $CAYMSfil$. For the shorter period for the best performing model although they are significantly better than the fixed benchmarks, in no case is the best performer significantly better than the next-best alternative. But for the full sample the TV measure is significantly better than the next-best alternative in all cases except for $h = 1$. We conclude that as for predictability, using the forecast performance criteria our proposed method is overall superior to the alternatives considered, in several cases by significantly and economically important margins.

lagged return).

Figure 11: DP - smoothed and filtered MS

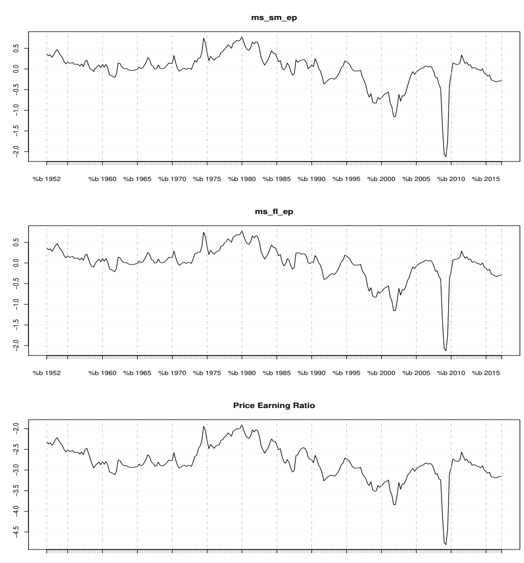

Figure 12: EP - smoothed and filtered MS in

Table 5: Long-horizon predictability: CAY variants

	1952 to 2007					1952 to 2019				
	h=1	h=4	h=8	h=12	h=16	h=1	h=4	h=8	h=12	h=16
CAYF	**0.65**	2.44	4.48	6.25	**8.40**	**0.64**	**2.66**	**5.57**	**7.59**	**8.76**
	(2.22)	(1.66)	(1.81)	(1.50)	(2.53)	(2.62)	(2.06)	(2.72)	(4.28)	(6.34)
	(2.29)	(2.52)	(2.63)	(3.33)	(4.37)	(2.7)	(3.59)	(4.87)	(5.33)	(5.82)
	[0.02]	[0.06]	[0.11]	[0.16]	[0.24]	[0.02]	[0.09]	[0.21]	[0.28]	[0.31]
CAYMSsm	**1.49**	**6.20**	**11.14**	**14.02**	**14.37**	**1.28**	**4.24**	**7.09**	**9.42**	**10.72**
	(4.04)	(4.26)	(4.31)	(4.10)	(4.83)	(3.39)	(2.73)	(3.81)	(4.33)	(4.63)
	(3.25)	(6.67)	(6.81)	(5.40)	(5.05)	(3.58)	(3.61)	(4.11)	(3.60)	(4.15)
	[0.04]	[0.18]	[0.33]	[0.34]	[0.30]	[0.04]	[0.16]	[0.18]	[0.19]	[0.20]
CAYMSfil	**1.51**	**6.22**	**11.15**	**14.05**	**14.39**	**1.30**	**4.24**	**7.21**	**9.52**	**10.83**
	(3.94)	(4.26)	(4.37)	(4.36)	(4.91)	(3.56)	(2.59)	(3.71)	(4.26)	(4.74)
	(3.39)	(6.695)	(6.89)	(5.46)	(5.23)	(3.74)	(3.79)	(4.11)	(3.81)	(4.28)
	[0.05]	[0.20]	[0.33]	[0.34]	[0.32]	[0.04]	[0.10]	[0.16]	[0.17]	[0.19]
CAYTVM	**1.91**	**7.62**	**13.01**	**15.39**	**16.09**	**1.25**	**5.02**	**9.85**	**12.33**	**12.93**
	(3.72)	(4.72)	(5.29)	(4.75)	(6.04)	(3.72)	(3.70)	(4.25)	(5.33)	(6.54)
	(3.80)	(5.47)	(6.83)	(5.89)	(6.16)	(3.95)	(4.34)	(6.53)	(6.42)	(6.27)
	[0.06]	[0.22]	[0.34]	[0.36]	[0.34]	[0.05]	[0.19]	[0.37]	[0.41]	[0.37]

Results from regressions of h-period-ahead cumulated CRSP-VW returns in excess of a 3-month Treasury-bill. We report OLS estimates of the regressors, Newey-West (1987) corrected t-statistics in parentheses, HAC corrected statistics in second parentheses (Zeileis 2006;weighted information sandwich variance estimators for parametric models fitted to time series data) and adjusted R^2 statistics in brackets. Significant coefficients at the 5% significance level are highlighted in bold face.

Table 6: Non-recursive forecast RMSE

	1952-2007					1952-2019				
Model	$h = 1$	$h = 4$	$h = 8$	$h = 12$	$h = 16$	$h = 1$	$h = 4$	$h = 8$	$h = 12$	$h = 16$
const	0.87	3.23	6.44	9.87	13.18	0.87	2.92	6.19	9.68	12.87
$r - r_f$	0.87	3.26	6.61	10.04	13.31	0.88	2.95	6.37	9.85	13.00
CAYF	0.77	3.05	5.49	7.65	9.28	0.77	2.82	5.34	8.20	11.41
CAYMSsm	0.71	2.89	3.39	4.85	7.84	0.69	2.66	5.02	7.79	10.58
CAYMSfil	0.71	2.72	**3.24**	**4.73**	**7.70**	0.68	2.65	4.99	7.81	10.57
CAYTVM	**0.65**	**2.66**	3.78	5.44	7.75	**0.65**	**2.40**	**3.43**	**5.36**	**8.63**

We report mean-squared forecast errors from out-of-sample h-period ahead forecasts of cumulated CRSP-VW returns in excess of a 3-month Treasury-bill. The single predictor variable is listed in the first column. The forecasting equation is estimated using a rolling window the size of the initial period 1952Q1 - 1980Q4). The CAY predictor variables are computed using the entire sample, while the *const* and $r_t - r_t^f$ apply only to the estimation window. Lowest RMSE are indicated in **bold.**

Table 7: Diebold and Mariano results for CAY non-recursive forecasts: all possible one-sided comparisons

	1952-2007						1952-2019						
Model	MSE	$const$	$r-r_f$	CAYF	CAYMSsm	CAYMSsm	MSE	$const$	$r-r_f$	CAYF	CAYMSsm	CAYMSsm	
$const$	0.872						0.873						h=1
$r-r_f$	0.873	0.29					0.879	0.23					
CAYF	0.766	0.13	0.09				0.774	0.16	0.11				
CAYMSsm	0.712	0.41	0.41	0.47			0.692	0.14	0.14	0.17			
CAYMSfil	0.719	0.47	0.47	0.52	0.79		0.687	0.13	0.13	0.15	0.39		
CAYSTVM	0.649	0.26	0.26	0.34	0.36	0.30	0.647	0.18	0.18	0.14	0.63	0.65	
$const$	3.225						2.918						h=4
$r-r_f$	3.259	0.21					2.947	0.77					
CAYF	3.050	0.21	0.24				2.818	0.29	0.26				
CAYMSsm	2.895	0.03	0.04	0.13			2.667	0.02	0.01	0.07			
CAYMSfil	2.727	0.00	0.01	0.04	0.06		2.659	0.01	0.01	0.05	0.31		
CAYSTVM	2.657	0.01	0.01	0.05	0.43	0.80	2.402	0.00	0.00	0.01	0.04	0.04	
$const$	6.437						6.191						h=8
$r-r_f$	6.610	0.29					6.373	0.83					
CAYF	5.492	0.06	0.07				5.343	0.08	0.06				
CAYMSsm	3.394	0.00	0.00	0.00			5.025	0.00	0.00	0.13			
CAYMSfil	3.246	0.00	0.00	0.00	0.07		4.995	0.00	0.00	0.12	0.17		
CAYSTVM	3.775	0.00	0.00	0.00	0.97	0.99	3.434	0.00	0.00	0.00	0.00	0.01	
$const$	9.866						9.683						h=12
$r-r_f$	10.039	0.51					9.848	0.89					
CAYF	7.648	0.01	0.01				8.204	0.06	0.03				
CAYMSsm	4.85	0.00	0.00	0.00			7.791	0.00	0.00	0.17			
CAYMSfil	4.733	0.00	0.00	0.00	0.13		7.818	0.00	0.00	0.18	0.68		
CAYSTVM	5.443	0.00	0.00	0.00	0.98	0.99	5.356	0.00	0.00	0.00	0.00	0.00	
$const$	13.181						12.874						h=16
$r-r_f$	13.307	0.22					12.998	0.51					
CAYF	9.278	0.00	0.00				11.410	0.11	0.10				
CAYMSsm	7.842	0.00	0.00	0.07			10.585	0.00	0.00	0.11			
CAYMSfil	7.703	0.00	0.00	0.06	0.16		10.574	0.00	0.00	0.11	0.53		
CAYSTVM	7.748	0.00	0.00	0.09	0.84	0.91	8.628	0.00	0.00	0.00	0.05	0.04	

We report the p values of the DM test applied on the results from Table 6. The null hypothesis is that the two competing forecasting models have the same predictive accuracy while the alternative is that accuracies differ. The results should therefore be read with reference to the respective reported RMSEs in the first column.

Table 8: Recursive RMSE results

	1952-2007					1952-2019				
Model	$h=1$	$h=4$	$h=8$	$h=12$	$h=16$	$h=1$	$h=4$	$h=8$	$h=12$	$h=16$
const	0.73	3.53	7.04	10.79	14.41	0.73	3.19	6.77	10.58	14.07
$r - r_f$	0.73	3.54	7.19	10.91	14.47	0.74	3.21	6.92	10.71	14.13
CAYF	0.72	3.28	5.91	8.24	9.99	0.72	3.04	5.75	8.83	12.29
CAYMSsm	0.75	3.08	3.61	5.17	8.35	0.73	2.84	5.35	8.30	11.21
CAYMSfil	0.75	2.90	**3.45**	**5.04**	8.20	0.72	2.83	5.32	8.33	11.2
CAYTVM	**0.69**	**2.82**	4.05	5.76	**8.21**	**0.69**	**2.54**	**3.63**	**5.68**	**9.14**

Note: The table provides the $RMSE$ for the recursive exercise. The initial window used is the same used as in the non-recursive version: 1952Q1 - 1980Q4.

Table 9: Diebold and Mariano results for CAY Recursive approach results

	1952-2007							1952-2019					
Model	MSE	$const$	$r-r_f$	CAYF	CAYMSsm	CAYMSsm		MSE	$const$	$r-r_f$	CAYF	CAYMSsm	CAYMSsm
$const$	0.734						h=1	0.732					
$r-r_f$	0.729	0.11						0.738	0.26				
CAYF	0.723	0.50	0.52					0.722	0.57	0.57			
CAYMSsm	0.749	0.47	0.47	0.54				0.735	0.16	0.16	0.19		
CAYMSfil	0.750	0.54	0.54	0.60	0.90			0.721	0.15	0.15	0.17	0.44	
CAYTVM	0.692	0.30	0.30	0.39	0.41	0.34		0.689	0.20	0.20	0.16	0.71	0.73
$const$	3.534						h=4	3.186					
$r-r_f$	3.540	0.24						3.214	0.87				
CAYF	3.281	0.24	0.27					3.037	0.33	0.29			
CAYMSsm	3.078	0.03	0.05	0.15				2.842	0.02	0.01	0.08		
CAYMSfil	2.911	0.00	0.01	0.05	0.07			2.828	0.01	0.01	0.06	0.35	
CAYTVM	2.823	0.01	0.01	0.06	0.49	0.92		2.545	0.00	0.00	0.01	0.05	0.05
$const$	7.039						h=8	6.767					
$r-r_f$	7.107	0.33						6.916	0.94				
CAYF	5.914	0.07	0.08					5.751	0.09	0.07			
CAYMSsm	3.615	0.00	0.00	0.00				5.350	0.00	0.00	0.15		
CAYMSfil	3.449	0.00	0.00	0.00	0.08			5.322	0.00	0.00	0.14	0.19	
CAYTVM	4.048	0.00	0.00	0.00	1.00	1.00		3.628	0.00	0.00	0.00	0.00	0.01
$const$	10.791						h=12	10.576					
$r-r_f$	10.911	0.58						10.707	1.00				
CAYF	8.236	0.01	0.01					8.832	0.07	0.03			
CAYMSsm	5.168	0.00	0.00	0.00				8.340	0.00	0.00	0.19		
CAYMSfil	5.044	0.00	0.00	0.00	0.15			8.329	0.00	0.00	0.20	0.77	
CAYTVM	5.762	0.00	0.00	0.00	1.00	1.00		5.68	0.00	0.00	0.00	0.00	0.00
$const$	14.415						h=16	14.069					
$r-r_f$	14.473	0.25						14.126	0.57				
CAYF	9.988	0.00	0.00					12.286	0.12	0.11			
CAYMSsm	8.348	0.00	0.00	0.08				11.214	0.00	0.00	0.12		
CAYMSfil	8.203	0.00	0.00	0.07	0.18			11.232	0.00	0.00	0.12	0.60	
CAYTVM	8.209	0.00	0.00	0.10	0.96	1.00		9.140	0.00	0.00	0.00	0.06	0.05

We report the p values of the DM test applied on the results from Table 4. The null hypothesis is that the two competing forecasting models have the same predictive accuracy while the alternative is that accuracies differ. The results should therefore be read with reference to the respective reported RMSEs in the first column.

Finally, Table 8 reports RMSE from the corresponding recursively estimated models, where all the models are estimated as described above. So these are genuinely out-of-sample forecasts. In the shorter sample the $CAYTVM$ is the best performer at $h = 1$, 4 and 16, CAYMS at 8 and 12. For the full sample, $CAYTVM$ is in all cases the best. Looking at the DM tests in Table 9, we see that the tests of best against next-best performer are insignificant at all horizons. However, $CAYTVM$ is significantly better than CAYMS at all horizons except 1.

We conclude that our smooth time variation method of estimating CAY delivers strong evidence of predictability and forecasting power. This evidence is particularly powerful for the genuinely recursive forecasting exercise.

5.2 Financial ratios

Tables 10 and 11 report the long-horizon predictability results from our variants. For $X= DP$, EP X is the unadjusted series, $XAdj$ the mean-adjusted, $\hat{\varepsilon}_{\bar{X}_{TV}}$ the time-varying mean adjusted, MS_X^{sm} the smoothed Markov switching and MS_X^{fil} the MS filtered.

Using the same metrics as previously, for DP our time varying method is the clear best performer over both the samples. For DP, the mean-adjusted measure performs well on significance and R^2 criteria, but is dominated by our time varying measure. The Markov-switching approach does less well, although better than the unadjusted series itself.

For EP over the shorter pre-crisis period our TV measure is the only serious contender. However, for the full sample including the post-crisis period, its performance collapses and the filtered MS version is the only reasonable performer. As we observed above (see eg Figure 9), the financial crisis had a dramatic, counter-intuitive (earnings moved in the opposite direction to dividends) and short lived effect on the EP data.

Table 12 reports the results for the recursive out-of-sample forecasts. For DP, our time varying method is best by a large margin for both time series. Moreover, as Table 14 shows, these improvements are overwhelmingly significant. These are strong results. For EP (also in Table 12) our method dominates in the period to 2007, and is again highly significant. But as for the predictive results, this is not the case for the full sample, where the filtered MS model is best except to horizon $h = 16$.

For the non-recursive rolling window forecasts, the time varying DP is best in all but one case (for the pre-crisis sample at $h = 16$ it is beaten by the adjusted measure by a small margin). For EP, in the shorter sample the TVM is on balance best although at horizons $h = 1$ and 16 the mean adjusted EP is best. In the extended sample the mean adjusted measure is best at all horizons and the TV measure worst, which we attribute to the crisis episode.

5.3 Relative forecast performance

Finally, we note that although the CAY variants significantly outperform the benchmark forecast at almost all horizons and that as observed above our mew time varying mean estimator is on balance the strongest performer, it is overwhelmingly dominated by the financial ratios, and especially by the dividend price ratio. But our point in this paper is not to run a forecast performance race, but to demonstrate that our method of allowing for location shifts is generally a strong contender.

Table 10: Long-horizon predictability: DP

	1952 to 2007					1952 to 2019				
	h=1	h=4	h=8	h=12	h=16	h=1	h=4	h=8	h=12	h=16
DP	0.02	0.09	**0.17**	**0.22**	**0.25**	0.02	0.08	**0.15**	**0.19**	**0.22**
	1.46	1.03	0.66	0.63	0.67	1.43	0.99	0.68	0.66	0.73
	1.52	1.69	2.26	2.69	2.85	1.46	1.86	2.07	2.05	1.96
	0.01	0.04	0.09	0.11	0.12	0.01	0.04	0.07	0.08	0.09
$DPAdj$	**0.14**	**0.49**	**0.76**	**0.87**	**0.67**	**0.12**	**0.45**	**0.71**	**0.80**	**0.75**
	4.00	3.49	2.50	2.65	1.98	3.63	3.61	3.41	3.25	2.93
	4.37	5.28	3.37	3.14	2.22	3.62	5.05	4.49	3.93	3.71
	0.07	0.21	0.27	0.26	0.13	0.07	0.22	0.28	0.27	0.19
$\hat{\varepsilon}_{\bar{DP}_{TV}}$	**0.16**	**0.63**	**1.04**	**1.22**	**1.22**	**0.13**	**0.55**	**0.93**	**1.09**	**1.14**
	4.33	4.18	4.05	3.96	4.62	3.43	4.18	3.71	3.69	5.64
	4.41	5.44	5.86	5.88	5.58	3.43	5.29	5.83	5.76	6.35
	0.08	0.31	0.46	0.47	0.40	0.06	0.25	0.38	0.39	0.35
MS_{DP}^{sm}	**0.07**	**0.29**	**0.48**	**0.53**	0.45	**0.06**	**0.27**	0.48	0.55	0.52
	2.76	1.98	1.47	1.04	0.87	2.20	2.15	1.77	1.51	1.43
	2.84	2.99	2.25	2.32	1.69	2.26	3.11	2.54	2.42	2.07
	0.04	0.16	0.23	0.20	0.12	0.02	0.12	0.21	0.20	0.15
MS_{DP}^{fil}	**0.05**	0.20	0.31	0.31	0.26	**0.06**	**0.25**	**0.43**	**0.48**	0.44
	2.02	1.16	0.76	0.52	0.41	2.03	2.05	1.45	1.20	0.96
	2.07	1.59	1.38	1.24	0.95	2.09	2.70	2.25	2.17	1.76
	0.02	0.08	0.10	0.07	0.04	0.02	0.11	0.17	0.16	0.11

Results from regressions of h-period-ahead cumulated CRSP-VW returns in excess of a 3-month Treasury-bill. We report OLS estimates of the regressors, Newey-West (1987) corrected t-statistics in parentheses, HAC corrected statistics in second parentheses (Zeileis 2006;weighted information sandwich variance estimators for parametric models fitted to time series data) and adjusted R^2 statistics in brackets. Significant coefficients at the 5% significance level are highlighted in bold face.

Table 11: Long-horizon predictability: EP

	1952 to 2007					1952 to 2019				
	h=1	h=4	h=8	h=12	h=16	h=1	h=4	h=8	h=12	h=16
EP	0.02	0.08	0.13	0.17	0.17	0.01	0.04	0.07	0.08	0.07
	1.13	0.89	0.67	0.62	0.54	0.52	0.54	0.45	0.43	0.28
	1.19	1.70	1.62	1.76	1.77	0.50	0.97	0.82	0.81	0.67
	0.00	0.03	0.05	0.06	0.05	0.00	0.01	0.01	0.01	0.01
$EPAdj$	**0.07**	**0.24**	0.27	0.33	0.22	0.02	0.08	0.07	0.08	0.02
	3.19	2.78	1.48	1.21	0.79	0.74	0.77	0.45	0.39	0.10
	3.06	3.29	1.94	1.69	1.00	0.78	0.85	0.52	0.48	0.14
	0.03	0.08	0.05	0.05	0.02	0.00	0.01	0.00	0.00	0.00
$\hat{\varepsilon}_{\bar{EP}_{TV}}$	**0.09**	**0.35**	**0.55**	**0.64**	**0.60**	0.02	0.11	0.15	0.17	0.12
	5.04	3.45	2.40	1.75	1.58	0.73	0.84	0.67	0.64	0.40
	2.96	5.23	3.18	2.72	2.43	0.76	0.97	0.79	0.74	0.52
	0.04	0.17	0.21	0.21	0.15	0.00	0.03	0.03	0.03	0.01
MS_{EP}^{sm}	0.04	0.14	0.15	0.17	0.10	0.01	0.04	0.06	0.08	0.06
	1.58	1.03	0.60	0.46	0.25	0.62	0.53	0.46	0.42	0.22
	1.52	1.86	1.43	0.86	0.37	0.54	0.80	0.73	0.66	0.41
	0.01	0.03	0.02	0.02	0.00	0.00	0.01	0.01	0.01	0.00
MS_{EP}^{fil}	0.03	0.11	0.14	0.16	0.09	0.01	0.04	0.06	0.07	0.06
	1.29	0.78	0.55	0.44	0.23	0.54	0.46	0.46	0.40	0.22
	1.27	1.40	1.25	0.79	0.34	0.48	0.73	0.70	0.62	0.38
	0.01	0.02	0.02	0.01	0.00	0.48	0.73	0.70	0.62	0.38

Results from regressions of h-period-ahead cumulated CRSP-VW returns in excess of a 3-month Treasury-bill. We report OLS estimates of the regressors, Newey-West (1987) corrected t-statistics in parentheses, HAC corrected statistics in second parentheses (Zeileis 2006;weighted information sandwich variance estimators for parametric models fitted to time series data) and adjusted R^2 statistics in brackets. Significant coefficients at the 5% significance level are highlighted in bold face.

Table 12: Recursive forecast RMSE, financial ratios

	1952-2007					1952-2019				
Model	$h=1$	$h=4$	$h=8$	$h=12$	$h=16$	$h=1$	$h=4$	$h=8$	$h=12$	$h=16$
DP	0.21	0.80	1.50	2.52	3.73	0.14	0.55	1.06	1.81	2.67
DP_{Adj}	0.18	0.78	1.70	2.60	3.36	0.18	0.79	1.69	2.49	3.12
$\hat{\varepsilon}_{\bar{DP}_{TV}}$	**0.12**	**0.48**	**0.98**	**1.52**	**2.34**	**0.11**	**0.38**	**0.63**	**1.00**	**1.58**
MS_{DP}^{sm}	0.18	0.87	1.62	2.17	3.30	0.18	0.74	0.97	1.58	2.15
MS_{DP}^{fil}	0.18	0.84	1.56	2.09	3.20	0.17	0.72	0.94	1.53	2.12
EP	0.21	0.77	1.58	2.40	3.22	0.19	0.70	1.46	2.27	3.13
EP_{Adj}	0.18	0.74	1.50	2.27	2.98	0.18	0.74	1.51	2.22	**2.85**
$\hat{\varepsilon}_{\bar{EP}_{TV}}$	**0.12**	**0.50**	**1.08**	**1.67**	**2.51**	0.18	0.73	1.48	2.20	2.92
MS_{EP}^{sm}	0.18	0.79	1.74	2.81	4.01	0.16	0.68	1.43	2.25	3.17
MS_{EP}^{fil}	0.18	0.79	1.72	2.78	3.96	**0.16**	**0.67**	**1.40**	**2.17**	3.04
$const$	0.73	3.53	7.04	10.79	14.41	0.73	3.19	6.77	10.58	14.07
$r-r_f$	0.73	3.54	7.19	10.91	14.47	0.74	3.21	6.92	10.71	14.13

Note: The table provides the $RMSE$ for the recursive exercise. The window used is 1952Q1 - 1980Q4. Minimum RMSE in bold.

Table 13: Non recursive rolling forecast RMSE

	1952-2007					1952-2019				
Model	$h=1$	$h=4$	$h=8$	$h=12$	$h=16$	$h=1$	$h=4$	$h=8$	$h=12$	$h=16$
DP	0.79	1.86	3.27	4.49	5.28	0.72	1.41	1.96	2.71	3.16
DP_{Adj}	0.82	2.01	3.06	3.18	**3.40**	0.75	1.74	2.42	2.76	3.12
$\hat{\varepsilon}_{\bar{DP}_{TV}}$	**0.79**	**1.11**	**2.25**	**3.16**	3.72	**0.59**	**1.30**	**1.70**	**2.42**	**2.51**
MS^{sm}_{DP}	0.80	1.97	3.32	4.28	5.22	0.73	1.42	2.09	2.63	3.21
MS^{fil}_{DP}	0.80	1.98	3.35	4.33	5.32	0.72	1.50	2.15	2.64	3.22
EP	0.80	1.89	3.05	3.97	4.41	0.73	1.50	1.96	2.53	2.77
EP_{Adj}	**0.78**	1.79	2.85	3.40	**3.60**	**0.72**	**1.37**	**1.92**	**2.36**	**2.48**
$\hat{\varepsilon}_{\bar{EP}_{TV}}$	0.79	**1.45**	**2.42**	**3.24**	3.64	0.73	1.71	2.23	2.89	3.35
MS^{sm}_{EP}	0.80	1.88	2.88	3.58	3.83	0.73	1.56	2.21	2.71	3.06
MS^{fil}_{EP}	0.80	1.87	2.92	3.62	3.86	0.73	1.55	2.26	2.72	3.06
$const$	0.87	3.23	6.44	9.87	13.18	0.87	2.92	6.19	9.68	12.87
$r - r_f$	0.87	3.26	6.61	10.04	13.31	0.88	2.95	6.37	9.85	13.00

In the top panel we report mean-squared forecast errors from out-of-sample h-period ahead forecasts of cumulated CRSP-VW returns in excess of a 3-month Treasury-bill. The single predictor variable is listed in the first column. The forecasting equation estimated using a rolling window the size of the initial period 1952Q1 - 1980Q4). The predictor variables are computed using the entire sample, excepting $const$ and $r_t - r_t^f$.

Table 14: Diebold and Mariano results for DP Recursive approach results

	1952-2007							1952-2019						
Model	MSE	$const$	$r-r_f$	DP	DP_{Adj}	MS_{DP}^{sm}	MS_{DP}^{fil}	MSE	$const$	$r-r_f$	DP	DP_{Adj}	MS_{DP}^{sm}	MS_{DP}^{fil}
h=1														
$const$	0.734							0.732						
$r-r_f$	0.729	0.11						0.738	0.26					
DP	0.209	0.00	0.00					0.143	0.00	0.00				
DP_{Adj}	0.175	0.00	0.00	0.97				0.180	0.00	0.00	0.94			
MS_{DP}^{sm}	0.183	0.00	0.00	0.92	0.45			0.175	0.00	0.00	0.89	0.44		
MS_{DP}^{fil}	0.176	0.00	0.00	0.91	0.88	0.63		0.169	0.00	0.00	0.88	0.85	0.61	
$\hat{\varepsilon}_{\bar{DP}_{TV}}$	0.116	0.00	0.00	0.02	0.00	0.00	0.00	0.114	0.00	0.00	0.02	0.00	0.00	0.00
h=4														
$const$	3.534							3.186						
$r-r_f$	3.540	0.24						3.214	0.87					
DP	0.796	0.00	0.00					0.546	0.00	0.00				
DP_{Adj}	0.775	0.00	0.00	0.94				0.794	0.00	0.00	0.84			
MS_{DP}^{sm}	0.874	0.00	0.00	0.89	0.44			0.739	0.00	0.00	0.80	0.39		
MS_{DP}^{fil}	0.841	0.00	0.00	0.88	0.86	0.61		0.718	0.00	0.00	0.79	0.76	0.55	
$\hat{\varepsilon}_{\bar{DP}_{TV}}$	0.476	0.00	0.00	0.02	0.00	0.00	0.00	0.375	0.00	0.00	0.02	0.00	0.00	0.00
h=8														
$const$	7.039							6.767						
$r-r_f$	7.107	0.33						6.916	0.94					
DP	1.499	0.00	0.00					1.056	0.00	0.00				
DP_{Adj}	1.696	0.00	0.00	0.92				1.686	0.00	0.00	0.89			
MS_{DP}^{sm}	1.619	0.00	0.00	0.87	0.43			0.974	0.00	0.00	0.84	0.41		
MS_{DP}^{fil}	1.558	0.00	0.00	0.86	0.83	0.60		0.940	0.00	0.00	0.83	0.81	0.58	
$\hat{\varepsilon}_{\bar{DP}_{TV}}$	0.980	0.00	0.00	0.02	0.00	0.00	0.00	0.628	0.00	0.00	0.02	0.00	0.00	0.00
h=12														
$const$	10.791							10.576						
$r-r_f$	10.911	0.58						10.707	1.00					
DP	2.525	0.00	0.00					1.815	0.00	0.00				
DP_{Adj}	2.598	0.00	0.00	0.89				2.487	0.00	0.00	0.86			
MS_{DP}^{sm}	2.165	0.00	0.00	0.85	0.41			1.576	0.00	0.00	0.82	0.40		
MS_{DP}^{fil}	2.091	0.00	0.00	0.84	0.81	0.58		1.531	0.00	0.00	0.81	0.78	0.56	
$\hat{\varepsilon}_{\bar{DP}_{TV}}$	1.520	0.00	0.00	0.02	0.00	0.00	0.00	1.000	0.00	0.00	0.02	0.00	0.00	0.00
h=16														
$const$	14.415							14.069						
$r-r_f$	14.473	0.25						14.126	0.57					
DP	3.729	0.00	0.00					2.668	0.00	0.00				
DP_{Adj}	3.355	0.00	0.00	0.87				3.122	0.00	0.00	0.91			
MS_{DP}^{sm}	3.302	0.00	0.00	0.82	0.40			2.146	0.00	0.00	0.87	0.42		
MS_{DP}^{fil}	3.201	0.00	0.00	0.81	0.79	0.56		2.121	0.00	0.00	0.86	0.83	0.59	
$\hat{\varepsilon}_{\bar{DP}_{TV}}$	2.340	0.00	0.00	0.02	0.00	0.00	0.00	1.580	0.00	0.00	0.02	0.00	0.00	0.00

We report the p values of the DM test applied on the results from Table 4. The null hypothesis is that the two competing forecasting models have the same predictive accuracy while the alternative is that accuracies differ. The results should therefore be read with reference to the respective reported RMSEs in the first column.

Table 15: Diebold and Mariano results for EP Recursive approach results

	1952-2007							1952-2019						
Model	MSE	$const$	$r-r_f$	EP	EP_{Adj}	MS_{EP}^{sm}	MS_{EP}^{fil}	MSE	$const$	$r-r_f$	EP	EP_{Adj}	MS_{EP}^{sm}	MS_{EP}^{fil}
h=1														
$const$	0.734							0.732						
$r-r_f$	0.729	0.11						0.738	0.26					
EP	0.209	0.00	0.00					0.186	0.00	0.00				
EP_{Adj}	0.179	0.00	0.00	0.97				0.175	0.00	0.00	0.94			
MS_{EP}^{sm}	0.184	0.00	0.00	0.92	0.45			0.162	0.00	0.00	0.89	0.44		
MS_{EP}^{fil}	0.182	0.00	0.00	0.91	0.88	0.63		0.161	0.00	0.00	0.88	0.85	0.61	
$\hat{\varepsilon}_{\bar{EP}_{TV}}$	0.116	0.00	0.00	0.02	0.00	0.00	0.00	0.175	0.00	0.00	0.02	0.00	0.00	0.00
h=4														
$const$	3.534							3.186						
$r-r_f$	3.540	0.24						3.214	0.87					
EP	0.765	0.00	0.00					0.702	0.00	0.00				
EP_{Adj}	0.737	0.00	0.00	0.94				0.744	0.00	0.00	0.84			
MS_{EP}^{sm}	0.794	0.00	0.00	0.89	0.44			0.683	0.00	0.00	0.80	0.39		
MS_{EP}^{fil}	0.789	0.00	0.00	0.88	0.86	0.61		0.670	0.00	0.00	0.79	0.76	0.55	
$\hat{\varepsilon}_{\bar{EP}_{TV}}$	0.500	0.00	0.00	0.02	0.00	0.00	0.00	0.731	0.00	0.00	0.02	0.00	0.00	0.00
h=8														
$const$	7.039							6.767						
$r-r_f$	7.107	0.33						6.916	0.94					
EP	1.580	0.00	0.00					1.458	0.00	0.00				
EP_{Adj}	1.501	0.00	0.00	0.92				1.513	0.00	0.00	0.89			
MS_{EP}^{sm}	1.737	0.00	0.00	0.87	0.43			1.434	0.00	0.00	0.84	0.41		
MS_{EP}^{fil}	1.722	0.00	0.00	0.86	0.83	0.60		1.396	0.00	0.00	0.83	0.81	0.58	
$\hat{\varepsilon}_{\bar{EP}_{TV}}$	1.078	0.00	0.00	0.02	0.00	0.00	0.00	1.480	0.00	0.00	0.02	0.00	0.00	0.00
h=12														
$const$	10.791							10.576						
$r-r_f$	10.911	0.58						10.707	1.00					
EP	2.396	0.00	0.00					2.274	0.00	0.00				
EP_{Adj}	2.266	0.00	0.00	0.89				2.222	0.00	0.00	0.86			
MS_{EP}^{sm}	2.811	0.00	0.00	0.85	0.41			2.246	0.00	0.00	0.82	0.40		
MS_{EP}^{fil}	2.780	0.00	0.00	0.84	0.81	0.58		2.170	0.00	0.00	0.81	0.78	0.56	
$\hat{\varepsilon}_{\bar{EP}_{TV}}$	1.672	0.00	0.00	0.02	0.00	0.00	0.00	2.201	0.00	0.00	0.02	0.00	0.00	0.00
h=16														
$const$	14.415							14.069						
$r-r_f$	14.473	0.25						14.126	0.57					
EP	3.216	0.00	0.00					3.132	0.00	0.00				
EP_{Adj}	2.975	0.00	0.00	0.87				2.854	0.00	0.00	0.91			
MS_{EP}^{sm}	4.014	0.00	0.00	0.82	0.40			3.173	0.00	0.00	0.87	0.42		
MS_{EP}^{fil}	3.962	0.00	0.00	0.81	0.79	0.56		3.039	0.00	0.00	0.86	0.83	0.59	
$\hat{\varepsilon}_{\bar{EP}_{TV}}$	2.506	0.00	0.00	0.02	0.00	0.00	0.00	2.920	0.00	0.00	0.02	0.00	0.00	0.00

We report the p values of the DM test applied on the results from Table 4. The null hypothesis is that the two competing forecasting models have the same predictive accuracy while the alternative is that accuracies differ. The results should therefore be read with reference to the respective reported RMSEs in the first column.

Table 16: Diebold and Mariano results for DP Non Recursive rolling results

	1952-2007							1952-2019						
Model	MSE	$const$	$r-r_f$	DP	DP_{Adj}	MS_{DP}^{sm}	MS_{DP}^{fil}	MSE	$const$	$r-r_f$	DP	DP_{Adj}	MS_{DP}^{sm}	MS_{DP}^{fil}
h=1														
$const$	0.872							0.873						
$r-r_f$	0.873	0.29						0.879	0.23					
DP	0.786	0.00	0.00					0.718	0.00	0.00				
DP_{Adj}	0.821	0.00	0.00	0.97				0.754	0.00	0.00	0.94			
MS_{DP}^{sm}	0.799	0.00	0.00	0.92	0.45			0.732	0.00	0.00	0.89	0.44		
MS_{DP}^{fil}	0.803	0.00	0.00	0.91	0.88	0.63		0.724	0.00	0.00	0.88	0.85	0.61	
$\hat{\varepsilon}_{\bar{DP}_{TV}}$	0.792	0.00	0.00	0.02	0.00	0.00	0.00	0.591	0.00	0.00	0.02	0.00	0.00	0.00
h=4														
$const$	3.225							2.918						
$r-r_f$	3.259	0.21						2.947	0.77					
DP	1.858	0.00	0.00					1.411	0.00	0.00				
DP_{Adj}	2.008	0.00	0.00	0.94				1.735	0.00	0.00	0.84			
MS_{DP}^{sm}	1.971	0.00	0.00	0.89	0.44			1.418	0.00	0.00	0.80	0.39		
MS_{DP}^{fil}	1.975	0.00	0.00	0.88	0.86	0.61		1.502	0.00	0.00	0.79	0.76	0.55	
$\hat{\varepsilon}_{\bar{DP}_{TV}}$	1.105	0.00	0.00	0.02	0.00	0.00	0.00	1.298	0.00	0.00	0.02	0.00	0.00	0.00
h=8														
$const$	6.437							6.191						
$r-r_f$	6.610	0.29						6.373	0.83					
DP	3.273	0.00	0.00					1.962	0.00	0.00				
DP_{Adj}	3.059	0.00	0.00	0.92				2.420	0.00	0.00	0.89			
MS_{DP}^{sm}	3.317	0.00	0.00	0.87	0.43			2.091	0.00	0.00	0.84	0.41		
MS_{DP}^{fil}	3.345	0.00	0.00	0.86	0.83	0.60		2.153	0.00	0.00	0.83	0.81	0.58	
$\hat{\varepsilon}_{\bar{DP}_{TV}}$	2.247	0.00	0.00	0.02	0.00	0.00	0.00	1.698	0.00	0.00	0.02	0.02	0.00	0.00
h=12														
$const$	9.866							9.683						
$r-r_f$	10.039	0.51						9.848	0.89					
DP	4.494	0.00	0.00					2.712	0.00	0.00				
DP_{Adj}	3.181	0.00	0.00	0.89				2.759	0.00	0.00	0.86			
MS_{DP}^{sm}	4.276	0.00	0.00	0.85	0.41			2.627	0.00	0.00	0.82	0.40		
MS_{DP}^{fil}	4.325	0.00	0.00	0.84	0.81	0.58		2.638	0.00	0.00	0.81	0.78	0.56	
$\hat{\varepsilon}_{\bar{DP}_{TV}}$	3.165	0.00	0.00	0.02	0.00	0.00	0.00	2.422	0.00	0.00	0.02	0.00	0.00	0.00
h=16														
$const$	13.181							12.874						
$r-r_f$	13.307	0.22						12.998	0.51					
DP	5.278	0.00	0.00					3.158	0.00	0.00				
DP_{Adj}	3.395	0.00	0.00	0.87				3.119	0.00	0.00	0.91			
MS_{DP}^{sm}	5.215	0.00	0.00	0.82	0.40			3.206	0.00	0.00	0.87	0.42		
MS_{DP}^{fil}	5.323	0.00	0.00	0.81	0.79	0.56		3.220	0.00	0.00	0.86	0.83	0.59	
$\hat{\varepsilon}_{\bar{DP}_{TV}}$	3.721	0.00	0.00	0.02	0.00	0.00	0.00	2.514	0.00	0.00	0.02	0.02	0.00	0.00

We report the p values of the DM test applied on the results from Table 4. The null hypothesis is that the two competing forecasting models have the same predictive accuracy while the alternative is that accuracies differ. The results should therefore be read with reference to the respective reported RMSEs in the first column.

Table 17: Diebold and Mariano results for EP Non Recursive rolling results

	1952-2007							1952-2019						
Model	MSE	$const$	$r-r_f$	EP	DP_{Adj}	MS_{EP}^{sm}	MS_{EP}^{fil}	MSE	$const$	$r-r_f$	EP	DE_{Adj}	MS_{DE}^{sm}	MS_{EP}^{fil}
h=1														
$const$	0.872							0.873						
$r-r_f$	0.873	0.29						0.879	0.23					
EP	0.802	0.00	0.00					0.732	0.00	0.00				
EP_{Adj}	0.777	0.00	0.00	0.97				0.715	0.00	0.00	0.94			
MS_{EP}^{sm}	0.803	0.00	0.00	0.92	0.45			0.734	0.00	0.00	0.89	0.44		
MS_{EP}^{fil}	0.797	0.00	0.00	0.91	0.88	0.63		0.732	0.00	0.00	0.88	0.85	0.61	
$\hat{\varepsilon}_{\bar{EP}_{TV}}$	0.792	0.00	0.00	0.02	0.00	0.00	0.00	0.734	0.00	0.00	0.02	0.00	0.00	0.00
h=4														
$const$	3.225							2.918						
$r-r_f$	3.259	0.21						2.947	0.77					
EP	1.888	0.00	0.00					0.702	0.00	0.00				
EP_{Adj}	1.790	0.00	0.00	0.94				0.744	0.00	0.00	0.84			
MS_{EP}^{sm}	1.883	0.00	0.00	0.89	0.44			0.683	0.00	0.00	0.80	0.39		
MS_{EP}^{fil}	1.874	0.00	0.00	0.88	0.86	0.61		0.670	0.00	0.00	0.79	0.76	0.55	
$\hat{\varepsilon}_{\bar{EP}_{TV}}$	1.452	0.00	0.00	0.02	0.02	0.00	0.00	0.731	0.00	0.00	0.02	0.02	0.02	0.00
h=8														
$const$	6.437							6.191						
$r-r_f$	6.610	0.29						6.373	0.83					
EP	3.045	0.00	0.00					1.501	0.00	0.00				
EP_{Adj}	2.852	0.00	0.00	0.92				1.373	0.00	0.00	0.89			
MS_{EP}^{sm}	2.882	0.00	0.00	0.87	0.43			1.560	0.00	0.00	0.84	0.41		
MS_{EP}^{fil}	2.916	0.00	0.00	0.86	0.83	0.60		1.546	0.00	0.00	0.83	0.81	0.58	
$\hat{\varepsilon}_{\bar{EP}_{TV}}$	2.422	0.00	0.00	0.02	0.02	0.00	0.00	1.712	0.00	0.00	0.02	0.02	0.00	0.00
h=12														
$const$	9.866							9.683						
$r-r_f$	10.039	0.51						9.848	0.89					
EP	3.967	0.00	0.00					2.526	0.00	0.00				
EP_{Adj}	3.402	0.00	0.00	0.89				2.362	0.00	0.00	0.86			
MS_{EP}^{sm}	3.582	0.00	0.00	0.85	0.41			2.707	0.00	0.00	0.82	0.40		
MS_{EP}^{fil}	3.621	0.00	0.00	0.84	0.81	0.58		2.722	0.00	0.00	0.81	0.78	0.56	
$\hat{\varepsilon}_{\bar{EP}_{TV}}$	3.238	0.00	0.00	0.02	0.02	0.00	0.00	2.887	0.00	0.00	0.02	0.02	0.00	0.00
h=16														
$const$	13.181							12.874						
$r-r_f$	13.307	0.22						12.998	0.51					
EP	4.412	0.00	0.00					2.770	0.00	0.00				
EP_{Adj}	3.598	0.00	0.00	0.87				2.481	0.00	0.00	0.91			
MS_{EP}^{sm}	3.829	0.00	0.00	0.82	0.40			3.057	0.00	0.00	0.87	0.42		
MS_{EP}^{fil}	3.863	0.00	0.00	0.81	0.79	0.56		3.063	0.00	0.00	0.86	0.83	0.59	
$\hat{\varepsilon}_{\bar{EP}_{TV}}$	3.635	0.00	0.00	0.02	0.02	0.02	0.00	3.351	0.00	0.00	0.02	0.02	0.00	0.00

We report the p values of the DM test applied on the results from Table 4. The null hypothesis is that the two competing forecasting models have the same predictive accuracy while the alternative is that accuracies differ. The results should therefore be read with reference to the respective reported RMSEs in the first column.

6 Robustness: predictability

As a final check, we apply the tests developed in Harvey et al. (2021)[17] which they show display attractive finite sample size control and power across a wide range of persistence and endogeneity levels for the predictor. Their approach is based on the standard regression t-ratio and a variant where the predictor is quasi-GLS demeaned. In the strongly persistent near-unit root environment, the limiting null distributions of these statistics depend on the endogeneity and local-to-unity parameters characterising the predictor. Analysis of the asymptotic local power functions of feasible implementations of these two tests, based on asymptotically conservative critical values, motivates a switching procedure between the two, employing the quasi-GLS demeaned variant unless the magnitude of the estimated endogeneity correlation parameter is small. Additionally, if the data suggests the predictor is weakly persistent, the test statistic switches to the standard t-ratio test. They show that the test robustly out-performs alternatives, including the IVX estimation method of Kostakis et al. (2015). As it is designed for single period prediction, we apply it to that case alone. The tables below report upper tail approximately normal tests against the null of non-predictability, together with the estimated serial correlation $\hat{\rho}_{yx}$, the selected tests τ_{hyb}= and the test statistic itself, τ_{hyb}.

We also present the mean of the tests from the recursive results. In all cases our method is significant at 5% or above, and moreover has a p-value that is less than any of the other methods.

[17]We are grateful to Rob Taylor for making their code available.

Table 18: Harvey *et al* 2021 tests for one-step predictability: CAY variants

	Full Sample			***Recursive***		
	$\hat{\rho}_{yx}$	τ_{hyb}=	τ_{hyb}	$\hat{\rho}_{yx}$	τ_{hyb}=	τ_{hyb}
1952-2007						
const	0.051	τ_{con}	0.530	0.055	τ_{con}	0.569
$r - r_f$	0.864	τ'_{con}	1.370	0.901	τ'_{con}	1.428
CAYF	0.670	τ_N	**1.920**	0.665	τ_N	1.905
CAYMS	0.682	τ_N	**1.930**	0.676	τ_N	1.913
CAYTVM	0.657	τ_N	**2.180**	0.659	τ_N	2.185
1952-2019						
const	0.050	τ_{con}	0.560	0.054	τ_{con}	0.980
$r - r_f$	0.849	τ'_{con}	1.550	0.885	τ'_{con}	1.470
CAYF	0.725	τ_N	**1.905**	0.720	τ_N	1.891
CAYMS	0.738	τ_N	**1.913**	0.732	τ_N	1.896
CAYTVM	0.711	τ_N	**2.185**	0.713	τ_N	2.166

The column labelled τ_{hyb}= states which of the constituent tests is selected in the hybrid test τ_{hyb}. Tests upper tail approximately normal tests against the null of non-predictability. Estimated serial correlation $\hat{\rho}_{yx}$, test statistic τ_{hyb}: bold indicates significance at 5% for full sample results. The recursive estimates report the average of the predictability tests performed fully recursively.

Table 19: Harvey *et al* 2021 tests for one-step predictability: DP variants

	Full Sample			***Recursive***		
	$\hat{\rho}_{yx}$	$\tau_{hyb}=$	τ_{hyb}	$\hat{\rho}_{yx}$	$\tau_{hyb}=$	τ_{hyb}
1952-2007						
$const$	0.051	τ_{con}	0.530	0.055	τ_{con}	0.569
$r - r_f$	0.864	τ'_{con}	1.379	0.901	τ'_{con}	1.428
DP	-0.980	τ'_{con}	0.884	-0.988	τ'_{con}	0.949
DP_{Adj}	-0.993	τ'_{con}	1.264	-0.994	τ'_{con}	1.318
$\hat{\varepsilon}_{\bar{DP}_{TV}}$	-0.994	τ_N	**2.080**	-0.994	τ_N	2.064
MS^{sm}_{DP}	-0.992	τ_N	**1.673**	-0.995	τ_N	1.658
MS^{fil}_{DP}	-0.992	τ_N	**1.661**	-0.99	τ_N	1.665
1952-2019						
$const$	0.050	τ_{con}	0.560	0.054	τ_{con}	0.980
$r - r_f$	0.849	τ'_{con}	1.550	0.885	τ'_{con}	1.470
DP	-0.973	τ'_{con}	0.877	-0.981	τ'_{con}	0.942
DP_{Adj}	-0.984	τ'_{con}	1.253	-0.990	τ'_{con}	1.306
$\hat{\varepsilon}_{\bar{DP}_{TV}}$	-0.985	τ_N	**2.062**	-0.992	τ_N	2.046
MS^{sm}_{DP}	-0.983	τ_N	**1.658**	-0.993	τ_N	1.644
MS^{fil}_{DP}	-0.983	τ_N	**1.646**	-0.993	τ_N	1.632

The column labelled $\tau_{hyb}=$ states which of the constituent tests is selected in the hybrid test τ_{hyb}. Tests upper tail approximately normal tests against the null of non-predictability. Estimated serial correlation $\hat{\rho}_{yx}$, test statistic τ_{hyb}: bold indicates significance at 5% for full sample results. The recursive estimates report the average of the predictability tests performed fully recursively.

Table 20: Harvey *et al* 2021 tests for one-step predictability: EP variants

	Full Sample			***Recursive***		
	$\hat{\rho}_{yx}$	$\tau_{hyb}=$	τ_{hyb}	$\hat{\rho}_{yx}$	$\tau_{hyb}=$	τ_{hyb}
1952-2007						
$const$	0.051	τ_{con}	0.530	0.055	τ_{con}	0.569
$r - r_f$	0.864	τ_{con}	1.370	0.901	$\tau_{con}^{'}$	1.428
EP	-0.662	$\tau_{con}^{'}$	0.679	-0.676	$\tau_{con}^{'}$	0.729
EP_{Adj}	-0.675	$\tau_{con}^{'}$	0.987	-0.680	$\tau_{con}^{'}$	1.029
$\hat{\varepsilon}_{\bar{EP}_{TV}}$	-0.688	$\tau_{con}^{'}$	**1.840**	-0.692	$\tau_{con}^{'}$	1.858
MS_{EP}^{sm}	-0.698	$\tau_{con}^{'}$	**1.651**	-0.729	$\tau_{con}^{'}$	1.636
MS_{EP}^{fil}	-0.698	$\tau_{con}^{'}$	**1.639**	-0.727	$\tau_{con}^{'}$	1.643
1952-2019						
$const$	0.050	τ_{con}	0.560	0.054	τ_{con}	0.980
$r - r_f$	0.849	$\tau_{con}^{'}$	1.550	0.885	$\tau_{con}^{'}$	1.470
EP	-0.685	$\tau_{con}^{'}$	0.673	-0.726	$\tau_{con}^{'}$	0.722
EP_{Adj}	-0.687	$\tau_{con}^{'}$	0.961	-0.708	$\tau_{con}^{'}$	1.002
$\hat{\varepsilon}_{\bar{EP}_{TV}}$	-0.718	τ_N	**2.046**	-0.699	τ_N	2.067
MS_{EP}^{sm}	-0.714	τ_N	**2.001**	-0.730	τ_N	2.006
MS_{EP}^{fil}	-0.713	τ_N	**1.987**	-0.729	τ_N	1.992

The column labelled $\tau_{hyb}=$ states which of the constituent tests is selected in the hybrid test τ_{hyb}. Tests upper tail approximately normal tests against the null of non-predictability. Estimated serial correlation $\hat{\rho}_{yx}$, test statistic τ_{hyb}: bold indicates significance at 5% for full sample results. The recursive estimates report the average of the predictability tests performed fully recursively.

7 Conclusions

Predictability of returns is widely believed to exist but is subject to instabilities. In a frequently cited paper that focuses on dividend price ratios, (Lettau and Nieuwerburgh (2008)) argue that these are endemic but that allowing for discrete shifts in the unconditional mean (location shifts) can capture them well and restore predictability. In a related paper Bianchi et al. (2016) established similar results in the context of the consumption-asset-income (CAY) relation. We re-examine this for US stock returns, allowing for smooth variation in location, which we suggest is more plausible than abrupt regime changes and discrete shifts. We find that the ratios we examine do not exhibit stationarity (cointegration) in fixed parameter relationships, but that there is much stronger evidence for stationarity using smooth parameter variation in the location parameter, more so than for the other methods we examine.

Using the new method we look for evidence of predictability in returns and forecasting ability. We find that our technique offers uniformly superior predictability and forecasting performance than methods using Markov-switching or pre-tested (Bai-Perron) mean shifts for CAY and the dividend price ratio, using data over two samples, one over the pre-crisis period between 1952 and 2007, and one over a longer sample to 2019 (pre-pandemic). The success with the recursive forecast is particularly impressive.

The evidence for the earnings price ratio is equally strong for the pre-crisis period, but the measure fails for the extended sample, which we put down to the extraordinary movement in the price-earnings relationship in the immediate crisis period after 2008.

For predictability, there are also econometric advantages. Our unadjusted predictors exhibit strong persistence and near or actual unit roots. The conventional solution of various HAC corrections have been criticised and IVX methods have become popular, following Kostakis et al. (2015). Our method removes the need for these methods. Moreover, the method using prior selection for strength of persistence due to Harvey et al. (2021) produces similar results to ours for one-step ahead returns (selecting weak persistence for our measures). Consequently, we advocate the use of our method on both financial theoretic and econometric grounds.

References

ANG, A. AND G. BEKAERT (2007): "Stock Return Predictability: Is it There?" *Review of Financial Studies*, 20, 651–707.

BAI, J. AND P. PERRON (1998): "Estimating and Testing Linear Models with Multiple Structural Changes," *Econometrica*, 66, 47–78.

——— (2003): "Computation and analysis of multiple structural change models," *Journal of Applied Econometrics*, 18, 1–22.

BARNETT, A., H. MUMTAZ, AND K. THEODORIDIS (2014): "Forecasting UK GDP growth and inflation under structural change. A comparison of models with time-varying parameters," *International Journal of Forecasting*, 30, 129 – 143.

BENATI, L. (2008): "The 'Great Moderation' in the United Kingdom," *Journal of Money, Credit and Banking*, 40, 121–147.

BIANCHI, F., M. LETTAU, AND S. C. LUDVIGSON (2016): "Monetary policy and asset valuation: evidence from a Markov switching *cay*," *NBER Working Paper 22572 (revised 2021).*

——— (2021): "Monetary policy and asset valuation," *Journal of Finance (forthcoming).*

BIERENS, H. AND L. MARTINS (2010): "Time-varying cointegration," *Econometric Theory*, 26, 1453–1490.

BOUDOUKH, J., M. RICHARDSON, AND R. WHITELAW (2005): "The Myth of Long-Horizon Predictability," NBER Working Papers 11841, National Bureau of Economic Research, Inc.

BRENNAN, M. J. AND Y. XIA (2005): "tay's as good as cay," *Finance Research Letters*, 2, 1 – 14.

CAMPBELL, J. AND N. G. MANKIW (1989): "Consumption, Income and Interest Rates: Reinterpreting the Time Series Evidence," in *NBER Macroeconomics Annual 1989, Volume 4*, National Bureau of Economic Research, Inc, 185–246.

CAMPBELL, J. Y. AND R. J. SHILLER (1988): "Stock Prices, Earnings, and Expected Dividends," *Journal of Finance*, 43, 661–676.

COCHRANE, J. H. (2008): "The Dog That Did Not Bark: A Defense of Return Predictability," *Review of Financial Studies*, 21, 1533–1575.

COGLEY, T., G. PRIMICERI, AND T. SARGENT (2010): "Inflation-Gap Persistence in the US," *American Economic Journal: Macroeconomics*, 2, 43–69.

Cogley, T. and T. J. Sargent (2001): "Evolving Post-World War II U.S. Inflation Dynamics," *NBER Macroeconomics Annual*, 16, 331–73.

——— (2005): "Drift and Volatilities: Monetary Policies and Outcomes in the Post WWII U.S," *Review of Economic Dynamics*, 8, 262–302.

Dahlhaus, R. (2000): "A likelihood approximation for locally stationary processes," *The Annals of Statistics*, 28, 1762–1794.

Dahlhaus, R. and W. Polonik (2006): "Nonparametric quasi maximum likelihood estimation for Gaussian locally stationary processes," *The Annals of Statistics*, 34, 2790–2824.

Elsby, M., B. Hobijn, and A. Sahin (2013): "The Decline of the U.S. Labor Share," *Brookings Papers on Economic Activity*, 44, 1–63.

Favero, C. A., A. E. Gozluklu, and A. Tamoni (2011): "Demographic Trends, the Dividend-Price Ratio, and the Predictability of Long-Run Stock Market Returns," *Journal of Financial and Quantitative Analysis*, 46, 1493–1520.

Fernandez-Corugedo, E., S. Price, and A. P. Blake (2007): "The dynamics of aggregate UK consumers' non-durable expenditure," *Economic Modelling*, 24, 453–469.

Giraitis, L., G. Kapetanios, and T. Yates (2014a): "Inference on stochastic time-varying coefficient models," *Journal of Econometrics*, 179, 46–65.

——— (2014b): "Inference on Stochastic Time-Varying Coefficient Models," *Journal of Econometrics*, 179, 46–65.

——— (2018): "Inference on multivariate heteroscedastic stochastic time varying coefficient models," *Journal of Time Series Analysis*, 39, 129–149.

Hamilton, J. D. (1996): "Specification testing in Markov-switching time-series models," *Journal of Econometrics*, 70, 127–157.

Harvey, D. I., S. J. Leybourne, and A. R. Taylor (2021): "Simple tests for stock return predictability with good size and power properties," *Journal of Econometrics*, 224, 198–214.

John Y. Campbell, R. J. S. (1988a): "The Dividend-Price Ratio and Expectations of Future Dividends and Discount Factors," *Review of Financial Studies*, 1, 195–228.

——— (1988b): "The Dividend-Price Ratio and Expectations of Future Dividends and Discount Factors," *Review of Financial Studies*, 1, 195–228.

Kapetanios, G., S. Millard, K. Petrova, and S. Price (2019): "Time varying cointegration and the UK Great Ratios," Staff Working Papers 789, Bank of England.

——— (2020): "Time-varying cointegration with an application to the UK Great Ratios," *Economics Letters*, 193.

Kostakis, A., T. Magdalinos, and M. P. Stamatogiannis (2015): "Robust Econometric Inference for Stock Return Predictability," *Review of Financial Studies*, 28.

Lettau, M. and S. Ludvigson (2001): "Consumption, Aggregate Wealth, and Expected Stock Returns," *Journal of Finance*, 56, 815–849.

——— (2004): "Understanding Trend and Cycle in Asset Values: Reevaluating the Wealth Effect on Consumption," *American Economic Review*, 94, 276–299.

Lettau, M. and S. V. Nieuwerburgh (2008): "Reconciling the Return Predictability Evidence," *Review of Financial Studies*, 21, 1607–1652.

Paye, B. S. and A. Timmermann (2006): "Instability of return prediction models," *Journal of Empirical Finance*, 13, 274–315.

Phillips, P., D. Li, and J. Gao (2017): "Estimating smooth structural change in cointegration models," *Journal of Econometrics*, 196, 180–195.

Price, S. and C. Schleicher (2005): "Returns To Equity, Investment And Q: Evidence From The UK," *Manchester School*, 73, 32–57.

Rapach, D. E. and M. Wohar (2006): "Structural Breaks and Predictive Regression Models of Aggregate U.S. Stock Returns," *Journal of Financial Econometrics*, 4, 238–274.

Rodrigues, P. M., M. Demetrescu, I. Georgiev, and A. M. R. Taylor (2019): "Testing for Episodic Predictability in Stock Returns," Tech. rep.

Rossi, B. (2013): *Advances in Forecasting under Instability*, Elsevier, vol. 2 of *Handbook of Economic Forecasting*, chap. 0, 1203–1324.

Sims, C. and T. Zha (2006): "Were There Regime Switches in U.S. Monetary Policy?" *American Economic Review*, 96, 54–81.

Stock, J. H. and M. W. Watson (1993): "A Simple Estimator of Cointegrating Vectors in Higher Order Integrated Systems," *Econometrica*, 61, 783–820.

——— (1996): "Evidence on Structural Instability in Macroeconomic Time Series Relations," *Journal of Business & Economic Statistics*, 14, 11–30.

Welch, I. and A. Goyal (2008): "A Comprehensive Look at The Empirical Performance of Equity Premium Prediction," *Review of Financial Studies*, 21, 1455–1508.

Wright, S. (2004): "Measures of stock market value and returns for the US nonfinacial corporate sector, 1900-2002," *Review of Income and Wealth*, 50, 561 – 84.