This is an Accepted Manuscript of an article published by Taylor & Francis Group in *The European Journal of Finance* on 29 Jan 2020, available online: http://www.tandfonline.com/10.1080/1351847X.2020.1719175.

Forecasting US Stock Returns

David G McMillan
Division of Accounting and Finance
University of Stirling

November 2018
Revised June 2019
Second Revision October 2019

Abstract

We forecast quarterly US stock returns using 25 predictor variables. We consider a breadth of forecast methods and metrics, including bi- and multi-variate regressions, linear and non-linear models, rolling and recursive techniques, forecast combinations and statistical and economic evaluation. In doing so, we extend existing research both in terms of the range of predictor series and the scope of the analysis. In common with much of literature, a broad view over the full set of predictor variables tends to indicate that such models are unable to beat the historical mean model. However, nuances to these results reveals forecast success varies according to how the forecasts are evaluated and over time. Notably, the results reveal that the term structure of interest rates consistently provides the preferred forecast performance, especially when evaluated using the Sharpe ratio. The purchasing managers index also consistently provides a strong forecast performance. Further results also reveal that forecast combinations over the full set of variables do not outperform the preferred single variable forecasts, while forecast combinations using an interest rate subset group do perform well. The success of the term structure and the purchasing managers index highlights the importance of, respectively, investor and firm expectations of future economic performance in providing valuable stock return forecasts. This is also consistent with asset pricing models that indicate movements in returns are conditioned by such expectations.

Keywords: Stock Returns, Forecasting, Time-Variation, Rolling, Recursive, Term Structure
JEL Codes: C22, G12

Address for Correspondence: Professor David McMillan,
Accounting and Finance Division, University of Stirling, FK9 4LA
Telephone: +44(0)1786-467309
Fax: +44(0)1786-467308
E-mail: david.mcmillan@stir.ac.uk

1. Introduction.

Stock return predictability remains elusive and much sought after. Stock return predictability ties together several strands within the asset pricing literature and so remains a key empirical research question. Evidence of predictability linked to specific financial or economic variables would advance our understanding of an underlying asset pricing model that argues current stock returns are linked to future movements in economic conditions. Moreover, it would improve our knowledge of the links between real and financial markets. Equally, evidence of predictability arising from the movement of past returns or characteristics not related to economic conditions would suggest a reassessment of our asset pricing models is required. Thus, evidence of predictability is important for academics and policy-makers in understanding the movements of stock returns. Further, regardless of the source of predictability, supportive evidence is of interest to investors in building portfolios and making market timing decisions.

While empirical research geared towards stock return predictability is a recurring theme, recent impetus to this area is given by the work of Campbell and Shiller (1988) and Fama and French (1988) both of whom argue that financial ratios exhibit predictive power for subsequent stock returns. Pesaran and Timmermann (1995, 2000) consider a wider range of economic variables and report supportive evidence of predictability. However, consistent evidence in favour of predictability is lacking. Notably, Ang and Bekaert (2007) and Welch and Goyal (2008) undertake comprehensive exercises that suggest limited evidence of predictability. An explanation for the lack of consistent evidence is provided by work that suggests the presence of regime shifts or non-linear dynamics within the predictive relation or that predictability itself is a temporary phenomenon. For example, McMillan (2003) argues that a non-linear model is required to uncover more supportive evidence of predictability. Paye and Timmermann (2006) suggest that breaks occur within the coefficient of the predictive regression, while Lettau and van Nieuwerburgh (2008) suggest the presence of breaks in the predictor variable. Timmermann (2008) argues that predictability only exists short-lived periods of time, while Campbell and Thompson (2008), Park (2010) and McMillan and Wohar (2013) equally argue that predictability is not constant over time. Henkel et al (2011) suggest that predictability only arises during economic downturns. More recently, Hammerschmid and Lohre (2018) provide evidence of predictability based on economic regimes, while Baltas and Karyampas (2018) highlight that forecast success is dependent upon identifying market regimes.

This paper focuses on the out-of-sample forecast ability of a range of 25 variables that include financial ratios, firm specific variables, macroeconomic variables and series that correspond to confidence and recent market behaviour. Thus, we include variables that can be regarded as indicators of fundamental economic conditions (such as GDP, inflation and consumption), indicators regarding expectation of future economic conditions (such as the term structure of interest rates and purchasing managers index) as well as stock market indicators (including financial ratios and a moving average). We consider these variables individually and in a multivariate regression setting and consider forecast combinations of the former. The modelling approach includes linear and non-linear models conducted using rolling and recursive approaches. The forecast evaluation utilises statistical and economic based measures and equally allows for regimes of behaviour to be identified according to both economic and market conditions.[1] Thus, we seek to provide a comprehensive evaluation of where forecast power occurs both in terms of predictor variables and across time and regimes of behaviour.

As noted, this research area is one for which an extensive literature exists. Within this literature we can identify several marquee papers, such as (but not limited to) Hjalmarsson

[1] Statistical based forecast evaluation (such as mean squatted error criteria) are conducted without reference to the specific context in which they are made. Thus, there is increasing use of economic based forecast measures, which are designed to be context relevant (e.g., Leitch and Tanner, 1991; Pesaran and Skouras, 2004). In the context of financial markets, this refers to the potential profitability of trading strategies based on the forecasts (see, for example, Campbell and Thompson, 2008; Maio, 2016).

(2008) and Welch and Goyal (2008), who demonstrate that evidence in support of predictability is limited across time and countries. Nonetheless, both papers point to the view that the term structure of interest rates does have greater forecast power compared to other variables. Building on this work, Rapach et al (2010) and Elliot et al (2013) argue that greater forecast power is revealed through forecast combinations. Moreover, Rapach et al (2010) note that the forecast combinations are linked to economic activity, which, they argue, enhances the reliability of the forecasts (also see the arguments by Cochrane, 2008). This reinforces the point noted above that evidence of predictability may also indicate support for the underlying asset pricing model. The link to economic activity is also argued by Henkel et al (2011) in leading to regime dependent predictive power.

The key issue, therefore, is how this paper extends this large literature. First, is the nature of the data we consider within the forecast exercise. The work of Welch and Goyal (2008) considers fifteen predictor variables, with all, bar two, financial market variables. The same data set is also utilised by Rapach et al (2010). The non-financial variables are inflation and the investment to capital ratio. Likewise, Hjalmarsson (2008) utilises four financial variables (stock price ratios and interest rates). This paper includes a broader set of variables covering both financial and macroeconomic variables. While financial market variables can be considered as including forward looking elements, for example, the dividend/price ratio weighs investor expectations of future against past performance, macroeconomic variables typically do not capture firms forward looking behaviour. Thus, we include a measure that proxies for firm confidence, the purchasing managers index, which has previously not been considered. Second, relates to the estimation of the in-sample models and conduct of the out-of-sample forecasts. Hjalmarsson (2008), Goyal and Welch (2008) and Rapach et al (2010) all use a linear forecast model. In contrast, Henkel et al (2011) allow for regimes, while, more widely there is evidence in favour of non-linear dynamics (e.g., Guidolin et al, 2009). Equally, Goyal and

Welch (2008) and Rapach et al (2010) use fixed in- and out-of-sample periods (albeit they consider alternative fixed periods), whereas Hjalmarsson (2008) allows for recursive estimates to generate the forecasts. Updating the parameter estimates would appear to be a setting more akin to that faced by an investor operating in real-time, i.e., using all available information. Thus, we provide a comprehensive view by allowing our models to take both linear and non-linear functional forms and to vary over both economic and market regimes. Moreover, all forecasts are estimated on both a rolling and recursive basis, while rolling forecast evaluations are also considered. The existing literature argues that evidence in favour of predictability is elusive, this paper seeks to show where such predictability exists.

Our results reveal several key features. Statistical based forecast results tend to support the historical mean baseline model. However, this broad view disguises several nuances to these results. An examination of mean squared error components reveals the failure of predictive models arises from large unsystematic errors. Economic based forecast evaluations reveal better performance for the predictive models. An evaluation of threshold model based forecasts as well as market and economic regimes also indicates the potential to identify periods where explicit forecast models outperform the historical mean. Equally, time-variation in calculating the forecast evaluation measures reveals periods of time where the predictor variables perform relatively better or worse compared to the historical mean. Notwithstanding these results, the term structure of interest rates (especially) and the purchasing managers index consistently exhibit a strong forecast performance. For example, across the individual forecast models, the term structure achieves the highest Sharpe ratio over the full forecast sample and is typically ranked first or second across when considering the forecasts across different regimes. The purchasing managers index is often ranked first when the term structure is not and otherwise typically achieves a top three performance on the Sharpe ratio across the different approaches. A further interesting result is that for the forecast combinations, a subset of interest rate variables typically outperforms the combinations across all variables, in contrast to the existing literature.

The term structure variable is an indicator of investor expectations of future economic conditions, notably, whether expected future output will grow, leading to higher future inflation and interest rates. The purchasing managers index is an indicator of firm expectations of future economic performance and whether firms are seeking to expand supply. The success of these measures highlights the view that movements in stock returns are determined by expectations of expected future economic performance.[2] This is supportive of the general asset pricing principle advocated, for example, by Campbell and Shiller (1988) and Lamont (1998) where movements in stocks depend upon expectations of future economic conditions. Further, the nature of these results is similar to that of Ang and Bekaert (2007), Welch and Goyal (2008) and Hjalmarsson (2010) in providing evidence that the term structure provides a superior forecast performance compared to, for example, the dividend-price and price-earnings ratios often preferred in the literature. A new result here, is the ability of the purchasing managers index to also provide forecast power.

This paper contributes to our knowledge by emphasising the forecast ability of the term structure for stock returns and, to a lesser extent, the purchasing managers index and then other interest rate and firm investment measures. Equally, that the other predictor variables do not exhibit such forecast power. The results also emphasise the different conclusions that can be reached according to whether statistical or economic forecast evaluation measures are used. Further, the results support greater (and lesser) evidence of predictability across different market and economic regimes and different time periods. These latter points indicate a key result that forecast power is not a constant but varies over time.

[2] Other interest rate measures and the investment to capital ratio also perform well and support the view that investor and firm expectations act as the best predictors for stock returns.

2. Empirical Methodology.

The basic forecast equation is given by:

(1) $$r_t = \alpha + \beta x_{t-1} + \varepsilon_t$$

Where r_t is the stock return, x_t the predictor variable and ε_t a white noise error term. In order to conduct the forecasts, we consider two schemes, a rolling and a recursive approach. The purpose of these approaches is designed to mimic investors in real time and thus updating all the available information, including the data and parameter estimates. A further advantage of these approaches over a fixed out-of-sample period is to allow for the presence of breaks to be captured by the forecast model through the updating of data and coefficient values. The two approaches differ in only how they treat older observations, either retaining them through the recursive scheme or dropping them in the rolling approach. In both cases, we begin by estimating the initial model over an in-sample ten-year window and then obtain the forecast for the first out-of-sample observation.[3] To obtain the second forecast, the end of the in-sample period is then rolled forward by one observation. Under the recursive scheme the starting point of the in-sample remains fixed such that the in-sample period expands, while under the rolling scheme, the starting point of the in-sample also moves forward by one observations such that the number of in-sample observations remains fixed. These respective processes continue through the rest of the sample period and we generate two forecast time series.

We obtain individual forecasts for each of the series noted in the next section. In addition, we consider joint forecasts from the predictor variables. First, we estimate multivariate models and second, we consider forecast combinations. For the multivariate models, we estimate a regression that includes all 25 predictor variables. Additionally, we

[3] The choice of an initial 10-year in-sample is inevitably ad hoc but is informed by the need to include a sufficient number of observations to obtain reasonable estimates, while retaining as much out-of-sample period as possible. Of interest Jordan et al (2014) use 60 observations as their initial sample, while Hammerschmid and Lohre (2018) use an initial 10-year period. We view 10-years (40 observations) as the minimum needed to obtain a reasonable initial estimate.

estimate multivariate regressions that group the predictor variables according to type. Thus, we estimate a multivariate regression with stock market, interest rate and macroeconomic predictor variables respectively. For the forecast combinations, we follow the complete subset regressions (CSR) approach of Elliot et al (2013). This approach equally weights the individual forecasts across different subsets of the models and has the advantage of diversifying across individual forecasts and thus reduces issues relating to model uncertainty and stability. As we consider 25 predictor variables, an analysis of the complete subset is not feasible (it would involve over several million subsets), therefore, we focus on a limited set. Following the terminology of Elliot et al (2013), we consider *k=1*, both were we include and exclude the historical mean.[4] We also consider *k=2* and *k=1* for the stock market, interest rate and macroeconomic predictor variable groups.[5,6]

To evaluate the forecasts, we consider measures based on the size and the sign of the forecast error and thus provide measures that have statistical and economic content. We first consider the mean squared error (MSE) and decompose this measure to consider different elements of the forecast performance. The MSE is given by:

(2) $$MSE = (\sum_{t=1}^{\tau}(r_t - r_t^f)^2)/\tau$$

where τ is the forecast sample size, r_t is the actual return and r_t^f represents the forecast series. The MSE can also be decomposed into elements that represent the forecast bias, the difference in the variance of the forecast and actual series and a component that represent unsystematic forecast errors. This decomposition is given by:

(3) $$MSE = \frac{\sum_{t=1}^{\tau}(r_t - r_t^f)^2}{\tau} = (\overline{r_t} - \overline{r_t^f})^2 + (\sigma_{r,t} - \sigma_{r,t^f})^2 + 2(1-\rho)\sigma_{r,t}\sigma_{r,t^f}$$

[4] The historical mean acts as a baseline model for our predictive variables.
[5] We consider alternative combination schemes based on in-sample significance, for example, Pesaran and Timmermann (1995) consider different measures of in-sample fit (such as the Akaike and Schwarz criterion). However, these results did not improve upon those reported, although we use some of the information obtained from this exercise to illustrate the changing nature of significance.
[6] We experimented with a higher value for *k* for both all predictor variables and the groups, but the results did not indicate any improvement on those reported.

where the first component represents the bias, $\frac{(\bar{r}_t-\overline{r_t^f})^2}{(\sum_{t=1}^{\tau}(r_t-r_t^f)^2)/\tau}$, and measures the difference between the mean of the forecast and actual series. The second component, $\frac{(\sigma_{r,t}-\sigma_{r,tf})^2}{(\sum_{t=1}^{\tau}(r_t-r_t^f)^2)/\tau}$, captures the difference in the variance of the forecast and actual series, where σ represents the standard deviation. The third component, $\frac{2(1-\rho)\sigma_{r,t}\sigma_{r,tf}}{(\sum_{t=1}^{\tau}(r_t-r_t^f)^2)/\tau}$, captures the covariance proportion of the MSE and measures the unsystematic forecast error, where ρ represents the correlation between the forecast and actual series. This decomposition allows us to identify the source of any forecast difference between the alternative models.

While the MSE produces a single value for each separate forecast model, we also use the out-of-sample R-squared approach of Campbell and Thompson (2008) and Welch and Goyal (2008), which provides a single value to compare a baseline forecast with an alternative. Moreover, the use of this measure has become increasingly popular in the stock return forecasting literature (for two recent examples, see, Baltas and Karyampas, 2018; Hammerschmid and Lohre, 2018). Additionally, we use the test of Clark and West (2007) to provide a measure of statistical significance for these values.

The out-of-sample R-squared measure is given by:

$$R_{oos}^2 = 1-\left(\frac{\sum_{t=1}^{\tau}(r_t - r_t^{f_2})^2}{\sum_{t=1}^{\tau}(r_t - r_t^{f_1})^2}\right) \tag{4}$$

again τ is the forecast sample size, r_t is the actual return and $r_t^{f_i}$ represents the forecasts. The out-of-sample R-squared test measures a baseline model, denoted f_1, against the predictor model, denoted f_2. When the R_{oos}^2 value is positive, this indicates that the predictor model has greater forecasting power than the baseline model, otherwise the baseline model is preferred. To provide some statistical robustness to this measure, we use the test of Clark and West (2007). This test considers whether the mean squared error of two competing forecasts are statistically different. The Clark and West test adds a simple adjustment to the difference in the MSE values to account for additional parameter estimation error in the larger model. Clark and West suggest generating the following time-series:

(5) $$CW = FE_1 - FE_2 + FE_3$$

Where FE_1 represents the forecast error for the forecast series generated from the baseline model, FE_2 is the forecast series generated from the predictive model and FE_3 is the difference between the baseline and predictive model forecasts. The generated CW series is then regressed on a constant, with associated the *t*-statistic providing the measure of significance. As we are primarily interested in whether the competing model outperforms the baseline model, the CW test is typically implemented as a one tailed test with the null hypothesis that CW is equal to or less than zero against the alternative that it is significantly positive.[7]

The above tests measure the size of the forecast error. However, it is equally important to measure the sign of the forecast error as this provides market trading signals. Arguably, it is preferable to accurately predict a rise or fall in subsequent stock returns rather than to have a forecast value that is close to the realised value. Therefore, while the statistical forecast measures above and the economic measures below complement each other, within a trading scenario, the latter are preferable. More generally, the literature on asset return predictability highlights the importance of considering economic based forecast evaluations. Such an evaluation, which is more closely aligned with investors, has often found greater support for predictor variables over the historical mean. For stock markets this includes, for example, Campbell and Thompson (2008) and Maio (2016), while other assets are also examined, including interest rates (e.g., Della Corte et al., 2008; Sirichand and Hall, 2016) and foreign exchange (e.g., Garratt and Lee, 2009; West et al., 1993). Leitch and Tanner (1991) and Pesaran and Skouras (2004) both argue that given forecasts are intended to inform investor decision-

[7] We thank the Associate Editor for making this point.

making they should be evaluated within this context, thus, highlighting a preference for economic based measures. Furthermore, both Leitch and Tanner (1991) and Cenesizoglu and Timmermann (2012) show that only a weak relation exists between the statistical and economic forecast measures. This supports the above cited literature in which predictive models often find forecast success on the basis of economic measures even when they are outperformed by the historical mean on the basis of statistical ones.

Given this, we include several economic forecast measures. We calculate the success ratio (SR), which measures the proportion of correctly forecast signs:

(6) $$SR = \sum_{t=1}^{\tau} s_t \text{ where } s_t = I(r_t r_t^{f_i} > 0) = 1;\ 0 \text{ otherwise.}$$

Therefore, a SR value of one would indicate perfect sign predictability and a value of zero would indicate no sign predictability. Hence, in assessing the performance of each forecast model, we consider which model produces the highest SR value.

To complement this measure, we also provide a trading-based forecast. While the SR measure provides some trading information with respect to buy and sell signals, we expand this by considering a simple trading rule. Here, if the forecast for the subsequent period return is positive then an investor buys the stock, while if the forecast for the next period return is negative, then the investor (short) sells the stock. From this process, we obtain a time series of returns that represent the outcome of the trading rule and denote this π. To provide market relevant information, we then use this series to generate the Sharpe ratio as such:

(7) $$SHARPE_i = \frac{\bar{\pi} - r_f}{\sigma}$$

Where the Sharpe ratio is calculated as the ratio of the mean trading profit ($\bar{\pi}$) minus a short-term (3-month) Treasury bill as the risk-free rate and the trading return standard deviation (σ). A model that produces a higher Sharpe ratio therefore has superior risk-adjusted returns.

Further to the Sharpe ratio and following Welch and Goyal (2008), Campbell and

Thompson (2008) and Maio (2016), we compute the certainty equivalence value (CEV). This measures the change in average utility between the two forecast approaches and represents the fee an investor would be willing to pay to invest in the active trading strategy, as given by the predictive model, as opposed to a passive strategy based on following the market, as given by the historical mean model. Returns to the active and passive trading strategies are generated as above, and following Maio (2016), the change in CEV is calculated as:

(8) $$CEV = E(R_t^{f_2}) - E(R_t^{f_1}) + \frac{\gamma}{2}[Var(R_t^{f_1}) - Var(R_t^{f_2})]$$

With $R_t^{f_2}$ the trading return obtained from the active predictive forecast model, $R_t^{f_1}$ the trading return from the passive historical mean model and γ is the coefficient of relative risk aversion, set to three following Campbell and Thompson (2008) and Maio (2016).

3. Data and Main Empirical Results.

Our variable of interest to be forecast is the S&P 500 composite index return (difference log of the price index series). Our analysis primarily focuses on the return of the price index, we also consider the total return index, but results are highly similar.[8] The data is sampled quarterly over the period 1960:1 to 2017:2. The data is obtained from Datastream, the St Louis Federal Reserve (FRED) database and the website of Amit Goyal.[9] We use quarterly data as we wish to include some variables that are only available over such a data frequency, notably but not only, GDP.[10] Moreover, while much research in this area uses monthly data, quarterly is not unique, see, Rapach et al (2010) and Elliot et al (2013).[11]

The predictor variables are selected from a list of commonly used variables (see, for

[8] The difference between these two series is that the latter includes dividends in the index.
[9] See, http://www.hec.unil.ch/agoyal/
[10] While other work uses industrial production as a measure of overall activity, in an economy such as the US, which is dominated by the service sector, this is not representative.
[11] Goyal and Welch (2008) consider both annual and monthly data.

example, Welch and Goyal, 2008; Hammerschmid and Lohre, 2018). We group our data as being stock market orientated, interest rate orientated or macroeconomic orientated variables. The stock market based predictor variables are the log dividend-price ratio, the log price-earnings ratio, the cyclically adjusted price-earnings ratio, the payout ratio, the Fed model, the size premium, the value premium, the momentum premium, the book/market ratio, stock return variance, equity allocation, equity issuance and a short stock return lagged moving average. The interest rate variables are the 10-year minus 3-month government treasuries term structure, the default yield (the difference between BAA and AAA rated corporate bonds) and the default return (the difference between long-term corporate and government bonds). The macroeconomic variables are the quarterly change in GDP, consumption, investment, the CPI and central government consumption and investment, Tobin's q-ratio, the purchasing managers index (PMI), the investment to capital (IK) ratio and the consumption, wealth and income ratio (of Lettau and Ludvigson, 2001). While no set of predictor variables can be exhaustive, the above selection is motivated by an attempt to cover a wide range of variable types, include financial price ratio variables, firm characteristic variables, interest rates variables and macroeconomic variables and covers measures of confidence and market behaviour, with the primary restriction being data availability.

Table 1 presents summary statistics for the data used in the forecast models, while Figure 1 presents the time plots. As data similar to this has been used in previous studies, we will only make a few salient observations. A key issue in the predictive equations concerns the time series properties and in particular, stationarity of the variables. The final column of Table 1 presents the DF-GLS test (Elliot et al, 1996) and reveals that five predictor variables (the log dividend-price ratio, CAPE, book/market ratio, the Q-ratio and the CAY ratio) exhibit non-stationarity and so any forecast results must be treated with caution.[12] Stationarity of the other 20 variables, however, supports their use in the forecast regressions. In terms of the graphical depiction of the variables, we can see notable events in terms of the dotcom bubble and financial crisis.

Full In-sample Estimates

Table 2 present the estimates of equation (1) over the full sample using both the price only and the total return indices to form the stock return series. Each predictor variable is estimated individually, and we report the coefficient value, with significance based on the Newey-West *t*-statistic and R-squared value. We also report the multivariate regressions (for the price index return series only). The multivariate regressions include, first, all variables and second, the variables according to their group (stock market, interest rate and macroeconomic variables). As these equations cover the full sample, they are not used in the forecast exercise but are intended to provide information with regard to any variables that exhibit such full (in-)sample predictive power.[13]

The results here show that only a limited number of variables exhibit statistical significance (including up to the 10% level). For the bivariate (single regressor) regressions, the variables that are significant across both the price and total return series are Fed model, the value premium, equity allocation, the default return, PMI, IK ratio and the CAY ratio while the dividend-price ratio and the q-ratio are additionally significant for the total return series only. For the full multivariate regression, the Fed model, size and value premiums, the default return,

[12] Non-stationarity implies a spurious regression problem and bias in the estimated coefficients. We propose no adjustment (e.g., taking first-differences) as we are primarily concerned with the out-of-sample forecasts rather than in-sample estimates. Furthermore, each of these variables are expected to be stationary asymptotically and are included in levels in previous work (see, for example, Campbell and Thompson, 2008; Welch and Goyal, 2008).

[13] There is an interesting debate regarding the power of full in-sample estimates against out-of-sample forecasts (see, for example, Diebold, 2015). While this is not the focus of this paper, it is of interest to note whether the significant full in-sample predictor variables are also those that achieve strong forecast performance.

GDP and government spending and investment growth and PMI are significant. Across the different groups of variables, for the stock market series, CAPE, size and value premiums and equity allocation are significant. For the interest rate variables, only the default return is significant, while for the macroeconomic series, GDP and government spending and investment growth and PMI are significant. Thus, broadly (but not exactly) the same variables are significant across the different exercises.

The fact that there is limited full sample significance is not surprising. Indeed, there is much evidence that stock return predictability is characterised by regimes of predictability, perhaps due to breaks or non-linearities. For example, Paye and Timmermann (2006) suggest that breaks may exist in the predictive relation, while Lettau and van Nieuwerburgh (2008) suggest breaks in the predictor variable. McMillan (2014, 2015) seeks to explicitly model time-variation within the predictive series (dividend-price ratio), while Timmermann (2008), Chen (2009) and McMillan and Wohar (2013) argue that returns predictability may only occur over short periods of time. This, therefore, further motivates the use of the rolling and recursive forecast schemes that can accommodate such patterns of behaviour. Given the broad similarity in the outcomes for the price only and total return series, the results below focus only on the former but results for the latter are available upon request (and again, highly similar in nature).

Forecast Results

Table 3 presents the rolling regression based forecast results for the MSE measure and its component parts. The historical mean (HM) forecast acts as the baseline measure. As with the forecast models, the historical mean forecast is obtained using a rolling and recursive scheme and thus allowing the value of the constant term to change. Forecasts are obtained for each individual predictor variable listed in the first column. Under the 'Groups' heading multivariate regressions are conducting for all variables and those listed in each group (stock market, interest rate or macroeconomic variables), while the results for the forecast combinations of individual predictor variables are noted in the final section of the table.

The results for the overall MSE measure show that the values (multiplied by 100 in the table) obtained by the historical mean forecast and the 25 individual predictor variables are very similar in value, with PMI the only single variable that achieves a lower value than the HM, while the stock return variance variable does notably worse. The multivariate forecasts perform particularly poorly in comparison to all forecasts. The All and Stock Market groups perform particularly poorly, while the Interest Rate group achieves a result more comparable with the individual forecasts. All the forecast combinations (except the stock market group) achieve a lower MSE value than the HM and the combination based on the interest rate variables achieves the best performance across all the forecasts.

Examining the components of the MSE, we see that in terms of the bias, i.e., on average how close are the model's forecasts compared to the actual series, eight individual predictors outperform the HM and seventeen are worse although, again, the values are similar. All the multivariate regressions perform worse than the HM, while all the forecast combinations perform better. The DP and PMI series achieve the lowest values, although as a group the forecast combinations perform well. The results based on the variance and covariance proportions of the MSE provide an interesting dichotomy. All the predictor models achieve a smaller difference between the variance of the forecast and actual return series compared to the HM series. In contrast, the HM series outperforms all the predictor models on the basis of the covariance component. Within these results, the forecast combinations generally perform worse on the variance measure and better on the covariance measure and this may reflect its diversification benefit. This latter forecast measure component captures unsystematic forecast errors and suggests that the failure of the predictor models to consistently outperform the HM does not lie in a systematic failure of the predictor but in (large) errors that arise from unexpected movements in returns.

Table 4 presents the same set of results for the forecasts obtained by the recursive modelling approach. The results here are broadly similar to those obtained under the rolling modelling scheme. The overall MSE values are very similar between the HM and predictor series. Of the individual forecasts, we again see the PMI series outperform the HM, while the same is true for the IK and CAY ratios. Further, all the forecast combinations outperform the HM, while again, the multivariate forecasts do notably poorly. In terms of the MSE components, sixteen of the individual predictors have a lower mean difference compared to the HM between the forecasts and actual series, as do all the multivariate and combined forecasts. As with the rolling forecasts, all the predictor models achieve a lower variance forecast error component and a higher covariance forecast error component compared to the HM. An exception to this is with the forecast combinations where $k=1$, here these forecasts reveal a higher variance and lower covariance forecast error compared to the HM.

The results of the MSE forecasts in Tables 3 and 4 suggest that, looking at the overall MSE values, there is little difference between the HM and the predictor models, although with very few exceptions, the HM performs better. However, this general result masks the view that several predictor variables achieve a better forecast based on a lower average forecast error and a lower variance forecast error value using either or both of the rolling and recursive techniques. Notably, this includes the dividend-price ratio (rolling), the cyclically-adjusted price-earnings ratio (both), the Fed model (rolling), the size premium (rolling), equity allocation (both), term structure (recursive), default yield (recursive), default return (both), GDP growth (recursive), consumption growth (recursive), investment growth (recursive), government consumption and investment (rolling). q-ratio (recursive), PMI (both), IK ratio (recursive) and CAY ratio (recursive). In addition, all the recursive multivariate and both rolling and recursive (with the noted exceptions) combined forecast models outperform the HM approach based on the mean and variance components. However, where the predictor models perform poorly in comparison to the HM is with respect to the unsystematic covariance component and thus the large unexpected movements in returns, resulting in an overall performance that is roughly equivalent between the benchmark and forecast model.

While the results in the above two discussed tables examine the MSE value for each forecast model whether individually, jointly or in combination, Table 5 presents the forecast results using the rolling and recursive schemes based on the out-of-sample R-squared value (OOS R^2) and the success ratio. The OOS R^2 essentially is a comparison of the MSE values between the forecasts based on the predictor models and the HM, as the baseline model. Given the MSE values in Tables 3 and 4, it is unsurprising that very few predictor models achieve a positive OOS R^2 value. For those that do exhibit a positive value, we conduct the Clark and West (2007) test, for which an asterisk(s) denotes statistical significance. For the rolling forecasts, only the PMI across the individual predictor variable forecasts achieves a positive OOS R^2 value, i.e., that its MSE is lower than the value for the HM, while all forecast combination models (except the stock market group) also achieve a positive value. Both the PMI and interest rate forecast combination group exhibits 10% statistical significance based on the Clark and West test. For the recursive approach, in addition to the PMI, the IK and CAY ratios now also achieve a positive OOS R^2 value, together with the same combination forecast groups. Here, only the CAY ratio is statistically significant using the Clark and West test, and at the 5% significance level.

While the above measures are based on the size of the forecast error, in the context of financial returns data, sign forecasting is, at least, of equal importance as it implies market timing signals. The success ratio, which measures the proportion of correctly forecast return signs, is reported in Table 5 and is more suggestive of reasonable forecast performance by the predictor variables. For the rolling forecasts, a success ratio higher than the HM is achieved by fourteen individual forecasts, the interest rate group multivariate forecast and all forecast combinations. Moreover, the highest success ratio is achieved by interest rate variables, as well as the PMI. For the recursive forecasts, there is less support, with the success ratio values very similar across most forecast models. Nonetheless, two of the interest rate series (the term structure and the default yield) and the PMI do achieve a higher success ratio, while several forecast series and most forecast combinations achieve a value equal to HM.

Table 6 provides further forecast measures based on trading indicators, notably the Sharpe ratio and the certainty equivalence value (CEV).[14] As with the success ratio results, we observe some difference between the rolling and recursive approaches, with greater evidence of superior forecast performance relative to the HM for the rolling approach. For the rolling forecasts, the majority of the predictor models achieve a higher Sharpe ratio and CEV, with only the CAPE, MOM, equity issuance and the q-ratio achieving lower values, as do the multivariate stock market and macroeconomic models. Thus, 21 of the individual predictor variables achieve improved trading based forecast performance compared to the HM, as do the interest rate and all-variable multivariate groups and all forecast combinations. Within this, the term structure achieves the highest values for the individual predictor variables, while the interest rate forecast combination achieves the highest values overall. For the recursive forecasts there is less success in terms of a higher Sharpe ratio and CEV. Here, the HM is only beaten by equity allocation, the term structure, default yield, PMI and IK ratio for the individual predictors and the interest rate forecast combination (the default yield also achieves a marginally higher value). The term structure achieves the highest set of values for the recursive forecasts, although this value is lower than the equivalent rolling forecast values for the term structure and the interest rate forecast combinations.

[14] As noted, we use a value of three for the coefficient of relative risk aversion, following Campbell and Thompson (2008) and Maio (2016). Hammerschmid and Lohre (2018) use a value of five, which we therefore also consider. The results remain consistent to those reported in Table 6, both in terms of coefficient sign and ordering of preferred model.

In comparing the rolling and recursive values across the MSE and OOS R^2 values, the recursive approach appears to be largely preferred and achieves improved values. The same is also broadly true with the success and Sharpe ratio values, where they are typically higher with the recursive forecasts although in comparison to the benchmark, the rolling forecasts are preferred. This latter point can be seen clearly with respect to the CEV values, where they are positive for 21 individual predictor variables, two multivariate forecast models and all the forecast combinations for the rolling forecasts. In contrast, for the recursive forecasts, the CEV is positive for only five individual predictor forecasts, no multivariate forecasts and only one forecast combination (for the interest rate group). Moreover, the highest Sharpe ratio is achieved by the rolling term structure from individual predictor models and the rolling interest rate forecast combinations. Thus, the best rolling forecast outperforms the best recursive forecast, although across the range of forecasts the recursive approach is more consistent.

In conducting the rolling and recursive regressions we can examine the statistical significance of the individual predictive variables to consider how such significance changes over the sample period. As noted above, the literature identifies the view that predictability may vary over time. Paye and Timmermann (2006) argue that changes can occur in parameter values, while Timmermann (2008) argues that predictability exists only in small sub-periods. To illustrate this, Figure 2 presents a set of graphs that shows the number of significant variables in each sample period for both the rolling and recursive approaches and for *t*-values equal to 1.96 (5% significance level) and 2.576 (1% significance level). Specifically, the line graph in Figure 2 represents the number of significant variables across the 25 individual predictive regressions across each period and significance level.[15]

Across the four scenarios, we can see, as expected, that there is more indicative

[15] As such, these results are based on marginal significance and the potential exists that global significance is overstated (see, for example, Inoue and Rossi, 2005), however, the illustrative nature of the results remains.

evidence of significance using the t=1.96 level and using the recursive approach. At the t=1.96 level, the average number of significant variables in any sample period is four and a half for the rolling approach and marginally over five for the recursive approach. For the rolling method, the maximum number of predictive variables in any given time period is twelve (1992Q3), while only a single significant variable is noted in 2008Q3 and 2009Q3. For the recursive method, the maximum number of significant predictive variables is ten, while at least two variables are significant at each sample period. Using the t=2.576 cut-off, for the rolling approach, the average number of significant predictor variables is just over two, while it is almost three for the recursive approach. For the rolling method, we see several periods where there is no predictability, and this is notably concentrated in 2001 and around the late 2000s and early 2010s. For the recursive approach, at least one variable is significant in each time period, while a lower degree of predictability is noted in the late 1970s, the early 2000s and the early 2010s.

Across the sample period, we see evidence of a greater number of predictor variables (particularly examining the recursive plots) over the periods of the first half of the 1970s, the second half of the 1980s and first half of the 1990s. Less predictability is observed during the late 1970s and early 1980s and towards the end of the sample period. Across individual series, while there are too many graphs to consider, notable variables that exhibit significance across the sample include the dividend-price ratio, the price-earnings ratio, the cyclically adjusted price-earnings ratio, the term structure of interest rates, the q-ratio and PMI. Nonetheless, all of the variables exhibit periods of significance and insignificance, supporting the view that predictability only occurs over sub-sample periods but that periods of significance occur more regularly for some variables than others.

4. Further Results.[16]

Time-Varying Forecast Models

The above analysis measures the performance of the forecasts obtained from the predictive models both individually and as a group, either through multivariate forecasts or forecast combinations. The use of rolling and recursive modelling approaches allows for time-variation to exist in the parameter values and the statistical significance of the regressions. However, the estimated model is nonetheless a linear one. As noted in the Introduction, there is evidence that forecasts may be improved through considering differing regimes of behaviour. For example, Hammerschmid and Lohre (2018) consider forecasts according to macroeconomic conditions using a Markov-switching approach, while threshold regressions are considered by McMillan (2001, 2003). Henkel et al (2011) argue that predictability only arises during recessionary periods, while Baltas and Karyampas (2018) examine forecast power across up and down market periods.

We consider the importance of regimes of behaviour in predictability and forecasting in two different ways. First, we examine forecast ability of the predictor variables according to whether the market is in a bull or bear phase and whether the economy is in a contractionary or expansionary state. Second, we estimate an explicit threshold regression (TR) model for each predictor variable. In the TR models we need to choose a threshold variable that determines the switch between regimes. We consider five alternative threshold variables, the predictor variable itself and four alternative variables designed to capture economic or market regimes of behaviour. To capture the general economic state, we include the term structure of interest rates, which has been shown to capture future economic conditions (e.g., Estrella and

[16] We also consider the Campbell and Thompson (2008) restrictions on non-negative forecasts. Arguably, this could also be seen as a restriction on short-selling. The results are largely consistent with those reported above in terms of those variables that achieve a preferred forecast performance. Notably, the HM is preferred using statistical forecast metrics, while the economic measures support the predictor models. Again, the term structure is identified as the best individual predictor. Given the qualitative similarity to the results above, these are not reported but are available upon request.

Hardouvelis, 1991; Harvey, 1997; Estrella and Mishkin, 1998; Lange, 2018). We also consider the composite leading indicator (CLI) obtained from the St Louis Federal Reserve (FRED) and the expansionary/contractionary and bull/bear states noted above and defined below.

Table 7 presents the OOS R^2 and Sharpe ratio results when we separate the forecast sample between bull and bear markets and expansionary and contractionary periods. We only report these two statistics and for the rolling regressions for space considerations, but these results highlight the key conclusions from this approach.[17] To define bull/bear market periods, we follow Cooper et al (2004) and use a three-year moving average of the stock index. Specifically, if the change in the moving average is positive then the market is characterised as a bull market, while if the change in the three-year moving average is negative, the market is in a bear phase. To define expansionary and contractionary regimes, we use output (GDP) growth over two consecutive quarters. Where this value is positive then we ascribe that to be an expansionary regime and a contractionary regime when it is negative.

In the bear market regime, we can see that the HM again outperforms the predictive models on the basis of the OOS R^2, although the values are close to zero, suggesting little difference in performance. Notwithstanding this, there are some exceptions where the OOS R^2 value is positive, namely, for the term structure, the CAY ratio and the interest rate group forecast combination (the value is also positive, but very marginally so, for the *k=1* all and macroeconomic group forecast combinations). A similar picture is seen in the bull market regime, with few instances of positive OOS R^2 values, but again all the values are small in magnitude. Notably, a positive value is reported for PMI and all the forecast combinations (except the interest rate group) and, more marginally, for investment growth. With regard to the Sharpe ratios, we see a greater distinction between the two regimes. In the bear market

[17] As noted, while we only report a subset of the results in order to limit the number of tables, the full suite of forecast measures is available upon request.

regime, the HM achieves the lowest (i.e., worst) Sharpe ratio and while most of the values are negative (given it is a bear market regime), for nine individual series, three multivariate regressions and three forecast combination, we observe a positive Sharpe ratio. Across the individual predictors, the term structure achieves the highest Sharpe ratio, while the interest rate forecast combination achieves the highest value overall. In the bull market regime, however, only four individual predictor models (term structure, investment growth, PMI and IK ratio) outperform the HM, while the forecast combinations (except the interest rate and macroeconomic groups) also outperform the HM. Again, the term structure forecast model performs well, although the PMI value is slightly higher. These results suggest a clear distinction in the ability of the predictive models (and the HM) across market regimes.

Examining the results for expansionary and contractionary regimes, we see a similar dichotomy as with the market regimes. Across both regimes, on the basis of the OOS R^2, the HM is typically preferred but all the values are small (with the exception some of the multivariate models). Of interest, with the forecast combinations, the $k=1$ all-variable and stock market and $k=2$ all-variable groups achieve a positive value in the expansionary regime, while the interest rate and macroeconomic groups achieve a positive value in the contractionary regime. In the expansionary regime, the HM and predictive models achieve a similar degree of success using the Sharpe ratio, with twelve predictor variables outperforming the HM. In addition, two of the multivariate models also achieve a higher Sharpe ratio than the HM, while all the forecast combinations do. The term structure predictive variable achieves the highest Sharpe ratio across the range of models. As with the bear market regime, in the contractionary regime, all the predictive models achieve a higher Sharpe ratio compared to the HM (except momentum). Moreover, while for many of the forecast models, the Sharpe ratio is negative (given the state of the economy), for fifteen individual predictors, all the multivariate models and all (except the stock market group) forecast combinations, the Sharpe ratio is positive. The term structure variable again produces a relatively high Sharpe ratio, although the values for PMI and CAY are slightly higher, while the interest group forecast combination achieves the highest value.

Overall, these results suggest that the HM approach typically outperforms the majority of the predictive models in bull markets and economic expansions, while, the predictive models perform well during bear market conditions and economic contractions. Notwithstanding this, across all regimes, the term structure as the individual predictive model and the interest rate group as the forecast combinations, consistently achieve a strong performance.

The OOS R^2 and Sharpe ratio results of the TR regressions are reported in Tables 8 and 9 respectively. In Table 8, which presents the OOS R^2 values, the results present a similar picture to that revealed earlier for the linear models in Table 5. Specifically, the OOS R^2 values are small and nearly all negative, indicating preference for the HM. Across the five different threshold variables, the term structure achieves a positive value once (it appears twice, but both values refer to the same regression), the PMI three times and the IK and CAY ratios once each. In terms of statistical significance, the Clark and West test is significant at the 1% level for the term structure predictor variable and at the 10% for PMI when using the term structure and CLI as the threshold variable. Table 9 presents the Sharpe ratio across the TR models. Of particular note, the PMI series has a Sharpe ratio higher than the HM across all five threshold models, while the term structure has a higher value for four of the models and the IK ratio for two. We can also observe that for a further six predictor series, a higher Sharpe ratio is obtained when using the term structure as the threshold variable.

Time-Varying Forecast Evaluation

The literature highlights the view that forecast success may only occur in pockets of time. The evidence reported in Figure 2 illustrates that in-sample predictive power of the variables varies over time, while the above analysis indicates that forecast success can vary with regimes of behaviour. Therefore, we would expect the forecast success of the predictive variables to change over time. Using the first set of linear based results reported in Tables 5 and 6, we calculate the OOS R^2 and Sharpe ratio on a rolling basis to consider how these values and thus the relative forecast success varies over time.[18] We only present the plots for four predictive models, the CSR *k=1* for the all-variables and interest rate group and the PMI and term structure predictor variables. We choose these forecast models as the above results indicate preference for them.[19]

Taking both figures, we see evidence where the forecasts models are preferred on both the OOS R^2 and Sharpe ratio measures. Although it is noticeable that the periods of success across these two measures do not coincide exactly. Looking at Figure 3 for the OOS R^2 plots, we can see a positive value indicating preference over the HM occurring during the early mid-1980s, the first half of the 1990s, the early to mid-2000s and, to a lesser extent, the mid-2010s. We also observe that this pattern is more clearly seen in the forecast combination and term structure graphs, while for the PMI model the periods of success are more transient and largely occur over the first half of the sample. For the Sharpe ratio plots in Figure 4, the periods of greater forecast success occur slightly after those indicated for the OOS R^2. Notably, the higher Sharpe ratios are seen during the later mid-1980s, the mid-late 1990s, the late mid-2000s and towards the end of the sample. Of interest, the forecast models perform poorly during the early 2000s, which coincides with the dotcom crash, and, to a lesser extent, from the financial crisis period, but not for the interest rate group forecast combination.

Overall, the results from the full set of empirical tests above suggest that, if choosing one predictor variable, the term structure of interest rates provides the best set of out-of-sample

[18] Similar to, for example, the cumulative sum of squared forecast errors and rolling Sharpe ratio graphs in Baltas and Karyampas (2018).

[19] Nonetheless, results, both rolling and recursive, for all forecasts are available upon request.

forecasts. The purchasing managers index provides the second best set of forecasts. The term structure reveals investor expectations of the future course of the economy. A steepening term structure indicates that investors expect higher future interest rates that will arise from higher future inflation and thus an expanding economy. The result that the term structure achieves the best forecast performance is similar to that reported by Welch and Goyal (2008) for their monthly results and Hjalmarsson (2010). While, a set of research seeks to emphasise the ability of ‘fundamental to price’ ratio series (beginning with Campbell and Shiller, 1988) as proxies for expected returns, the results here suggest that a more explicit predictor of future economic conditions provides a better forecast performance.

In considering the results with respect to the previous literature, we can highlight two key distinctions. As noted above, the work of Welch and Goyal (2008) and Hjalmarsson (2010) finds some favour for the term structure. We support this, but also find support for the PMI, a variable not considered in these papers or indeed in the wider stock return predictability literature. The PMI provides an indicator of firm confidence, who will expand orders should they expect an upturn in economic conditions, and a subsequent rise in the stock market. Thus, we argue that the PMI measure should be included in any forecast set. In the work of Rapach et al (2010) and Elliot et al (2013), they argue that forecast combinations outperform single variable forecasts and that a combination across a wide range of variables is preferred. In the terminology of Elliot et al (2013), Rapach et al (2010) only consider *k=1*, while Elliot et al (2013) indicate a preference for *k=2* or *3* according to their Tables 3 and 5. The results here are at variance with this work in two respects. First, any benefit of the combined forecast over a single variable forecast is marginal. For example, the term structure variable outperforms the forecast combinations on the overall Sharpe ratio, while the combinations outperform the term structure on the overall OOS R^2. Second, and equally pertinent, a narrower set of forecast combination variables, the interest rate group, outperforms the larger, 25 variable set. Thus, adding more variables to the forecast combinations does not necessarily improve performance.

4. Summary and Conclusions.

Using quarterly US data from 1960 to the end of 2017, we conduct ten-year rolling and recursive forecasts for a range of 25 financial and economic predictor variables. The forecasts are generated from individual regressions, multivariate regressions and forecast combinations. We use both statistical and economic evaluations of the forecasts that are based on linear and threshold models and are considered over economic and market cycles and calculated over the out-of-sample period as an average and on a rolling basis.

The results present several interesting conclusions that both compliment and contrast with the existing literature; however, the overriding takeaway point is with regard to single predictor variables, the term structure of interest rates (10-year Treasury bond minus 3-month Treasury bill) and (to a lesser extent) the purchasing managers index provide consistent forecast performance that is superior to the HM across different forecasting approaches and regimes of behaviour. Using linear and non-linear models, rolling and recursive approaches and allowing for regimes according to economic and market conditions, these two variables are consistently the best forecast models when using the economic based (Sharpe ratio and CEV) forecast measures, which are most relevant for investors. A second takeaway point is that in contrast to the existing literature, the forecast combinations of a small set of interest rate variables outperforms combinations based on the full set of variables. Moreover, forecast combination are not unequivocally preferred to single variable forecasts.

The historical mean model outperforms that vast majority of the predictor variables and models. This most noticeably occurs using the mean squared error based measures of forecasting ability, however, the numerical difference in values is typically small. One point of interest is that a decomposition of the mean squared error reveals that the forecast models typically outperform the historical mean model in terms of forecast bias and the volatility of forecasts but are subject to large unsystematic forecast errors resulting in an overall poorer performance. In contrast, the economic based measures, the Sharpe ratio and certainty equivalence values, show forecast improvement over the historical mean.

We consider whether these results vary when separating the forecast evaluations between periods of bull and bear market behaviour and economic expansion and contraction (we also consider but do not report results when imposing short selling restrictions). The results indicate the forecast models are more accurate during bear markets and economic contractions. This is supportive of the view that fundamentals are more important in periods of market stress. Further, we extend the analysis of regimes of behaviour by considering an explicit non-linear threshold model using a range of threshold variables. These results confirm the success of the forecast models based upon economic evaluations, while the HM is still largely preferred on the basis of statistical measures.

The use of rolling and recursive approaches for both the forecast models and the forecast evaluation, allows us to identify both periods of in-sample predictive significance and out-of-sample forecast performance relative to the historical mean model, which are otherwise masked by examining statistics over the whole period. Time periods during each of the 1980s, 1990s, 2000s and 2010s reveal evidence of in-sample predictability and out-of-sample forecast power and equally periods where such evidence is lacking. The pertinent point from this exercise is that it highlights the temporary nature of predictability. Thus, either examining only in-sample predictive results or computing single out-of-sample forecast statistics can fail to reveal the periods of forecast power that exist. This supports the view emerging view in the literature that predictive and forecast power for stock returns is a temporary phenomenon and modelling should account for this time-variation.

Cutting across the alternative modelling and forecasting approaches, and alternative forecast evaluations, the term structure of interest rates and the purchasing managers index achieve consistently strong forecast performance, especially (but not only) when assessed according to the Sharpe ratio measure. These two variables are based on either investor or firm expectations of future economic performance i.e., do investors expect higher future inflation and interest rates or firms expect an increase in orders, as the economy expands. The concluding point of this paper is that quarterly US stock returns can be forecast and that while forecast performance is variable, these two series provide as consistent a performance as is likely to occur. Beyond, individual predictor variables, a forecast combination of interest rate variables also provides a strong set of results and is superior to a larger set of predictor variable forecast combinations. It remains to be seen whether a similar result will be repeated across alternative markets.

References

Ang, A. and Bekaert, G. (2007). Stock return predictability: Is it there? Review of Financial Studies, 20, 651-707.

Baltas, N. and Karyampas, D. (2018). Forecasting the equity risk premium: The importance of regime-dependent evaluation. Journal of Financial Markets, 38, 83-102.

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

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

Cenesizoglu, T. and Timmermann, A. (2012). Do return prediction models add economic value? Journal of Banking and Finance, 36, 2974-2987.

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

Clark, T.E. and West, K.D. (2007). Approximately normal tests for equal predictive accuracy in nested models. Journal of Econometrics, 138, 291-311.

Cochrane, J.H. (2008). Financial markets and the real economy. In Mehra, R. (ed.) Handbook of the Equity Risk Premium Elsevier, 237-325. Elsevier: Amsterdam.

Cooper, M., Gutierrez Jr, R. and Hameed, A. (2004). Market states and momentum. Journal of Finance, 59, 1345-1365.

Della Corte, P., Sarno, L. and Thornton, D.L. (2008). The expectation hypothesis of the term structure of very short-term rates: Statistical tests and economic value. Journal of Financial Economics, 89, 158-174.

Diebold, F.X. (2015). Comparing predictive accuracy, twenty years later: A personal perspective on the use and abuse of Diebold–Mariano tests. Journal of Business and Economic Statistics, 33, 1-9.

Elliott, G., Gargano, A. and Timmermann, A. (2013). Complete subset regressions. Journal of Econometrics, 177, 357.373.

Elliott, G., Rothenberg, T.J. and Stock, J. (1996). Efficient tests for an autoregressive unit root. Econometrica, 64, 813-836.

Estrella, A. and Hardouvelis, G. A. (1991). The term structure as a predictor of real economic activity. Journal of Finance, 46, 555-576.

Estrella, A. and Mishkin, F. S. (1998). Predicting U.S. recessions: Financial variables as leading indicators. Review of Economics and Statistics, 80, 45-61.

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

Garratt, A. and Lee, K. (2009). Investing under model uncertainty: Decision based evaluation of exchange rate forecasts in the US, UK and Japan. Journal of International Money and Finance, 29, 403-422.

Guidolin, M., Hyde, S. McMillan, D.G. and Ono, S. (2009). Non-linear predictability in stock and bond returns: When and where is it exploitable? International Journal of Forecasting, 25, 373-399.

Hammerschmid, R. and Lohre, H. (2018). Regime shifts and stock return predictability. International Review of Economics and Finance, 56, 138-160.

Harvey, C.R. (1997). The relationship between the term structure of interest rates and Canadian economic growth. Canadian Journal of Economics, 30, 169-193.

Henkel, S.J., Martin, J.S. and Nardari, F. (2011). Time-varying short-horizon predictability. Journal of Financial Economics, 99, 560-580.

Hjalmarsson, E. (2010). Predicting global stock returns. Journal of Financial and Quantitative Analysis, 45, 49-80.

Inoue, A. and Rossi, B. (2005). Recursive predictability tests for real-time data. Journal of Business and Economic Statistics, 23, 336-345.

Jordan, S.J., Vivian, A.J., Wohar, M.E. (2014). Forecasting returns: New European evidence. Journal of Empirical Finance, 26, 76-95.

Lange, R.H. (2018). The predictive content of the term premium for GDP growth in Canada: Evidence from linear, Markov-switching and probit estimations. North American Journal of Economics and Finance, 44, 80-91.

Leitch, G. and Tanner, J.E. (1991). Economic forecast evaluation: Profits versus the conventional error measures. American Economic Review, 81, 580-590.

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

Lettau, M. and Van Nieuwerburgh, S. (2008). Reconciling the return predictability evidence. Review of Financial Studies, 21, 1607-1652.

Maio, P. (2016). Cross-sectional return dispersion and the equity premium. Journal of Financial Markets, 29, 87-109.

McMillan, D.G. (2001). Non-linear predictability of stock market returns: Evidence from non-parametric and threshold models. International Review of Economics and Finance, 10, 353-368.

McMillan, D.G. (2003). Non-linear predictability of UK stock market returns. Oxford Bulletin of Economics and Statistics, 65, 557-573.

McMillan, D.G. (2014). Modelling time-variation in the stock return-dividend yield predictive equation. Financial Markets, Institutions and Instruments, 23, 273-302.

McMillan, D.G. (2015). Time-varying predictability for stock returns, dividend growth and consumption growth. International Journal of Finance and Economics, 20, 362–373.

McMillan, D.G. and Wohar, M.E. (2013). A panel analysis of the stock return dividend yield relation: predicting returns and dividend growth. Manchester School, 81, 386-400.

Park, C. (2010). When does the dividend-price ratio predict stock returns? Journal of Empirical Finance, 17, 81-101.

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

Pesaran, M.H. and Skouras, S. (2004). Decision-based methods for forecast evaluation. In Clements, M.P. and Hendry, D.F. (eds) A Companion to Economic Forecasting, Ch11, 241-267. Blackwell: Oxford.

Pesaran, M.H. and Timmermann, A. (1995). Predictability of stock returns: Robustness and economic significance. Journal of Finance, 50, 1201-1228.

Pesaran, M.H. and Timmermann, A. (2000). A recursive modelling approach to predicting UK stock returns. Economic Journal, 110, 159-191.

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

Sirichand, K. and Hall, S. (2016). Decision-based forecast evaluation of UK interest rate predictability. Journal of Forecasting, 35, 93-112.

Timmermann, A. (2008). Elusive return predictability. International Journal of Forecasting, 24, 1-18.

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

West, K., Edison, H. and Cho, D. (1993). A utility-based comparison of some models of exchange rate volatility. Journal of International Economics, 35, 23-45.

Table 1. Summary Statistics

Variables	Mean	Median	SD	Skew	Kurt	UR
Stock Market Variables						
Returns	0.016	0.026	0.080	-0.944	4.900	-12.544
DP	1.016	1.083	0.396	-0.206	2.339	-1.524*
PE	2.844	2.880	0.437	0.743	6.213	-3.780
CAPE	19.924	20.291	7.797	0.641	3.651	-1.181*
DE	-0.747	-0.762	0.320	2.852	18.677	-3.646
FED	1.173	0.995	0.619	2.141	8.908	-2.518
SMB	0.397	0.250	2.765	-0.486	11.147	-5.103
HML	0.244	0.340	2.410	-0.183	4.994	-5.764
MOM	1.544	1.430	3.550	0.194	5.928	-9.100
BM†	0.503	0.458	0.258	0.800	2.844	-1.321*
SVAR†	0.623	0.365	1.041	7.109	65.766	-8.887
EQ Alloc	0.345	0.358	0.070	-0.081	1.879	-1.839
Net Eq Is†	0.011	0.014	0.020	-0.743	3.408	-2.282
MA1Yr	0.016	0.023	0.040	-0.860	4.170	-4.481
Interest Rate Variables						
TS	1.570	1.630	1.180	-0.175	2.569	-4.253
Def Yield†	1.017	0.900	0.449	1.793	7.804	-3.635
Def Ret†	0.065	0.148	2.423	0.317	14.240	-7.166
Macroeconomic Variables						
GDP	0.746	0.753	0.830	-0.265	4.561	-2.531
Cons	1.613	1.551	0.942	-0.357	5.950	-3.326
Inv	1.567	1.868	4.014	-0.581	5.172	-2.991
Infl	0.921	0.846	0.922	-0.006	6.826	-2.742
Gov C&I	0.461	0.373	0.995	0.480	4.166	-3.062
Q-Ratio	0.748	0.772	0.269	0.250	2.800	-1.573*
PMI	52.898	53.550	6.578	-0.543	4.285	-5.351
IK†	0.036	0.036	0.004	0.234	2.438	-2.498
CAY†	-0.016	0.196	2.505	-0.211	2.104	-0.708*

Notes: Entries are summary statistics (mean, median, standard deviation skewness and kurtosis values) for our variables. The final column is the DF-GLS unit root test, where an asterisk denoted non-stationarity (using the 10% significance level). The list of variables is: stock returns (difference log) multiplied by 100. The explanatory variables are grouped by type as stock market variables: DP (log dividend-price ratio), PE (log price-earnings ratio), CAPE (cyclically adjusted PE ratio), DE (log dividend-earnings ratio), FED (earnings yield dividend by the 10-year Treasury bond), SMB (the return premium to small firms over large firms), HML (the return premium to value firms over growth firms), MOM (the return premium to past winner firms over loser firms), BM (the book-to-market ratio), Svar (stock market volatility), Eq Alloc (equity market allocation, market value of stocks divided by the sum of market value of stocks and investor holdings of cash and bonds), Net Eq Is (net equity issuance), MA1Yr (a one-year lagged moving average of stock returns); interest rate variables: TS (the difference between the yield on a 10-year Treasury bond and 3-month Treasury bill), Def Yield (default yield as the difference between the yield on BAA and AAA rated corporate bonds), Def Ret (default return as the difference between the return on long-term corporate and government bonds); macroeconomic variables: GDP (the period growth rate of real GDP), Cons (the period growth rate of personal consumption), Inv (the period growth rate of investment), Infl (the period change in CPI), Gov C&I (central government consumption and investment), Q-ratio (Tobin's Q), PMI (the purchasing managers index), IK (the investment to capital ratio), CAY(the consumption-wealth ratio). The variables marked with a † are taken from Welch and Goyal (2008), other variables are obtained from the Federal Reserve FRED website and from Datastream.

Table 2. Full Sample Predictability Estimates

	Bivariate Regressions						Multivariate Regressions	
Predict Vars	Price Index			Total Return			ALL	By Groups
	α	β	R^2	α	β	R^2	β	β
Stock Market Variables								
DP	-0.40	1.98	0.010	-0.34	2.66*	0.017	1.05	4.74
PE	5.11	-1.23	0.004	7.29	-1.73	0.009	7.06	4.01
CAPE	3.28**	-0.08	0.007	4.67***	-0.12	0.013	0.55	0.50**
DE	2.15	0.72	0.001	2.98**	0.83	0.001	-0.66	-0.20
FED	0.26	1.15*	0.008	1.06	1.12*	0.007	3.34**	0.78
SMB	1.71***	-0.27	0.009	2.47***	-0.27	0.009	-0.57**	-0.35*
HML	1.69***	-0.33*	0.010	2.44***	-0.32*	0.009	-0.67**	-0.52**
MOM	1.39***	0.15	0.004	2.61***	-0.16	0.005	-0.21	-0.22
BM	0.73	1.75	0.003	0.97	2.76	0.008	3.38	4.16
Svar	1.60***	1.46	0.000	2.37***	-0.45	0.000	-0.16	-0.52
Eq Alloc	8.54***	-20.1***	0.031	10.28***	-22.9***	0.040	-31.40	-52.5***
Net Eq Is	1.77**	-14.32	0.001	2.50***	-12.24	0.001	-12.19	-1.06
MA1Yr	1.66**	-0.03	0.000	2.42***	-0.02	0.000	-9.38	-1.98
Interest Rate Variables								
TS	0.86	0.48	0.005	1.65*	0.45	0.005	0.34	0.21
Def Yield	0.42	1.20	0.005	0.94	1.44	0.007	0.60	0.99
Def Return	1.64***	0.43*	0.017	2.39***	0.43*	0.017	0.60***	0.41*
Macroeconomic Variables								
GDP	1.45*	0.12	0.000	2.29***	0.10	0.000	0.33**	0.36**
Cons	1.57	0.02	0.000	2.14	0.14	0.000	-0.44	0.15
Inv	1.79***	-0.11	0.003	2.53***	-0.11	0.003	0.37	-0.38
Infl	1.83**	-0.24	0.001	2.44***	-0.09	0.000	-0.30	0.13
Gov C&I	1.71***	-0.22	0.001	2.45***	-0.19	0.001	-0.13**	-0.12*
Q-Ratio	3.53**	-2.59	0.008	4.98***	-3.51*	0.014	-4.93	-1.36
PMI	13.59***	-0.23**	0.035	14.61***	-0.23**	0.036	-0.3***	-0.4***
IK	10.03**	-2.34*	0.011	10.65**	-2.30*	0.011	1.08	1.90
CAY	1.69***	0.45**	0.020	2.34***	0.45**	0.020	0.42	0.24

Notes: Entries are the coefficient estimates, Newey-West adjusted *t*-statistics and R-squared values from equation (1). The bivariate regressions include only a single predictor variable. The column 'ALL' is a multivariate regression including all variables. The column 'By Group' is a set of multivariate regressions for each of the groupings (Stock Market Variables; Interest Rate Variables' Macroeconomic Variables). For interest, the Adjusted R-square values are: 0.12 (ALL); 0.04 (Stock Market); 0.01 (Interest Rates); 0.06 (Macroeconomic). The increasing number of asterisks refer to statistical significance at the 10%, 5% and 1% levels. The variables are listed in Table 1.

Table 3. MSE and Components – Rolling Forecasts

Predictor Vars	MSE*100	Bias*100	Variance	Covariance
HM	0.696	0.034	0.722	0.283
Stock Market Variables				
DP	0.761	0.001	0.500	0.506
PE	0.785	0.349	0.386	0.616
CAPE	0.751	0.013	0.544	0.461
DE	0.770	0.210	0.286	0.717
FED	0.747	0.030	0.357	0.648
SMB	0.720	0.026	0.633	0.372
HML	0.718	0.100	0.576	0.428
MOM	0.717	0.079	0.621	0.383
BM	0.758	0.042	0.485	0.520
Svar	0.927	0.114	0.068	0.936
Eq Alloc	0.734	0.031	0.452	0.553
Net Eq Iss	0.723	0.116	0.409	0.595
MA1Yr	0.735	0.284	0.528	0.474
Interest Rate Variables				
TS	0.707	0.057	0.529	0.475
Def Yield	0.746	0.228	0.260	0.743
Def Return	0.741	0.020	0.422	0.584
Macroeconomic Variables				
GDP	0.712	0.183	0.589	0.414
Cons	0.725	0.196	0.532	0.472
Inv	0.718	0.171	0.559	0.445
Infl	0.736	0.280	0.527	0.475
Gov C&I	0.706	0.017	0.593	0.412
Q-Ratio	0.751	0.046	0.516	0.489
PMI	0.691	0.001	0.498	0.508
IK	0.709	0.574	0.515	0.484
CAY	0.735	0.061	0.441	0.564
Multivariate Regression Groups				
Stock Mkt	1.812	0.118	0.078	0.926
IR	0.796	0.089	0.197	0.807
Macro	0.957	1.567	0.060	0.929
All	3.546	0.035	0.249	0.756
Combined Forecasts (CSR)				
CSR – k=1	0.691	0.016	0.699	0.307
CSR – k=1 (ex HM)	0.691	0.015	0.695	0.310
CSR – k=2	0.694	0.010	0.606	0.399
CSR – SM k=1	0.700	0.017	0.658	0.348
CSR – IR k=1	0.683	0.018	0.570	0.435
CSR – Macro k=1	0.691	0.011	0.707	0.298

Notes: Entries are the MSE (mean squared error) and its components as identified in equations (2)-(3). The explanatory variables are given in Table 1. The multivariate regressions contain the variables for each group listed under the sub-heading. CSR are the forecast combination of Elliot et al (2013).

Table 4. MSE and Components – Recursive Forecasts

Predictor Vars	MSE*100	Bias*100	Variance	Covariance
HM	0.685	0.221	0.902	0.101
Stock Market Variables				
DP	0.716	0.340	0.588	0.414
PE	0.732	0.541	0.641	0.358
CAPE	0.715	0.064	0.743	0.262
DE	0.714	1.011	0.595	0.400
FED	0.709	0.368	0.617	0.385
SMB	0.692	0.289	0.796	0.206
HML	0.693	0.286	0.778	0.225
MOM	0.694	0.207	0.849	0.154
BM	0.729	0.114	0.676	0.328
Svar	0.812	0.097	0.233	0.772
Eq Alloc	0.697	0.018	0.662	0.343
Net Eq Iss	0.706	0.015	0.610	0.395
MA1Yr	0.706	0.360	0.794	0.208
Interest Rate Variables				
TS	0.693	0.095	0.619	0.386
Def Yield	0.702	0.043	0.674	0.331
Def Return	0.691	0.147	0.659	0.345
Macroeconomic Variables				
GDP	0.696	0.218	0.855	0.148
Cons	0.695	0.215	0.812	0.191
Inv	0.692	0.170	0.797	0.206
Infl	0.704	0.493	0.677	0.324
Gov C&I	0.689	0.241	0.869	0.134
Q-Ratio	0.707	0.095	0.720	0.285
PMI	0.677	0.021	0.569	0.436
IK	0.683	0.027	0.765	0.240
CAY	0.678	0.022	0.675	0.330
Multivariate Regression Groups				
Stock Market	0.965	0.183	0.071	0.933
IR	0.722	0.184	0.501	0.502
Macro	0.713	0.048	0.450	0.555
All	1.228	0.214	0.001	1.002
Combined Forecasts (CSR)				
CSR – k=1	0.683	0.108	0.910	0.094
CSR – k=1 (ex HM)	0.683	0.104	0.908	0.096
CSR – k=2	0.684	0.041	0.841	0.164
CSR – SM k=1	0.691	0.156	0.870	0.133
CSR – IR k=1	0.680	0.002	0.782	0.223
CSR – Macro k=1	0.680	0.115	0.880	0.124
Notes: As for Table 3.				

Table 5. OOS R-Squared and Success Ratio

Predictor Vars	Rolling Forecasts		Recursive Forecasts	
	OOS R2	Success Ratio	OOS R2	Success Ratio
HM	-	0.60	-	0.65
Stock Market Variables				
DP	-0.095	0.59	-0.044	0.56
PE	-0.128	0.65	-0.067	0.65
CAPE	-0.080	0.56	-0.043	0.63
DE	-0.107	0.62	-0.042	0.62
FED	-0.074	0.60	-0.034	0.61
SMB	-0.035	0.62	-0.010	0.63
HML	-0.032	0.62	-0.011	0.64
MOM	-0.032	0.57	-0.012	0.65
BM	-0.090	0.61	-0.064	0.64
SVar	-0.333	0.61	-0.185	0.64
Eq Alloc	-0.055	0.57	-0.017	0.65
Net Eq Iss	-0.040	0.55	-0.030	0.63
MA1Yr	-0.056	0.58	-0.030	0.63
Interest Rate Variables				
TS	-0.016	0.66	-0.011	0.67
Def Yield	-0.073	0.65	-0.024	0.66
Def Return	-0.066	0.61	-0.008	0.62
Macroeconomic Variables				
GDP	-0.023	0.58	-0.015	0.63
Cons	-0.043	0.58	-0.014	0.63
Inv	-0.032	0.63	-0.009	0.65
Infl	-0.058	0.59	-0.027	0.63
Gov C&I	-0.015	0.62	-0.005	0.64
Q-Ratio	-0.080	0.58	-0.031	0.61
PMI	0.007*	0.66	0.013	0.67
IK	-0.020	0.65	0.003	0.65
CAY	-0.056	0.61	0.010**	0.58
Multivariate Regression Groups				
Stock Market	-1.606	0.54	-0.408	0.49
IR	-0.144	0.63	-0.054	0.62
Macro	-0.376	0.53	-0.040	0.57
All	-4.098	0.53	-0.792	0.55
Combined Forecasts (CSR)				
CSR – k=1	0.007	0.64	0.017	0.65
CSR – k=1 (ex HM)	0.007	0.64	0.017	0.65
CSR – k=2	0.002	0.66	0.002	0.65
CSR – SM k=1	-0.007	0.64	-0.008	0.65
CSR – IR k=1	0.018*	0.66	0.007	0.65
CSR – Macro k=1	0.006	0.64	0.008	0.64
Notes: Entries are the out-of-sample (OOS) R-squared values from equation (4) and the success ratio of equation (6). The asterisk(s) for the OOS R-squared test indicates significance based on the Clark and West (2007) test of equation (5). A single asterisk represents 10% and a double represents 5% significance.				

Table 6. Sharpe Ratio and CEV

Predictor Vars	Rolling Forecasts		Recursive Forecasts	
	Sharpe Ratio	CEV	Sharpe Ratio	CEV
HM	0.074	-	0.209	-
Stock Market Variables				
DP	0.081	0.155	0.111	-3.954
PE	0.182	3.695	0.166	-1.919
CAPE	0.053	-0.476	0.155	-2.379
DE	0.187	3.924	0.178	-1.420
FED	0.089	0.360	0.063	-5.217
SMB	0.109	0.956	0.142	-2.890
HML	0.186	3.898	0.179	-1.358
MOM	0.017	-1.039	0.197	-0.565
BM	0.116	1.188	0.166	-1.938
SVar	0.094	0.507	0.162	-2.089
Eq Alloc	0.096	0.572	0.226	0.871
Net Eq Iss	0.040	-0.703	0.180	-1.348
MA1Yr	0.102	0.737	0.140	-2.959
Interest Rate Variables				
TS	0.271	8.317	0.264	2.908
Def Yield	0.201	4.596	0.209	0.016
Def Return	0.176	3.468	0.133	-3.224
Macroeconomic Variables				
GDP	0.100	0.669	0.178	-1.429
Cons	0.079	0.116	0.157	-2.312
Inv	0.132	1.719	0.176	-1.499
Infl	0.087	0.326	0.180	-1.345
Gov C&I	0.134	1.774	0.179	-1.387
Q-Ratio	0.037	-0.754	0.132	-3.261
PMI	0.215	5.268	0.262	2.837
IK	0.244	6.762	0.227	0.904
CAY	0.148	2.296	0.157	-2.298
Multivariate Regression Groups				
Stock Market	0.059	-0.348	-0.055	-6.319
IR	0.194	4.249	0.165	-1.957
Macro	0.029	-0.880	0.091	-4.548
All	0.119	1.278	0.125	-3.485
Combined Forecasts (CSR)				
CSR – k=1	0.173	3.304	0.199	-0.484
CSR – k=1 (ex HM)	0.173	3.304	0.199	-0.484
CSR – k=2	0.239	6.496	0.199	-0.484
CSR – SM k=1	0.153	2.485	0.199	-0.484
CSR – IR k=1	0.288	9.304	0.217	0.427
CSR – Macro k=1	0.237	6.417	0.189	-0.918

Notes: Entries are the Sharpe Ratio of equation (7) and CEV of equation (8).

Table 7. Forecast Results According to Market and Economic Conditions – Rolling Forecasts

Pred Vars	Bull vs Bear				Expansionary vs Contractionary			
	OOS-R^2		Sharpe Ratio		OOS-R^2		Sharpe Ratio	
	Bear	Bull	Bear	Bull	Expan	Contract	Expan	Contract
HM	-	-	-0.451	0.326	-	-	0.155	-0.250
Stock Market Variables								
DP	-0.047	-0.139	-0.059	0.153	-0.067	-0.154	0.109	-0.030
PE	-0.170	-0.089	0.009	0.278	-0.035	-0.332	0.239	-0.024
CAPE	-0.039	-0.119	-0.132	0.144	-0.069	-0.103	0.109	-0.158
DE	-0.188	-0.030	0.022	0.280	-0.030	-0.275	0.222	0.065
FED	-0.038	-0.108	-0.140	0.205	-0.070	-0.083	0.159	-0.191
SMB	-0.039	-0.030	-0.257	0.297	-0.017	-0.073	0.194	-0.224
HML	-0.013	-0.050	0.043	0.267	-0.023	-0.052	0.208	0.116
MOM	-0.028	-0.035	-0.397	0.207	-0.026	-0.044	0.094	-0.315
BM	-0.075	-0.105	-0.004	0.181	-0.069	-0.129	0.139	0.0151
SVar	-0.113	-0.539	-0.267	0.276	-0.437	-0.104	0.114	0.020
Eq Alloc	-0.079	-0.032	-0.183	0.239	-0.009	-0.153	0.187	-0.244
Net Eq I	-0.026	-0.053	-0.279	0.194	-0.070	0.028	0.041	0.034
MA1Yr	-0.097	-0.018	-0.277	0.294	-0.044	-0.084	0.137	-0.033
Interest Rate Variables								
TS	0.024	-0.055	0.128	0.359	-0.044	0.045	0.281	0.265
Def Yd	-0.090	-0.057	0.011	0.309	-0.092	-0.032	0.171	0.352
Def Ret	-0.023	-0.106	0.003	0.273	-0.076	-0.044	0.187	0.150
Macroeconomic Variables								
GDP	-0.027	-0.019	-0.201	0.253	-0.014	-0.043	0.106	0.077
Cons	-0.009	-0.075	-0.102	0.171	-0.057	-0.009	0.125	-0.103
Inv	-0.072	0.005	-0.225	0.320	-0.016	-0.068	0.152	0.064
Infl	-0.055	-0.064	-0.162	0.214	-0.065	-0.041	0.094	0.063
Gov C&I	-0.012	-0.017	-0.272	0.347	-0.041	0.042	0.163	0.029
Q-Ratio	-0.031	-0.126	0.024	0.045	-0.067	-0.109	0.105	-0.246
PMI	-0.018	0.030	-0.042	0.362	-0.022	0.069	0.206	0.274
IK	-0.019	-0.021	0.090	0.335	-0.058	0.063	0.267	0.178
CAY	0.016	-0.124	0.049	0.203	-0.118	0.077	0.093	0.403
Multivariate Regression Groups								
Stock Mkt	-1.022	-2.155	0.128	0.029	-1.812	-1.122	0.163	0.033
IR	-0.082	-0.203	0.231	0.185	-0.197	-0.029	0.161	0.354
Macro	-0.223	-0.519	-0.056	0.071	-0.516	-0.056	0.009	0.124
All	-1.863	-6.198	0.257	0.061	-5.272	-1.487	0.100	0.226
Combined Forecast (CSR)								
CSR – k=1	0.002	0.012	-0.126	0.337	0.012	-0.005	0.219	0.009
CSR – k=1 (ex HM)	0.001	0.012	-0.126	0.337	0.013	-0.005	0.219	0.009
CSR – k=2	-0.005	0.009	0.072	0.337	0.013	-0.022	0.253	0.210
CSR – SM k=1	-0.021	0.007	-0.190	0.337	0.018	-0.060	0.236	-0.158

CSR – IR k=1	0.047	-0.009	0.266	0.313	-0.014	0.087	0.260	0.452
CSR – Macro k=1	0.004	0.008	0.096	0.321	-0.005	0.031	0.202	0.424
Notes: Entries are the out-of-sample R-squared values and Sharpe Ratios obtained for the rolling forecasts during bull and bear market phases and expansionary and contractionary economic conditions.								

Table 8. OOS R-Squared– Threshold Model

Predictor Variables	Threshold Variable				
	Pred Var	TS	CLI	Exp./Con.	Bull/Bear
Stock Market Variables					
DP	-0.156	-0.044	-0.079	-0.096	-0.234
PE	-0.061	-0.329	-0.198	-0.067	-0.097
CAPE	-0.729	-0.229	-0.065	-0.084	-0.325
DE	-0.170	-0.053	-0.172	-0.053	-0.153
FED	-0.121	-0.059	-0.064	-0.042	-0.485
SMB	-0.019	-0.016	-0.078	-0.013	-0.066
HML	-0.011	-0.020	-0.091	-0.011	-0.020
MOM	-0.011	-0.014	-0.031	-0.018	-0.036
BM	-0.123	-0.328	-0.254	-0.064	-0.115
SVar	-0.236	-0.497	-0.263	-0.239	-0.203
Eq Alloc	-0.080	-0.327	-0.036	-0.058	-0.206
Net Eq Iss	-0.040	-0.071	-0.029	-0.030	-0.242
MA1Yr	-0.045	-0.026	-0.052	-0.030	-0.054
Interest Rate Variables					
TS	0.012***	0.012***	-0.153	-0.032	-0.119
Def Yield	-0.110	-0.024	-0.065	-0.024	-0.081
Def Return	-0.028	-0.034	-0.048	-0.013	-0.024
Macroeconomic Variables					
GDP	-0.029	-0.045	-0.037	-0.065	-0.015
Cons	-0.061	-0.047	-0.025	-0.036	-0.014
Inv	-0.031	-0.009	-0.029	-0.009	-0.004
Infl	-0.007	-0.050	-0.030	-0.029	-0.025
Gov C&I	-0.009	-0.048	-0.023	-0.033	-0.044
Q-Ratio	-0.045	-0.241	-0.047	-0.090	-0.300
PMI	-0.001	0.014*	0.013*	0.004	-0.014
IK	0.003	-0.009	-0.023	0.003	-0.013
CAY	-0.011	0.003	-0.062	-0.023	-0.035

Notes: Entries are the out-of-sample (OOS) R-squared values from equation (4), with the forecasts now obtained from a threshold model. The alternative threshold variables are: the lag of the predictor variable, the lag of the term structure variable, the lag of the Composite Leading Indicator, the lag of two period GDP growth and the lag of the three-year stock market moving average. Asterisks relate to the Clark and West (2007) test.

Table 9. Sharpe Ratio – Threshold Model

Predictor Variables	Threshold Variable				
	Pred Var	TS	CLI	Exp./Con.	Bull/Bear
HM	0.209	0.209	0.209	0.209	0.209
Stock Market Variables					
DP	0.142	0.111	0.135	0.059	0.132
PE	0.174	0.216	0.070	0.166	0.186
CAPE	0.148	0.217	0.122	0.123	0.157
DE	0.087	0.178	0.128	0.178	0.159
FED	0.082	0.076	0.033	0.063	0.052
SMB	0.148	0.178	0.106	0.142	0.084
HML	0.187	0.216	0.118	0.179	0.188
MOM	0.197	0.200	0.172	0.197	0.205
BM	0.146	0.237	0.109	0.166	0.147
SVar	0.152	0.214	0.129	0.147	0.165
Eq Alloc	0.160	0.196	0.140	0.201	0.206
Net Eq Iss	0.124	0.151	0.180	0.180	0.156
MA1Yr	0.122	0.203	0.091	0.140	0.145
Interest Rate Variables					
TS	0.324	0.324	0.178	0.260	0.243
Def Yield	0.179	0.209	0.174	0.209	0.188
Def Return	0.096	0.167	0.090	0.133	0.117
Macroeconomic Variables					
GDP	0.169	0.213	0.143	0.153	0.178
Cons	0.073	0.151	0.160	0.153	0.208
Inv	0.147	0.176	0.157	0.176	0.195
Infl	0.159	0.191	0.085	0.180	0.200
Gov C&I	0.179	0.152	0.151	0.170	0.179
Q-Ratio	0.124	0.116	0.142	0.107	0.133
PMI	0.235	0.264	0.262	0.267	0.268
IK	0.227	0.198	0.163	0.227	0.235
CAY	0.125	0.169	0.150	0.157	0.085

Notes: Entries are the Sharpe Ratio of equation (7), with the forecasts now obtained from a threshold model. The threshold variables are the same as for Table 8.

Accepted for publication in *European Journal of Finance* published by Taylor and Francis.

Figure 1. Time Series Plots

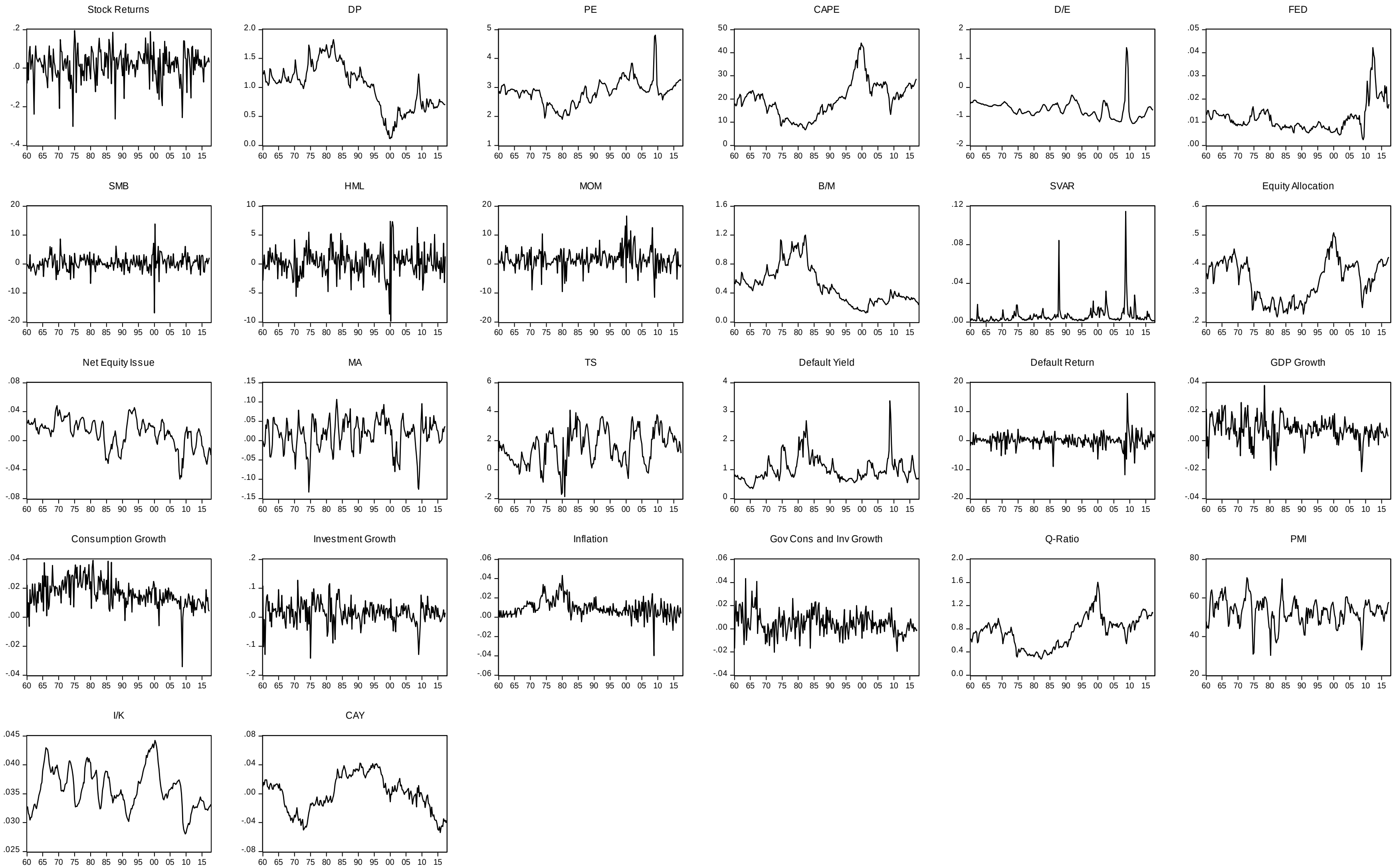

Figure 2. Number of Significant Predictor Variables

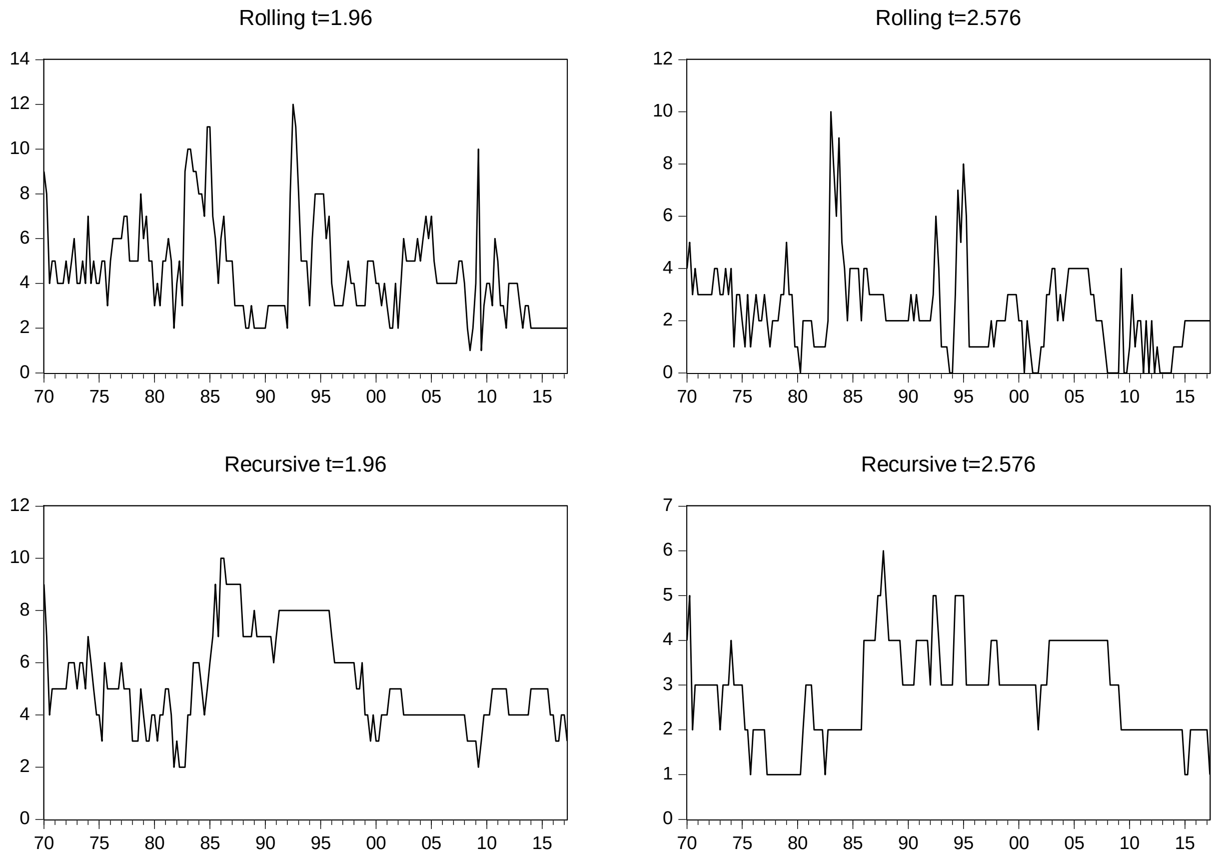

Figure 3. Rolling OOS R^2

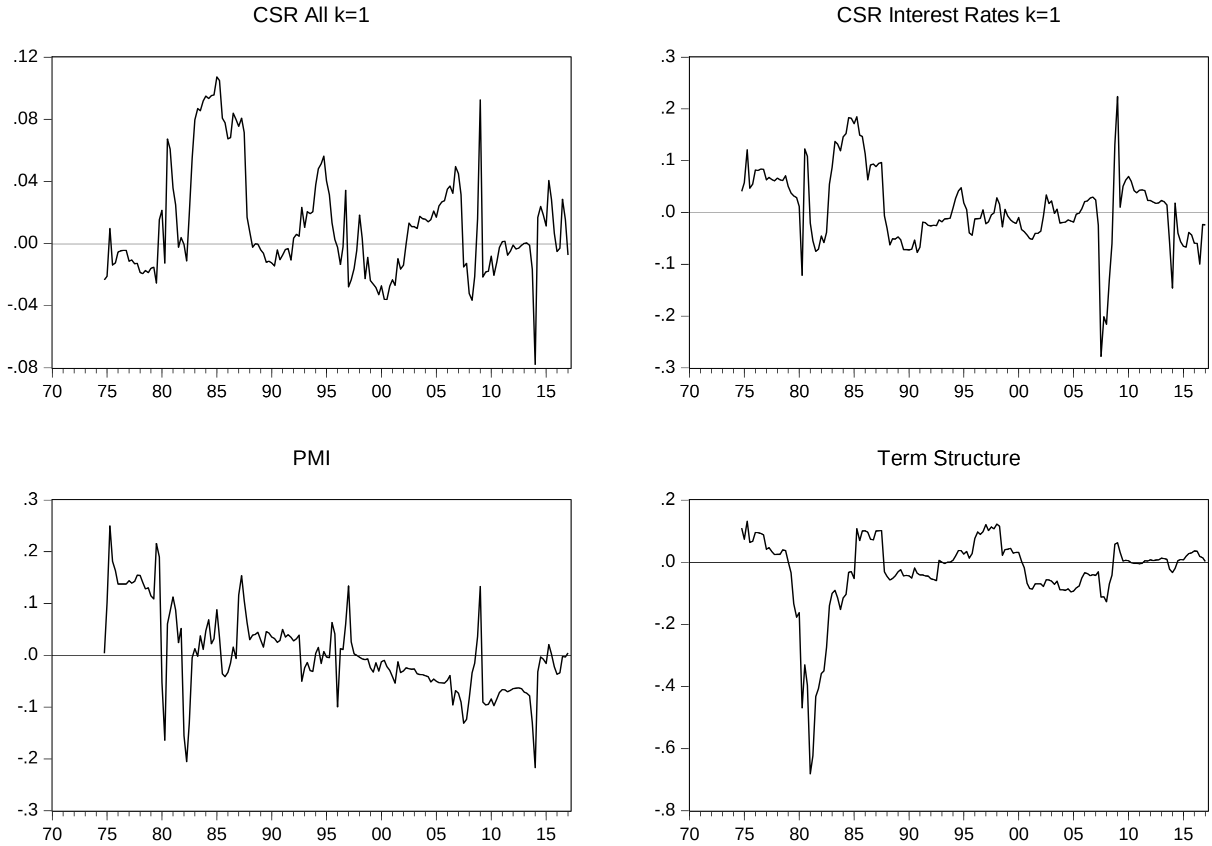

Figure 4. Rolling Sharpe Ratio