

Out-of-Sample Equity Premium Predictability in South Africa: Evidence from a Large Number of Predictors*

Rangan Gupta*, Mampho P. Modise** and Josine Uwilingiye***

Abstract

This paper uses a predictive regression framework to examine the out-of-sample predictability of South Africa's equity premium, using a host of financial and macroeconomic variables. We employ various methods of forecast combination, bootstrap aggregation (bagging), diffusion index (principal component) and Bayesian regressions to allow for a simultaneous role of the variables under consideration, besides individual predictive regressions. We assess both the statistical and economic significance of the individual predictive regressions, combination methods, bagging, principal components and Bayesian regressions. Our results show that forecast combination methods and principal component regressions improve the predictability of the equity premium relative to the benchmark autoregressive model of order one (AR(1)). However, the Bayesian predictive regressions are found to be the standout performers with the models outperforming the individual regressions, forecast combination methods, bagging and principal component regressions, both in terms of statistical (forecasting) and economic (utility) gains.

Keywords: Equity Premium; Predictive Regressions; Forecast Combinations; Bagging; Principal Component Regressions; Bayesian Predictive Regressions.

JEL Codes: C11, C13, C22, C33, C53, G11, G12.

1. Introduction

Forecasting stock market behaviour has received great attention in recent years from both academics and policy-makers. The current uncertainties regarding the economic performance of the major global economies (especially the United States and the Euro zone) and the likelihood that the global economy may experience a double-dip recession has continued to emphasise the importance of predicting the behaviour of leading indicators (including stock returns) accurately. Stock and Watson (2003) and Forni *et al.* (2003), amongst others, show that stock prices act as leading indicators in helping predict the behaviour of output and inflation in the economy. In this regard, recently, Gupta and Hartley (2011) have highlighted similar abilities of stock prices for South Africa.

Literature proposes numerous financial and macroeconomic variables as possible predictors of stock markets behaviour including valuation ratios, such as price-earnings ratio (Campbell and Shiller 1988, 1998) and price-dividend ratio (Fama and French 1988, 1989); payout ratio (Lamont, 1998); interest rates (Ang and Bekaert, 2001; Campbell, 1987); the term spread

* We would like to thank two anonymous referees for many helpful comments. However, any remaining errors is solely ours.

(Campbell, 1987); stock returns of South Africa's major trading partners (Rapach *et al.*, 2010a); the inflation rate (Fama, 1981); money stock (Geske and Roll, 1983); industrial production and the employment rate (Rapach *et al.*, 2005); world oil production and the refiner acquisition cost of imported crude oil (Peersman and Van Robays, 2009); as well as industrial and financial stock returns (Jiang *et al.*, 2009; Rapach *et al.*, 2011; Neely *et al.*, 2011). Although most studies focus on in-sample tests and conclude that there is significant evidence of return predictability, Rapach *et al.*, (2005) and Goyal and Welch (2008) show that these potential predictors are unable to deliver consistently superior out-of-sample forecasts of equity premium relative to a benchmark; autoregressive model of order one or a random walk model respectively. To improve out-of-sample equity premium based on these variables, we propose four approaches – bagging forecasts, combination of model forecasts, principal component and Bayesian regressions – based on monthly data with the in-sample covering the period from 1990:01 to 1996:12, while the out-of-sample covering the period from 1997:01 to 2010:12. While, the starting date and the end point of the data sample is contingent on data availability, the choice of the out-of-sample period is driven by the fact that this period encompasses a host of domestic and global events that are likely to have affected the South African stock market. The out-of-sample period covers events such as the Asian financial crisis, South Africa's decision to move to an inflation targeting regime in 2000, the currency crisis in late 2001, and finally the US sub-prime crisis.

The first approach we use applies bootstrap aggregating (bagging) to a general-to-specific procedure based on a general dynamic linear regression model with the 23 possible predictors. Following Rapach and Strauss (2010), we construct the bagging forecasts using a moving-block bootstrap. The second approach is to combine individual forecasts using a number of different methods proposed in recent financial literature. There is evidence (see Bates and Ganger, 1969 and Rapach *et al.*, 2009) showing that combining individual forecasts tends to outperform the individual forecasts themselves. Forecast combination methods are proven to generate consistent and significant out-of-sample gains and link out-of-sample predictability to the real economy (Rapach *et al.*, 2009 and Kong *et al.*, 2009). We analyse combination forecasts formed as weighted averages of the 23 individual predictive regression model forecasts for a period starting from 1997:01. The methods we consider include: simple averages, discounting (Stock and Watson, 2004), clusters (Aiolfi and Timmermann, 2006), and principal components (Neely *et al.*, 2011). Note, in addition to forecast combination via principal components, we also look at forecasting capabilities of predictive regressions based on principal components extracted from the entire data set. Our last approach is to assess the out-of-sample predictability of equity premium of South Africa using the Bayesian regression methods under the Gaussian and double-exponential priors used by De Mol *et al.* (2008). De Mol *et al.* (2008) show that forecasts produced by Bayesian regression models are highly correlated with principal component forecasts and perform equally well, if not better, for a wide range of prior choices.

To test the out-of-sample forecasts, we employ the out-of-sample R^2 statistic, R_{OS}^2 , developed by Campbell and Thompson (2008), which measures the reduction in the mean squared forecast error (MSFE) for a predictive regression forecast relative to a benchmark forecast (Rapach *et al.*, 2009 and Kong *et al.*, 2009). Because the R_{OS}^2 does not explicitly account for the risk borne by an investor over the out-of-sample period, we follow Marquering and Verbeek (2004), Campbell and Thompson (2008), Goyal and Welch (2008), and Wachter and Warusawitharana (2009) and compute realised utility gains for a mean-variance investor on a real-time basis. Goyal and Welch (2008) show that out-of-sample criteria are crucial in assessing equity premium predictability. Additionally, we provide statistical explanations for the relatively good out-of-sample

performance of forecast combination regarding the equity premium. Using forecast encompassing tests we are able to explain the econometric sources of the benefits of forecast combination. In our analysis, however, we only compare the forecasts from our best model (which happens to be the Lasso: LARS) with forecasts from individual regressions and other combination models.

For individual regressions, our results emphasise the importance of interest rate variables in explaining the behaviour of equity premium, relative to the bench mark AR(1) model. The interest rate variable that exhibit highest utility gains is the term spread, which is around 12 per cent at an annualised rate. Barring the inflation rate, no other variable show significant forecasting gains over the out-of-sample relative to the AR(1) model. As expected, combining information across individual regressions outperforms individual forecasts themselves and our results show the following; firstly, constructing principal components using the original data and combining those principal components improves the out-of-sample predictability for equity premium in South Africa. Secondly, various combining methods also provide significant out-of-sample gains relative to the benchmark random walk model – with the cluster combining methods and principal component combining methods outperforming other combining methods that we consider. Also interestingly, even though the performance of the bagging model is quite poor, when we take the mean of bagging forecasts and principal component forecast combination methods, the performance markedly improves to the extent that, the model outperforms all the various forecast combination methods. Thirdly, predictive regressions based on the second principal component perform better than forecast combination methods. But, the Bayesian regression forecasts outperforms the individual regression forecasts, the bagging model, the alternative combining methods and principal component regressions. The utility gains for the Bayesian regression forecasts are significantly higher than for the other combination model forecasts and the individual regression models – with the LASSO:Landweber having the highest utility gain of 65.35 percent at an annualised rate. The forecast encompassing test results further substantiate the importance of the Bayesian regressions in explaining South African equity premium behaviour. The remainder of the paper is structured as follows: The econometric models are described in Section 2; Section 3 provides the data and discusses the results obtained from the different models; and Section 4 concludes.

2. Econometric methodology

2.1 Predictive regression

We analyse the South African equity premium using a standard predictive regression framework, expressed as:

$$r_{t+1} = \alpha + \beta z_t + \gamma r_t + \mu_{t+1} \quad (1)$$

where r_{t+1} is the equity premium, z_t is the variable whose predictive ability is of interest and μ_{t+1} is the disturbance term. The variable z_t has predictive power when $\beta \neq 0$. We include the lagged equity premium as a control variable when testing the predictive ability of z_t since the estimated value of $\gamma = 0.93$, with a p -value of 0.00. We further divide the total sample of T observation for r_t and z_t into an in-sample portion comprising the first m observations (1990:01

to 1996:12) and an out-of-sample portion made up of the last q observation (1997:01 to 2010:12). The initial out-of-sample forecast of the equity premium based on the predictor \mathbf{z}_t is given by:

$$\hat{r}_{i,m+1} = \hat{\alpha}_{i,m} + \hat{\beta}_{i,m} \hat{\mathbf{z}}_{i,m} + \hat{\gamma}_{i,m} \hat{r}_{i,m} \quad (2)$$

where $\hat{\alpha}_{i,m}$, $\hat{\beta}_{i,m}$ and $\hat{\gamma}_{i,m}$ are the OLS estimates of α , β and γ in equation 1. The period is then updated by using data available through $m + 1$ in order to generate a second set of forecasts, given by:

$$\hat{r}_{i,m+2} = \hat{\alpha}_{i,m+1} + \hat{\beta}_{i,m+1} \hat{\mathbf{z}}_{i,m+1} + \hat{\gamma}_{i,m+1} \hat{r}_{i,m+1} \quad (3)$$

This process is repeated through to the end of the out-of-sample period, generating a series of q out-of-sample forecasts of the equity premium based on $\mathbf{z}_{i,t}$, $\{\hat{r}_{i,t+1}\}_{t=m}^{T-1}$.

There has emerged a consensus amongst financial economists suggesting that equity premium tends to be unpredictable and, as a result, could be approximated by a random walk model (Pesaran, 2003). Consequently, our random walk model is defined similar to Campbell and Thompson (2008) and Goyal and Welch (2008) as the historical average of the equity premium. The historical average that serves as a natural benchmark forecast model corresponding to a constant expected equity premium is defined as follows: $r_{t+1}^{RW} = 1/t \sum_{j=1}^t r_j$. But given the high and significant persistence of the equity premium as discussed above and as suggested by one of the referees, we use the AR(1) model obtained by setting $\beta = 0$ in (1) as our benchmark:

However, following the literature, we report, in the Appendix, results from all the models relative to the historical average as well.

2.2 Bagging forecasts

To define the bagging forecasts we follow a procedure in Inoue and Kilian (2008) and Rapach and Strauss (2010). We specify the bagging model of a one-month-ahead forecast horizon as:

$$\mathbf{r}_{t+1} = \alpha + \gamma \mathbf{r}_t + \sum_{i=1}^n \beta_i \mathbf{z}_{i,t} + \mu_{t+1} \quad (4)$$

We estimate \mathbf{r}_{t+1} via OLS, using equation 4 with data from 1990:01 through to time t and compute the t -statistics corresponding to each predictor. The $\mathbf{z}_{i,t}$ variables with t -statistics less than 1.645 in absolute value are dropped from equation 4 and the model is re-estimated. The forecast of \mathbf{r}_{t+1} is obtained by regressing only the included $\mathbf{z}_{i,t}$ variables value into the re-estimated equation 4 and setting the disturbance term to its expected value of zero.

We construct the bagging forecasts by means of a moving-block bootstrap. Basically we generate a large number B of pseudo-samples of size t by randomly drawing blocks of size s (with replacement) from the observations of the equity premium and possible predictors from 1990:01 through to time t . We estimate equation 4 using the pretesting procedure to determine the predictors to include in the forecasting model. As specified earlier, \mathbf{r}_{t+1} is forecast by adding

only the included $\mathbf{z}_{i,t}$ variables and values into the re-estimated version of the forecasting model. Even with the moving-block bootstrap, the disturbance term is set to its expected value of zero. The bagging model forecast corresponds to the average of the B (which is set to 100) forecast for the bootstrap pseudo-sample. The out-of-sample period comprises of a series of p recursive simulated out-of-sample forecasts using the bagging procedure. The recursive out-of-sample forecast therefore takes the form: $\{\hat{r}_{BA,t+1}\}_{t=R}^{T-1}$, where R corresponds to the in-sample for the entire data set of T observations.

2.3 Combination forecasts

Following Bates and Granger (1969) and Rapach and Strauss (2010), we use information across individual forecasts via forecast combining methods since combining individual forecasts is known to outperform the individual forecasts themselves. We consider a number of combining methods, and some of these models require a hold-out period to calculate the weight used to combine the individual regression forecasts. For the hold-out period, we use the first P_0 out-of-sample observations (the first five years in our case). With the exception of the Bayesian models, the combination forecasts of r_{t+1} made at time t , $\hat{r}_{CB,t+1}$, are a linear combination of the individual regressions constructed above, meaning:

$$\hat{r}_{CB,t+1} = \sum_{i=1}^n w_{i,t} \hat{r}_{i,t+1} \quad (5)$$

where $\sum_{i=1}^n w_{i,t} = 1$.

The weights are constructed using the start of the hold-out period to time t . We use a post hold-out period for each of the combining methods (with the exception of the Bayesian models and the simple combining methods), meaning we have a $P_1 = P - P_0$ combination forecast for evaluation. Below is a full discussion on each of the combining forecast models.

2.3.1 Simple combining methods

Following Stock and Watson (2003 and 2004), we look at three combining methods which tend to work well in forecasting using a large number of potential predictors. We assess the mean, median and the trimmed mean. The mean is defined as: $w_{i,t} = \frac{1}{n}$ ($i = 1, \dots, n$) in equation 5. The median combining method is simply defined as the sample median of $\{\hat{r}_{i,t+1}\}_{i=1}^n$. We define the trimmed mean combination forecast as $w_{i,t} = 0$ for the individual forecast with the smallest and largest forecasts at time t and $w_{i,t} = \frac{1}{(n-2)}$ for the remaining individual forecasts in equation 5.

2.3.2 Discount MSFE combining methods

In a discount MSFE, the weights in equation 5 are a function of recent historical forecasting performance of the individual regression models (Rapach and Strauss, 2010 and Stock and Watson, 2004) and are defined as: $w_{i,t} = \rho_{i,t}^{-1} / \sum_{j=1}^n \rho_{j,t}^{-1}$, where

$$\rho_{i,t} = \sum_{s=R}^{t-1} \psi^{t-1-s} (r_{s+1} - \hat{r}_{i,s+1})^2 \quad (6)$$

In this model the discount factor is given by ψ . When $\psi = 1$, there is no discounting and equation 5 produces the optimal combination forecast for the case where the individual forecasts are uncorrelated. A discount factor that is less than 1 places greater importance to the recent forecasting accuracy of the individual regressions. We follow Rapach and Strauss (2010) in selecting the value of the discount factors as 1.0 and 0.9.

2.3.3 Cluster combining methods

Cluster combining methods incorporate persistence in forecasting performance. The procedure we follow was developed by Aiolfi and Timmermann (2006) and used in Rapach and Strauss (2010). The initial combination forecast: $r_{(R+P_0)+1}$ is computed by grouping the individual regression forecasts over the initial hold-out out-of-sample period: $\{\hat{r}_{i,s+1}\}_{s=R}^{R+(P_0-1)}$ ($i = 1, \dots, n$) into H equal-sized clusters based on MSFE. The first cluster will include those individual regressions with the lowest MSFE values. The second cluster will have the next lowest MSFE values and so on. To construct the first combination forecast, we average the individual regression forecasts of $r_{(R+P_0)+1}$ in the first cluster. The average of the individual regression forecasts of $r_{(R+P_0)+1}$ in the first cluster will be the first combination forecast. We compute the MSFE for the individual regression forecasts $\{\hat{r}_{i,s+1}\}_{s=R+1}^{R+P_0}$ ($i = 1, \dots, n$) to form the second combination forecast. The individual regression forecasts are then grouped into H clusters. The average of the individual regression forecasts of $r_{(R+P_0)+2}$ included in the first cluster becomes the second combination forecast. We do this procedure through the end of the available out-of-sample period. Basically we form clusters by computing MSFE using a rolling window. Since the number of clusters serves to define the size of the first cluster, as none of the clusters are used in generating the forecasts and that the greater the number of clusters, the smaller the size of the first cluster, we select $H = 2$ and $H = 3$ (following Rapach and Strauss, 2010 and Aiolfi and Timmermann, 2006).

2.3.4 Principal components combining methods

Another forecasting combination method that we use involves generating a combination forecast using the first x principle components of the individual regressions out-of-sample forecasts. The first x principal components of the uncentred second moment matrix of the individual regression forecasts are denoted by; $\hat{F}_{1,s+1}, \dots, \hat{F}_{x,s+1}$ for $s = R, \dots, t$. The combination forecast of r_{t+1} at time t based on the fitted principal components is given by the following regression:

$r_{t+1} = \phi_1 \hat{F}_{1,s+1} + \dots + \phi_x \hat{F}_{x,s+1} + \mu_{t+1}$ with $s = R, \dots, t - 1$. The combination forecast is given by $\hat{\phi}_1 \hat{F}_{1,s+1} + \dots + \hat{\phi}_x \hat{F}_{x,s+1}$, where $\hat{\phi}_1, \dots, \hat{\phi}_x$ are the OLS estimates of ϕ_1, \dots, ϕ_x . Bai and Ng (2002) developed the IC_{p_3} information criterion to select the number of principal components. We use this criterion since other familiar information criteria such as the Akaike Information Criterion and the Schwarz Information Criterion do not always estimate the correct number of factors consistently (see Bai and Ng, 2002).

2.4 Diffusion index (principal component) regression and Bayesian regressions¹

The use of dynamic factor analysis enables us to effectively summarise information from the 23 variables in our analysis to a small number of principal components. This helps with the problem of in-sample over-fitting when using a large number of variables (Neely *et al.*, 2011 and Ludvigson and Ng, 2007, 2009). We therefore consider the entire sample period (1990:01 to 2010:12) to construct principal components, and in turn, use these principal components individually and together instead of individual predictors (\hat{z}) in equation 1. This model differs from the principal components combining method discussed earlier, whereby, we combine out-of-sample forecasts obtained from individual predictive regressions.

To select the number of factors to include in our analysis, we use an information criterion described and used in Alessi *et al.* (2010). It is crucial to select the correct number of factors, since we need the factors to be relatively small to avoid the problem of in-sample over-fitting, but not too small, thereby neglecting important information in the 23 individual predictors. The procedure that we follow selects the number of factors by correcting for the tendency of traditional criteria to overestimate the true number of factors. It is quite well-known that when the number of variables, from which principal components are to be extracted, are small relative to the number of data points, the Bai and Ng (2002) criteria does not have a minimum, and hence, cannot lead to an optimal choice of the number of factors. Based on Alessi *et al.* (2010), we extract two factors, which are found to be sufficient and efficient in summarising the information contained in the 23 possible predictors.

In order to identify these two factors, Figures 1 and 2 plots the marginal R^2 of the bivariate regression involving the two factors on each of the 23 predictors. Figures 1 and 2 show that the first constructed factor contains more information from the financial variables, whereas the second principal component contain information from the macroeconomic variables. Hence, the principal component 1 can be dubbed a financial factor, while the principal component 2 can be called the macroeconomic factor.

¹ Please refer to De Mol *et al.*, (2008), Belmonte *et al.*, (2011) and Korobilis (2011) for technical details on these methods.

Figure 1: Principal component 1 – representing financial variables

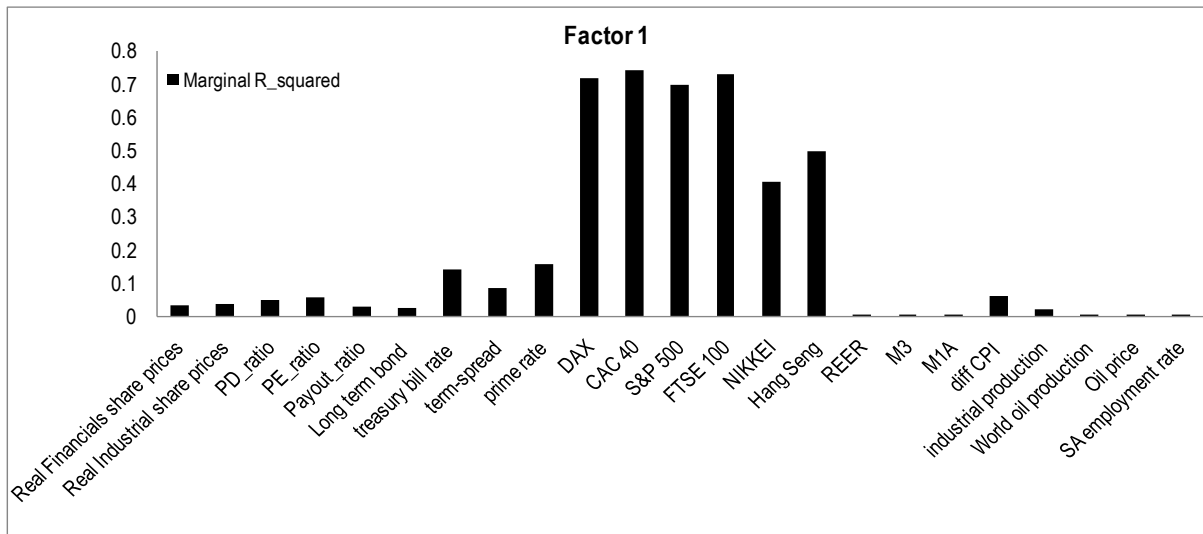
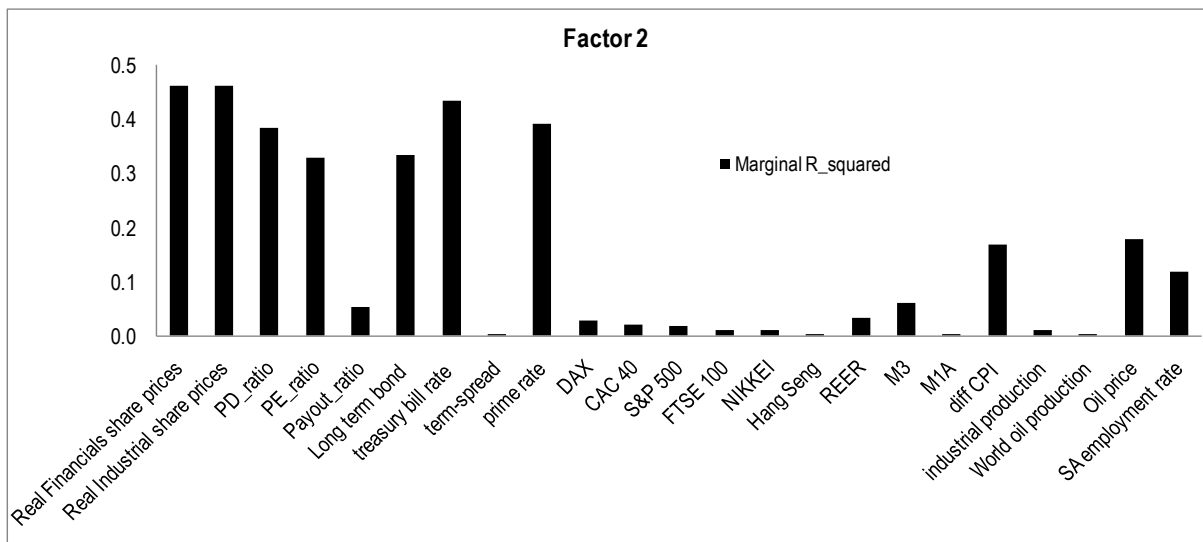


Figure 2: Principal component 2 – representing macroeconomic variables



In addition to analysing principal components based on predictive regressions, we also consider Bayesian regressions to assess the out-of-sample predictability of equity premium using all of the 23 variables together in equation 1. The use of Bayesian regressions to summarise large information sets is becoming widely used in financial literature. Pettenuzzo *et al.* (2008) use Bayesian regressions to forecast out-of-sample equity premium and find that these regressions tend to produce better forecasts than standard models. Other studies that consider the use of Bayesian regressions to forecast equity premium, although not always focusing on the out-of-sample predictive ability, include Stambaugh (1999), Avramov (2002), Cremers (2002) and Dangel and Halling (2007). Following De Mol *et al.* (2008), we choose two types of priors for the Bayesian regressions corresponding to the cases of variable aggregation and variable selection. In the case of the Gaussian prior, the maximized posterior distribution generates coefficients implying that all the predictors, including the lagged excess returns, in the panel are given non-zero coefficients. Unlike the principal component regressions, which involve regressors that are linear combinations of all variables in the panel with unit weight to the dominant ones and zero to the others, the Gaussian prior gives decreasing weight to the ordered eigenvalues of the covariance matrix of the data. On the other hand, the double-exponential prior puts more mass

near zero and in the tails, and thus, induces a tendency of the coefficients maximizing the posterior distribution to be either large or zero. As a result, it leads to a sparse specification with a recovery of a few large coefficients instead of many small ones and truly zero rather than small values, resulting in variable selection rather than in variable aggregation.

Under the Gaussian prior, it is relatively simple to compute the maximiser of the posterior density, since, with independent and identically distributed (*i.i.d.*) regression coefficients, the solution amounts to solving a penalised least-squares of the coefficients (the Ridge regression problem). The double-exponential prior, on the other hand, does not have an analytical form for the maximiser of the posterior density, but under the prior of *i.i.d.* regression coefficients, the solution boils down to a Least Absolute Shrinkage and Selection Operator (Lasso) regression problem. Following De Mol *et al.* (2008) we also consider two algorithms for the Lasso regression, the least angle regression (LARS) and the iterative Landweber scheme with soft-thresholding at each iteration. Lasso regression combines variable selection and parameter estimation, with the estimator depending in a non-linear manner on the variable to be predicted. Literature shows that, although the Gaussian and the double-exponential priors are estimated differently, an out-of-sample evaluation for these methods produces similar means squared errors.

2.5 Forecast evaluation

To evaluate the out-of-sample forecasts for different models (individual regressions, combination models, bagging regression, principal component and Bayesian regressions), we use the out-of-sample R^2 statistic, R_{OS}^2 . The R_{OS}^2 was suggested by Campbell and Thompson (2008) and used by Rapach and Strauss (2010), Rapach *et al.* (2009), as well as Jiang *et al.* (2009) and it compares \hat{r}_{t+1} (which can either be the individual regressions, combination, bagging, diffusion index and Bayesian regressions forecasts) and the AR(1) model forecasts, \bar{r}_{t+1} . The R_{OS}^2 is generated by:

$$R_{OS}^2 = 1 - \frac{(\sum_{k=q_0+1}^q r_{m+k} - \hat{r}_{m+k})^2}{(\sum_{k=q_0+1}^q r_{m+k} - \bar{r}_{m+k})^2} \quad (7)$$

The R_{OS}^2 measures the reduction in the MSFE for the individual regressions, combination, bagging, diffusion index and Bayesian regressions forecasts relative to the AR(1) model for the equity premium. This means that the \hat{r}_{t+1} forecast outperforms the AR (1) when $R_{OS}^2 > 0$, while an $R_{OS}^2 < 0$ suggests that the walker(1) outperforms the other models.

We further test whether the individual regressions, combination, bagging, diffusion index and Bayesian regressions forecasts has a significantly lower MSFE than the benchmark AR(1) model forecast. The null hypothesis in this case becomes $R_{OS}^2 \leq 0$ against the alternative hypothesis of $R_{OS}^2 > 0$. We use the MSFE-adjusted statistic developed by Diebold and Mariano (1995) and West (1996) which generates asymptotically valid inferences when comparing forecasts from nested linear models and is defined as:

$$f_{t+1} = (r_{t+1} - \bar{r}_{t+1})^2 - [(r_{t+1} - \hat{r}_{t+1})^2 - (\bar{r}_{t+1} - \hat{r}_{t+1})^2] \quad (8)$$

We then regress $\{f_{t+1}\}_{s=m+q_0}^{T-1}$ on a constant and calculating the t -statistics corresponding to a one-sided (upper tail) test – with the standard normal distribution.

2.6 Forecasting encompassing test

We further use a forecasting encompassing test² to compare the information content in the best performing model, in terms of the highest R_{OS}^2 value, with other models. To construct the forecasting encompassing test, we start by forming an optimal composite forecast of r_{t+1} as a convex combination of the forecast from models i and j , which takes the following form:

$$\hat{r}_{t+1}^* = (1 - \lambda)\hat{r}_{i,t+1} + \lambda\hat{r}_{j,t+1} \quad (9)$$

where λ is between 0 and 1. If $\lambda = 0$ (the null hypothesis) then model i forecast encompasses the model j forecast, as model j does not contain any useful information beyond that already contained in model i . However, if $\lambda > 0$ (the one sided alternative hypothesis) then model j encompasses the model i forecast. To test the null hypothesis we use a test statistic developed by Harvey *et al.* (1998). Firstly we define:

$$d_{t+1} = (\hat{\mu}_{i,t+1} - \hat{\mu}_{j,t+1})\hat{\mu}_{i,t+1} \quad (10)$$

$$\text{where } \hat{\mu}_{i,t+1} = r_{t+1} - \hat{r}_{i,t+1} \text{ and } \hat{\mu}_{j,t+1} = r_{t+1} - \hat{r}_{j,t+1} \quad (11)$$

We then define $\bar{d} = \left[\frac{1}{q-q_0}\right] \sum_{k=q_0+1}^q d_{R+k}$. We use the modified version of the test statistic, which is defined as:

$$MHLN = \left[\frac{q-q_0-1}{q-q_0}\right] [\hat{V}(\bar{d})^{-\frac{1}{2}}] \bar{d} \quad (12)$$

$$\text{with } \hat{V}(\bar{d}) = (q - q_0)^{-1} \hat{\phi}_0 \text{ and } \hat{\phi}_0 = (q - q_0)^{-1} \sum_{k=q_0+1}^q (d_{R+k} - \bar{d})^2.$$

In essence, our results only show the null hypothesis whether the best forecasting model encompasses the other models with an the alternative hypothesis that the forecast from the best performing model does not encompasses the other models.

² Forecast encompassing provides a means for comparing the information content in different forecasts. Recent research by Corradi and Swanson (2006) demonstrates the size of the in-sample period relative to the out-of-sample period, type of estimation window (fixed, rolling, or recursive), and whether the models are nested or non-nested can affect the asymptotic distribution of the test statistic. Strictly speaking, all of the conditions required for the validity of the asymptotic distribution may not be met in our case, hence, our inferences based on the MHLN statistic is intended to serve as a rough guide to statistical significance.

2.7 Utility gains

In line with Rapach et al (2010b) and Rapach and Zhou (2012), we analyse the equity premium forecasts with profit- or utility-based metric, which provide more direct measures of the value of forecasts to economic agents. A leading utility-based metric for analysing equity premium forecasts is the average utility gain for a mean-variance investor. The first step is to compute the average utility for a mean-variance investor with relative risk aversion θ^3 who allocates his portfolio between stocks and risk-free bills based on the equity premium predictive regression forecasts. This requires the investor to forecast the variance of the equity premium. Following Campbell and Thompson (2007) and Rapach and Zhou (2012), we assume that the investor allocates the following share of his portfolio to equities during $t + 1$

$$a_{i,t} = \frac{1}{\theta} \left(\frac{\hat{r}_{i,t+1}}{\hat{\sigma}_{i,t+1}^2} \right) \quad (13)$$

where $\hat{\sigma}_{i,t+1}^2$ is a forecast of the variance of the equity premium. The average utility level realised by the investor over the out-of-sample period is given by:

$$\hat{v}_i = \hat{\mu}_i - 0.5\theta\hat{\sigma}_i^2 \quad (14)$$

where $\hat{\mu}_i$ and $\hat{\sigma}_i^2$ are the sample mean and variance of the portfolio formed on the basis of $\hat{r}_{i,t+1}$ and $\hat{\sigma}_{i,t+1}^2$ over the out-of-sample forecast evaluation period. If the investor instead relies on the benchmark AR(1) model of the equity premium, he allocates the portfolio share as:

$$a_{0,t} = \frac{1}{\theta} \left(\frac{\bar{r}_{t+1}}{\hat{\sigma}_{t+1}^2} \right) \quad (15)$$

to equity during $t + 1$ and he will realise an average utility level of

$$\hat{v}_0 = \hat{\mu}_0 - 0.5\theta\hat{\sigma}_0^2 \quad (16)$$

where $\hat{\mu}_0$ and $\hat{\sigma}_0^2$ are the sample mean and variance over the out-of-sample period formed on the basis of \bar{r}_{t+1} and $\hat{\sigma}_{t+1}^2$. The difference between equation (14) and (16) represents the utility gain accruing to using the predictive regression forecast of the equity premium in place of the AR(1) forecast in the asset allocation decision. The utility gain is basically the portfolio management fee that an investor is willing to pay to have access to the additional information available in a predictive regression model or combination, bagging, diffusion index and Bayesian regressions relative to the information in the AR(1) model alone.

³ Following Rapach and Zhou (2012), we report the utility gains for $\theta = 3$ since the results are qualitatively similar for other reasonable θ values.

3 Empirical results

The results obtained from different models are discussed in this section and reported in Tables 1 and 2. We begin by discussing the data used in the analysis.

3.1 Data

We use monthly data from 1990:01 to 1996:12 for the in-sample period and 1997:01 to 2010:12 as the out-of-sample period for the equity premium and the possible predictors. The variables are discussed below:

Equity premium: Nominal return on a stock market index (All-share index) in excess of the risk-free interest rate (the Treasury bill rate);

Financials share prices:⁴ Real stock returns for the financial sector in South Africa, computed as the first difference in the log-levels of real Financial Stock Index;

Industrial share prices:⁵ Real stock returns for the industries in South Africa, computed as the first difference in the log-levels of real Industrial Stock Index;

Price-dividend ratio (log-level): One-year moving sum of the ratio of nominal dividend to nominal stock prices;

Price-earnings ratio (log-level): One-year moving sum of the ratio of nominal earnings to nominal stock prices;

Payout ratio (log-level): The ratio of price-earnings to price-dividend;

Relative long-term bond yield: Difference between the long-term government bond yield and a 12-month backward-looking moving average;

Relative 90 days Treasury bill rate: Difference between the 90-day Treasury bill rate and a 12-month backward-looking moving average;

Term spread: Difference between long-term government bond yield and the 90-day Treasury bill rate;

Relative money market rate: Difference between the prime rate and the 12-month backward-looking moving average;

DAX (log-level): The real stock returns for Germany, computed as the first difference of the real DAX (Deutscher Aktien-Index) – a blue chip stock market index consisting of the 30 major German companies trading on the Frankfurt Stock Exchange;

CAC (log-level): The real stock returns for France, computed as the first difference of the real CAC 40 (the benchmark French stock market index);

⁴ Jiang *et al.* (2009), Neely *et al.* (2011), Rapach *et al.* (2011), amongst others, suggests that sub-sectors of the overall share index (such as financial and industrial stock prices) are also possible predictors of equity premium.

⁵ See above footnote.

S&P 500 (log-level): The real stock returns for the US, computed as the first difference of the real S&P 500, which is the free-float capitalisation-weighted index of the prices of 500 large-cap common stocks;

FTSE 100 (log-level): The real stock returns for the United Kingdom, computed as the first difference of the real FTSE 100 all-share index, which is a capitalisation-weighted index of around 100 companies traded on the London Stock Exchange;

NIKKEI (log-level): The real stock returns for Japan, computed as the first difference of the real Nikkei 225 stock index for the Tokyo Stock Exchange;

Hang-Seng (log-level): The real stock returns for Hong Kong, computed as the first difference of the real Hang Seng Index, which is a free float-adjusted market capitalisation-weighted stock market index;

Real effective exchange rate: First difference in log-levels of real effective exchange rate index;

Broad money supply growth rate: First difference in the log-levels of real broadly defined money stock;

Narrow money supply growth rate: First difference in the log-levels of real narrowly defined money stock;

The inflation rate: First difference in the log-levels of the consumer price index;

Industrial production growth rate: First difference in the log-levels of industrial production;

Employment growth rate: First difference in the log-levels of employment;

World oil production growth rate: First difference in the log-levels of the world oil production; and

Crude oil price growth rate: Refiner acquisition cost of imported crude oil growth rate in real terms. To obtain the rand denominated price, we use the rand/dollar exchange rate, and then deflate the nominal value using the consumer price index to obtain the real crude oil price.

We used monthly data obtained from the South African Reserve Bank, Statistics South Africa, Bloomberg and the US Energy Information Administration. Further, barring the Treasury bill rate and the inflation rate, for which we use the first difference, all the other variables were found to be stationary based on standard unit roots tests.⁶ Following Rapach *et al.* (2005), we measure interest rate variables as deviations from a backward-moving average. This is because, if real interest rates play a crucial role in determining stock returns, then measuring the interest rate as deviations from a backward-looking moving average tends to make the nominal interest rate effectively a real interest rate. That is, the behaviour of expected inflation is such that most of the fluctuations in the relative nominal interest rate reflect movements in the relative real component. We also use growth rates for the other variables, all in an effort to have variables that are stationary.

3.2 Out-of-sample equity premium predictability

Table 1 reports the out-of-sample R_{OS}^2 , for each of the individual predictive regression models, combining methods, bagging regression, principal component and Bayesian regression models relative to the benchmark AR(1) model. For R_{OS}^2 statistics greater than zero, the statistical significance is assessed with the Clark and West (2007) MSFE-adjusted statistics discussed earlier. The results for the utility

⁶ The unit root tests are available upon request from the authors.

gains for these models are also reported in Column 3 of Table 1. From Table 1, we see that only 5 (price-dividend ratio, relative Treasury bill rate, term spread, relative money market rate and the inflation rate) out of the 23 predictors used in this paper, produce better forecasts than the AR(1) model. Out of these 5 predictors, the R_{OS}^2 is insignificant for the price dividend ratio, while the same is significant at 1 percent level for the other three interest rate based predictors, and at 10 percent level for the inflation rate. Barring financial share prices, for which there is no forecasting gains relative to the AR(1) model, the remaining 18 predictors are outperformed by the AR(1) model. When we compare these results with the random walk (historical average) used as the benchmark (reported in Table A1 in the Appendix), instead of the AR(1) model, we find that 18 of the 23 individual predictive regressions have positive R_{OS}^2 , two of which are less than or equal to 0.26 per cent (making them statistically insignificant). The significant R_{OS}^2 (significant at least at 10 percent level of significance) vary from 0.47 per cent for the FSTE 100 stock returns to 9.26 per cent for relative money market rate. The individual regressions for the payout ratio, DAX returns, the real effective exchange rate, the oil price and the employment growth are outperformed by the benchmark random walk model. Overall, the interest rate variables and the stock returns for some of South Africa's major trading partners exhibit some out-of-sample forecast – emphasising the importance of these variables when predicting equity premium for South Africa. Also, we find some out-of-sample predictive power from other variables, with only the employment being outperformed by the benchmark random walk model. Our results suggest that most of the variables included in our analysis contain important information for explaining the behaviour of the equity premium in South Africa, when using the historical average as the benchmark. So, we observe evidence of relatively more predictability when using a weaker benchmark (the random walk model) in some sense, instead of the AR(1) model, which accounts for high and significant persistence in the equity premium.

As far as the economic significance of the results are concerned, 12 out of 23 predictors produce utility gains relative to the AR(1) model, implying that even though some of the predictors might not produce statistical gains in terms of forecasting, implying that an investor is willing to pay a portfolio management fee to have access to the additional information available in a predictive regression model than can be obtained from the AR(1) model. Positive utility gains are obtained for financial and industrial share price, the valuation ratios, relative Treasury bill rate, term spread and relative money market rate, the stock returns based on the DAX, CAC 40 and S&P 500, oil price and employment growth. Interestingly, the highest utility gains are obtained from the interest rate variables, which in turn, also produce the highest forecasting gains. The inflation rate, however, fails to produce utility gains even after producing significant forecasting gains.

Apart from the alternative combination model forecasts and bagging, that we discuss in section 2.3, we also use predictive regressions based on the two principal components that were extracted using the entire sample period from 1990:01 to 2010:12. We looked at these two principal components individually and also considered them simultaneously in the predictive regression model and compared the forecasts with the benchmark AR(1), as well as the random walk model. All the combination model forecasts outperform the benchmark AR(1) model significantly at least at the 10 percent level of significance – with the R_{OS}^2 varying from 0.39 per cent (trimmed mean) to 3.08 per cent (PC_IC_p3). The diffusion index approach based on the principal components yields the following results: while the AR(1) model outperforms the predictive regression based on the first principal component, forecasting gains obtained from the second principal component yields statistical gains at one percent level of significance. This is not surprising since the second principal component mainly represents the interest

rate variables, while the first principal component captures the international stock returns. Forecasting gains are even higher (and also greater than all the combination methods) when we actually use both the principal components, suggesting that information contained in the first principal component can add value to forecasting when used simultaneously with the second principal component.⁷ Interestingly, even though the bagging model performs worse than the AR(1) model on its own, when we average the bagging forecasts and the forecasts obtained from the principal component combination method (IC_p3) since it performs the best among combination methods used, the model performs better than the AR(1) model.⁸ In fact, this model produces the highest R_{os}^2 amongst all the forecast combination methods, but cannot beat the performance of the diffusion index model based on both the principal components. As can be seen from Table A1, barring that the predictive regression model based on the first principal component also outperforms the random walk model when used as the benchmark, the remaining results are qualitatively exactly the same for the diffusion index approach, bagging and combination methods as obtained and discussed above with the AR(1) model as the benchmark.

The utility gains, reported in column 7 of Table 1, are, in general, in line with the forecasting performances of the combination, diffusion index and bagging methods. The glaring exception is the bagging model, which produces substantial utility gains of 37.16 percent at an annualised rate, even when it performed worse than the AR(1) model in terms of forecasting. Otherwise, high utility gains are obtained from the principal component forecast combination, and simple average of the principal component forecast combination and bagging methods.

.When analysing the results obtained for Bayesian regressions reported in Table 1, it is important to note the large R_{os}^2 generated by each Bayesian specifications. All the R_{os}^2 are greater than 13 per cent, with the Lasso: LARS having the highest R_{os}^2 of 21.70 per cent – and all the R_{os}^2 are significant at 1 per cent level of significance. Further, the R_{os}^2 statistics for the Bayesian regressions are greater than the largest R_{os}^2 amongst the individual predictive regressions (8.07 percent with the relative money market rate) and the combination, bagging and diffusion index approaches (3.97 per cent for the model that combines the two constructed principal components). Amongst the Bayesian regressions, the Ridge regression has the lowest R_{os}^2 value of 13.04 per cent, but is still significantly larger than all the non-Bayesian models. Again (as seen from Table A1), as with the results under the diffusion index approach, bagging and combination methods, the results for the Bayesian regressions with the random walk model used as the benchmark is qualitatively exactly the same as obtained and discussed above with the AR(1) model as the benchmark. Note that for the Bayesian (Ridge) regression, we run the regression using the first estimation sample 1990:1-1996:12 for a grid of priors. We then choose the priors for which the in-sample fit explains a given fraction of the variance of the excess returns. In our case, the ridge regression produced the lowest MSFE when 90 per cent of the variance of the excess stock return was explained.

For the double-exponential prior under the two alternative algorithms, we select the prior that delivers a given number of non-zero coefficients at each estimation step in the out-of-sample evaluation period. We look at the cases of 1 to 24 (we now also include the lagged value of the equity premium as one of the regressors) non-zero coefficients. We found the 5 non-zero coefficients produced the minimum MSFE under the double exponential prior for both the LARS and Landweber algorithms. We therefore

⁷ A potential drawback of the diffusion index model is that the estimated factors are designed to explain the covariation among the individual predictors themselves, without explicitly taking into account the relationship between the predictors and the targeted variable that we want to forecast. Kelly and Pruitt (2011) develop a three-pass regression filter (3PRF) to estimate the factors that are the most relevant for forecasting the target. When we used this approach in forecasting the equity premium based on two factors, we obtained significant (at 10 percent) forecasting gains ($R_{os}^2=0.83$) relative to the AR(1) model, which in turn, is way less than the forecasting gains obtained from the standard diffusion index approach discussed in the main text. The details of these results are available upon request from the authors.

⁸ Rapach and Strauss (2010) also obtained a similar result when forecasting employment growth for the US.

Table 1: One-month ahead forecasting and encompassing test results for the individual regressions, combining methods, bagging, principal component and Bayesian regressions:

	R ² _{OS} (per cent)	Utility gains Annual percent	Best model (LARS) Encompasses other model (p-values)		R ² _{OS} (per cent)	Utility gains Annual percent	Best model (LARS) Encompasses other model (p-values)
<i>Individual forecasts</i>				<i>Combination forecasts</i>			
Financials share prices	0	1.37	0.83	Principal component 1	-0.44	-3.22	0.86
Industrial share prices	-0.31	1.62	0.89	Principal component 2	3.76***	9.79	0.76
Price dividend ratio	0.03	0.50	0.90	BA model	-3.55	37.16	0.99
Price earnings ratio	-0.56	0.85	0.86	Both Principal Components 1 and 2	3.97***	3.99	0.72
Payout ratio	-1.49	-17.94	0.97	Mean	1.43**	0.50	0.84
Relative long term bond yield	-0.11	-14.79	0.87	Median	0.50*	-0.37	0.82
Relative treasury bill rate	6.76***	4.76	0.63	Trimmed mean	0.39*	0.50	0.95
Term spread	3.02***	11.89	0.79	DMSFE (delta = 1.0)	1.45**	0.42	0.88
Relative money market rate	8.07***	3.81	0.62	DMSFE (delta = 0.90)	1.61**	-0.65	0.87
DAX	-1.39	1.68	0.97	C(2,PB)	2.13***	-0.48	0.89
CAC 40	-1.04	2.37	0.95	C(3,PB)	2.86***	-2.11	0.76
S&P 500	-1.27	1.25	0.92	PC(IC_p3)	3.08***	35.90	0.74
FTSE 100	-0.83	-0.54	0.98	BA and PC(IC_p3)	3.16***	40.36	0.72
NIKKEI	-0.41	-2.94	0.83	Ridge	13.04***	58.39	0.21
Hang Seng	-0.35	-1.53	0.90	Lasso: LARS	21.70***	58.39	
Real effective exchange rate	-1.34	-2.04	0.96	Lasso: Landweber	20.93***	65.35	0.13
Broad money supply	-0.65	-3.09	0.97				
Narrow money supply	-0.41	-1.16	0.88				
Inflation	1.20*	-5.52	0.83				
Industrial production	-0.6	-1.07	0.92				

World oil production	-0.32	-0.15	0.94
Oil price	-2.83	4.15	0.91
Employment growth rate	-1.49	1.45	0.95

R²_{OS} is the Campbell and Thompson (2008) out-of-sample R², with ***, ** and * indicating significance at 1, 5 and 10 percent levels respectively. The *p*-values correspond to the forecast encompassing test with the null hypothesis that the best model (Lasso: LARS) forecast encompasses the forecasts from all the other models individually.

examine the five variables selected at the beginning (1997:01) and at the end (2010:11) of the sample out-of-sample evaluation period. There are a number of results that emerges from Table 2. Firstly, the variables that are selected by both the Lasso: LARS and the Lasso: Landweber are inconsistent across time horizons and across models. Secondly, the choice of variable selection is not entirely in line with the performance of each variables when using the R_{os}^2 to evaluate the out-of-sample predictability. Thirdly, for both the Lasso algorithms and time periods, the variables selected include both the financial and macroeconomic variables. The above results suggest that the variables included in these algorithms are not necessarily consistent through time as a result of collinearity and hence we have temporal instability (De Mol *et al.*, 2008).

Just like the forecasting gains, the Bayesian models also produce very high utility gains when compared to the individual predictive regressions, the diffusion index approach, bagging and combination methods. The Lasso: Landweber based Bayesian regression produces the largest utility gains.

	1997:01 (First out-of-sample period)	2010:11 (Last out-of-sample period)
Variables included in the Lasso: LARS model	Real Financials share prices	Real Financials share prices
	Price dividend ratio	Money market rate
	Payout ratio	S&P 500
	DAX	Industrial production
	Employment rate	Lagged Excess Returns
Variables included in the Lasso: Landweber model	Real Industrial share prices	Payout ratio
	Payout ratio	Term spread
	Money market rate	CAC 40
	NIKKEI	Oil price
	Oil price	Employment rate

We provide statistical explanations for the relatively good out-of-sample performance of the Lasso: LARS with respect to the equity premium. Through the forecast encompassing tests, we are able to show that the Lasso: LARS incorporates useful forecasting information from the macroeconomic and financial variables included in our analysis. Table 1 also reports p -values for the MHLN statistic applied to the out-of-sample forecasts. Each entry in the table corresponds to the null hypothesis that the Lasso: LARS forecast encompasses the forecasts for the individual regressions, the diffusion index approach, bagging and combination methods.. From Table 1, it is visible that the Lasso: LARS forecast encompasses all models we consider – suggesting that it is important to combine information from individual variables using the Lasso: LARS model specification to incorporate additional information thus explaining the out-of-sample gains corresponding to the Lasso: LARS model. Not surprisingly, this is also the case when we use the historical average as the benchmark, as observed from Table A1.

Following Goyal and Welch (2008) and Rapach and Zhou (2012), Figure 3 presents the cumulative difference in squared forecast errors for the AR(1) forecast relative to the predictive regression forecast. While, Figure 4 does the same relative to the diffusion index approach, bagging, combination methods and Bayesian regressions. This is an informative tool that provides a graphical representation of the consistency and volatility of the forecasting performances of these models over time. When the curves in Figures 3 and 4 is in the positive (negative) plane then the predictive regression model, the diffusion

Figure 3: Cumulative square predictive error for the AR(1) minus the cumulative square predictive error for the individual regressionsbg3

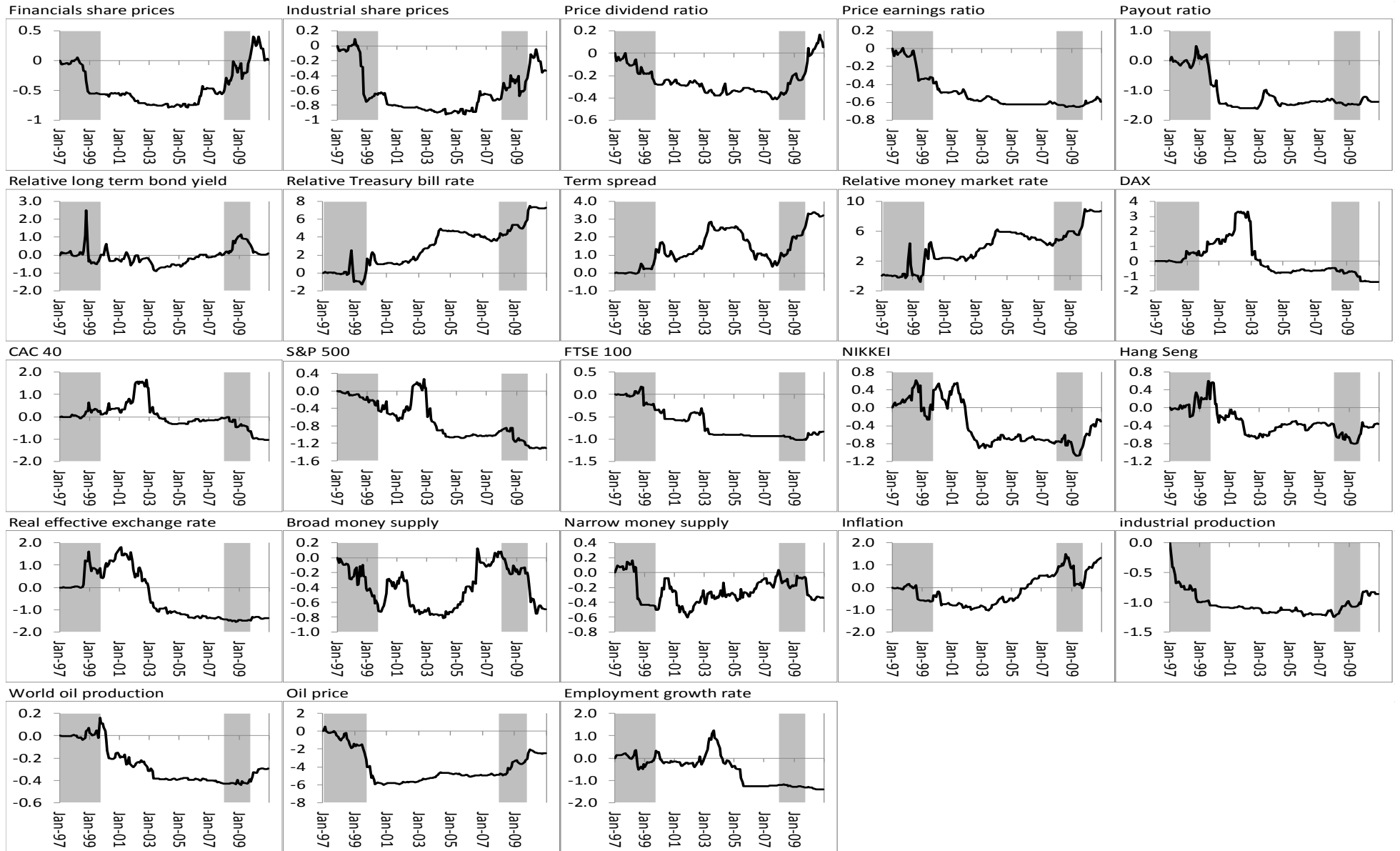
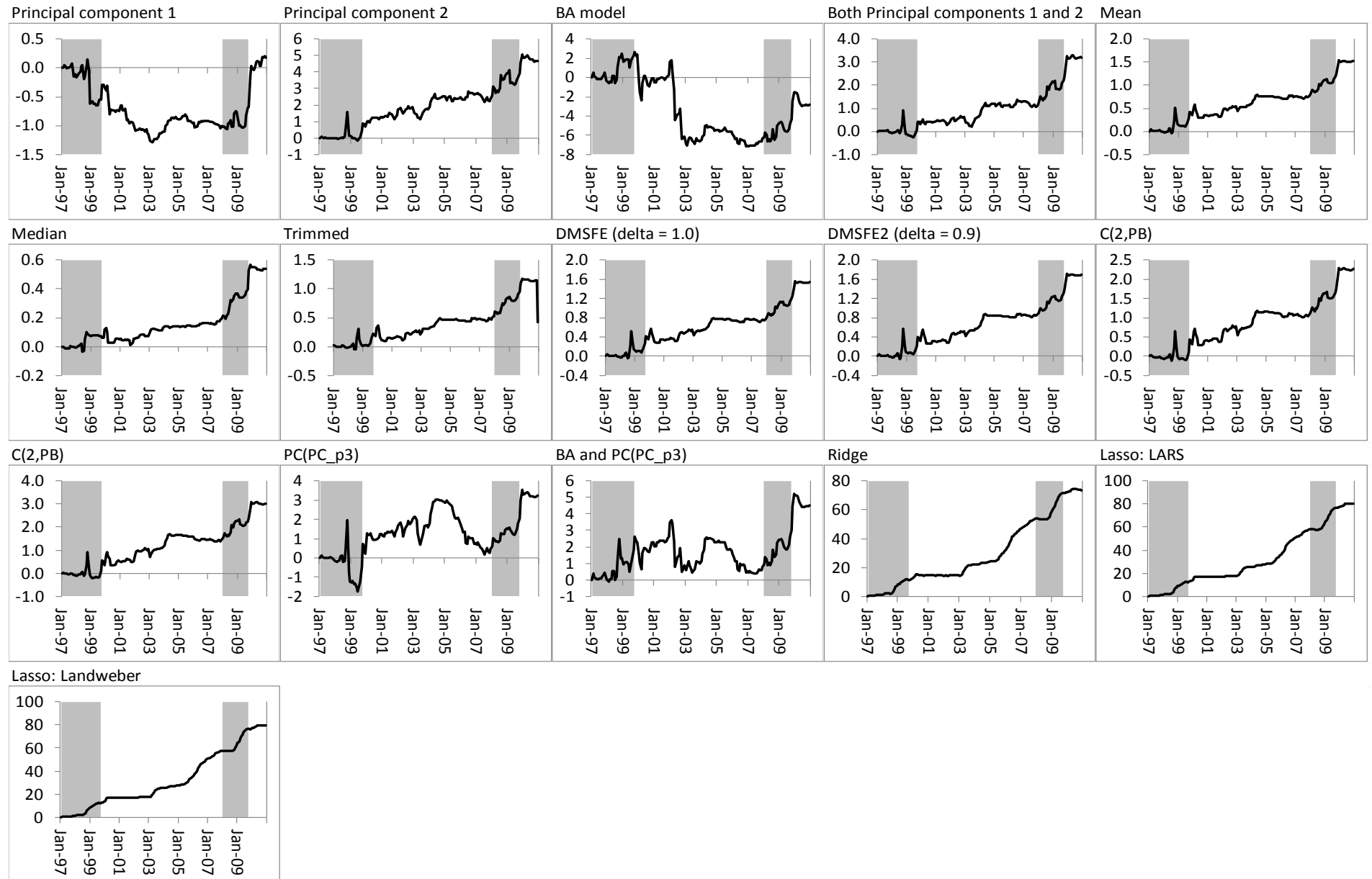


Figure 4: Cumulative square predictive error for the AR(1) minus the cumulative square predictive error for the combination models



index approach, bagging, combination methods and Bayesian regressions outperforms (is outperformed by) the benchmark AR(1) model. As pointed out in Goyal and Welch (2007), and reiterated in Rapach *et al.*, (2008), the units on the plots are not intuitive; these plots are, however, useful in determining the how the predictive regressions, diffusion index approach, bagging, combination methods and Bayesian regressions perform in terms of MSFE relative to the benchmark AR(1) model at each point of time over the out-of-sample horizon.

The figures echo the same story as indicated by R_{os}^2 reported in Table 1. In other words, amongst the 4 variables that show significant forecasting gains, for the relative Treasury bill rate, term spread, relative money market rate, the cumulative difference in squared forecast errors for the AR(1) forecast relative to the predictive regression forecast is consistently positive. As far as the inflation rate is concerned, the graph remains in the negative plane for most of the out-of-sample period, but high forecasting gains are registered during the “Great Recession”. The high forecasting gains form the price dividend ratio during and after the financial crisis produces the positive, but insignificant R_{os}^2 . Interestingly, for many of the other predictors which produces negative R_{os}^2 , forecasting gains are observed based on the predictors relative to the AR(1) model during the recessionary period of 1997:1-1999:8, but not during the recession of 2007:12-2009:08. In general, the graphs are more volatile during the recession, emphasising the difficulty in predicting equity premium during downturns relative to the benchmark.

As with individual predictive regression models, the cumulative difference in squared forecast errors for the AR(1) forecast relative to the diffusion index approach, bagging, combination methods and Bayesian regressions forecast is consistently positive for all the cases, barring the bagging approach and the predictive regression based on the first principal component (financial variable) for which $R_{os}^2 < 0$. But again, for these two cases, forecasting gains are observed based relative to the AR(1) model during the recessionary period of 1997:1-1999:8, but not during the recession of 2007:12-2009:08. The graphs depict minimal volatility, irrespective of whether the economy is in an upturn or a downturn for the Bayesian regressions; thus highlighting their superiority in terms of forecasting performances.

4 Conclusion

The aim of this study is to analyse the predictability of South Africa equity premium considering financial and macroeconomic variables using monthly data from 1990:01 to 2010:12 for the out-of-sample period of 1997:01 to 2010:12 period. The literature suggest that combining individual forecasts is known to outperform the individual forecasts themselves, we therefore use a number of combination model forecasts, besides, bootstrap aggregation (bagging), principal component and Bayesian regressions to simultaneously incorporate information from 23 possible predictors. We find that only four (the relative money market rate, relative treasury bill rate, the term spread and the inflation rate) of the 23 predictors significantly outperforms the benchmark AR(1) model. However, in terms of economic significance, 12 of the 23 predictors produce positive utility gains, with term spread producing the highest utility gains of 11.89 percent amongst the individual predictors.

As suggested in the literature, all the combination model forecasts outperform the benchmark AR(1) walk model and have R_{os}^2 that are statistically significant at least at 10 per cent level of significance. Though the bagging model performs way poorer than the benchmark, when combined with the principal component based forecast combination method (the best performing forecast combination

method), the combination outperforms the AR(1) model and ends up beating all the other forecast combination methods in terms of R_{OS}^2 . The principal component based predictive regressions, based on the second principal component (depicting macroeconomic variables) and when involving both principal components also perform significantly better than the AR(1) model. The Bayesian regressions (Ridge, Lasso: LARS and Lasso: Landweber) are, however, found to be the standout performers, with these models outperforming forecast combination methods, bagging, the diffusion index approach and all individual predictive regressions. All the R_{OS}^2 are greater than 13 per cent, with the Lasso: LARS having the highest R_{OS}^2 of 21.70 per cent – and all the R_{OS}^2 are significant at 1 per cent level of significance. When using forecast encompassing tests to provide statistical explanations for the relatively good out-of-sample performance of the Lasso: LARS, we find that this model forecast encompasses all other model specifications. This means that the Lasso: LARS incorporates useful forecasting information from the macroeconomic and financial variables included in our analysis. The utility gains are, in general, in line with the forecasting performances of the combination, diffusion index, bagging methods and the Bayesian regressions. The benefits of predicting the equity premium using Bayesian models are also evident in the utility gains for these models. The models have the highest utility gains compared to all other models, with the Lasso: Lanweber producing the highest utility gain of 65.35 percent at an annualised rate.

References

- Aiolfi, M. & Timmermann, A. 2006. Persistence in Forecasting Performance and Conditional Combination Strategies. *Journal of Econometrics*, 135: 31–53.
- Alessi, L., Barigozzi, M. & Capasso, M. 2010. Improved Penalization for Determining the Number of Factors in Approximate Factor Models. *Statistics & Probability Letters*, 80: 1806–1813.
- Ang, A. & Bekaert, G. 2001. Stock Return Predictability: Is it there? *National Bureau of Economic Research, Working Paper* No. 8207.
- Avramov, D. 2002. Stock Return Predictability and Model Uncertainty. *Journal of Financial Economics* 64: 423–458.
- Bai, J. & Ng, S. 2002. Determining the Number of Factors in Approximate Factor Models. *Econometrica*, 70: 191–221.
- Bates, J.M. & Granger, C.W.J. 1969. The Combination of Forecasts. *Operational Research Quarterly*, 20: 451–468.
- Belmonte, M.A.G., Koop, G. & Korobilis, D. 2011. Hierarchical Shrinkage in Time-varying Parameter Models. *Université catholique de Louvain, Center for Operations Research and Econometrics, Core Discussion Papers* No. 2011036.
- Campbell, J.Y. & Shiller, J.R. 1988. Stock Prices, Earnings, and Expected Dividends. *Journal of Finance*, 43: 661–676.
- Campbell, J.Y. & Shiller, J.R. 1998. Valuation Ratios and the Long-Run Stock Market Outlook. *Journal of Portfolio Management*, Winter: 11–26.

- Campbell, J.Y. & Thompson, S.B. 2008. Predicting the Equity Premium Out of Sample: can Anything Beat the Historical Average? *Review of Financial Studies*, 21: 1509–1531.
- Campbell, J.Y. 1987. Stock Returns and the Term Structure. *Journal of Financial Economics*, 18: 373–399.
- Clark, T.E. & West, K.D. 2007. Approximately Normal Tests for Equal Predictive Accuracy in Nested Models. *Journal of Econometrics*, 138: 291–311.
- Corradi, V. & Swanson, N.R. 2006. Predictive Density Evaluation. In: Elliott, G., Granger, C. W. J., Timmermann, A., eds. *Handbook of Economic Forecasting*. Amsterdam: Elsevier, 197–284.
- Cremers, K.J.M. 2002. Stock Return Predictability: A Bayesian Model Selection Perspective. *Review of Financial Studies* 15: 1223–1249.
- Dangl, T. & Halling, M. 2007. Predictive Regressions with Time-Varying Coefficients. *Working Paper, University of Utah*.
- De Mol, C., Giannone, D. & Reichlin, L. 2008. Forecasting Using a Large Number of Predictors: Is Bayesian Regression a Valid Alternative to Principal Components? *Journal of Econometrics*, 146: 318–328.
- Diebold, F.X. & Mariano, R.S. 1995. Comparing Predictive Accuracy. *Journal of Business and Economic Statistics*, 13: 253–263.
- Fama, E.F. & French, R.K. 1988. Dividend Yields and Expected Stock Returns. *Journal of Financial Economics*, 22: 3–25.
- Fama, E.F. & French, R.K. 1989. Business Conditions and Expected Returns on Stocks and Bonds. *Journal of Financial Economics*, 25: 23–49.
- Fama, E.F. 1981. Stock Returns, Real Activity, Inflation and Money. *American Economic Review*, 71: 545–565.
- Forni, M., Hallin, M., Lippi, M. & Reichlin, L. 2003. Do Financial Variables Help Forecasting Inflation and Real Activity in the Euro Area? *Journal of Monetary Economics*, 50: 1243–1255.
- Geske, R. & Roll, R. 1983. The Fiscal and Monetary Linkages between Stock Returns and Inflation. *Journal of Finance*, 38: 1–34.
- Goyal, A. & Welch, I. 2008. A Comprehensive Look at the Empirical Performance of Equity Premium Prediction. *Review of Financial Studies*, 21: 1455–508.
- Gupta, R. & Hartley, F. 2011. The Role of Asset Prices in Forecasting Inflation and Output in South Africa. *University of Pretoria, Department of Economics, Working Paper No. 2011–15*.
- Harvey, D.I., Leybourne, S.J, & Newbold, P. 1998. Tests for Forecast Encompassing. *Journal of Business and Economic Statistics*, 16: 254–59.
- Inoue, A. & Kilian, L. 2008. How Useful is Bagging in Forecasting Economic Time Series? A Case Study of U.S. CPI Inflation. *Journal of the American Statistical Association*, 103: 511–522.

- Jiang, F., Rapach, D.E., Strauss, J.K. & Tu, J. 2009. How Predictable is the Chinese Stock Market? *Dissertations and Theses Collection*, Paper 57.
- Kelly, B., Pruitt, S. 2011. The three-pass regression filter: a new approach to forecasting using many predictors. *University of Chicago Booth School of Business Working Paper No. 11-19*.
- Kong, A., Rapach, D.E., Strauss, J.K, Tu, J. & Zhou, G. 2009. How Predictable are Components of the Aggregate Market Portfolio? *University of St. Louis*, Mimeo.
- Korobilis, D. 2011. Hierarchical Shrinkage Priors for Dynamic Regressions with Many Predictors. *Université catholique de Louvain, Center for Operations Research and Econometrics, Core Discussion Papers No. 2011021*.
- Lamont, O. 1998. Earnings and Expected Returns. *Journal of Finance*, 53: 1563-1587.
- Ludvigson, S.C. & Ng, S. 2007. The Empirical Risk–Return Relation: A Factor Analysis Approach. *Journal of Financial Econometrics* 83: 171–222.
- Ludvigson, S.C. & Ng, S. 2009. Macro Factors in Bond Risk Premia. *Review of Financial Studies*, 22: 5027–5067.
- Marquering, W. & Verbeek, M. 2004. The economic value of predicting stock index returns and volatility. *Journal of Financial and Quantitative Analysis*, 39: 407–429.
- Neely, C.J., Rapach, D.E., Tu, J. & Zhou, G. 2011. Forecasting the Equity Risk Premium: The Role of Technical Indicators. *Federal Reserve Bank of St. Louis Working Paper No. 2010-008C*.
- Peersman, G. & Van Robays, I. 2009. Oil and the Euro Area Economy. *Economic Policy*, 24: 603–651.
- Pesaran, M.H. 2003. Market Efficiency and Stock Market Predictability. *Mphil Subject 301*.
- Pettenuzzo, D., Timmermann, A., Valkanov, R. & Wu, R. 2008. Return Predictability under Equilibrium Constraints on the Equity Premium. *Working Paper, University of California at San Diego*.
- Rapach, D.E. & Strauss, J.K. 2010. Bagging or Combining (or Both)? An Analysis Based on Forecasting U.S. Employment Growth. *Econometric Review*, 29: 511–533.
- Rapach, D.E. & Zhou, G. Forthcoming. Forecasting Stock Returns. *Forthcoming in the Handbook of Economic Forecasting, Volume 2*.
- Rapach, D.E., Strauss, J.K. & Zhou, G. 2009. Out-of-Sample Equity Premium Prediction: Combination Forecasts and Links to the Real Economy. *The Review of Financial Studies*, 23: 821-862.
- Rapach, D.E., Strauss, J.K. & Zhou, G. 2010a. International Stock Return Predictability: What is the Role of the United States? *St. Louis University, Department of Economics*, Mimeo.
- Rapach, D.E., Strauss, J.K. & Zhou, G. 2010b. Out-of-sample equity premium prediction: combination forecasts and links to the real economy. *Review of Financial Studies*, 23: 821–862.
- Rapach, D.E., Strauss, J.K., Tu, J. & Zhou, G. 2011. Out-of-Sample Industry Return Predictability: Evidence from a Large Number of Predictors. *St. Louis University, Department of Economics*, Mimeo.

- Rapach, D.E., Wohar, M.E & Rangvid, J. 2005. Macroeconomic Variables and International Stock Return Predictability. *International Journal of Forecasting*, 21: 137–166.
- Stambaugh, R.F. 1999. Predictive Regressions. *Journal of Financial Economics*, 54: 375–421.
- Stock, J.H. & Watson, M.W. 2003. Forecasting Output and Inflation: The Role of Asset Prices. *Journal of Economic Literature*, 41: 788–829.
- Stock, J.H. & Watson, M.W. 2004. Combination Forecasts Of Output Growth in A Seven-Country Data Set. *Journal of Forecasting*, 23: 405–430.
- Wachter, J.A. & Warusawitharana, M. 2009. Predictable returns and asset allocation: should a skeptical investor time the market? *Journal of Econometrics*, 148: 162–178.
- West, K.D. 1996. Asymptotic Inference about Predictive Ability. *Econometrica*, 64: 1067–1084.

APPENDIX:

Table A1: One-month ahead forecasting and encompassing test results for the individual regressions, combining methods, bagging, principal component and Bayesian regressions:

	R ² _{OS} (per cent)	Best model (LARS) Encompasses other model (p-values)		R ² _{OS} (per cent)	Best model (LARS) Encompasses other model (p-values)
<i>Individual forecasts</i>			<i>Combination forecasts</i>		
Financials share prices	1.29**	0.82	Principal component 1	0.85**	0.85
Industrial share prices	0.98**	0.88	Principal component 2	5.00***	0.76
Price dividend ratio	1.31**	0.89	BA model	-2.21	0.98
Price earnings ratio	0.74*	0.85	Both Principal Components 1 and 2	5.21***	0.72
Payout ratio	-0.18	0.96	Mean	2.70***	0.83
Relative long term bond yield	1.18**	0.86	Median	1.78***	0.81
Relative treasury bill rate	7.96***	0.63	Trimmed mean	1.67**	0.94
Term spread	4.27***	0.78	DMSFE (delta = 1.0)	2.72***	0.87
Relative money market rate	9.26***	0.62	DMSFE (delta = 0.90)	2.88***	0.86
DAX	-0.09	0.96	C(2,PB)	3.39***	0.88
CAC 40	0.26	0.94	C(3,PB)	4.11***	0.76
S&P 500	0.04	0.91	PC(IC_p3)	4.33***	0.74
FTSE 100	0.47*	0.97	BA and PC(IC_p3)	4.41***	0.72
NIKKEI	0.89**	0.82	Ridge	14.16***	0.21
Hang Seng	0.95**	0.89	Lasso: LARS	22.71***	
Real effective exchange rate	-0.03	0.95	Lasso: Landweber	21.95***	0.13
Broad money supply	0.65*	0.96			
Narrow money supply	0.89**	0.87			
Inflation	2.48***	0.82			
Industrial production	0.69*	0.91			
World oil production	0.97**	0.93			
Oil price	-1.5	0.9			
Employment growth rate	-0.18	0.94			

R²_{OS} is the Campbell and Thompson (2008) out-of-sample R², with ***, ** and * indicating significance at 1, 5 and 10 percent levels respectively. The p-values correspond to the forecast encompassing test with the null hypothesis that the best model (Lasso: LARS) forecast encompasses the forecasts from all the other models individually.