Combination Forecasts of Integrated Information

Wanqian Zhang [*]

author

Bart F. Diris

Supervisor

Erasmus University Rotterdam, The Netherlands

Bachelor Thesis

June 28, 2011

[*]Address correspondence to Wanqian Zhang, Erasmus School of Economics, Erasmus University Rotterdam, The Netherlands; Email: tony.wanqian.zhang@gmail.com

Abstract

This paper aims to achieve gains in forecasting performance resulting from combination by incorporating information of the economic variables. The objective is to improve forecast combinations by carefully integrating the information from the explanatory variables. An important approach of incorporating information from economic variables is factor analysis, involves extracting a relatively small number of common factors from a large number of variables and using these factors in a single forecasting model. Next to the Principle Component Analysis (PCA) for achieving this goal we also propose a new approach by dividing the whole set of candidate predictors into subsets before extracting the principal components. This is called combination forecast of integrated information. The idea of incorporating and dividing information subsequently combining the forecast intuitively might work since this methodology pays both attention to which forecasts to include and how to weight them. The diversification benefits through combining forecasts is known and especially, this methodology attempts to incorporate relevant information from the observed variables at the same time.

Keywords: Combination Forecasts, Combination of Information, Principle Component Analysis, Integrating of Information, Diversification Gains.

1 Description

Following Bates and Granger (1969), forecast combination has proven to be a highly successful forecasting strategy and has come to be viewed as a simple and effective way to improve the forecasting performance over the standard methodology of selecting one or more predictive variables and regressing future stock returns on these predictors. Combination of predictions is more accurate out-of-sample than predictions of individual models themselves. A first explanation for this is that ex ante equally probable models produce enormous differences in predictions. Choosing one model and neglecting the other would lead to waste of information. By combining forecasts from several models we implicitly acknowledge that more than one model could provide good forecasts and we guard against misspecification by not putting all the weight on one single model. Secondly, combining models leads to less uncertainty/variability/estimation errors in predictions similar to diversification applied by combining individual stocks in large portfolios to reduces risk. Both explanations contribute to more (expected) predictive power 'out-of-sample'. As a result, predicting stock returns using forecast combination is now extensively used in both the financial sector, from central banks to private forecasters, and the academic world in research and studies. While the literature on forecast combination is extensive, examples of formal evaluations of forecast methods see Stock and Watson (2004) include focus on macroeconomic forecasting, where forecast combination has performed well. Recent work of Rapach et al. (2010), showing combining forecasts deliver statistically and economically significant out-of-sample gains relative to the historical average consistently over time. See also Timmerman (2006) for recent theoretical contributions.

However, there are also arguments against using forecast combinations. Timmerman (2006) for instance states that estimation errors that contaminate the combination weights are known to be a serious problem for many combination techniques especially when the sample size is small relative to the number of forecasts. Another disadvantage is that the estimated coefficients and standard errors are hard to interpret and that multicollinearity is to some extent unobservable. Adding multiple individual forecasts without knowing whether they are of significance can result in less accurate forecasts or a loss of efficiency. In situations where the information sets underlying the individual forecasts are unobserved, most would agree that forecast combinations can add value. Though, when the information sets underlying the individual forecasts are observed, one can obtain forecasts of interest by combining the individual forecasts or using information underlying the econometric model.

Forecast combinations require deciding both which forecasts to include and how to weight them. To pool forecasts and to pool information are directly linked. Since the beginning of forecast combination, many studies can be found for assigning weights to the included models and relatively little attention is paid to which forecasts need to be included. As mentioned although not explicitly investigated in seminar research report by Ibisevic et al. (2011), the standard methodology of selecting one or more predictive variables and regressing future stock returns on these predictors seems very simplistic. Model instability has come to be an issue. This paper aims to achieve gains in forecasting performance resulting from combination by incorporating information of the economic variables. The objective is to improve the forecast combination by carefully integrating the information from the explanatory variables.

In the literature, an important approach of incorporating information from economic variables is factor analysis, involves extracting a relatively small number of common factors from a large number of variables and using these factors in a single forecasting model. This is called 'combination of information' by Huang and Lee (2009). This is based on the idea that variations in a large number of economic variables can be modeled by a small number of reference variables. Central to both the theoretical and the empirical validity of factor model is, as stated in Bai and Ng (2002), the correct specification of the number of factors. The work of Bai and Ng (2002) provides a formal statistical procedure that consistently estimate the number of factors from observed data. This method is relevant and can be partly used in this paper. In summary, the main applications of factor analytic techniques are: (1) Data reduction, to reduce the number of variables and (2) Structure detection, to detect structure in the relationships between variables. Regression-based factor models are commonly used to forecast the expected return and the risk of a portfolio. The expected return on each asset in the portfolio is approximated as a weighted sum of the expected returns to several market risk factors. The weights are called $factorsensitivities$, or, more specifically, factor betas and are estimated by regression, as described by Alexander (2008). Recent work by Stock and Watson (2002), they shed light on forecasting a single time series using a large number of predictor series and estimated by principal components approach and show that factor estimates are consistent in an approximate factor model with idiosyncratic errors that are serially and cross-sectionally correlated.

The outline of the chapter is as follows. Section 2 extracts main conclusions from the empirical literature. Section 3 discusses the classical principal component methodology and also an extended methodology motivated by Stock and Watson (2002), which can be seen as the starting position of this paper. The following section provides an intuitive approach based on the principal component methodology, which is related to the objective of this paper. Section 5 proceeds with discussing various combination forecast frameworks and section 6 covers the evaluation of the forecasts. Data description is provided in section 7 and empirical results are presented in section 8. At last, section 9 concludes.

2 Empirical Literature

The empirical investigation of this paper is related to several high-quality articles and this section summarizes their main contributions in the economic literature.

2.1 Combining Forecasts or Combining Information

When the objective is to forecast a variable of interest with many explanatory variables available, there are generally two possible directions one can proceed: Combination of Forecast from simple models each incorporating a part of the whole information set or Combination of Information that aims to focus on the entire information set into a super model. The forecasting literature on both combining forecast or combining information is extensive and a large part concerned with methodology has tended to improve the forecast. This paper has tried to link both and it is therefore important to extract meaningful conclusions from both forecasting strategies beforehand in the empirical literature.

In the forecasting literature, many studies argue that one should try to combine the information that goes into the model and not to combine forecasts that come out the model. Diebold (1989) for instance finds that time is better spent improving specification while pooling forecasts only takes a short-run perspective and by mechanically combining forecasts of possible incorrect models does not lead to truly understanding of the problem. When the information sets underlying the individual forecasts are observed, one can obtain forecasts of interest combining information by identifying and assuming of the existence of a single ”super” model.

However, Diebold (1989) further points out, "...it should be clear that the two approaches are highly complementary, and that forecast combination can be viewed as a key link between short-run, real-time forecast production process, and the long run." Diebold and Pauly (1990) recommend that in a world in which information set can be instantaneously combined, then combination of information is always preferred. However, particularly in real time, pooling information set is either impossible or costly. Timmerman (2006) discovers that there are many similarities between forecast combination problem and the standard problem of constructing a single econometric specification. First, for the economic variable to be explained, one can find a long list of possible explanatory variables and the question is which candidate predictors in case of combining information or which individual forecasts in case of combining forecasts have to be selected. For both cases the omission of relevant variables leads to biased estimates while carelessly selecting the input could lead to wrong models consisting of redundant variables and a loss of efficiency as a result. Diebold and Pauly (1990) find that estimation errors can contaminate the combination weights and this is a serious problem for many combination techniques especially when the sample size is small relative to the the number of forecasts. However, in a data where there are many relevant input variables possible, combining information may suffer from dimensionality. Second, for both cases the choice of functional form for mapping the information as well as the choice of estimation method have to be determined, c.f. Timmerman (2006). As Hendry and Clements (2002) point out, misspecification due to complicated dynamics and nonlinearly might occur for constructing a "super" model.

Forecast combination was first introduced by Bates and Granger (1969), motivated from the idea of portfolio diversification. Many high-quality surveys, like the work of Clemen and Winkler (1986) and Clemen (1989), proceed developing this approach and report that forecast accuracy can be substantially improved through the combination of multiple individual forecasts and in many cases one can make dramatically improvements by using simple combination methods. Using combining one can guard forecasts against structure breaks that possible affect the individual forecasts, c.f. Makridakis (1989), Hendry and Clements (2002), Stock and Watson (1998) and Timmerman (2006). At the same time, combining forecast can guard against possible misspecification biases and measurement errors of the individual forecasts since we acknowledge that more than one model could provide good forecasts for different points in time. Much of the current popular literature have found out that simple combination forecasts can dominate many of the combination of information schemes. Huang and Lee (2009) find that the in-sample fit of combining information is better than combining forecasts but this is no longer valid for out-of-sample analysis. They further point out that combining forecasts can be superior to the forecasts by combining information in case of misspecified models as well as correctly specified models. More arguments and advantages of using combining forecasts can be found in many surveys, however relatively little attention is paid in the combining of both approaches. Recent work of Aiolfi et al. (2010) explain that additional improvements can be gained by using a simple combination method and model based forecasts. They successfully shed light on the design of a universe of forecasts including both subjective survey forecasts and forecasts from time-series model to improve the forecasting performance.

2.2 Forecasting using Principle Components

The methodology used for this paper is related to several economic literature about factor analysis. Stock and Watson (2002) for instance provide both theoretical arguments and empirical evidence to prove that consistent estimates of the common factors can be constructed by principle component analysis, even in the presence of time variation in the factor model. Stock and Watson (2002) and Stock and Watson (2004) find that feasible forecasts are shown to be asymptotically efficient, because the difference between the feasible and infeasible forecasts constructed using the actual values of the factors converges in probability to 0 as both N and T are large. When macroeconomic forecasts are composed from many individual variables over a long period, using common factors can provide robustness against weakly cross-sectionally dependence and small temporal instability which is a plague to many simple regressions. A explanation for this is that instabilities can be ruled out and reduced by the construction of common factors if the instabilities are clearly different between the related series. As a result, the instability does not affect the consistency of the principle components estimator of the common factors.

Stock and Watson (2002) further show that predictions using PCA approach based on the estimated factors of large data sets are significantly improved compared to the predictions of low-dimensional forecasting regressions. They have applied factor models and principle components to forecast several macroeconomic variables, and successfully obtained forecasting gains. One of the experiments is for instance the forecasting of the Federal Reserve Board's

Index of Industrial Production using 149 monthly macroeconomic variables (e.g., production, price inflation, interest rates) as explanatory variables. Also noteworthily from this experiment is that the simulated out-of-sample results suggest that nearly all the forecasting gains comes from first two or three factors and this discovery also account in the choice of some preliminaries for improving the forecasting performance in this paper. Furthermore, Ludvigson and Ng (2007) have applied factor analysis build on an even larger data set of macroeconomic and financial variables $\mathbf{X}$. Based on the argument that small number of predictive variables can result to an omitted-information bias an therefore they try to strike this problem by employing a methodology for incorporating a large amount of conditioning information, and they find notable in-sample predictive ability for the factors. In the paper of Stock and Watson (2002), Ludvigson and Ng (2007) or other economic literature of factor analysis, many methodologies used are related to the advantage of using factor models that the information in a large data set can be effectively summarized by a relatively small number of estimated factors.

3 Classical principal component methodology

As described in section 1, in data sets with many variables, groups of variables often move together. Therefore, reduction or simplification of many explanatory variables to a few principal components provides some useful contributions in forecasting. Principle Component Analysis (PCA) is defined as a mathematical procedure that uses an orthogonal transformation and has come to be viewed as a quantitatively rigorous method for achieving this simplification. The aim is to convert a set of observations of possibly correlated variables into a set of values of uncorrelated variables called principal components. The number of principal components is less than or equal to the number of original variables. Each principal component is a linear combination of the original variables and all the principal components are orthogonal to each other, so there is no redundant information. The principal components as a whole form an orthogonal basis for the space of the data by orthogonal transformation, which are ordered so that the first few retain most of the variation present in all of the original variables. Moreover, the first principal components has as high a variance as possible and each succeeding component in turn has the highest variance possible under the constraint that it be orthogonal to the preceding components.

Take a look back in the history of techniques of multivariate analysis, PCA can be viewed as the oldest and best known and it was first introduced by Pearson (1901), later developed by Hotelling (1933). Until the advent of computers, PCA was not widely used, but it is now entrenched in every statistical computer package. PCA is a very powerful statistical tool that works best on a highly collinear system and it has great impact in many financial applications, in particular to term structures of interest rates, forwards, futures or volatility and also modelling hedge funds. This is because there are only a few important sources of information in the data, which are common to all the variables. The PCA is classified as an useful technique because it allows one to extract these key sources of variation from the data.

This section continues as follows. Section 3.1 provides some basic derivations of the principle components. Since the main interest of this paper is not the technical or mathematical results of PCA, we summarize the important concepts of PCA without much attention to technical details. Section 3.2 proceeds with the methodology of forecasting using principle components based on existing articles and section 3.3 describes the procedure of determining the number of factors which can be seen as an important part of the PCA.

3.1 Basic Derivations of PCA

PCA is based on the eigenvalue-eigenvector decomposition of a variables correlation or covariance matrix. Consider a set of N variables (In section 3.2, denoted as candidate predictors. Data description see section 7) with time series data in a $T \times N$ matrix $\mathbf{X}$. Let $\mathbf{V}$ be the covariance matrix (or correlation matrix) of $\mathbf{X}$. The principle components of $\mathbf{V}$ are the columns of the $T \times N$ matrix $\mathbf{P}$ defined by

$$\mathbf{P} = \mathbf{X}\mathbf{W} \tag{1}$$

where $\mathbf{W}$ is the $N \times N$ orthogonal matrix of eigenvectors of $\mathbf{V}$. Therefore (1) shows that the original correlated variables $\mathbf{X}$ has been transformed into a system of orthogonal variables $\mathbf{P}$, i.e. the system of principle components. By the fact that $\mathbf{W}$ is a orthogonal matrix, by using the definition of orthogonality, $\mathbf{W}^{-1} = \mathbf{W}'$ and so

$$\mathbf{X} = \mathbf{P}\mathbf{W}' \tag{2}$$

The idea now is try to 'truncate' (2) at some point $r < N$, in such a way that the first r factors represent the original variables $\mathbf{X}$. This boils down to use the first r principle components of the covariance matrix of $\mathbf{X}$. For this purpose $\mathbf{W}$ is ordered where the first column of $\mathbf{W}$ is the eigenvector corresponding to the largest eigenvalue of $\mathbf{V}$, the second column of $\mathbf{W}$ is the eigenvector corresponding to the second largest eigenvalue of $\mathbf{V}$, and so on. The mth principal component is the mth column of $\mathbf{P}$ which is derived from the mth column of $\mathbf{W}$. The sum of squares of the elements in the mth principle component is the mth largest eigenvalue of $\mathbf{V}$, denoted λ_m. The total variation in $\mathbf{X}$ is the sum of the eigenvalues of $\mathbf{V}$, $\lambda_1 + ... + \lambda_n$, so the first r principle components of the original variables capture a proportion

$$\frac{\lambda_1 + ... + \lambda_r}{\lambda_1 + ... + \lambda_n} \tag{3}$$

of the total variation in the system. The question remains how to determine the number of factors r. This is described in section 3.3.

When the first r column of $\mathbf{P}$ are used as the column of a $T \times r$ matrix $\mathbf{P}^*$, we can adjust (2) into an approximation of the original variables, in terms of the first r principal components only:

$$\mathbf{X} \approx \mathbf{P}^* \mathbf{W}^{*\prime} \tag{4}$$

where $\mathbf{W}^*$ is the $N \times r$ matrix whose r columns are given by the first r eigenvectors.

3.2 Forecasting using Principle Components

In the literature, forecasting using principal components is motivated by the work of e.g. Stock and Watson (2002), Bai and Ng (2002), and Huang and Lee (2009). In this paper, we take the classical principal component methodology of forecast as benchmark and it can be summarized as follows:

Let y_t be the scalar time series variable to be forecast and let X_t be a N-dimensional multiple time series of candidate predictors. It is assumed that (X_t, y_{t+h}) admit a factor model representation with r common latent factors F_t,

$$X_t = \Lambda F_t + e_t \tag{5}$$

where Λ is $N \times r$ and F_t is $r \times 1$. In equation (5), by applying the classical principal component methodology, the latent common factors $F = (F_1 F_2 \cdots F_T)'$ is solved by:

$$\widehat{F} = X \hat{\Lambda} / N \tag{6}$$

where N is the size of x_t, $X = (x_1 x_2 \cdots x_T)'$, and factor loadings $\hat{\Lambda}$ is set to $\sqrt{N}$ times the eigenvectors corresponding to the r largest eigenvalues of $\mathbf{X}'\mathbf{X}$, and,

$$y_{t+h} = \beta' F_t + \varepsilon_{t+h} \tag{7}$$

Therefore, the forecast can be constructed after $\hat{\beta}'$ is obtained by regression of y_t on $(1 \ \hat{F}'_{t-1})$ $(t = 1, 2, ..., T)$ and,

$$\hat{y}_{T+h} = (1 \quad F'_T)\hat{\beta}_T \tag{8}$$

In Stock and Watson (2002), by including observed variables in equation (7), they have obtained,

$$y_{t+h} = \beta' F_t + \gamma' \omega_t + \varepsilon_{t+h} \tag{9}$$

where e_t is a $N \times 1$ vector idiosyncratic disturbances, h is the forecast horizon, ω_t is a $m \times 1$ vector of observed variables (e.g., lags of y_t), that together with F_t are useful for forecasting y_{t+h}, and ε_{t+h} is the resulting forecast error. Data are available for $\{y_t, X_t, \omega_t\}_{t=1}^{T}$, and the goal is to forecast y_{T+h}. Additionally, we consider fixing h a priori at a value equals to 1 which refers to the one-step ahead predictions.

To construct forecasts of y_{T+h}, principal components of $\{X_t\}_{t=1}^{T}$ serve as estimates of the factors. For any given time T of forecast, we estimate the factors using all available observations till T. The estimated factors, together with ω_t, are then used in (9) to estimate the regression coefficients. The forecast is constructed as $\hat{y}_{T+h} = \hat{\beta}'\hat{F}_t + \hat{\gamma}'\omega_t$, where $\hat{\beta}$, $\hat{\gamma}$, and $\hat{F}_T$ are the estimated coefficients and factors. By recursion we obtain series of q out-of-sample forecasts of the equity premium. If the idiosyncratic disturbances e_t in (5) were crosssectionally independent and temporally iid, then (5) is the classic factor analysis model. However, in a macroeconomic forecasting application, these assumptions are unlikely to be satisfied. By the work of Stock and Watson (2002), under general conditions, the principal components of X_t are asymptotically consistent estimators of the true latent factors. Consistency requires that both N and $T \to \infty$. See Stock and Watson (2002) for more detailed explanations.

Unlike Stock and Watson (2002), this paper considers forecasting single time series using a finite number of predictor series based on Rapach et al. (2010), Welch and Goyal (2008) and seminar research report Ibisevic et al. (2011). This is more relevant for forecasting which is our main interest. In the literature, Stock and Watson (2004) for instance compared the out-of-sample performance of a dynamic factor model using a relatively small number of N

and several combination forecasts. They discover that the dynamic factor performs relatively poor and it might suffer from this small N. However, this finding merits further study.

3.3 Determination of the Number of Factors *(r)*

An important part to the principle component analysis is the correct specification of the number of factors r. In the literature, there are several methods for consistently estimating the number of factors. By the work of Bai and Ng (2002) for instance, they state that r should be determined that best captures the variations in X, using k observed informative factors before estimating the unobserved factor loadings. They shed light on this unresolved issue and they successfully constructed estimators of r based on the penalized versions of the minimized values of

$$V(k, F^k) = (NT)^{-1} \sum_{i=1}^{N} \sum_{t=1}^{T} (X_{i,t} - \lambda_i^{k'} F_t^k)^2 \tag{10}$$

the sum of squared residuals from regressions of X_i on the k factors for all i, F^k is the matrix of k factors. Note that a model with $k+1$ factors can fit no worse than a model with k factors, but efficiency is lost as more factor loadings are being estimated. Therefore, a loss function $V(k, F^k) + kg(N, T)$, where $g(N, T)$ is the penalty function for overhitting, can be used to estimate r. However, one remark of this method by Bai and Ng (2002) is it focuses on the information set X and the observed F^k, given by equation (5). This approach can be seen as the backline but it is not appropriate for forecasting in equation (7), which is our main interest and more relevant for forecasting. For this paper, we follow the information criteria used in Huang and Lee (2009) for which estimated number of factors k is selected by

$$\min_{1 \leq k \leq k_{max}} IC_k = ln(SSR(k)/T) + g(T)k \tag{11}$$

where $SSR(k)$ is the sum of squared residuals from the forecasting model in (7) for all k, $1 \leq k \leq k_{max}$ and k_{max} is set at 15. The penalty function is denoted by $g(T)$ which is $2/T$ for AIC and lnT/T for BIC based[1]. The determination of the number of factors r used for this paper is a dynamic approach and based on the expanding in-sample period. Therefore, we allow r varies over time. For example, suppose the sample period is 1947Q1 - 2005Q4, the out-of-sample starts from 1965Q1. The forecast on time 1970Q1 is based on the in-sample

[1]No remarkable differences are observed after obtaining the results between the two criteria, therefore we only report AIC in our results.

period till this moment which is 1947Q1 - 1969Q1 and for this in-sample period the information criteria AIC and BIC give $r = 2$ as optimum. Therefore the forecasting model with 2 principle components is used for the forecasting on time 1970Q1. For 1970Q2, we use observations 1947Q1 - 1970Q1 to determine r and AIC and BIC give in this case $r = 1$, therefore this forecasting model with 1 principle component is used for the forecasting on time 1970Q2. Proceeding in this manner through the end of the out-of-sample period, we generate a series of r. Allowing r varies over time means we acknowledge that the optimal r changes and this dynamic approach has the advantage that we do not exclude the possibility of different models providing good forecasts. Further, note that the selection of principle components applied in this paper and also in many more researches using PCA are based on the importance or the so called fraction of variance explained. In other words, when the number of factors r is determined, then the first r principle components are used for estimating and forecasting the equity premium while less important components are ignored. It is therefore questionable to exclude the possibility that minor components as 'good' predictors. In the literature of PCA, Qian et al. (1994) for instance proposed a principle component selection by the criterion of the minimum mean difference of complexity. This criterion considers the lost information due to the reduction of the parameters as well as the observed variables. They further point out the the principle components turn out to be identical to the classical principle components under the assumption of normality by minimization of a index. For more details, we refer to their paper.

4 Combination forecasts of integrated information

As mentioned before, the focus of this paper is incorporating information subsequently applied by different existing combining frameworks and therefore attempting to achieve gains in forecasting performance resulting from combination. Therefore in the spirit of improving the forecast combination by carefully integrating the information from the explanatory variables, this paper propose a different approach and can be intuitively explained as follows.

Suppose X_t be a N-dimensional multiple time series of candidate predictors, it is possible to divide the $N - dimensional$ X_t into k subsets of candidate predictors before extracting the principal components, consisting of $\{n_1, n_2, n_3, ..., n_k\}$ subsets and $n_1 + n_2 + n_3 + ... +$

$n_k = N$. By applying the same procedure as described in equation (5), (6), (7), (8) using principal components methodology, now on different subsets of candidate predictors, then different forecasts based on the corresponding subsets of candidate predictors can be obtained. Data set used for this paper consist 15 candidate predictors, that is $N = 15$. Dividing the whole set N for instance into 5 subsets n_1, n_2, n_3, n_4, n_5 where each subset consist of 3 candidate variables in order that 5 forecasts can be obtained and used for combination forecasts. We can rewrite equation (7) and (8), where i represents the ith principal components model

$$y_{i,t+h} = \beta_i' F_{i,t} + \varepsilon_{i,t+h} \tag{12}$$

where the latent common factors of each ith model $F = (F_1 F_2 \cdots F_T)'$ is solved by equation (6) applied by ith subset of X_t, and,

$$y_{i,T+h} = (1 \quad F_{i,T}')\beta_{i,T} \tag{13}$$

The information criteria AIC and BIC can be applied for specification of r in (12), as discussed in section 3.3. Again we determine the number of factors r dynamically and allow r varies over time, now on each principle component model based on different subsets using a expanding window of the in-sample.

The next step is to combine the principal component forecast obtained from each subsets of candidate variables to find optimal weights for every model i using various combining frameworks (section 5).

An unresolved issue is how to divide data set consisting of N candidate predictors into subsets. Suppose that M denotes the number of subsets, for $M = 2$, the first subset consists 8 candidate predictors and the second subset 7. For $M = 3$, each subset consists 5 candidate predictors and so on. In order to make the correct choice of choosing the best division of N, for this report we consider the economic relevance of the candidate predictors and in each subset the correlation between the variables. We are aware of the fact that there are many combinations possible and we are convinced of the existence of a better division than the chosen one for this report. In any event, it merits further study and research in the future.

The simple idea of incorporating and dividing information subsequently combining the forecast intuitively might work since this methodology pays both attention to which forecasts to include and how to weight them. The diversification benefits through combining forecasts is known and especially, this methodology attempts to incorporate relevant information from the observed variables at the same time.

5 Combining Frameworks

This section provides a quick summary of the various combination forecast methods used in this paper: simple combination forecasts; discounted MSFE forecasts. Combining forecasts is concerned with the objective to find optimal weights $w_{i,t}$ for every model i. Otherwise said, suppose there are N models and the prediction of model i for the stock returns in period $t+1$ is $\hat{r}_{i,t+1}$. This means that we need to find the optimal weights in:

$$\hat{r}_{c,t+1} = \sum_{i=1}^{N} w_{i,t} \cdot \hat{r}_{i,t+1} \tag{14}$$

such that $\hat{r}_{t+1}$ is a good estimator for r_{t+1}. To choose the weights, we consider simple averaging schemes (mean, median) and also discounted MSFE forecasts combination based on Stock and Watson (2004), where the combining weights formed at time t are functions of historical forecasting performance of the individual models over the holdout out-of-sample period. Moreover, addition to the usual setting of an in-sample and out-of-sample period, we also define a holdout period consisting of $q0$ observations from the end of the in-sample period. This means that the weights are based on previous predictive performance where models which performed more accurate get higher weights, their discount mean square prediction error (DMSPE) combining method employs the following weights, equivalent:

$$w_{i,t} = \frac{(\phi_{i,t})^{-1}}{\sum_{j=1}^{N} (\phi_{j,t})^{-1}} \tag{15}$$

$$\phi_{i,t} = \sum_{m}^{t-1} \theta^{t-1-m} (r_{m+1} - \hat{r}_{i,m+1})^2 \tag{16}$$

where θ is less than or equal to one. θ smaller than 1 means that more recent observations have more influence in calculating the weights. For this paper, we consider fixing θ a priori at a value equals to 0.9 [2].

[2]For more detailed description of the choice of theta, see Ibisevic et al. (2011) section 5

6 Forecast Evaluation

To evaluate our forecasts we exactly follow the forecast evaluation method as described by Rapach et al. (2010) and Ibisevic et al. (2011). In the following we summarize this approach. To evaluate the out-of-sample predictions we use the R^2_{OOS} suggested by Campbell and Thompson (2008). This statistic measures the mean squared prediction errors(MSPE) for both models and compares them with each other. The R^2_{OOS} is computed by:

$$R^2_{OOS} = 1 - \frac{\sum_{t=q_0+1}^{q}(r_t - \hat{r}_t)^2}{\sum_{t=q_0+1}^{q}(r_t - \overline{r}_t)^2} \tag{17}$$

We sum from q_0 to q (the out-of-sample without the training period). To test which model performs better, we compare the fitted value obtained through a combined method ($\hat{r}_t$) with the predictions of the alternative model, our benchmark ($\overline{r}_t$). This is equivalent to the following hypothesis:

$$H_0 : R^2_{OOS} \leq 0$$

Under the null hypothesis the combined method ($\hat{r}_t$) does not outperform the predictions of the alternative model ($\overline{r}_t$).

$$H_1 : R^2_{OOS} > 0$$

Under the alternative hypothesis the combined method ($\hat{r}_t$) outperforms the predictions of the alternative model ($\overline{r}_t$). The significance of the R^2_{OOS} is based on the out-of-sample MSPE-adjusted statistic that is developed for comparing forecasts of nested models (Clark and West, 2007). This is necessary as the statistic of a nested model follows a non-standard distribution (Clark and McCracken (2001) and McCracken (2007)). Before we calculate the MSPE-adjusted statistic we first need to define:

$$f_{t+1} = (r_{t+1} - \overline{r}_{t+1})^2 - [(r_{t+1} - \hat{r}_{t+1})^2 - (\overline{r}_{t+1} - \hat{r}_{t+1})^2] \tag{18}$$

After obtaining $\{f_{s+1}\}_1^{T-1}$ we regress the set on a constant to test whether we can reject the null hypothesis (meaning $H_1 : R^2_{OOS} > 0$ is true). When the R^2_{OOS} is significant larger than zero we still have to consider whether it is economically meaningful to use the combining method. In Campbell and Thompson (2008) is said that a small positive R^2_{OOS}, like 1% for quarterly data, has a markable increment of the annual portfolio returns for a mean-variance investor. According to Campbell and Thompson (2008), Welch and Goyal (2008), Marquering and Verbeek (2004) and Wachter and Warusawitharana (2009) the R^2_{OOS} does not explicitly account the risk of the investment over the out-of sample period. Like Rapach et al. (2010) state the investor may not acknowledge the increment of the returns when they account the risk that is involved. To cope for this we calculate the average utility for a mean-variance investor with a risk aversion parameter γ. Similar to Campbell and Thompson (2008) we forecast the mean and the variance of the stock returns using a rolling window. We use the predicted returns of the alternative model to obtain the proportion of stocks and risk-free bonds that a portfolio of the mean-variance investor should contain.

$$w_{0,t} = (\frac{1}{\gamma})(\frac{\overline{r}_{t+1}}{\hat{\sigma}^2_{t+1}}) \tag{19}$$

The $\hat{\sigma}^2_{t+1}$ is estimated with the rolling window of the variance of the stock and the portfolio weight ($w_{0,t}$) is constrained between 0%($w_{0,t} = 0$)and 150%($w_{0,t} = 1.5$). Thereafter the investor obtain an average utility level of:

$$\hat{v}_0 = \hat{\mu}_0 - (\frac{1}{2})\gamma\hat{\sigma}^2_0 \tag{20}$$

The $\hat{\mu}_0$ and $\hat{\sigma}^2_0$ are the sample mean and the variance of the return on the portfolio that is formed by using forecasts of the alternative model. After obtaining the utility of the mean-variance investor, we calculate the utility of the same investor that made use of the predicted returns of the combined method. His equity proportion equals:

$$w_{j,t} = (\frac{1}{\gamma})(\frac{\hat{r}_{t+1}}{\hat{\sigma}^2_{t+1}}) \tag{21}$$

and realizes an average utility level of:

$$\hat{v}_j = \hat{\mu}_0 - (\frac{1}{2})\gamma\hat{\sigma}^2_0 \tag{22}$$

Respectively, the $\hat{\mu}_0$ and $\hat{\sigma}^2_0$ are the sample mean and the variance of the return on the portfolio that is formed by using forecasts of the combined method. We can interpret the utility gain as a fee that the investor is willing to pay to get additional information. We take the difference of the utilities (22) - (20) and multiply it by 400 to express it in annualized percentage return and we report our results with a risk aversion(γ) of 3 like in Rapach et al. (2010).

7 Data

In this section we provide a quick glim of the variables used throughout the report. We consider variables from Welch and Goyal (2008) which quarterly data are available (1947Q1

- 2008Q4) as candidate predictors for predicting the equity premium. In this paper, some variables were transformed by taking the natural logarithms. Note that there are more data transformation possible for the given data (e.g. interpolating of outliers, adjusting seasonality, taking first differences). We only take the natural logarithms and season adjustments as data transformation for comparison purposes since this is also applied in the seminar report by Ibisevic et al. (2011). We analyze the sample period 1947Q1 till 2008Q4 with four out-of-sample periods starting in 1965Q1, 1976Q1, 2000Q1 till 2005Q4 and the last one 2005Q1 till 2008Q4.

7.1 Assets

- Real risk free rate, Real Rf; log(1+Rf) - log(1+Infl): with Rf = riskfree rate, with Infl = inflation. This is the difference between log of risk free rate and the log of inflation.
- Excess stock return, ExStock; log(1+Rs) - log(1+Rf): with Rs = stock return. This is the difference between log of stock return and the log of risk free rate. Stock returns are measured as continuously compounded returns on the S&P500 index, including dividends.
- Excess bond return, ExBond; log(1+Rb) - log(1+Rf): with Rb = long-term government bond return. This is the difference between the log of long-term government bond return and the log of risk free rate.

7.2 Candidate Predictors

- Dividend-price ratio, log(D/P): with D = dividend and P = price of S&P500 Index. This is the difference between the log of dividends (12 month moving sums) paid on the S&P500 index and log of stock prices (S&P500).
- Dividend yield, log(D/Y): with Y = lagged Price of S&P500 Index. This is the difference between the log of dividends and log of lagged stock prices.
- Earnings-price ratio, log(E/P): with E = earnings. This is the difference between the log of earnings and the log of prices.
- Dividend-payout ratio, log(D/E): This is the difference between the log of dividends and log of earnings.

- Stock variance, log(SVAR): This is the log of sum of squared daily returns on the S&P500 index.

- Book-to-market ratio, log(B/M): with B = book value of Dow Jones and M = Market value of Dow Jones. This is the log of ratio of book value to market value for the Dow Jones Industrial Average.

- Net equity expansion, NTIS: This is the ratio of 12 month moving sums of net issues by NYSE-listed stocks to total end-of-year market capitalization of NYSE stocks.

- Treasure bill rate, log(1+TBL): with TBL = annualized interest on 3-month Treasure bill rate.

- Long-term yield, log(1+LTY): with LTY = annualized long-term government bond yield.

- Long-term return, log(1+LTR): with LTR = long-term government bond return

- Term spread, log(1+LTY) - log(1+TBL): This is difference between the long term yield on government bonds and the Treasury-bill.

- Default yield spread, DFY; log(1+BAA) log(1+AAA): with BAA = annualized yield on BAA rated bonds and AAA = annualized yield on AAA rated bonds. This is the difference between BAA- and AAA-rated corporate bond yields.

- Default return spread, DFR; log(1+LTCR) - log(1+LTR): with LTCR = long-term corporate bond return. This is the difference between long-term corporate bond and long-term government bond returns.

- Inflation, log(1+INFL): Since inflation is released in the following month, note that it is lagged by one period.

- Investment-to-capital ratio, I/K: This is the ratio of aggregate investment to aggregate capital for the entire economy.

8 Empirical Results

This section examines the empirical performance of the principle component approach and the combination forecasts constructed using the 15 candidate predictors. Moreover, we begin by briefly summarizing the forecasting performance of the classical principle component methodology of whole data set. Then we proceed discussing the combination forecasting performance of the principle component models corresponding to the subsets of the whole data set. The focus will lay on the out-of-sample performances and the comparison between the different approaches evaluated by using two benchmarks: (i) Historical average, Table 1 and (ii) The mean of the panel of forecasts obtained from the individual predictors, Table 2.

We first discuss the principle component model based on the whole data set consisting of 15 well-known economic variables from the predictability literature as candidate predictors. As briefly mentioned before, we consider this as the basic model and we want to investigate whether incorporating information using PCA approach can achieve gains in forecasting. Before reporting the results of the four out-of-sample periods we first explain the determination of the number of factors r. As discussed in section 3.3, we allow r varies over time and this dynamic r takes a value between $1, 2, ..., 15$. For any given time T of forecast, we determine the optimal r using all available observations till T and by recursion we have obtained a series of r for this single PCA model. However, for this paper we consider a priori letting r varies between small values like $1, 2, 3$. According to the results of Stock and Watson (2002), they report that nearly all of the forecasting gains come from the first two or three factor. The results suggest that the first three principle component are more than adequate for the PCA factor model.

Table 1 and Table 2 report the results of the four out-of-sample periods where different PCA models performances are compared with the historical average benchmark and combined mean respectively. The number of PCA models is denoted by M, therefore our basic PCA model using whole data set corresponds with $M = 1$. For $M = 2$, the whole data set is divided into two subsets and therefore two principle component forecasts are then obtained and combined through simple combinations. For $M = 3$, three principle component models based on three subsets are used and so on. Furthermore, as discussed in section 4, for the choice of choosing the best division of N we consider in this report the economic relevance of the candidate predictors and in each subset the correlation between the variables. For instance when we consider 4 principle component models based on 4 subsets, then the division is as follows: (i) DP, DY, BM and EP. (ii) DE, DFY, DFR and INFL. (iii) IK, NTIS, TBL and LTY. (iv) LTR, TMS and SVAR. DP, For instance in (i), DY and EP are valuation ratios and all three are highly correlated with BM. For (ii), DFY and DFR are corporate bond returns and correlated with inflation. We further discover that in (iii), TBL and LTY are highly related and in (iv), LTR and TMS are provided from the same source, the *Ibbotsons Stocks, Bonds, Bills and Inflation Yearbook*. Note that there are many combinations possible (2^{15}), therefore a better division of the 15 variables is highly possible. Furthermore, the chosen criterium selecting the division depends on the forecaster. For different whole set of explanatory variables, then different criterium can be applied. However, the focus of this paper is to achieve possible gains in dividing whole set into subsets. In any event, further study is necessary. For this paper, the divisions for all M are presented in section 10.

Table 1 an 2 also report the number of factors used for forecasting. In case of $r \leq 2$, we allow r varies between 1 and 2 and for $r \leq 3$, r varies between the first three components. Figure 2 for instance illustrates the varying r for $M = 1$ model and out-of-sample starts from 1965. Not surprisingly, most of the times $r = 1$ is chosen resulting from the AIC criterium and especially during the start period of the out-of-sample 1965-1970 and also during the 90's. We further discover that r varies more during relatively instable periods, like in the 70's and 80's, (Large negative peak around 1975 caused by the oil crisis followed by the Saving and Loan crisis and the Black Monday that occurs in the 80's).

8.1 Integrating Benefits of Forecasting using PCA

Turn to Table 1 and 2 for $M = 1$, the basic model, several results emerge. First, for three out four out-of-sample periods our basic model outperform the historical average. Evidently forecasting benefits can be obtained by incorporating information through PCA. Second, we notice that a varying r does not necessarily perform better than a static factor regression, for instance $r = 1$. Thus, on the one hand a dynamic approach for determining r has the advantage that we do not restrict the model for a varying optimal r through time, on the other it has the disadvantage that there exists a possibility of a loss of efficiency and higher uncertainty due to the time-varying parameter. Third, although many of the improvements of the basic model are substantial relative to the historical benchmark, the gains are modest in case of combined mean benchmark as reported in Table 2. Fourth, when the number of forecasts is small relative to the sample size, as in the case of the third and fourth out-of-sample period, significant integrating benefits of forecasting are obtained by forecasting using PCA on the total information set. Fifth, PCA approach based on the estimated factors of total data sets are significantly improved compared to the predictions of low-dimensional individual forecasting regressions, as reported in Rapach et al. (2010).

8.2 Combination Forecast of Integrated Information

We next shed light on the combination forecasting performance of the principle component models associated with the subsets of the whole information set. The results are reported in Table 1 and 2 for $M = 2, 3, 4, 5, 7$, and for each M three simple combinations methods, denoted as cMean, cMedian and cPhi are used for combining the forecasts of subsets. Note that the same forecasting procedure is applied for each sub principle model. Therefore, the number of factors r are also dynamically determined for each sub model before forecasting and combining. For instance, for any T two optimal r have to be determined in case of 2 sub principle component models and 7 in case of $M = 7$.

Table 1 and 2 show several notable results for $M > 1$ models. First, we notice that for all four out-of-sample periods $M > 1$ models significantly outperform the historical average in terms of R^2_{OOS} statistics. One of the striking improvement emerges for instance by evaluating the performances of the second out-of-sample period, where cMedian for $M = 5$ or cMedian for $M = 7$ significantly outperform the historical benchmark at 10% level. This remarkable result can be stressed by the findings of Rapach et al. (2010) and Welch and Goyal (2008), where they prove that the out-of-sample predictive ability of many individual economic variables deteriorates over the $1976Q1 - 2005Q4$ out-of-sample period. Therefore, the results suggest that using common factors can provide robustness against temporal instability which is a problem to many simple regressions of individual macroeconomic predictors.

Second, as already mentioned before, dynamic factor regression does not necessarily outperform a static factor regression. For $M > 1$ models, it is interesting to question whether the performance increases when the number of models increases. However, at the same time, the number of time-varying parameters r increases as well. Moreover, on the one hand we account diversification benefits, and on the other higher possibility of a loss of efficiency. It turns out that in case of increasing M, the forecasts have quite variable performance. For the first and third out-of-sample period, small M models perform slightly better dan higher M models. Turning to the results of relatively instable out-of-sample period $1976Q1 - 2005Q4$ and $2005Q4-2008Q4$, we see that higher M models clearly perform better. The results suggest that during these periods a dynamic approach might be more capable to guard against possible instabilities than a static factor regression, and this outcome generally confirms the findings of a dynamic parameter reported in Ibisevic et al. (2011).

Third, many of the improvements of the $M > 1$ models are of a significant margin relative to the historical benchmark, in some cases even at a 5% level. However, in case of combined mean as benchmark, the results also yielding potential gains although less substantial.

8.3 Comparison between $M = 1$ and $M > 1$ models

We next compare the results of section 8.1 and 8.2. It is of interest to question whether improvement of forecasts performance can be gained from combining, and if this is indeed the case, then we might also examine the forecasting environment. First, we find that all $M > 1$ models perform better in the first and second out-of-sample periods than $M = 1$ model while for relatively small out-of-sample periods the opposites holds. The results suggest that when the in-sample size is small related to the total sample size, possible gains can be obtained resulting from combining. As mentioned in the introduction, Timmerman (2006) states that estimation errors that contaminate the combination weights can emerge when the sample size is small relative to the number of forecasts. Therefore, our results does not confirm Timmerman (2006). However, to make a comparison between the two different findings here is questionable since the forecasting performance are evaluated by taking the historical average or combined mean as benchmark. Moreover, the results are not only related to the number of forecasts but also the periods where the results are evaluated. Therefore, it is desirable to study the characteristics of the different time periods. In any event, this merits further studies.

Proceeding discussing the characteristics of time periods, we notice secondly in the second out-of-sample period a clearly difference in forecasting performance between $M = 1$ and $M > 1$ models. As already mentioned, $M = 5$ and $M = 7$ perform well during this period while $M = 1$ based on the whole information performance poor in terms of R^2_{OOS} statistics. Although forecasting performance by incorporating information from PCA approach is improved compared to the individual forecasts, it still does not succeed in outperform the historical average. Moreover, instability can be 'average out' through the construction of common factors, but $M = 1$ still suffers from this that plagues the individual predictors. This is illustrate in figure 3, where the betas are presented over time corresponding to the first principle component for $M = 1 and r = 1$ model, starting from 1965 out-of-sample period.

In turn, combining the forecasts can still reduce the instability after the incorporating of information using PCA. In summary, this outcome suggest that in case when the out-of-sample period is relatively instable, then gains in forecasting performance resulting form combination by incorporating information using PCA can be obtained. Hence, by putting attention to both deciding which forecasts to include and how to weight them and by linking both, one can possibly obtain improved out-of-sample results.

9 Conclusion

Combining forecasts is widely used for dealing with the misspecification biases that affect individual forecasting models. Since the beginning of forecast combination, many studies can be found for assigning weights to the included models and relatively little attention is paid to which forecasts need to be included. Model instability has come to be an issue since individual models may be biased. This paper aims to achieve gains in forecasting performance resulting from combination by incorporating information of the economic variables. An important approach of incorporating information from economic variables is factor analysis, involves extracting a relatively small number of common factors from a large number of variables and using these factors in a single forecasting model. We also show that the correct specification of the number of factors r has come to be a important part of the research and by allowing r varies over time means we acknowledge that the optimal r changes. This dynamic approach has the advantage that we do not exclude the possibility of different models providing good forecasts.

Next to the Principle Component Analysis (PCA) we also propose a new approach by dividing the whole set of candidate predictors into subsets before extracting the principal components. This is called combination forecast of integrated information. The idea of incorporating and dividing information subsequently combining the forecast is in the spirit of this paper since this methodology pays both attention to which forecasts to include and how to weight them.

The empirical analysis in this paper yields 3 main conclusions. First, the principle component model based on the whole data set consisting of 15 candidate predictors perform well. The results suggest that forecasting gains can be obtained by incorporating information through PCA. Note that here the first three principle components are adequate for the PCA factor model. The results further point out that a varying r does not necessarily perform better than a static factor regression. On the one hand a dynamic approach for determining r has the advantage that we do not restrict the model for a varying optimal r through time, on the other it has the disadvantage that there exists a possibility of a loss of efficiency and higher uncertainty due to the time-varying parameter.

Second, turn to the combination forecasting performance of the principle component models associated with the subsets of the whole information set, a notable finding is that using common factors can provide robustness against temporal instability which is a problem to many simple regressions of individual macroeconomic predictors. Furthermore, the results suggest that during relatively instable periods a dynamic approach might be more capable to guard against possible instabilities than a static factor regression. We show that in case of higher number of sub principle models, forecasting gains increase resulting from higher number of dynamic r. However, there is some room for improvement for this approach by using different divisions of N candidate predictors into subsets. We are aware of the fact that there are many combinations possible (2^{15}) therefore it merits further study and research in the future.

Third, during relatively instable out-of-sample periods, although forecasting performance by incorporating information from PCA approach is improved compared to the individual forecasts, it still does not succeed in outperform the historical average. Moreover, instability can be 'average out' through the construction of common factors, but $M = 1$ still suffers from this that plagues the individual predictors. However, the results suggest that by combining the forecasts obtained from the sub principle component models can still reduce the instability after the incorporating of information using PCA. Hence, by putting attention to both deciding which forecasts to include and how to weight them and by linking both, one can possibly obtain improved out-of-sample results.

10 Tables

Table 1 to Table 4 present the equity premium out-of-sample forecasting results for several PCA models, where for each table there are 4 Panels A to D corresponding to the 4 different out-of-sample periods. R^2_{OOS} is the Campbell and Thompson (2008) out-of-sample R^2 statistic in column (2) or (7). Utility gain Δ is the portfolio management fee (in annualized percentage return) that an investor with mean-variance preferences and risk aversion coefficient of three would be willing to pay to have access in column (3) or (8) to the forecasting model given in column (1) or (6) (relative to the historical average (combination mean) benchmark forecasting model. Statistical significance for the R^2_{OOS} statistic is based on the p-value in column (5) or (10) for the Clark and West (2007) out-of-sample $MSPE$-adjusted statistic. The divisions of subsets are as follows: For $M = 2$: (i) DP, DY, BM, EP, DE, DFY, DFR, INFL. (ii) IK, NTIS, TBL, LTY, LTR, TMS, SVAR. For $M = 3$: (i) DP, DY, BM, EP, DE. (ii) DFY, DFR, INFL, IK, NTIS. (iii) TBL, LTY, LTR, TMS, SVAR. For $M = 4$: (i) DP, DY, BM, EP. (ii) DE, DFY, DFR, INFL. (iii) IK, NTIS, TBL, LTY. (iv) LTR, TMS, SVAR. For $M = 5$: (i) DP, DY, BM. (ii) EP, DE, DFY. (iii) DFR, INFL, IK. (iv) NTIS, TBL, LTY. (v) LTR, TMS, SVAR. For $M = 7$: (i) DP, DY. (ii) BM, EP. (iii) DE, DFY. (iv) DFR, INFL. (v) IK, NTIS. (vi) TBL, LTY. (vii) LTR, TMS, SVAR.

Table 1: Principle Component models with historical average as benchmark

	R^2_{OOS}(%)	Δ(%)	*t-stat*	*p-value*		R^2_{OOS}(%)	Δ(%)	*t-stat*	*p-value*
A. Out-of-sample period: 1965 Quarter 1 - 2005 Quarter 4									
PCA $M = 1$					*PCA $M = 4$ and $r \leq 2$*				
$r = 1$	3.21	3.86	2.19	0.04	*cMean*	4.20	3.10	2.08	0.05
$r \leq 2$	4.86	3.66	2.29	0.18	*cMedian*	3.22	2.54	1.82	0.08
$r \leq 3$	3.79	3.66	2.27	0.14	*cPhi*	4.06	3.05	2.01	0.05
PCA $M = 2$ and $r \leq 3$					*PCA $M = 5$ and $r \leq 2$*				
cMean	4.55	3.69	2.30	0.03	*cMean*	5.26	3.59	2.30	0.03
cMedian	4.55	3.69	2.30	0.03	*cMedian*	5.49	3.36	2.45	0.02
cPhi	4.08	3.66	2.15	0.04	*cPhi*	5.22	3.54	2.24	0.03
PCA $M = 3$ and $r \leq 3$					*PCA $M = 7$ and $r \leq 2$*				
cMean	3.87	3.28	2.01	0.05	*cMean*	4.05	3.30	2.23	0.03
cMedian	5.86	3.37	2.44	0.02	*cMedian*	4.30	3.33	2.40	0.02
cPhi	3.75	3.11	1.95	0.06	*cPhi*	3.87	3.19	2.11	0.04
B. Out-of-sample period: 1976 Quarter 1 - 2005 Quarter 4									
PCA $M = 1$					*PCA $M = 4$ and $r \leq 2$*				
$r = 1$	-1.50	2.21	0.71	0.31	*cMean*	-0.24	0.91	1.03	0.23
$r \leq 2$	-2.11	1.74	0.95	0.39	*cMedian*	-1.03	0.28	0.73	0.31
$r \leq 3$	-3.66	1.76	1.00	0.37	*cPhi*	-0.22	0.92	1.04	0.23
PCA $M = 2$ and $r \leq 3$					*PCA $M = 5$ and $r \leq 2$*				
cMean	-0.64	1.76	1.19	0.20	*cMean*	1.52	1.70	1.47	0.14
cMedian	-0.64	1.76	1.19	0.20	*cMedian*	2.10	1.50	1.67	0.09
cPhi	-0.80	1.79	1.13	0.21	*cPhi*	1.53	1.68	1.48	0.13
PCA $M = 3$ and $r \leq 3$					*PCA $M = 7$ and $r \leq 2$*				
cMean	-1.09	1.26	0.96	0.25	*cMean*	0.86	1.31	1.33	0.17
cMedian	-2.45	0.77	0.53	0.35	*cMedian*	2.05	1.48	1.70	0.09
cPhi	-1.20	1.27	0.91	0.26	*cPhi*	0.77	1.25	1.30	0.17

	R^2_{OOS}(%)	*Δ(%)*	*t-stat*	*p-value*		R^2_{OOS}(%)	*Δ(%)*	*t-stat*	*p-value*
C. Out-of-sample period: 2000 Quarter 1 - 2005 Quarter 4									
PCA M = 1					*PCA M = 4 and r ≤ 2*				
r = 1	8.33	14.71	1.30	0.17	*cMean*	3.82	4.61	1.15	0.21
r ≤ 2	9.85	7.81	1.39	0.13	*cMedian*	2.35	2.62	0.91	0.26
r ≤ 3	9.45	7.83	1.34	0.13	*cPhi*	4.03	4.85	1.18	0.20
PCA M = 2 and r ≤ 3					*PCA M = 5 and r ≤ 2*				
cMean	4.47	5.78	1.01	0.24	PC_Mean	3.38	4.12	0.95	0.25
cMedian	4.47	5.78	1.01	0.24	PC_Medi	3.87	3.41	1.17	0.20
cPhi	4.71	5.88	1.03	0.24	PC_Phi	3.69	4.42	0.99	0.24
PCA M = 3 and r ≤ 3					*PCA M = 7 and r ≤ 2*				
cMean	2.62	4.07	0.97	0.25	*PC_Mean*	3.22	4.15	1.09	0.22
cMedian	1.08	1.35	0.66	0.32	*PC_Medi*	3.19	3.65	0.96	0.25
cPhi	2.89	4.32	1.01	0.24	*PC_Phi*	3.33	4.31	1.10	0.22
D. Out-of-sample period: 2005 Quarter 1 - 2008 Quarter 4									
PCA M = 1					*PCA M = 4 and r ≤ 2*				
r = 1	10.65	2.26	1.58	0.12	*cMean*	1.16	0.05	0.7	0.31
r ≤ 2	8.81	1.21	1.34	0.15	*cMedian*	1.09	0.05	0.77	0.30
r ≤ 3	7.34	1.21	1.31	0.17	*cPhi*	1.03	0.07	0.72	0.31
PCA M = 2 and r ≤ 3					*PCA M = 5 and r ≤ 2*				
cMean	5.91	0.48	1.15	0.21	*cMean*	7.33	0.94	1.39	0.15
cMedian	5.91	0.48	1.15	0.21	*cMedian*	12.33	1.87	1.40	0.15
cPhi	5.75	0.48	1.15	0.21	*cPhi*	7.47	1.01	1.42	0.15
PCA M = 3 and r ≤ 3					*PCA M = 7 and r ≤ 2*				
cMean	9.11	1.23	1.48	0.13	*PC_Mean*	4.23	0.51	1.28	0.18
cMedian	11.53	1.78	1.41	0.15	*PC_Medi*	4.90	0.45	0.90	0.27
cPhi	9.18	1.28	1.49	0.13	*PC_Phi*	4.24	0.55	1.29	0.17

Table 2: Principle Component models with combining mean as benchmark

	$R^2_{OOS}(\%)$	*Δ(%)*	*t-stat*	*p-value*		$R^2_{OOS}(\%)$	*Δ(%)*	*t-stat*	*p-value*
A. Out-of-sample period: 1965 Quarter 1 - 2005 Quarter 4									
PCA M = 1					*PCA M = 4 and r ≤ 2*				
r = 1	-0.34	0.26	1.02	0.24	*cMean*	0.69	0.57	1.14	0.21
r ≤ 2	1.38	1.13	1.59	0.38	*cMedian*	-0.33	0.01	0.77	0.30
r ≤ 3	0.27	1.13	1.54	0.36	*cPhi*	0.54	0.52	1.06	0.23
PCA M = 2 and r ≤ 3					*PCA M = 5 and r ≤ 2*				
cMean	1.06	1.16	1.48	0.13	*cMean*	1.80	1.06	1.44	0.14
cMedian	1.06	1.16	1.48	0.13	*cMedian*	2.03	0.83	1.58	0.11
cPhi	0.64	2.17	0.99	0.25	*cPhi*	1.75	1.01	1.39	0.15
PCA M = 3 and r ≤ 3					*PCA M = 7 and r ≤ 2*				
cMean	0.35	0.75	1.14	0.21	*cMean*	0.54	0.77	1.11	0.22
cMedian	2.41	0.84	1.75	0.09	*cMedian*	0.80	0.80	1.30	0.17
cPhi	0.23	0.58	1.08	0.22	*cPhi*	0.35	0.66	0.98	0.25
B. Out-of-sample period: 1976 Quarter 1 - 2005 Quarter 4									
PCA M = 1					*PCA M = 4 and r ≤ 2*				
r = 1	-2.55	0.64	0.49	0.35	*cMean*	-1.28	0.06	0.61	0.33
r ≤ 2	-3.17	0.89	0.65	0.38	*cMedian*	-2.08	-0.57	0.19	0.39
r ≤ 3	-4.74	0.91	0.66	0.39	*cPhi*	-1.26	0.07	0.60	0.33
PCA M = 2 and r ≤ 3					*PCA M = 5 and r ≤ 2*				
cMean	-1.69	0.91	0.86	0.28	*cMean*	0.50	0.84	1.17	0.20
cMedian	-1.69	0.91	0.86	0.28	*cMedian*	1.08	0.65	1.42	0.15
cPhi	-1.84	0.94	0.80	0.29	*cPhi*	0.51	0.83	1.17	0.20
PCA M = 3 and r ≤ 3					*PCA M = 7 and r ≤ 2*				
cMean	-2.14	0.41	0.57	0.34	*cMean*	-0.17	0.45	0.98	0.25
cMedian	-3.52	-0.08	0.17	0.39	*cMedian*	1.03	0.63	1.54	0.12
cPhi	-2.25	0.41	0.53	0.35	*cPhi*	-0.26	0.40	0.93	0.26

	R^2_{OOS}(%)	Δ(%)	*t-stat*	*p-value*		R^2_{OOS}(%)	Δ(%)	*t-stat*	*p-value*
C. Out-of-sample period: 2000 Quarter 1 - 2005 Quarter 4									
PCA M = 1					*PCA M = 4 and r ≤ 2*				
r = 1	5.64	4.96	1.03	0.24	*cMean*	1.00	1.90	0.56	0.34
r ≤ 2	7.20	5.10	1.18	0.17	*cMedian*	-0.52	-0.09	0.08	0.40
r ≤ 3	6.80	5.12	1.13	0.18	*cPhi*	1.21	2.14	0.62	0.33
PCA M = 2 and r ≤ 3					*PCA M = 5 and r ≤ 2*				
cMean	1.67	3.07	0.60	0.33	*cMean*	0.54	1.41	0.41	0.37
cMedian	1.67	3.07	0.60	0.33	*cMedian*	1.06	0.70	0.63	0.33
cPhi	1.92	3.17	0.63	0.33	*cPhi*	0.87	1.71	0.49	0.35
PCA M = 3 and r ≤ 3					*PCA M = 7 and r ≤ 2*				
cMean	-0.23	1.36	0.24	0.39	*cMean*	0.39	1.44	0.37	0.37
cMedian	-1.82	-1.36	-0.31	0.38	*cMedian*	0.35	0.94	0.28	0.38
cPhi	0.04	1.61	0.33	0.38	*cPhi*	0.50	1.59	0.41	0.37
D. Out-of-sample period: 2005 Quarter 1 - 2008 Quarter 4									
PCA M = 1					*PCA M = 4 and r ≤ 2*				
r = 1	9.68	1.19	1.46	0.14	*cMean*	-0.27	-0.18	0.08	0.40
r ≤ 2	7.82	1.25	1.18	0.20	*cMedian*	-0.95	-0.31	-0.20	0.39
r ≤ 3	6.34	1.25	1.15	0.22	*cPhi*	-0.25	-0.16	0.16	0.39
PCA M = 2 and r ≤ 3					*PCA M = 5 and r ≤ 2*				
cMean	4.89	0.51	1.04	0.23	*cMean*	6.32	0.98	1.42	0.15
cMedian	4.89	0.51	1.04	0.23	*cMedian*	11.38	1.91	1.40	0.15
cPhi	4.73	0.52	1.04	0.23	*cPhi*	6.47	1.04	1.45	0.14
PCA M = 3 and r ≤ 3					*PCA M = 7 and r ≤ 2*				
cMean	8.12	1.26	1.46	0.14	*cMean*	3.39	0.43	1.49	0.13
cMedian	10.57	1.82	1.39	0.15	*cMedian*	3.80	0.45	0.95	0.26
cPhi	8.20	1.31	1.46	0.14	*cPhi*	3.33	0.47	1.49	0.13

Table 3: Correlation matrix of candidate predictors 1965 Quarter 1 - 2005 Quarter 4

	DP	DY	BM	EP	DE	DFY	DFR	INFL	IK	NTIS	TBL	LTY	LTR	TMS	SVAR
DP	1.00	0.98	0.90	0.88	0.29	0.48	0.02	0.51	-0.19	-0.02	0.66	0.68	0.02	-0.18	-0.11
DY	0.98	1.00	0.88	0.86	0.30	0.52	0.05	0.50	-0.21	-0.05	0.64	0.67	0.06	-0.16	-0.18
BM	0.90	0.88	1.00	0.90	0.04	0.56	0.01	0.63	-0.03	0.03	0.67	0.63	-0.01	-0.28	-0.08
EP	0.88	0.86	0.90	1.00	-0.19	0.44	-0.03	0.63	0.00	-0.17	0.74	0.68	0.00	-0.32	-0.10
DE	0.29	0.30	0.04	-0.19	1.00	0.10	0.09	-0.22	-0.42	0.29	-0.15	0.01	0.03	0.29	-0.06
DFY	0.48	0.52	0.56	0.44	0.10	1.00	0.09	0.35	-0.33	-0.20	0.47	0.65	0.28	0.15	0.14
DFR	0.02	0.05	0.01	-0.03	0.09	0.09	1.00	-0.15	-0.14	0.05	-0.09	0.04	-0.43	0.21	-0.07
INFL	0.51	0.50	0.63	0.63	-0.22	0.35	-0.15	1.00	0.11	0.01	0.51	0.39	0.07	-0.35	0.03
IK	-0.19	-0.21	-0.03	0.00	-0.42	-0.33	-0.14	0.11	1.00	-0.06	0.23	-0.14	-0.13	-0.65	0.00
NTIS	-0.02	-0.05	0.03	-0.17	0.29	-0.20	0.05	0.01	-0.06	1.00	-0.23	-0.26	-0.21	0.02	-0.07
TBL	0.66	0.64	0.67	0.74	-0.15	0.47	-0.09	0.51	0.23	-0.23	1.00	0.83	-0.10	-0.55	-0.02
LTY	0.68	0.67	0.63	0.68	0.01	0.65	0.04	0.39	-0.14	-0.26	0.83	1.00	0.00	0.00	0.03
LTR	0.02	0.06	-0.01	0.00	0.03	0.28	-0.43	0.07	-0.13	-0.21	-0.10	0.00	1.00	0.17	0.16
TMS	-0.18	-0.16	-0.28	-0.32	0.29	0.15	0.21	-0.35	-0.65	0.02	-0.55	0.00	0.17	1.00	0.08
SVAR	-0.11	-0.18	-0.08	-0.10	-0.06	0.14	-0.07	0.03	0.00	-0.07	-0.02	0.03	0.16	0.08	1.00

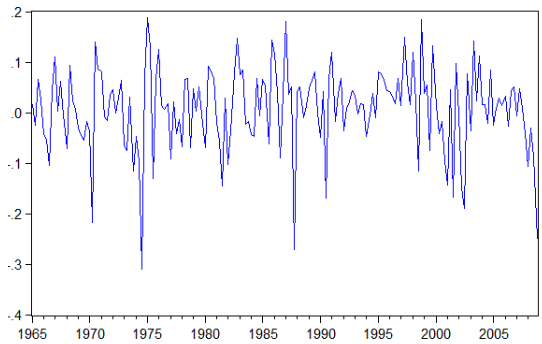

Figure 1: *Returns of exStock (sample 1965Q1-2008Q4)*

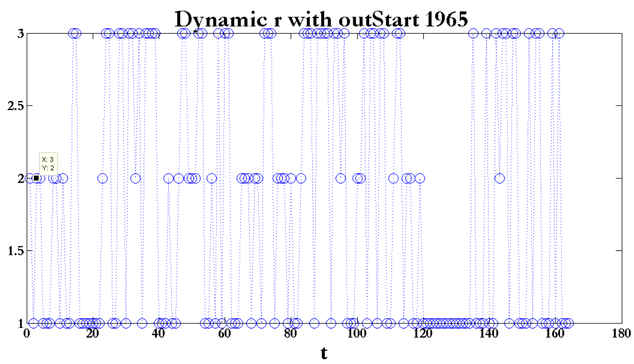

Figure 2: *Number of factors; Dynamic r (sample 1965Q1-2005Q4)*

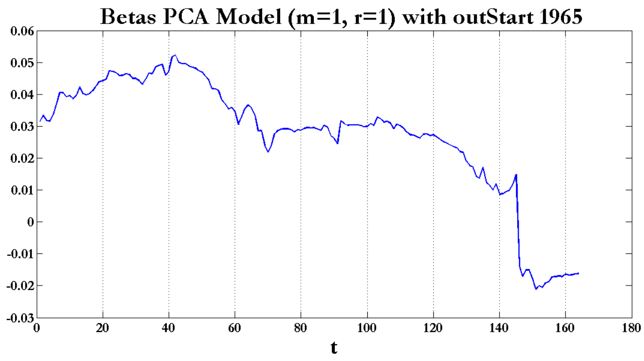

Figure 3: *Betas PCA Model (sample 1965Q1-2005Q4)*

References

M. Aiolfi, C. Capistran, and A. Timmermann. Forecast combination. *Banco de Mexico, Working Papers*, (4), 2010.

C. Alexander. *Market Risk Analysis Volume II.* John Wiley and Sons Ltd, 2008.

J. Bai and S. Ng. Determining the numbetr of factors in approximate factor models. *Econometrica*, 70, 2002.

J.M. Bates and C.W.J. Granger. The combination of forecasts. *Operational Research Quarterly*, 20:451–468, 1969.

John Y. Campbell and Samuel B. Thompson. Predicting the equity premium out of sample: can anything beat the historical average? *Review of the Financial Studies*, 21:1509–31, 2008.

Todd E. Clark and Michael W. McCracken. Tests of equal forecast accuracy and encompassing for nested models. *Journal of Econometrics*, 105:85–110, 2001.

Todd E. Clark and Kenneth D. West. Approximately normal tests for equal predictive accuaracy in nested models. *Journal of econometrics*, 138:291–311, 2007.

R. Clemen and R. Winkler. Combining economic forecasts. *Journal of Business and Economic Statistics*, 4, 1986.

Robert T. Clemen. Combining forecasts: A review and annotated bibliography. *International Journal of Forecasting*, 5(4):559–583, 1989.

F. Diebold. Forecast combination and encompassing: Reconciling two divergent literatures. *International journal of forecasting*, 5, 1989.

F. Diebold and P. Pauly. The use of prior information in forecast combination. *International journal of forecasting*, 6, 1990.

D. Hendry and M. Clements. Pooling of forecasts. *Econometrics Journal*, 5, 2002.

H. Huang and T. Lee. To combine forecasts or to combine information. *Econometric Reviews*, 29(5-6):534–570, 2009.

S. Ibisevic, K. Lai, M.A. Snijder, and W. Zhang. Combining forecasts: Improved approach and analysis of various frameworks. April 2011.

S. Ludvigson and S. Ng. The empirical risk-return relation: a factor analysis approach. *Financial Economics*, pages 171–222, 2007.

S. Makridakis. Why combining works? *International Journal of Forecasting*, 5, 1989.

W. Marquering and M. Verbeek. The economic value of predicting stock index returns and volatility. *Journal of Financial and Quantitative Analysis*, 39:407429, 2004.

Michael W. McCracken. Asymptotics for out of sample test of granger causality. *journal of econometrics*, 140:719–752, 2007.

G. Qian, G. Gabor, and R.P. Gupta. Principle components selection by the criterion of the minimum mean difference of complexity. *Journal of Multivariate Analysis*, 49, 1994.

David E. Rapach, Jack K. Strauss, and Guofu Zhou. Out-of-sample equity premium prediction: Combination forecasts and links to the real economy. *The Review of Financial Studies*, 23 (2):821–862, 2010.

J.H. Stock and M.W. Watson. A comparison of linear and nonlinear univariate models for forecasting macroeconomic time series. *NBER Working Paper*, 6607, 1998.

J.H. Stock and M.W. Watson. Forcasting using principle components from a large number of predictors. *Journal of the American Statistical Association*, 97(460):1167–1179, 2002.

J.H. Stock and M.W. Watson. Combination forecasts of output growth in a seven-country data set. *Journal of Economic Literature*, 23:405–430, 2004.

Allan Timmerman. *Handbook of Economic Forecasting*, chapter 4, pages 135–196. Elsevier B.V., 2006.

J.A. Wachter and M. Warusawitharana. Predictable returns and asset allocation: Should a skeptical investor time the market? *Journal of Econometrics*, 148:162178, 2009.

Ivo Welch and Amit Goyal. A comprehensive look at the empirical performance of equity premium prediction. *The Review of Financial Studies*, 21(4):1455–1508, 2008.