Forecasting Bond Risk Premia with Forecast Combinations using Many Predictors

MASTER OF SCIENCE THESIS

For the degree of Master of Science in Quantitative Finance at
Erasmus University Rotterdam

Author
D.W.B. Bender
347544

Academic supervisor
prof.dr. D. van Dijk

Co-reader
dr. B.F. Diris

November 17, 2015

Erasmus School of Economics (ESE) · Erasmus University Rotterdam

Abstract

In this report the application of forecast combinations is examined for construction of excess bond return forecasts using a large set of predictor variables. The analysis is done by applying a large number of forecast combination methods to forecast one-year U.S. government excess bond returns for maturities ranging from two to five years. For every maturity 753 forecast sets are constructed using different forecast combinations methods, predictor variables and cluster methods. These are then compared using the model confidence set procedure. The forecast combination methods are subsequently subjected to a significance test and to a test over different time periods. The conclusion that this report draws is that two sets of forecast sets outperform the other forecast sets as well as the historical average benchmark and achieve a higher out-of-sample R^2 then was previously found in the literature. These are the forecast sets that are constructed when either the recursive OLS weighting scheme or the complete subset regression method is applied to the set of forecast sets which are constructed with the macroeconomic predictors and which are first clustered according to their economic background. These two forecast combination methods continue to outperform the others for every maturity and also if the time period is changed.

Keywords: Forecast combinations - Excess bond returns - Large datasets

Table of Contents

1. Introduction

Forecasting bond returns by linking them to macroeconomic fundamentals is a subject which has gained more attention from the academic community over the past three decades. There are numerous papers regarding the subject, which mostly attempt to obtain factors from the macroeconomic predictor variables or use financial variables, such as forward rates and yield spreads, to construct bond return forecasts.

Bonds are used for many different reasons such as long-term planning and diversification as well as saving and managing interest rate risk. So if it were possible to construct accurate forecasts of the excess bond returns, that would help enormously when timing the best moment to buy bonds and deciding whether to sell the bond prior to its maturity date. This would not only be of interest to individual investors but also to pension funds and monetary and fiscal policy makers.

Technological advances over the last few decades have led to impressive gains not only in computational power, but also in the quantity of available financial and macroeconomic data. This provides the opportunity to exploit a much richer base of information than was conventionally possible. But this abundance of information also poses new challenges. With many predictors, high dimensionality and high estimation error can become a problem.

One specific method of dealing with a large number of predictors has been of particular interest in the literature for almost half a century. Combining different forecasts to construct a new forecast, a method known as 'Forecast Combinations', was first proposed by Bates and Granger [1969]. The advantage of this method is that it is not necessary to use a complex model which attempts to construct forecasts from a large set of predictor variables, which can be both time-consuming and computationally difficult. Instead a simple model is used to construct a separate forecast for each predictor variable, and these forecasts are then combined, using a certain weighting scheme, to construct a single forecast.

Papers that have examined the use of large sets of macroeconomic variables for forecasting bond returns have focussed on extracting factors from these variables which have subsequently been used to construct forecasts. This has been done with models such as principal components analysis by Ludvigson and Ng [2009] or a two-step adaptive group lasso procedure by Huang and Shi [2010]. But constructing individual forecasts from each of these variables and combining these into a single bond return forecast is a subject which has not been investigated so far. This paper studies the application of forecast combinations to construct forecasts of one-year U.S. government excess bond returns. A large set of macroeconomic and financial predictor variables each use a linear regression model to construct out-of-sample forecast sets of the excess bond returns. Different methods of forecast combinations are then applied to these forecast sets. To examine the differences in information contained in the macroeconomic and financial predictor variables, forecast sets are constructed using both of the predictors sets separately as well as simultaneously.
Even though forecast combinations is proposed as a solution to the dimensionality problems that arise when forecasts are constructed from large sets of predictor variables, the number of forecasts that need to be combined is still too large for some of the models to deal with. Using the 'clustering' method, the number of forecasts to be combined can be reduced significantly. This was first proposed by Timmermann [2006], who stated that the method is motivated by the assumption of a common factor structure underlying the forcasting models. The idea is that before the forecast sets are combined, they are divided into clusters based on specific criteria. The forecast sets in each cluster are then combined using equal weights to create a single forecast set for each cluster. As a result the number of forecast sets is reduced and the weighting schemes are applied to the clustered forecast sets.

In this report the forecast sets are clustered according to two different criteria: performance and category. To determine clusters based on performance, the k-means algorithm is applied. This method was proposed by Aiolfi and Timmermann [2006] and is based on the past mean squared forecast error (MSFE) performance of the forecast sets. By applying this method it is possible to combine the forecast sets that are similar and this can change the results of simple weighting schemes as well as make it possible to apply more complex weighting schemes. The assignment of clusters in this case varies over time as it is updated for each observation. The category cluster method is based solely on the macroeconomic and financial properties of the predictor variables, and as a result is time-invariant.

Three predictor variables sets, macroeconomic, financial and both macroeconomic and financial, are used to construct three sets of forecast sets, and three methods, i.e. the no cluster method, the category cluster method and the performance cluster method, were applied to each of these three sets. This creates nine different sets of forecast sets. Twelve forecast combinations methods are applied to these nine forecast sets. These methods have been proposed by Bates and Granger [1969], Stock and Watson [2006] and Elliott et al. [2013]. In addition the effect of shrinkage towards equal weights, as suggested by Diebold and Pauly [1990], is investigated by applying it to a selection of the forecast combination methods. This results in a final total of 753 combined forecast sets for each maturity.

In order to discover which of these forecast combination methods constructs the most accurate forecast sets, they all need to be compared for each maturity, and then evaluated to determine which forecast set(s) perform(s) the best. Given the number of forecast sets this is no trivial matter. However the model confidence set procedure, introduced by Hansen [2011], provides a solution. A model confidence set is a set of models which is constructed such that, for a given significance level, it will contain the *best* model(s) of the original set of models. The model confidence set procedure is applied to forecast sets of the excess bond returns which are constructed in this paper. For each maturity, a model confidence set is determined at a significance level of 95%.

The main conclusion of this report is that two methods of forecast combinations, namely the recursive OLS weighting scheme and the complete subset regression method, are able to construct forecasts of one-year U.S. government excess bond returns, which consistently outperform both the historical average benchmark and the other methods of forecast combinations. The recursive OLS weighting scheme is able to construct the best performing forecast sets when either 25% or 50% shrinkage towards equal weights is applied, which indicates that adding shrinkage towards equal weights to weighting schemes can improve their performance. The complete subset regression method produces the best performing forecast sets when the number of predictor variables, k, that are used is either the maximum number that are available or close to that maximum. Both methods construct these forecast sets when they use the macroeconomic predictor variables. These forecast sets have a higher out-of-sample R^2 (R^2_{OS}), which is a measure that was first proposed by Campbell and Thompson [2008], then has been previously reported in the literature. The main conclusion is based on the result that out of a total of 753 different forecast sets, these forecast sets significantly outperform the historical mean benchmark as well as the other forecast sets consistently for each maturity. The model confidence set procedure produces sets which almost exclusively contain these forecast sets. When the forecast combination methods are subjected to robustness checks, these two forecast combination methods continue to construct forecast sets that outperform the benchmark and the other forecast sets.

The results of all the other forecast sets are very similar to each other in terms of MSFE. Differences in the parameters levels and the addition of different levels of shrinkage seem to make very little difference. For most of those forecast methods even different sets of predictor variables make only a slight difference to the results, which indicates that these forecast combination methods are not succesful in extracting useful information from the predictor variables.

In Huang and Shi [2010] forecasts are constructed of the same excess bond returns for the period 1985 till 2007. The resulting R^2_{OS} were between 0.36 and 0.39 for the 5 maturities. In the case of Cieslak and Povala [2011], the results varied in the range of 0.20 to 0.24 for the time period 1971-2009 for their forecasts which are constructed using a cycle-related factor. Finally Cooper and Priestley [2009] construct forecasts for the period 1965-2003 using the output gap. They report R^2_{OS} between 0.20 and 0.30. All these forecasts were constructed from the same excess bond returns with maturities ranging from 2 to 5 years, and the same measure for performance was used, namely the R^2_{OS}. These are also the same excess bond returns and out-of-sample measure that are used in this report. The forecasts sets that are included in the model confidence sets for each maturity have an R^2_{OS} in the range from 0.34 to 0.71 for the time period 1975 until 2011. The time spans are different but they all cover at least the period from 1985 to 2003. Comparing the results, it can be concluded that the forecasts obtained in this report perform better than those found in the literature. The comparison between the results of Huang and Shi [2010] and the results in this paper is especially interesting, since the same macroeconomic predictor variables are used in both cases. Additionally this report examines the effect financial variables, but in the case where just the macroeconomic variables are used, the best performing methods result in an R^2_{OS} that is up to 0.35 higher than than those found by Huang and Shi [2010].

A second conclusion that can be drawn from the results in this report is that the macroeconomic predictor variables provide useful information when constructing excess bond return forecasts. By contrast, the sets of financial predictor variables do not produce useful information when constructing forecasts, either with solely these predictors, or in conjunction with the macroeconomic predictor variables. The forecast sets that are constructed with the macroeconomic predictor variables easily outperform those that were constructed with the financial predictor variables, and the forecast sets that were constructed with both sets of predictor variables are, at best, a worse version of those constructed with solely the macroeconomic variables. The fact that a large set of macroeconomic variables is able to produce accurate forecasts of bond returns is in line with what has previously been stated in the literature. However several papers have also stated that there is valuable information, and even complementary information to macroeconomic variables, in financial variables. This is in contrast to the conclusion of this report; however this report examined the information that could be extracted from a large set of financial variables, whereas the other papers only used a small amount of financial variables to predict bond returns.

The clustering of the forecast sets according to two cluster methods also produced results from which a conclusion can be drawn. This is that clustering the forecast sets according to their past performance does not significantly improve the performance of the resulting combined forecast sets, and the same is true for most of the forecast sets that were clustered according to the category method. However there are a number of combined forecast sets that are constructed using the category cluster method whose performance is significantly better than that of the other combined forecast sets. At each maturity the final model confidence set consists solely of these combined forecast sets that were clustered by category. It is possible that predictor variables that fall in the same category contain overlapping information and by combining these, this overlap can be reduced.

The rest of the paper is constructed as follows: Section 2 presents an overview of the related literature. Section 3 describes the sets of predictor variables and how the excess bond return data is constructed. Section 4 further examines the forecast combination methods and the model confidence set procedure and gives an overview of the combined forecast sets of the excess bond returns. Section 5 first presents the results of the forecast sets as well as the results of the model confidence set procedures and compares them for the different maturities. Secondly the forecast combination methods are subjected to two robustness analyses, one on the significance level of the model confidence set and one on the period of the observations. And finally Section 6 presents the conclusions.

2. Literature

In this Section first the findings of different articles on the use of macroeconomic variables and financial variables for the forecasting of bond premia are summarised, and secondly a brief overview is given of a number of articles on the subject of forecast combinations. Literature about the specific methodology that is used in this report can be found the Section 4.

In the literature the use of macroeconomic variables to predict bond returns has already been quite successful. Several papers report high in- and R^2_{OS} results when applying various methods to macroeconomic variables. The application of financial variables for predicting bond returns has received slightly less attention is the past, but has still also been shown to deliver positive results.

Cochrane and Piazzesi [2005] construct a new financial variable which is able to forecast excess bond returns across a range of maturities. They investigate one-year excess returns of n-year U.S. government bonds and run regressions of bond returns at time $t+1$ on five forward rates, with different maturities, at time t. As prices, yields and forward rates are linear functions of each other, the forecasts are the same for each of these variables. They find that a linear combination of five forward spreads explains a significant amount of the variation in the following year's excess returns on bonds with maturities from two to five years. Based on this linear combination they construct their single-factor model. The excess bond returns that Cochrane and Piazzesi [2005] first calculated have become a constant element in the literature on forecasting bond premia. Practically every paper on the subject attempts to forecast these excess bond returns for 2- to 5- year maturities.

A single macroeconomic variable that has been shown to be able to produce accurate forecasts of bond returns is the output gap. By examining this single production-based macroeconomic variable, Cooper and Priestley [2009] exclude both the level of asset prices and also any consumption data, and are able to assess whether this specific business cycle variable is able to predict bond returns and is thus an important predictor of bond risk premia. They examine the quality of the forecasts by using the R^2_{OS}, which was proposed by Campbell and Thompson [2008] and find R^2_{OS} ranging from 0.25-0.31 for the period January 1975 - December 2003. From these results they conclude that the notion of bond return predictability is a rational response to changing business conditions rather than market inefficiency. Cieslak and Povala [2011] decompose long-term yields into a persistent component and maturity-related cycles. They apply predictive regressions to the one-year excess bond returns on a common factor constructed from these cycles. They determine the R^2_{OS} for different time periods, 1971-2009 and 1985-2009, which ranges from 0.18 to 0.37.
Both of these papers illustrate the fact that macroeconomic variables contain information which can be used to construct forecasts of bond returns. By using a very small number of these variables they are able to construct forecasts which perform well when measured by the R^2_{OS}.

Other papers focus on the possibility of extracting factors from large sets of macroeconomic variables which can be used to construct forecasts of excess bond returns. Ludvigson and Ng [2009] investigate whether macroeconomic sources contain information for bond predictability using a dynamic factor analysis on a set of 132 monthly macroeconomic variables. They obtain a macroeconomic factor and construct forecasts of the excess bond returns. They apply a 'simple' principal components analysis to extract factors and reduce the dimensionality. For these forecasts they find an R^2_{OS} of 26% for the two-year bond. They also add the Cochrane and Piazzesi [2005]'s single factor model, which henceforth will be referred to as CP_t, to their model with macroeconomic variables. The result is an R^2_{OS} of 44% for the same

bond. This is a 13% improvement over just using the CP_t and an 18% improvement over the dynamic factor model. This implies that the macroeconomic variables contain information which is complementary to CP_t. Huang and Shi [2010] use a two-step adaptive group lasso procedure to construct excess bond return forecasts and use the same macroeconomic predictor variables as Ludvigson and Ng [2009]. They group 21 of these macroeconomic predictors together, based on four factors, corresponding to employment, housing, financial and inflation factors. These four factors are summed and this is the single macroeconomic factor which they refer to as the SAFLasso factor. They find that this factor can predict excess bond returns on 2- to 5- year maturity bond with an R^2_{OS} of 0.36-0.39 for the period January 1985 - December 2007. They conclude that their SAGLasso factor contains information about future excess bond returns beyond that which is covered by either the CP_t or the Ludvigson and Ng [2009] factors.
Ludvigson and Ng [2009] and Huang and Shi [2010] both conclude that factors can be extracted from large sets of macroeconomic variables which are able to produce forecasts of excess bond returns that perform well when measured with the R^2_{OS}. The first also illustrates that using both macroeconomic and financial variables to construct a forecast can yield better results than using either of them seperately. Results from Huang and Shi [2010] indicate that clustering predictors together according to their economic background can lead to factors that are able to contruct accurate forecasts.

The literature on using large sets of financial variables to predict (excess) bond returns is lacking. However Ludvigson and Ng [2007] do use a large set of financial predictors to make predictions about the stock market. They use the financial predictors to forecast one-quarter-ahead excess stock market returns and volatility. They conclude that three new factors; 'volatility', 'risk premium' and 'real', contain information which more commonly used predictor variables lack.

The method of forecast combinations was first proposed by Bates and Granger [1969]. They concluded that a combination of separate simple forecasts could yield lower MSFE results than either individual forecasts. Subsequent empirical studies have found that the combined forecasts are able to produce better forecasts than other individual forecasts. Timmermann

[2006] continues on this subject and investigates different methods of determining weights for combining forecasts. He concludes that simple combinations, which ignore correlations between forecast errors, often perform better than more complex weighting schemes which are designed to determine theoretically optimal weights. Elliott et al. [2013] introduce another method of combining forecasts. They propose combining the forecasts, for a given set of predictor variables, that are constructed from all possible linear regression models for a certain number of predictors. This is the complete subset regression method. They examine the trade-off between model complexity and model fit, and conclude that the subsets can give more accurate forecasts than standard equally weighted forecast combinations. They apply their method to a set of twelve possible predictors and also discuss the potential computational issues if the number of predictors becomes large. Elliott et al. [2015] analyse the complete subset regression method for situations where the number of predictors is large relative to the sample size. They conclude that it can offer a favorable bias-variance trade-off in the presence of many weak predictor variables. In addition they conclude that the complete subset regression method can construct out-of-sample point forecasts of U.S. employment, GDP growth and inflation which are more accurate than those constructed by a dynamic factor approach or univariate regressions, which do not exploit the information contained in the cross-section of the predictors. Many forecast combination methods have been researched in the literature and they have shown good results when dealing with large numbers of predictors.

In the literature, the results regarding bond premia indicate that there is indeed information in macroceconomic variables which can be used to construct accurate forecasts of bond returns. The use of financial variables to produce forecasts of bond returns has been done by, for instance, Cochrane and Piazzesi [2005], with their single factor model, but using large sets of financial variables to predict bond returns has not yet been attempted. The literature on forecast combinations argues that when forecasting with large sets of predictors, forecast combination methods are able to construct forecasts, even when the number of predictors is large relative to the sample, which are more accurate than individual forecasts. However using forecast combination methods to forecast (excess) bond returns is a subject that has not yet been attempted in the literature.

3. Data

Before discussing the methods used in this paper, the one-year excess bond returns and the predictor variable sets will be examined more closely.

3-1 One-year Excess Bond Returns

The U.S. government bond return data is taken from the Fama-Bliss dataset which is available from the Center for Research in Security Prices (CRSP). The dataset contains monthly 1- through 5-year zero-coupon U.S. Treasury bond prices and this is obtained for the period that spans from January 1964 till December 2011.

The one-year excess bond returns are then constructed by borrowing at the one-year rate, buying a long-term bond and selling it one year later. This is done as follows:

$$p_t^{(n)} = \text{log price of n-year discount bond at time } t. \tag{3-1}$$

where n is given in years and indicates the maturity and t is given in months. The log yield then is the defined as follows:

$$y_t^{(n)} \equiv -\frac{1}{n} p_t^{(n)} \tag{3-2}$$

The one-year bond return from buying a n-year bond at time t and selling it at a year later at time $t+12$ is defined as:

$$r_{t+12}^{(n)} \equiv p_{t+12}^{(n-1)} - p_t^{(n)} \tag{3-3}$$

The one-year excess bond return is then determined as:

$$rx_{t+12}^{(n)} \equiv r_{t+12}^{(n)} - y_t^{(1)} \tag{3-4}$$

This is applied to the entire Fama-Bliss dataset and this results in four sets of one-year excess bond returns; $rx^{(2)}$, $rx^{(3)}$, $rx^{(4)}$ and $rx^{(5)}$, these correspond to the 2, 3, 4 and 5 year maturity respectively.

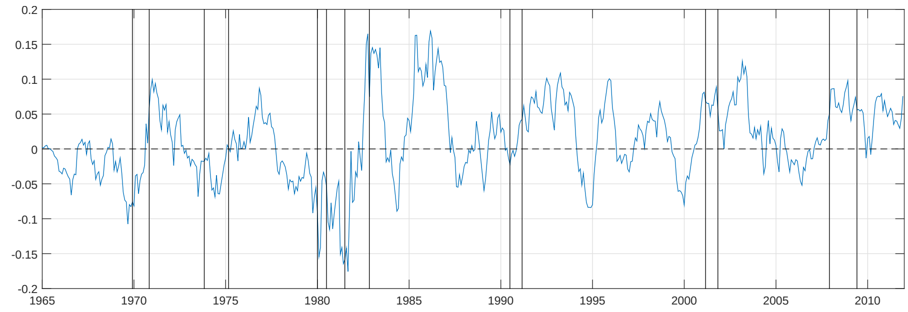
3-1-1 Properties

First the excess bond returns for the 5 year maturity will be discussed in this Section. Second these results will be compared to those for the 2, 3 and 4 year maturity. The complete results of these can be found in Appendix A.

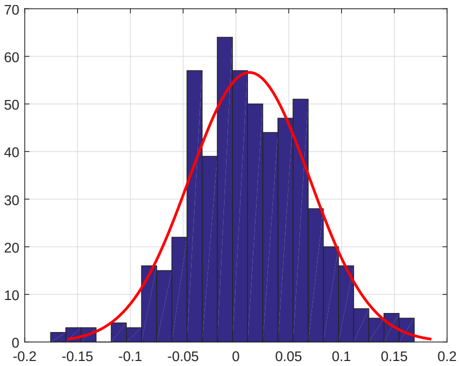
Figure 3-1a shows the constructed set of one-year excess bond returns for the 5 year maturity. The bond returns are high at troughs and low at peaks, which indicate that they have a business cycle pattern.

Figure 3-1b shows the histogram of the same bond returns. When comparing it to a normal distribution it seems to fit quite well. This suggests that the bond returns might not reject the null hypothesis of normality.

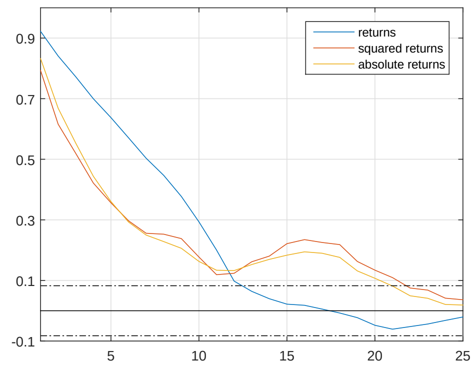
In figure 3-1c the autocorrelations for the sets of returns, squared returns and absolute returns are presented. It is clear the returns show high autocorrelations which slowly decline. It is logical that there is strong autocorrelation for the first twelve lags, as these lags are in months and excess bond returns are constructed with bond prices which are a year apart. The squared and absolute returns also start high but decline slightly faster, until they rise again to a peak around 16 and start declining again. This indicates that volatility clustering occurs in the return sets.



(a) One-year excess bond returns. The vertical lines represent the peaks and troughs of the US Business cycle.[1]



(b) Histograms of the returns



(c) Autocorrelations of returns, squared returns and absolute returns

Figure 3-1: Properties of the one-year excess bond returns for the 5 year maturity

The excess bond returns for the 5 year maturity are also compared to those of the 2, 3 and 4 year maturity. These three sets show very similar results to the 5 year maturity, though an increased bond maturity results in higher absolute excess returns; but this would seem logical. Table 3-1 presents the cross-correlation coefficients between each of the excess bond return sets. The cross-correlation is very high, between each of the maturities. When the difference in maturity is one year the lowest correlation is 0.982. Even when the largest difference in maturities is taken, that between 2 and 5 years, the correlation is still 0.943. Table 3-2 presents the statistics each of the sets of bond returns. As was suspected for the 5 year maturity, the skewness is very small and the kurtosis is very close to three. The Jarque-Bera statistic as a result is 2.10, which does not reject the null hypothesis of normality. The

[1] Data of the business cycle was obtained from the National Bureau of Economic Research, http://www.nber.org/cycles/cyclesmain.html

Table 3-1: Cross-correlation coefficients of the one-year excess bond return sets

	$rx^{(2)}$	$rx^{(3)}$	$rx^{(4)}$	$rx^{(5)}$
$rx^{(2)}$	1.000	-	-	-
$rx^{(3)}$	0.982	1.000	-	-
$rx^{(4)}$	0.962	0.990	1.000	-
$rx^{(5)}$	0.943	0.979	0.994	1.000

results for the other three sets are similar, and for each of them the null hypothesis is not rejected. The properties of the excess bond returns for each of the bond return sets are not

Table 3-2: Statistics of the one-year excess bond return sets

	$rx^{(2)}$	$rx^{(3)}$	$rx^{(4)}$	$rx^{(5)}$
Mean	0.005	0.009	0.012	0.013
Standard deviation	0.018	0.033	0.046	0.057
Skewness	0.02	-0.05	-0.01	-0.02
Kurtosis	3.31	3.37	3.32	3.30
Jarque-Bera	2.29	3.57	2.36	2.10

completely in line with the general 'stylized facts' of asset returns. These include non-normal distributions, no significant autocorrelations and volatility clustering. Each of the return sets however shows a normal distribution and high and slowly declining autocorrelations. The autocorrelations in the squared and absolute returns do indicate that there are periods of large returns which alternate with periods of small returns, which is characterised as volatility clustering. The forecast models that are designed in this paper must be able to capture the normal distribution of the excess returns as well as their autocorrelations and the volatility clustering.

3-2 Predictor Variables

Two different sets of predictor variables will be used to construct forecast sets, a macroeconomic and a financial set. As was discussed in Section 2, in the literature there are articles which have succesfully been able to link macroecomic variables to the predictability of bond

returns. Therefore the 132 macroeconomic predictor variables that are provided by James Stock and Mark Watson, who use the data from Stock and Watson [2004], are used in this report as the first set of predictor variables.
Additionally some articles report that combining financial factors such as the CP_t factor with factors constructed from macroeconomic variables, can produce better forecasts than either of them seperately. To test whether financial variables contain complementary information to the macroeconomic variables, the set of 147 financial predictor variables, which are obtained from Ludvigson and Ng [2007], are used as the second set of predictors in this report. All of these data series are available for the period spanning January 1960 till December 2011[2]. A detailed overview of all the data and their transformations can be found in Appendix A. All the raw data of the predictor variables are standardized before they are used for estimation.

The two sets of predictor variables can each be categorised into smaller sets. The macroeconomic variables are categorised into 14 categories and the financial variables into 4 categories. An overview of these categories and amount of predictor variables in each is given in Table 3-3.
In the financial categories, the factors include the three risk factors from Fama and French [1993], namely the excess return on the market, the small-minus-big and high-minus-low factors. It also includes the momentum factor, the bond risk premia factor of Cochrane and Piazzesi [2005] and the consumption-wealth variable of Lettau and Ludvigson [2001]. The category Size/BM portfolios is composed of 93 returns on stock portfolios sorted by size and book-to-market ratio, a method which was proposed by Fama and French [1992].

[2]Both datasets are obtained from the homepage of Sydney Ludvigson http://www.econ.nyu.edu/user/ludvigsons/

Table 3-3: Overview of the categories for the macroeconomic and financial variables

Macroeconomic variables		
Category	Description	# of predictors
Out	Real Output & Income	18
EMP	Employment & Hours	30
RTS	Real Retail, Manufactering & Trade Sales	2
Mon	Money & Credit Quantity Aggregates	11
PCE	Consumption	1
HSS	Housing Starts & Sales	10
Inv	Real Inventories & Inventory-sales Ratio	3
Ord	Orders & Unfilled Orders	6
AHE	Average Hourly Earnings	3
Pri	Price Indexes	21
Int	Interest Rates and Spreads	17
SPr	Stock Prices	4
FX	Exchange Rates	5
Oth	Miscellaneous	1
	Total	132
Financial variables		
Category	Description	# of predictors
PYD	Prices, Yields & Dividends	5
RF	Risk Factors	6
I	Industries	43
S&BM	Size/BM portfolios	93
	Total	147

4. Methods

In this section the method of forecast combination is first described in more detail. Subsequently the sets of forecast sets that are constructed from the predictor variables are described. Following this the weighting schemes and forecast combinations methods used in this report are described. Finally the two methods of comparing the combined forecast sets are described, namely the model confidence sets procedure and the R^2_{OS}.

4-1 Forecast Combinations

Forecast combinations construct a forecast from a combination of two or more different forecasts. The procedure for a model with N predictors and which uses a single regressor is as follows:

1. Take a standard regression model with a single regressor $x_{i,t}$

$$y_{t+h} = \alpha_i + \beta_i x_{i,t} + \varepsilon_{i,t} \tag{4-1}$$

   where h is the forecast horizon.

2. Construct forecast $\hat{y}_{i,\tau+h}$ using:

$$\hat{y}_{i,\tau+h} = \hat{\alpha}_i + \hat{\beta}_i x_{i,\tau} \tag{4-2}$$

where $\hat{\alpha}_i$ and $\hat{\beta}_i$ are constructed by regressing $\{y_t\}_{t=\tau+h-1-rw}^{\tau+h-1}$ on $\{x_{i,t}\}_{t=\tau-1-rw}^{\tau-1}$, where rw is the size of some rolling window.

3. Step 2 is repeated for $i = 1, ..., N$

4. The combined forecast for $t = \tau + h$ is constructed as:

$$\hat{y}^c_{\tau+h} = \sum_{i=1}^{N} \omega_i \hat{y}_{i,\tau+h}, \tag{4-3}$$

for certain weights ω_i.

It is also possible to construct $\hat{y}_{i,\tau}$ with multiple regressors, let the weights vary over time or even to include forecasts which are not obtained from regression models, but from other sources such as surveys or expert opinions.

If a simple weighting scheme such as equal weights is applied, it is straightforward to combine a large number of forecasts. In that case it is a matter of estimating individual forecasts based on a single predictor (as in Equation 4-2) or a set of predictors and subsequently weighing these according to the simple scheme to obtain a single prediction. It is however a different case when the weights are constructed by some other more complex weighting scheme. Complications can arise when attempting to construct these weights when the number of predictors is large relative to the number of observations.

For each maturity this report constructs a forecast set of the one-year excess bond returns with each of the predictor variables that have been discussed in the previous Section. These forecast sets are then combined using both simple and complex weighting schemes. All the forecasts that are constructed and examined have a fixed forecast horizon of twelve, i.e. $h = 12$, which corresponds to one-year ahead forecasts. Two cluster methods are applied to the forecast sets to reduce the amount of forecast sets before they are combined, which makes it possible to apply more complex weighting schemes. The construction of the forecast sets and the application of the cluster methods is decribed in Section 4-2 and the forecast combination methods are discussed in Sections 4-3 and 4-4.

4-2 Forecast Sets

4-2-1 Individual

The total number of predictor variables is 279, 132 macroeconomic and 147 financial, and is denoted as N. The predictor variables are denoted as $x_{i,t}$, where $i = 1, ..., N$, $t = 1, ..., T$ and T is the number of observations for which a forecast set is constructed. With each predictor variable a set of out-of-sample forecasts is constructed. To do this, first a standard forecast regression model is taken:

$$rx_{t+12}^{(n)} = \alpha_i^{(n)} + \beta_i^{(n)} x_{i,t} + \varepsilon_{i,t} \tag{4-4}$$

where n denotes the maturity. The out-of-sample forecasts of the one-year excess bond returns are constructed using a recursive rolling estimation window. A rolling window of 120 monthly observations is taken. The first out-of-sample forecast for predictor $x_{i,t}$ is given by:

$$\widehat{rx}_{m+13}^{(n),ind} = \hat{\alpha}_{i,m+1}^{(n)} + \hat{\beta}_{i,m+1}^{(n)} x_{i,m+1} \tag{4-5}$$

where $\hat{\alpha}_{i,m+1}^{(n)}$ and $\hat{\beta}_{i,m+1}^{(n)}$ are the OLS estimates of $\alpha_i^{(n)}$ and $\beta_i^{(n)}$ respectively, which are constructed by regressing $\{rx_t^{(n)}\}_{t=13}^{m+12}$ on $\{x_{i,t}\}_{t=1}^{m}$. The second out-of-sample forecasts is then given by:

$$\widehat{rx}_{m+14}^{(n),ind} = \hat{\alpha}_{i,m+2}^{(n)} + \hat{\beta}_{i,m+2}^{(n)} x_{i,m+2} \tag{4-6}$$

where $\hat{\alpha}_{i,m+2}^{(n)}$ and $\hat{\beta}_{i,m+2}^{(n)}$ are estimated by regressing $\{rx_t^{(n),}\}_{t=14}^{m+13}$ on $\{x_{i,t}\}_{t=2}^{m+1}$. Continuing in this manner throughout the whole sample results in a out-of-sample forecast set based on $x_{i,t}$ and containing forecasts $\{\widehat{rx}_{i,t+1}^{(n),ind}\}_{t=m+12}^{T}$.

The result of this is N forecast sets at every maturity, one for each predictor variable. These forecast sets are denoted as $\widehat{rx}_i^{(n),ind}$, where $i = 1, ..., N$ and $n = 2, ..., 5$.

The sets of forecast sets that are constructed from the macroeconomic and financial predictors contain 132 and 147 forecast sets respectively. This report examines the results of applying forecast combininations to each of these sets seperately, but also the results when forecast combinations are applied to the set containing them both. The latter set contains all 279 forecast sets. These three sets of forecast sets are referred to as the individual forecast sets.

Several of the weighting schemes that are applied are not able to combine such large numbers of forecast sets. Therefore the forecast sets are clustered and combined, such that each cluster has a single forecast set, before the weighting schemes are applied. This reduces the number of forecast sets to which the weighting schemes are applied.
Two cluster methods are used in this report. The first of these starts by defining categories which are defined by some economic property. The forecast sets are then assigned to a category based on the economic property of the predictor variable with which it was constructed. The forecast sets in each cluster are then combined and this results in a new clustered forecast set for each of these categories. This cluster method is referred to as the category cluster method. A reason for using categories as a cluster method is that when using a set of predictor variables, it is likely that predictors which have similar economic backgrounds contain similar information and by clustering those that fall in the same category before the combined forecasts are constructed this overlap in information can be filtered.
The second cluster method that is applied is assigning forecast sets to clusters according to their past MSFE and is referred to as the performance cluster method. The assumption that is made is that the forecast sets that are very similar will produce similar MSFEs, and by clustering these forecast sets accordingly, the amount of forecast sets can be reduced greatly without discarding too much information.

4-2-2 Category

Constructing the clusters by category is done by categorizing the predictor variables based on their economic backgrounds, which have been discussed in Section 3, and combining the corresponding forecast sets. These category clusters are time-invariant.

The macroeconomic predictor variables are divided into 14 clusters and the financial predictor variables into 4 clusters. The sets $\widehat{rx}_i^{(n),ind}$ that are in cluster j are denoted by $\widehat{rx}_{l,j}^{(n)}$, where $l = 1, ..., N_j$ and N_j is the number of forecast sets in cluster j. The combined forecast set of cluster j is denoted by $\widehat{rx}_j^{(n),cat}$ and is constructed by taking the equally weighted combination of the forecast sets within the cluster, as follows:

$$\widehat{rx}_j^{(n),cat} = \frac{1}{N_j} \sum_{l=1}^{N_j} \widehat{rx}_{l,j}^{(n)} \tag{4-7}$$

The forecast combination methods are applied to the set of forecast sets constructed from each of the macroeconomic and financial predictor variables, as well as to the set containing them both. For each of the three sets the result is a new set. These contain 14, 4 and 18 forecast sets respectively, which are significantly less than the individual forecast sets, and are referred to as the category forecast sets.

4-2-3 Performance

The second cluster method that is applied in this report is clusters based on past performance. Constructing these clusters is slightly more complicated than constructing them by category. The performance of each forecast set, $\widehat{rx}_i^{(n),ind}$, is measured with a recursive rolling window. The performance of each forecast set varies over time and therefore the performance clusters have to be updated at each observation. To estimate the performances of the forecast sets a rolling window of $m_p = 60$ observations is taken.

It must be noted that only the performance of an already constructed forecast can be measured. Therefore the performance clusters can only be determined after the first m observations of the full sample T, and the period for which out-of-sample forecasts can be generated which are based on performance clusters is $t = m + m_p + 1, ..., T$.

The performance of the forecasts is measured in terms of MSFE. First the MSFE is determined for the initial rolling window.

$$\text{MSFE}_{i,m+m_p+1}^{(n)} = \frac{1}{m_p} \sum_{t=m+2}^{m+m_p+1} \left(rx_t^{(n)} - \widehat{rx}_{i,t}^{(n),ind} \right)^2 \tag{4-8}$$

Second the rolling window is shifted one observation and the MSFE is recalculated:

$$\text{MSFE}_{i,m+m_p+2}^{(n)} = \frac{1}{m_p} \sum_{t=m+3}^{m+m_p+2} \left(rx_t^{(n)} - \widehat{rx}_{i,t}^{(n),ind} \right)^2 \tag{4-9}$$

This is continued until the $\text{MSFE}_{i,t}^{(n)}$ for the complete sample has been determined. For each maturity the resulting sets are $\left\{\text{MSFE}_{i,t}^{(n)}\right\}_{m+m_p+1}^{T}$ for $i = 1, ..., N$. Each forecast set now has a measure of its performance, for every observarion in the period $t = m + m_p + 1, ..., T$.

The forecast sets are assigned to a cluster but these assignments are updated at every observation, which means that the composition of each cluster may be completely different for each observation. To determine which forecast sets are assigned to which clusters the k-means algorithm is applied. CLustering forecasts according to using this algorithm was applied by Aiolfi and Timmermann [2006]. This algorithm assigns the forecast sets to clusters based on their MSFE. A number of clusters, k, is specified and the algorithm assigns each forecast set to a cluster. There is no given number of sets that has to be included in each cluster as long as each cluster contains at least one set. The advantage of this method is that forecast sets with very similar performances will always end up in the same cluster, whereas other methods which specify a fixed number of sets that each cluster must contain cannot guarantee this. For each observation the k-means algorithm is implemented as follows:

1. Collect $\text{MSFE}_{i,t}^{(n)}$ for $i = 1, ..., N$ in $\mathcal{V}_t^{(n)}$ and let v_i be the i-th element of $\mathcal{V}_t^{(n)}$.
2. Set the amount of desired clusters to k.
3. Determine k initial cluster centers (centroids). This is done with the following sub-algorithm:
4. Assigning the elements to the clusters is done in two steps. First with a batch update,

and second with a online update.

A more detailed description of this algorithm can be found in Appendix B.

The k-means algorithm is applied at each observation and this generates the performance clusters for each observation. The forecasts $\widehat{rx}_{i,t}^{(n),ind}$ that are in cluster r are denoted by $\widehat{rx}_{p,r,t}^{(n)}$, where $p = 1, ..., N_{r,t}$ and $N_{r,t}$ is the amount of forecasts in cluster r at time t. The forecast set of each cluster is constructed by taking the equally weighted average of the forecasts in the cluster at each observation.

$$\widehat{rx}_{r,t}^{(n),per} = \frac{1}{N_{r,t}} \sum_{p=1}^{N_{r,t}} \widehat{rx}_{p,r,t}^{(n)} \tag{4-10}$$

Before creating the clusters based on performance, the number of different clusters has to be selected. To determine the optimal number of clusters, the performance of the combined forecast sets for different amounts of clusters is examined. The aim of clustering is to significantly reduce the amount of predictor variables. Therefore the combined forecast sets are constructed with 5, 10 and 15 clusters.
The MSFEs of the forecast sets that were constructed using 5 performance clusters were on average more than 10% lower than those that were constructed using 10 and 15 clusters, when either the macroeconomic or financial forecast sets were used. Therefore it was decided that 5 clusters would be used when constructing performance clusters for either the macroeconomic or financial forecast sets. When the performance clusters are determined for the set that contains both the macroeconomic and financial forecast sets, 10 clusters are used.

This cluster method is applied to both the sets of forecast sets that are constructed with the macroeconomic and financial predictor variables, and this results in five forecast sets each. When the cluster method is applied to the set of forecast sets constructed from both sets of predictor variables, ten clusters and thus ten forecast sets are the result. The three sets of forecast sets are referred to as the performance forecast sets.

4-2-4 Overview

The previous sections describe how the initial forecast sets are constructed as well as how the two cluster methods are applied to combine these forecast sets into new forecast sets. An overview of these forecast sets is provided in Table 4-1. All nine sets of forecast sets will be combined according to the weighting schemes that are described in the next Section.

Table 4-1: An overview of the amount of constructed forecast sets

	Individual			Category			Performance		
	Mac	*Fin*	*All*	*Mac*	*Fin*	*All*	*Mac*	*Fin*	*All*
Forecast sets	132	147	279	14	4	18	5	5	10

The number represents the amount of forecast sets that are constructed with a given cluster method and a given set of predictor variables. E.g. using the category cluster method and just the set of financial predictor variables, four forecast sets are constructed.

4-3 Weighting Schemes

The sets of individual, category and performance forecast sets that were constructed in the previous section are combined using various weighting schemes. Each combined forecast is constructed as:

$$\widehat{cr}_t^{(n)} = \frac{1}{N} \sum_{i=1}^{N} \omega_{i,t} \widehat{rx}_{i,t}^{(n)} \tag{4-11}$$

where N is the number of forecasts and $\omega_{i,t}$ are the weights that are determined by the weighting schemes.

Not every weighting schemes that is described can be applied to each set of forecasts. Therefore it will be made apparent which weighting scheme is applied to which set.

4-3-1 Simple weights

Simple weights are, as the name suggests, the most basic weighting schemes. Simple forecasts are less prone to error maximization and to test their forecasting power this paper examines five different simple weighting schemes. Each of the five is applied to each of the sets of individual, category and performance forecast sets.

Equal weights. The weights in this scheme are simply set to $\omega_{i,t}^{ave} = \frac{1}{N}$. The weights are time-invariant so $\omega_{i,t}^{ave} = \omega_{i}^{ave}$.

Trimmed equal weights. This scheme determines new weights at each observation. Therefore the forecasts for which the weights have to be determined are examined at each observation. The weights are set to $\omega_{i,t}^{trim\_ave} = 0$ for the forecasts that have the 10% highest and lowest values, and to $\omega_{i,t}^{trim\_ave} = \frac{1}{0.8N}$ for the remaining forecast.

Median combination forecast. In this scheme the weights are set to $\omega_{i,t}^{med} = 0$ for all the forecasts except for the median of $\{\widehat{rx}_{i,t}^{(n)}\}_{i=1}^{N}$, for which the weight is set equal to 1. This is done at each observation.

The last two simple weighting schemes are based on the past performance of the forecast sets, but apply simple weights once the past performances have been determined.

Best performance. At observation τ the MSFE of each forecast set is determined for the previous 60 observations, i.e. for $t = \tau - 60, ..., \tau - 1$. For the forecast sets with the 10% lowest MSFEs the weight is set to one, $\omega_{i,\tau}^{BP} = \frac{1}{0.1N}$. For all the others it is set to zero, $\omega_{i,\tau}^{BP} = 0$.

Exclude worst. In this scheme the same MSFEs are examined as with *Best performance*. This time the forecast sets with the 10% highest MSFEs are excluded. Their weights are set to zero, $\omega_{i,\tau}^{EW} = 0$, while for the other forecasts it is set to $\omega_{i,\tau}^{EW} = \frac{1}{0.9N}$.

4-3-2 Recursive OLS weights

Recursive OLS weights are based on regressing the bond returns on the forecasts of these returns. The linear regression model is:

$$rx_t^{(n)} = \zeta RX_t^{(n)} + \varepsilon_t \tag{4-12}$$

where $RX_t^{(n)} = \left\{\widehat{rx}_{1,t}^{(n)}, ..., \widehat{rx}_{N,t}^{(n)}\right\}$ and ζ is a $1 \times n$ coefficient vector. To determine the OLS estimate of ζ, a rolling estimation window of size m_r is used. At observation τ, $\{rx_t^{(n)}\}_{\tau-m_r}^{\tau-1}$ is regressed on $\{RX_t^{(n)}\}_{\tau-m_r}^{\tau-1}$. The weight $\omega_{i,\tau}^{OLS}$ is the i-th element of vector $\hat{\zeta}_{i,\tau}$. No intercept is included in the model and no coefficients are restricted.

Two different window sizes are used, namely 36 and 60 months. These should be large enough to capture the characteristics of the returns, but not so long that they can't adapt to changes in the observations.
As the recursive OLS weights need to perform a regression at each observation, this is only applied to the category and performance forecast sets, as they have been reduced in size.

4-3-3 Adaptive Updating weights

Applying schemes where the weights depend inversely on the historical performance of individual forecasts was first proposed by Bates and Granger [1969]. Miller et al. [1992] modified this weighting scheme to give more recent data more weight in the estimation, by adding a discount factor. This section presents five weighting schemes based on these two papers.
The weighting schemes recursively examine the performance of the forecast sets and assign the weights based on the performance up to that point. Schemes $BG1$ and $BG2$ use a rolling window of size v to determine weights based on the past performance of the forecast sets. Scheme $BG3$ also applied a rolling window of size v but it also smooths its weights towards their values the previous observation and this means that $BG3$ actually uses an expanding window. Schemes $BG4$ and $BG5$ use an expanding window to do this. The last two apply exponential discounting to put more emphasis on recent performance. By using an expanding window these schemes make sure they always take all the available information into account when determining weights. The schemes using a rolling window only use the information that is contained in the last v observations. The weighting schemes use the forecast error as a measure, which is defined as:

$$e_{i,t}^{(n)} = rx_t^{(n)} - \widehat{rx}_{i,t}^{(n)} \tag{4-13}$$

The first weighting scheme, $BG1$, uses the relative performance of each forecast set over the past v observations to determine the weights. The weights correspond to the inverse of the sum of the MSFE over the past v observations of the each set scaled such that the sum of all weights is equal to one.

$$\omega_{i,\tau}^{BG1} = \frac{\left(\sum_{t=\tau-v}^{\tau-1} e_{i,t}^{(n)2}\right)^{-1}}{\sum_{j=1}^{N}\left(\sum_{t=\tau-v}^{\tau-1} e_{j,t}^{(n)2}\right)^{-1}} \tag{4-14}$$

Scheme $BG3$ uses the same relative performance of each forecast set over the past v observations as $BG1$, but smooths the weights $\omega_{i,\tau}^{BG3}$ towards the weights at the previous observation $\omega_{i,\tau-1}^{BG3}$ by a factor $\alpha \in (0,1)$. Scheme $BG1$ therfore is a special case of the weighting scheme $BG3$ where $\alpha = 0$.

$$\omega_{i,\tau}^{BG3} = \alpha\omega_{i,\tau-1}^{BG3} + (1-\alpha)\frac{\left(\sum_{t=\tau-v}^{\tau-1} e_{i,t}^{(n)2}\right)^{-1}}{\sum_{j=1}^{N}\left(\sum_{t=\tau-v}^{\tau-1} e_{j,t}^{(n)2}\right)^{-1}} \tag{4-15}$$

When α gets closer to one, in general the evolution of the weights becomes smoother. To test the effects of this parameter α, this weighting scheme is applied to the forecast sets with $\alpha = 1/4$, $1/2$ and $3/4$.

The last weighting scheme that uses a rolling window is scheme $BG2$. This weighting schemes determines weights based on the covariance matrix of the forecast errors. The covariance matrix includes the squared forecast error of each set that schemes $BG1$ and $BG3$ rely on, but adds the terms which measure the correlation between the forecast errors of the different forecast sets. This covariance matrix is denoted by $\widehat{\mathcal{U}}_t$ and has dimensions $N \times N$. At observation τ the elements of $\widehat{\mathcal{U}}_\tau$ are determined by:

$$\widehat{\mathcal{U}}_\tau[i,j] = v^{-1}\sum_{t=\tau-v}^{\tau-1} e_{i,t}^{(n)} e_{j,t}^{(n)} \tag{4-16}$$

The inverse of the covariance matrix is then taken and the weight of each forecast set corresponds to the sum of its inverted variance and covariance with the other sets, scaled so the sum of all weights is equal to one.

$$\boldsymbol{\omega}_\tau^{BG2} = \frac{\widehat{\mathcal{U}}_\tau^{-1}\boldsymbol{\iota}}{\boldsymbol{\iota}'\widehat{\mathcal{U}}_\tau^{-1}\boldsymbol{\iota}} \tag{4-17}$$

where ι is a $N \times 1$ vector of ones. The weights $\omega_{i,\tau}^{BG2}$ is the i-th elements of $\boldsymbol{\omega}_{\tau}^{BG2}$.
Each of the weighting schemes, *BG1*, *BG2* and *BG3*, uses a rolling estimation window v. For each scheme three windows sizes are applied, namely 12, 36 and 60 months. These should be long enough to capture the characteristics of the errors, but not so long that they can't adapt to changes in the forecast performances.

The last two weighting schemes discussed in this section use an expanding window instead of a rolling window. The main difference is that only the past v observations are taken into account when using a rolling window. The information that these observations contain is considered equally when constructing weights. When an expanding window is applied, each available past observation is used to construct the weights. So for the entire sample $t = 1, ..., \tau, ..., T$, when constructing $\omega_{i,\tau}$ the observations from the period $t = 1, ..., \tau - 1$ are used. However not every observation is given the same weight. The idea is that emphasis is put on more recent observations. This is done by exponentially discounting the past observations by λ^t.
Weighting scheme $BG4$, is very similar to scheme $BG1$. They both determine the weights based on the inverse of the MSFE. The difference being that $BG1$ looks at the past v observations, while $BG4$ looks at all past observations. As a results $BG4$ is fairly similar to $BG3$.

$$\omega_{i,\tau}^{BG4} = \frac{\left(\sum_{t=1}^{\tau-1} \lambda^t e_{i,t}^{(n)^2}\right)^{-1}}{\sum_{j=1}^{N} \left(\sum_{t=1}^{\tau-1} \lambda^t e_{j,t}^{(n)^2}\right)^{-1}} \tag{4-18}$$

Setting the parameter $\lambda = 1$ corresponds to putting equal weights on each past observation. Any higher values of λ correspond to putting relatively more weight on more recent observations. The last of these weighting schemes is $BG5$. As scheme $BG4$ was very similar to scheme $BG1$, so is scheme $BG5$ very similar to $BG2$. The scheme also determines weights by looking at a covariance matrix but again the difference is that the second uses a rolling window and the first an expanding window. The covariance matrix $\widehat{\mathcal{W}_t}$ is constructed with the expanding window and has dimensions $N \times N$. At each observation τ the elements of

$\widehat{\mathcal{W}}_\tau$ are determined by:

$$\widehat{\mathcal{W}}_\tau[i,j] = \sum_{t=1}^{\tau-1} \lambda^t e_{i,t}^{(n)} e_{j,t}^{(n)} \tag{4-19}$$

The elements $\widehat{\mathcal{W}}_\tau[i,i]$ are squared forecast errors which are used for scheme $BG4$, so as was the case between $BG2$ and $BG1$, $BG5$ includes the squared forecast errors and adds the correlation terms of the different forecast sets. This covariance matrix is then used to construct the weights.

$$\boldsymbol{\omega}_\tau^{BG5} = \frac{\widehat{\mathcal{W}}_\tau^{-1}\boldsymbol{\iota}}{\boldsymbol{\iota}'\widehat{\mathcal{W}}_\tau^{-1}\boldsymbol{\iota}} \tag{4-20}$$

where $\boldsymbol{\iota}$ is a $N \times 1$ vector of ones. The weight $\omega_{i,\tau}^{BG5}$ is the i-th elements of $\boldsymbol{\omega}_\tau^{BG5}$.
For both of the schemes $BG4$ and $BG5$, a λ of 1.05 and 1.10 will be chosen, as this will allow for more weight on recent observations, but still allow 'older' observations to add information.
By applying these five relatively similar weighting schemes several effects can be examined; using either a rolling window or an expanding window, using an entire covariance matrix or ignoring the correlation, and looking only at the performance of each forecast set individually; smoothing the weights towards the weights at the previous observation.
Weighting schemes $BG1$, $BG3$ and $BG4$ are applied to all of the constructed forecast sets. Weighting schemes $BG2$ and $BG5$ on the other hand are only applied to the forecast sets that were constructed after a cluster method was applied. The reason for this is that both these weighting schemes construct a covariance matrix at each observation, and this causes a problem if it is applied to a set of forecast sets that have not yet been reduced by a cluster method as the number of forecasts is too large.

4-3-4 Shrinkage

Eleven different weighting schemes have now been discussed; five simple, one recursive OLS and five adaptive updating. In addition to determining the weights based solely on these weighting schemes, shrinkage is applied to a number of them. Shrinkage is a method where the weights $\omega_{i,t}$, that are determined by some weighting scheme, are shrunk towards a value which is imposed a priori. The amount of shrinkage that is applied is controlled by the parameter γ. This takes the form:

$$\omega_{i,t}^{(*)\_shr} = \gamma\omega_{i,t}^{(*)} + (1-\gamma)\frac{1}{N} \tag{4-21}$$

In this case shrinkage towards equal weights is applied to the adaptive updating weighting schemes as well as the recursive OLS weights. To study the effects of different amounts of shrinkage several values of γ are applied; $\gamma = \frac{1}{4}, \frac{1}{2}, \frac{3}{4}$.

4-4 Complete Subset Regression

When constructing the combined forecasts in the previous sections, first forecast sets were constructed and then these were either combined, or clustered and then combinied. Each forecast set was constructed by regressing the excess bond returns on a single predictor variable set. These forecast sets can of course be constructed in any number of ways. One other method of constructing forecasts is applied in this paper and the only weighting scheme that is applied when combining the forecasts is equal weights.
This method is called the complete subset regression method and is proposed by Elliott et al. [2013]. Like all the other forecast combination methods that have already been discussed in the report, the complete subset regression method applies averaging over multiple forecasts, rather then selecting a single prediction model. The weighting schemes that are discussed in the previous sections, are all applied to the same forecast sets that have already been constructed using the predictor variables. The complete subset regression method constructs different forecast sets itself, using more than one predictor variable, and then applies an equal weights weighting scheme. Applying this method in addition to the weighting schemes adds an extra forecast combination method, and also one that constructs its combined forecasts from different initial forecasts.

The idea of the complete subset regression method is that given a total of K predictor variables for $k = 1,...,K$ a forecast set is constructed for each subset of k predictor variables. This

means that for each k there are $M_k = \frac{K!}{k!(K-k)!}$ forecast sets. For each k a single forecast set is constructed by combining the forecast sets which were constructed with the same amount of predictor variables. These forecast sets are combined using equal weights.
For a given value of k, $Z_{j,t,k}$ contains the j-th subset of k predictor variables, where $j = 1,...,M_k$. The forecasts are constructed using a rolling window of 120 observations. Forecast $\widehat{rx}^{(n),cSubsR}_{j,\tau,k}$ is constructed by:

$$\widehat{rx}^{(n),cSubsR}_{j,\tau,k} = \hat{\theta}_{j,\tau,k} Z_{j,\tau,k} \tag{4-22}$$

where $\hat{\theta}_{j,\tau,k}$ is the OLS estimate of $\theta_{j,k}$, which in turn is constructed by regressing $\{rx^{(n)}_t\}^{\tau-1}_{t=\tau-u}$ on $\{Z_{j,t,k}\}^{\tau-2}_{t=\tau-u-1}$, according to the following regressing model:

$$rx^{(n)}_{t+1} = \theta_{j,k} Z_{j,t,k} + \varepsilon_{j,t,k} \tag{4-23}$$

For each value k, M_k forecast sets $\widehat{rx}^{(n),cSubsR}_{j,\tau,k}$ are constructed, where $\widehat{rx}^{(n),cSubsR}_{j,k}$ is the set of forecasts $\widehat{rx}^{(n),cSubsR}_{j,\tau,k}$. A single forecast set is constructed for each value of k with an equal weights weighting scheme.

$$\widehat{cx}^{(n),cSubsR}_{k} = \frac{1}{M_k} \sum_{j=1}^{M_k} \widehat{rx}^{(n),cSubsR}_{j,k} \tag{4-24}$$

Appying the complete subset regression method for $k = 1$ is equivalent to applying the equal weights weighting scheme to the forecast sets, therefore this will not be done.
When the number of predictor variables that is used by the complete subset regression method is large, the computation time becomes very large. Therefore the complete subset regression method is only applied to the individual forecast sets for $k = 2$.
The complete subset regression method is applied to the predictor variables when they are clustered by categories. In this case it is applied for $k = 2,...,K$, where K corresponds to 14, 4 and 18 for the three sets of predictor variables respectively.
For the construction of the forecast sets, a rolling window of 120 observations is taken. However this means that this method cannot be applied to the performance sets, because a large

window is already needed to create the performance clusters. This means that if the complete subset regressions method would be applied, there would not be enough data left to verify the forecast sets.

4-5 Overview

In the previous sections a number of weighting schemes and methods of constructing forecasts were discussed. The total time sample, for which all of the data is available, spans the period from January 1964 till December 2011. The different weighting schemes, cluster methods and parameter settings each require a sample of observations to create the initial forecast. However the size of this sample is not the same in each case. Therefore the forecast sets that are constructed have different starting dates. E.g. the forecast set that is created with the weighting scheme equal weights for a set from the individual forecast set, spans the period January 1975 till December 2011, whereas the forecast set that is created using weighting scheme $BG4$ for a set from the performance forecast sets, spans the period February 1979 till December 2011.
This report aims to make a comparison of the forecasting power of the different methods. Therefore the forecast sets are all constructed for the same period, where the starting date is chosen such that each forecast method can be included. This spans the period February 1985 till December 2011, consisting of 323 forecasts. Twelve forecast combination methods have been described; five simple, the recursive OLS, five adaptive updating and the complete subset regression. These twelve methods are used to construct forecasts of one-year excess U.S. government bond returns. The methods are applied with different parameter settings and also with different amount of shrinkage applied. On top of that the forecasts are constructed using two sets of predictor variables, macroeconomic and financial, and the forecast combination methods are applied to each of these sets of forecast sets individually as well as combined. Finally the forecasts are also clustered according to two different methods. As a result, a large number of different forecast sets have been constructed. In total 753 forecast sets of the one-year excess U.S. government bond returns have been constructed for each of

Table 4-2: An overview of the combined forecast sets

	Individual			Category			Performance		
	Mac	*Fin*	*All*	*Mac*	*Fin*	*All*	*Mac*	*Fin*	*All*
Forecast sets	132	147	279	14	4	18	5	5	10
Simple									
Mean	*1*	*1*	*1*	*1*	*1*	*1*	*1*	*1*	*1*
Trimmed mean	*1*	*1*	*1*	*1*	*1*	*1*	*1*	*1*	*1*
Median	*1*	*1*	*1*	*1*	*1*	*1*	*1*	*1*	*1*
Exclude worst	*1*	*1*	*1*	*1*	*1*	*1*	*1*	*1*	*1*
Best performance	*1*	*1*	*1*	*1*	*1*	*1*	*1*	*1*	*1*
Scheme total	5	5	5	5	5	5	5	5	5
Recursive OLS									
Scheme total	0	0	0	8	8	8	8	8	8
Adaptive updating									
BG1	*12*	*12*	*12*	*12*	*12*	*12*	*12*	*12*	*12*
BG2	*0*	*0*	*0*	*12*	*12*	*12*	*12*	*12*	*12*
BG3	*36*	*36*	*36*	*36*	*36*	*36*	*36*	*36*	*36*
BG4	*8*	*8*	*8*	*8*	*8*	*8*	*8*	*8*	*8*
BG5	*0*	*0*	*0*	*8*	*8*	*8*	*8*	*8*	*8*
Scheme total	56	56	56	76	76	76	76	76	76
Complete subset regression									
Scheme total	1	1	1	13	3	17	0	0	0
Total	62	62	62	102	92	106	89	89	89

The nummber in italic represent the number of parameter settings for a given weighting scheme. E.g. the 147 forecast sets that were constructed from the *Individual Financial* sets are used to construct 36 new combined forecast sets using 36 different parameter settings of the weighting scheme *BG3*.

the four maturities, 2, 3, 4 and 5 years. An overview of these forecast sets is given in Table 4-2.

From each of the three individual forecast sets, 62 forecast sets are constructed. These sets are referred to as the combined individual sets. From the category forecast sets, three sets of forecast sets of 102, 92 and 106 have been constructed, for the three sets of predictor variables respectively. The reason for the difference in number of forecast sets is that the complete subset regression methods is applied for $k = 2, ..., K$ and K differs depending on which set of predictor variables is chosen. The constructed sets are referred to as the combined category sets. From the performance forecast sets, 89 combined forecast sets are constructed for each predictor variables set. These are referred to as the combined performance sets.

For each maturity n, the total set containing all three of the combined forecast sets is referred to as $\mathcal{N}^{0,(n)}$ and contains all 753 combined forecast sets.

4-6 Model Confidence Set

Comparing the results of the 753 forecast sets is certainly not a trivial matter. The model confidence set algorithm, which was introduced by Hansen [2011], provides a solution to this problem. For a given set of forecast sets $\mathcal{M}^{0,(n)}$ and significance level α, the model confidence set algorithm constructs a new set $\widehat{\mathcal{M}}^{*,(n)}_{1-\alpha}$, which contains the *best* forecast set(s) of the original set, such that $\widehat{\mathcal{M}}^{*,(n)}_{1-\alpha} \subseteq \mathcal{M}^{0,(n)}$. This set $\widehat{\mathcal{M}}^{*,(n)}_{1-\alpha}$ is called the model confidence set.

The model confidence set algorithm evaluates each forecast set based on its relative performance to the other forecast sets. This relative performance is measured as follows: Consider the initial set $\mathcal{M}^{0,(n)}$, which contains m_0 forecast sets of T observations. These forecast sets are evaluated by means of a loss function. The squared forecast error is applied as loss function, such that:

$$L^{(n)}_{i,t} = \left(rx^{(n)}_t - \widehat{cx}^{(n)}_{i,t}\right)^2 \tag{4-25}$$

where $i = 1, ..., m_0$ and $t = 1, ..., T$. The relative performance of two forecasts is defined as:

$$d_{ij,t}^{(n)} \equiv L_{i,t}^{(n)} - L_{j,t}^{(n)} \quad \text{for all} \quad i, j \in \mathcal{M}^{0,(n)} \tag{4-26}$$

The vector $d_{ij}^{(n)}$ consists of the elements $d_{ij,t}^{(n)}$ for $t = 1, ..., T$. The assumption is made that $\mu_{ij}^{(n)} \equiv E(d_{ij,t}^{(n)})$ is finite and independent of t, for all $i, j \in \mathcal{M}^{0,(n)}$. The alternatives are ranked in terms of expected loss, so that the alternative i is preferred to alternative j if $\mu_{ij}^{(n)} < 0$. The objective of the MCS algorithm is to determine $\mathcal{M}^{*,(n)}$, where $\mathcal{M}^{*,(n)}$ is defined as:

$$\mathcal{M}^{*,(n)} \equiv \{i \in \mathcal{M}^{0,(n)} : \mu_{ij}^{(n)} \leq 0 \text{ for all } j \in \mathcal{M}^{0,(n)}\} \tag{4-27}$$

To achieve this goal, a series of significance tests was performed to determine whether any of the forecast sets are significantly inferior to the other forecast sets of $\mathcal{M}^{0,(n)}$. If it is found that an forecast set is inferior, it is eliminated from the set. The hypothesis that is tested to determine this takes the form:

$$H_{0,\mathcal{M}^{(n)}} : \mu_{ij}^{(n)} = 0 \quad \text{for all } i, j \in \mathcal{M}^{(n)} \tag{4-28}$$

where $\mathcal{M}^{(n)} \subseteq \mathcal{M}^{0,(n)}$.
The model confidence set algorithm is based on an *equivalence test*, $\delta_{\mathcal{M}}$, and an *elimination rule*, $e_{\mathcal{M}}$. The equivalence test is used to test the hypothesis $H_{0,\mathcal{M}^{(n)}}$ for any $\mathcal{M}^{(n)} \subseteq \mathcal{M}^{0,(n)}$, and in the event that $H_{0,\mathcal{M}^{(n)}}$ is rejected, the elimination rule identifies the forecast set in $\mathcal{M}^{(n)}$ that is to be removed from $\mathcal{M}^{(n)}$. Here $\delta_{\mathcal{M}} = 0$ and $\delta_{\mathcal{M}} = 1$ correspond to the cases where $H_{0,\mathcal{M}^{(n)}}$ is accepted or rejected respectively. The algorithm works as followed:

1. Initially set $\mathcal{M}^{(n)} = \mathcal{M}^{0,(n)}$

2. Test $H_{0,\mathcal{M}^{(n)}}$ using $\delta_{\mathcal{M}}$ at level α

3. If $H_{0,\mathcal{M}^{(n)}}$ is accepted, define $\widehat{\mathcal{M}}_{1-\,alpha}^{*,(n)} = \mathcal{M}^{(n)}$; otherwise, use $e_{\mathcal{M}}$ to eliminate an object from $\mathcal{M}^{(n)}$ and repeat the procedure from Step 2

The model confidence set $\widehat{\mathcal{M}}^{*,(n)}_{1-\alpha}$ is the set of surviving objects.

The applied equivalence test is based on t-statistics. First the relative sample loss statistic is defined as $\bar{d}^{(n)}_{ij} \equiv \frac{1}{T}\sum_{t=1}^{T} d^{(n)}_{ij,t}$. This measures the relative sample loss between the ith and the jth forecast set. Second an estimate is made of the variance of $\bar{d}^{(n)}_{ij}$ using a bootstrap. This is defined as $\widehat{\text{var}}\left(\bar{d}^{(n)}_{ij}\right) \equiv \frac{1}{B}\sum_{b=1}^{B}\left(\bar{d}^{*,(n)}_{b,ij} - \bar{d}^{(n)}_{ij}\right)^2$. From these variables the following t-statistic can be constructed as:

$$t^{(n)}_{ij} = \frac{\bar{d}^{(n)}_{ij}}{\sqrt{\widehat{\text{var}}\left(\bar{d}^{(n)}_{ij}\right)}} \quad \text{for } i,j \in \mathcal{M}^{(n)} \tag{4-29}$$

This t-statistic is associated with the null hypothesis that $\mu^{(n)}_{ij} = 0$ and this null hypothesis can be tested with the test statistic:

$$T^{(n)}_{R,\mathcal{M}} \equiv \max_{i,j\in\mathcal{M}^{(n)}} |t^{(n)}_{ij}| \tag{4-30}$$

The natural elimination rule for this test statistic is $e^{(n)}_{R\mathcal{M}} = \arg\max_{i\in\mathcal{M}^{(n)}} \sup_{j\in\mathcal{M}^{(n)}} t^{(n)}_{ij}$ because this model is such that $t^{(n)}_{e_{R,\mathcal{M}}j} = T^{(n)}_{R,\mathcal{M}}$ for some $j \in \mathcal{M}^{(n)}$.

The model confidence set p-value for forecast set $e^{(n)}_{\mathcal{M}_j} \in \mathcal{M}^{0,(n)}$ is defined by $\hat{p}^{(n)}_{e_{\mathcal{M}_j}} \equiv \max_{i\leq j} P^{(n)}_{H_{0,\mathcal{M}_i}}$, where $P^{(n)}_{H_{0,\mathcal{M}_i}}$ denotes the p-value that is associated with the null hypothesis $H_{0,\mathcal{M}^{(n)}_i}$. As $\mathcal{M}^{(n)}_{m_0}$ consists of a single forecast set, the null hypothesis, $H_{0,\mathcal{M}^{(n)}_{m_0}}$, states that the last surviving forecast set is as good as itself, making the $P^{(n)}_{H_{0,\mathcal{M}_{m_0}}} = 1$.

The implementation of the model confidence set algorithm in this report uses a block-bootstrap procedure of B resamples, which is set to 10000. The block length l should be long enough to capture any autocorrelation found in the loss terms. Autocorrelation tests are performed and the conclusion of the tests is that a block length of 18 observations captures the autocorrelation in all of the loss terms for each set of combined forecast sets. The significance level which is applied when constructing the model confidence set is 95%, which corresponds to an α of 0.05.

The entire model confidence set procedure that is implemented in this report can be described in two steps. In the first step, the sample and bootstrap statistic are determined. And in the second step, sequential testing is performed. The details of this implementation are discussed in Appendix B.

The goal of the model confidence set algorithm is to compare all 753 forecast sets for each maturity and determine whether one or more forecast sets outperforms the rest of the forecast sets. For a given maturity however, performing the model confidence set procedure on 753 forecast sets at once is not computationally possible. Therefore the procedure is first applied to all the combined forecast sets that were constructed using macroeconomic predictors. Secondly it is applied to the combined forecast sets that were constructed using the financial predictors. And lastly the procedure is applied to the combined forecast sets that were constructed using both the macroeconomic and financial predictors. This procedure results in a model confidence set for each of these sets, which will be referred to as $\widehat{\mathcal{CM}}^{*,(n)}_{1-\alpha}$, $\widehat{\mathcal{CF}}^{*,(n)}_{1-\alpha}$ and $\widehat{\mathcal{CA}}^{*,(n)}_{1-\alpha}$ respectively. These three model confidence sets are then merged into a single set, $\widehat{\mathcal{N}}^{*,(n)}_{1-\alpha}$, and the model confidence procedure is then applied to this set. The goal is then achieved, as for each maturity a single model confidence set, $\widehat{\mathcal{M}}^{*,(n)}_{1-\alpha}$, remains.

4-7 Out-of-sample R^2

After the model confidence set procedure has been applied the set of surviving models is compared. One measure that is used is the R^2_{OS}. This was first proposed by Campbell and Thompson [2008] and is calculated as

$$R^2_{OS} = 1 - \frac{\text{MSFE}_{model}}{\text{MSFE}_{benchmark}} \tag{4-31}$$

In this report the historical mean forecasting model is taken as the benchmark. This is constructed at each observation t as:

$$\widehat{rx}^{(n)}_{t+1} = \frac{1}{t}\sum_{\tau=1}^{t} rx^{(n)}_{\tau} \tag{4-32}$$

If the R^2_{OS} has a positive value, this indicates that the forecast set has a lower MSFE than that of the benchmark.

5. Results

This Section presents the results of the combined forecast sets which were constructed with the forecast combination methods. First the resulting combined forecast sets for the 5 year maturity are presented, and the results of applying the model confidence set procedure to the combined forecast sets which were constructed using the same predictor variable sets, are also presented and discussed. Secondly the model confidence set procedure is applied to the three model confidence sets that resulted from each predictor set, and the results of this are discussed. In addition a selection of the 'best performing' forecast sets is made based on their respective MSFE, and these forecast sets are further examined. The two selection criteria, model confidence set procedure and MSFE selection, are also compared. Thirdly the results of the forecast sets for the 2, 3, and 4 year maturity are presented and the results of all four maturities are compared. Lastly two robustness checks are performed on the forecast sets. First the model confidence set results are examined for different significance levels and secondly the results are examined when the period of the observations is changed.

5-1 5 year maturity

In Section 4-5 it was concluded that for each maturity a total of 753 combined forecast sets would be constructed. 253 using macroeconomic predictor variables, 243 using financial predictors variables and 257 using both macroeconomic and financial variables. Firstly the MSFE results as well as the model confidence set results of the forecast sets constructed from each of these three sets are presented for the 5 year maturity. Subsequently the three resulting model confidence sets are merged and results of performing the model confidence set procedure on this new set is presented. The results of this are compared to the results when a selection of the forecast sets is made by ranking them based on their MSFE.

5-1-1 Results by predictor variable type

Figure 5-1 presents the results of each of the combined forecast sets for a 5 year maturity. The results of the model confidence set procedures are also presented in the figure, and these are summarised in Table 5-1.

The results of the combined forecast sets constructed with macroeconomic predictors are shown in Figure 5-1a. A large number of the forecast sets have very similar results, both in terms of MSFE and bias. These are all clustered around the point (2.5e-3,-7.5e-3), which in terms of MSFE is about the same as that of the benchmark, but in terms of absolute bias is slightly smaller. If this is examined more closely, it can be seen that in particular the forecast sets that were constructed with the performance cluster method have almost identical results, as the MSFE ranges between 2.51e-3 and 2.52e-3 for almost all of them. The forecast sets that were constructed without a cluster method are also almost all included in the point cluster of results. Their results vary more than those of the performance forecast sets, but they are still very similar. A number of the forecast sets that fall outside of the point cluster have fairly similar MSFE results and differ mainly in bias, but there are also a number of forecast sets that perform worse in terms of MSFE. However there is also a group of forecast sets which seems to outperform all the other forecast sets in terms of MSFE. These are 17 forecast sets, each constructed with the cateogry cluster method. The first eight of these are all the forecast sets that were constructed with the recursive OLS weighting scheme and the other seven were constructed with the complete subset regression forecast combination method. The forecast sets that were constructed with the category cluster method but with the other weighting schemes are also mostly contained in the cluster of results.

Out of a total of 253 forecast sets, the model confidence set procedure includes just four in the 95% model confidence set, $\widehat{\mathcal{CM}}^{*,(5)}_{95\%}$. These four forecast sets are those that were constructed using the recursive OLS weighting scheme and to which either 25% or 50% shrinkage towards equal weights was applied. One forecast set is excluded even though it has a very similar MSFE and a smaller bias than two of the forecast sets that are included. Figure 5-1b presents the results of the combined financial forecast sets. It can be observed

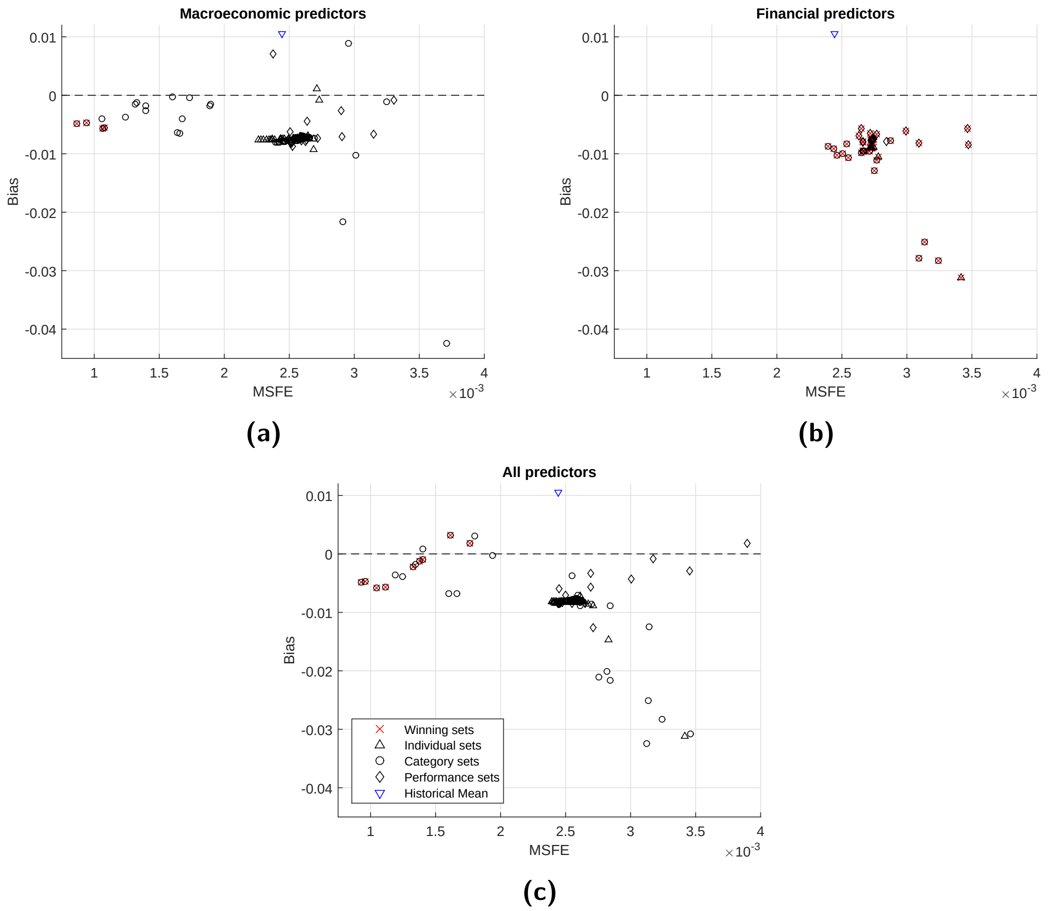


Figure 5-1: MSFE and bias (both displayed in base points) of each combined forecast set for the 5 year maturity. The subfigures show the results of the sets *Combined Macroeconomic*, *Combined Financial* and *Combined All* and their respective model confidence sets, $\widehat{\mathcal{CM}}^{*,(5)}_{95\%}$, $\widehat{\mathcal{CF}}^{*,(5)}_{95\%}$ and $\widehat{\mathcal{CA}}^{*,(5)}_{95\%}$, are indicated by the red marks.

that, with the exception of a few, all the forecast sets have a very similar performance in

Table 5-1: A breakdown of the model confidence sets

		SIM	ROLS	BG1	BG2	BG3	BG4	BG5	CSR	Total
Mac	Ind	0/5	-	0/12	-	0/36	0/8	-	0/1	0/62
	Cat	0/5	4/8	0/12	0/12	0/36	0/8	0/8	0/13	4/102
	Per	0/5	0/8	0/12	0/12	0/36	0/8	0/8	-	0/89
Fin	Ind	5/5	-	12/12	-	36/36	8/8	-	1/1	62/62
	Cat	5/5	8/8	7/12	11/12	0/36	7/8	8/8	3/3	49/92
	Per	5/5	7/8	12/12	12/12	36/36	8/8	8/8	-	88/89
All	Ind	0/5	-	0/12	-	0/36	0/8	-	0/1	0/62
	Cat	0/5	4/8	0/12	0/12	0/36	0/8	0/8	5/17	9/106
	Per	0/5	0/8	0/12	0/12	0/36	0/8	0/8	-	0/89
	Total	15/45	23/48	31/108	23/72	72/324	23/72	16/48	9/36	212/753

Every three rows represent the breakdown by forecast combination method of forecast sets which are included in the three model confidence sets.

terms of both MSFE and bias. On closer examination it is apparent that for each of the three cluster methods most of the forecast sets are clustered in a single point. Each of these three points has a slightly worse MSFE than that of the benchmark. In contrast to the results of the macroeconomic forecast sets there are no results which can be considered better in terms of MSFE than the benchmark. There are however 5 forecast sets that seems to perform slightly better than the other forecast sets and like with the macroeconomic results, these forecast sets are constructed with the recursive OLS weighting scheme.

In this case the model confidence set, $\widehat{\mathcal{CF}}^{*,(5)}_{95\%}$, includes 199 out of an initial 243 forecast sets. This means that over 80% of the initial forecast sets are included in the model confidence set. On closer examiniton it can be seen that the forecast sets that are constructed without a cluster method or those constructed with the performance cluster method, with the exception of one, are all included in the model confidence set. Out of the forecast sets that were constructed with the category cluster method and are in the point cluster, 43 are excluded from the model confidence set while the rest is included.

Another point of interest is that the forecast set that has the worst performance in terms of MSFE is still included in the model confidence set, while its MSFE is about two and half times as large as that of the 43 forecast sets that were excluded from the point cluster. Both these observations feel slightly counterintuitive.

The last result set is that of the combined forecast sets which are constructed using both macroeconomic and financial predictor variables. These are presented in Figure 5-1c. Again there is a point cluster, but it can be observed that these results are somewhere in between the results of the macroeconomic sets and the financials sets, as this cluster for the macroeconomic sets had a slightly lower MSFE and was more varied, whereas the cluster of the financial sets had a slightly higher MSFE and was less varied. Again forecast sets constructed with each of the three cluster methods are included in the point cluster and there are a number of forecast sets from each of them that lie outside of this cluster and perform worse in terms of MSFE. There is another similarity to the results of the macroeconomic forecast sets, and that is that again there are 17 forecast sets that are constructed using the category cluster method that seem to outperform both the rest of the forecast sets and the benchmark. These 17 forecast sets are exactly the same 17 as was the case with the macroeconomic forecast sets, these are again all eight of the recursive OLS weighting scheme forecast sets as well as seven of the complete subset regression method forecast sets.
As with the macroeconomic forecast sets, the model confidence set $\widehat{\mathcal{CA}}^{*,(5)}_{95\%}$ only includes forecast sets that were constructed using the category cluster method. This time nine forecast sets are included. The same four recursive OLS forecast with 25% and 50% shrinkage are included but this time five complete subset regression forecast sets are also included. These are the forecast sets that were constructed with 11, 13, 15, 16 and 18 regressors.

The results of applying the model confidence set to the three subsets of forecast sets based on their predictor variables sets are that for both the macroeconomic and the macroeconomic and financial forecast sets, only a handful of forecast sets are included. These forecast sets are all constructed using just two of the twelve forecast combination methods and indeed seem to outperform most of the other forecast sets and the benchmark in terms of MSFE. In the case of the financial forecast sets 199 out of 243 forecast sets are included and it seems that none of the forecast combination methods are able to construct a forecast set that significantly outperforms the other forecast sets or the benchmark.

5-1-2 Overall results

The three model confidence sets that are constructed with the forecast sets of each of the predictor variable sets are merged into the new set $\widehat{\mathcal{N}}^{*,(5)}_{95\%}$, i.e. $\widehat{\mathcal{N}}^{*,(5)}_{95\%} = \left\{\widehat{\mathcal{CM}}^{*,(5)}_{95\%}, \widehat{\mathcal{CF}}^{*,(5)}_{95\%}, \widehat{\mathcal{CA}}^{*,(5)}_{95\%}\right\}$. This new set consists of 212 of the initial 753 forecast sets. The model confidence set procedure is applied to this new set, and the results of this are presented in Figure 5-2. It is

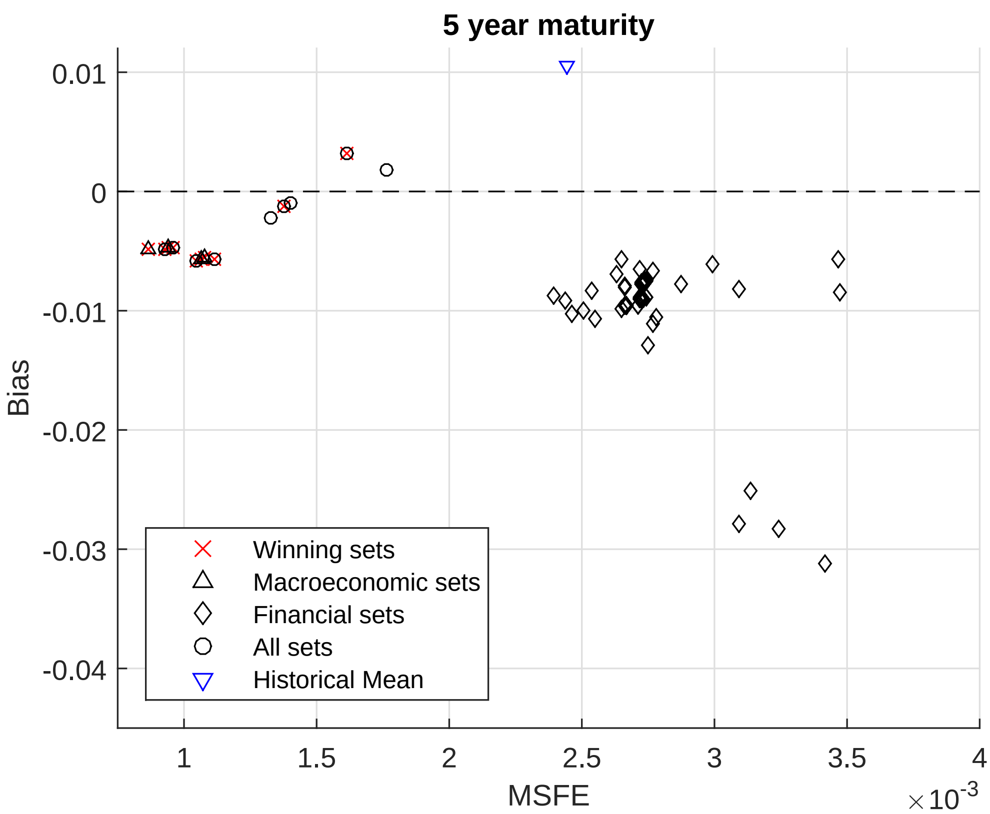


Figure 5-2: MSFE and bias of the forecast sets that are in the model confidence sets models of each of the cluster methods for the forecast sets of the 5 year maturity bonds. The model confidence set, $\widehat{\mathcal{M}}^{*,(5)}_{95\%}$, of the new set of forecast sets, $\widehat{\mathcal{N}}^{*,(5)}_{95\%}$, is represented by the red marks

clear that the forecast sets that were constructed with financial predictors perform much worse than the forecast sets that were constructed with either the macroeconomic or the macroeconomic and financial sets of predictors. The model confidence set confirms this, as it excludes all 199 of these forecast sets. All eight of the recursive OLS forecast sets are again included in the model confidence set, four from the macroeconomic and four from the macroeconomic and financial sets. Of the complete subset regression forecast sets that were included in the $\widehat{\mathcal{CA}}^{*,(5)}_{95\%}$, only two are included in the model confidence set, while the other three are excluded. The result of this is that in the final model confidence set for the 5 year maturity ten forecast sets remain. Table 5-2 gives a breakdown of the final model confidence set in terms of predictor variable, cluster method and forecast combination method.

Table 5-2: A breakdown of the model confidence sets

		SIM	ROLS	BG1	BG2	BG3	BG4	BG5	CSR	Total
Mac	Ind	0/5	-	0/12	-	0/36	0/8	-	0/1	0/62
	Cat	0/5	4/8	0/12	0/12	0/36	0/8	0/8	0/13	4/102
	Per	0/5	0/8	0/12	0/12	0/36	0/8	0/8	-	0/89
Fin	Ind	0/5	-	0/12	-	0/36	0/8	-	0/1	0/62
	Cat	0/5	0/8	0/12	0/12	0/36	0/8	0/8	0/3	0/92
	Per	0/5	0/8	0/12	0/12	0/36	0/8	0/8	-	0/89
All	Ind	0/5	-	0/12	-	0/36	0/8	-	0/1	0/62
	Cat	0/5	4/8	0/12	0/12	0/36	0/8	0/8	2/17	6/106
	Per	0/5	0/8	0/12	0/12	0/36	0/8	0/8	-	0/89
Total		0/45	8/48	0/108	0/72	0/324	0/72	0/48	2/36	10/753

The final row represents the breakdown by the forecast combination method of forecast sets which are included in the final model confidence set for the 5 year maturity.

According to the model confidence set results for the 5 year maturity it can be concluded that two of the forecast combination methods are able to construct forecast sets which can outperform both the historical mean benchmark and the other forecast sets. Moreover, these two forecast combination methods are only able to construct these when the initial forecast sets are constructed using either the macroeconomic predictor variables or both the macroeconomic and financial predictor variables.

Another observation that can be made is that all the forecast sets that are included in the final model confidence set were constructed with the forecast sets that were first combined according to the category cluster method. So reducing the number of forecast sets to which the weighting schemes are applied, can improve the performance, but does not necessarily do so, as can be concluded from the fact that none of the forecast sets that were constructed using the performance cluster method remain in the model confidence set. Applying the category cluster method reduces the number of forecast sets but also clusters forecast sets that were constructed using predictor variables containing similar information. Each of the forecast sets that are constructed with recursive OLS are subjected to three levels of shrinkage, 25%, 50% and 75% as well as being constructed without any shrinkage. The eight forecast sets included in the model confidence set and are constructed with the recursive OLS weighting scheme all include a level of shrinkage towards equal weights. To four of them 25% and to four of them 50% shrinkage towards equal weight is applied, whereas their corresponding forecast sets that received no shrinkage are not included in the model confidence set.

In the results presented in Figure 5-2, it can be seen that the model confidence set includes a number of forecast sets, but also excludes some sets that have very similar results in terms of MSFE and bias. The same was also the case when the model confidence set was applied, in Section 5-1-1, to the three sets of forecast sets for each predictor variable set. Therefore it would also be good to compare the forecast sets by some other measure. In Figure 5-3 the results of all the forecast sets are presented again. It is clear straightaway that, if the selection of the best performing forecast sets is made by only looking at the MFSE, that only forecast sets that are constructed using the recursive OLS weighting scheme or the complete subset regression method would be selected. The recursive OLS forecast sets seem to outperform the complete subset regression forecast sets in terms of MSFE, but the biases of the latter seem to be smaller than those of the recursive OLS forecast sets. This is also the case when the best performing foreast sets are selected by the model confidence set. However if the selection is made solely on the MSFE, more forecast sets are included. To verify whether excluding those forecast sets that the model confidence set excludes has merit, all of the forecast sets which are contained in the dotted box in Figure 5-3 are further examined.

Figure 5-4 gives the cumulative squared forecast error of the benchmark minus the cumulative squared forecast error of the forecast sets which are selected for further examination. In addition, the R^2_{OS} is given for each set, and the forecast sets which are included in the model confidence set are marked with an asterisk.

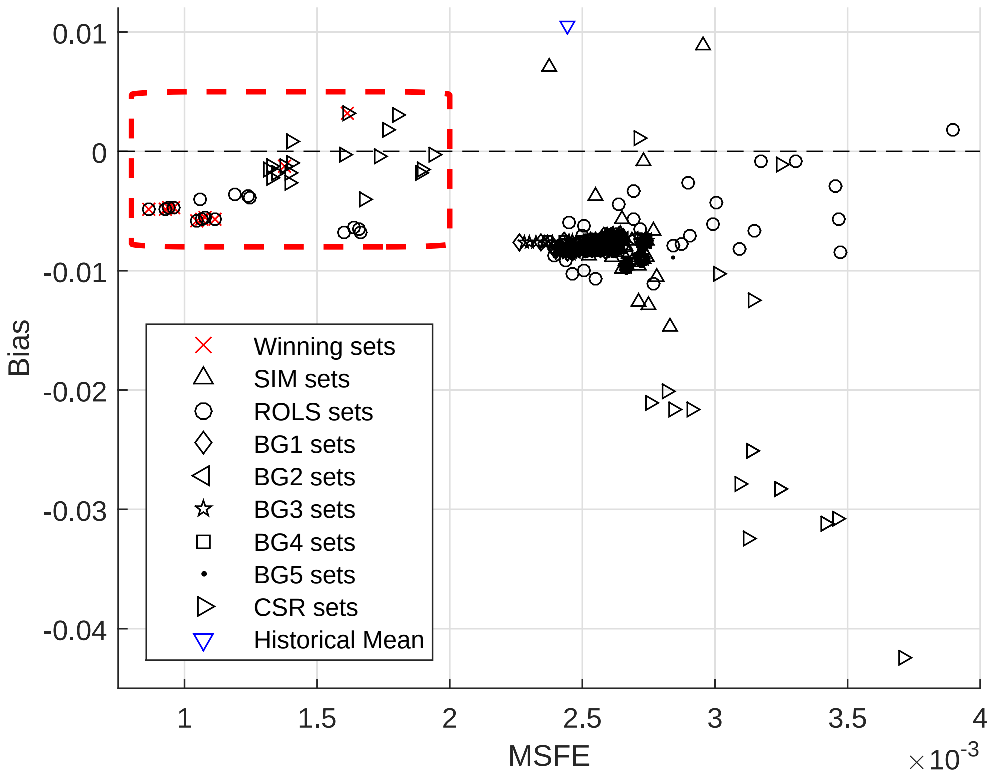


Figure 5-3: MSFE and bias of all the forecast sets of the 5 year maturity bonds. The model confidence set, $\widehat{\mathcal{M}}^{*,(5)}_{95\%}$ is represented by the red marks

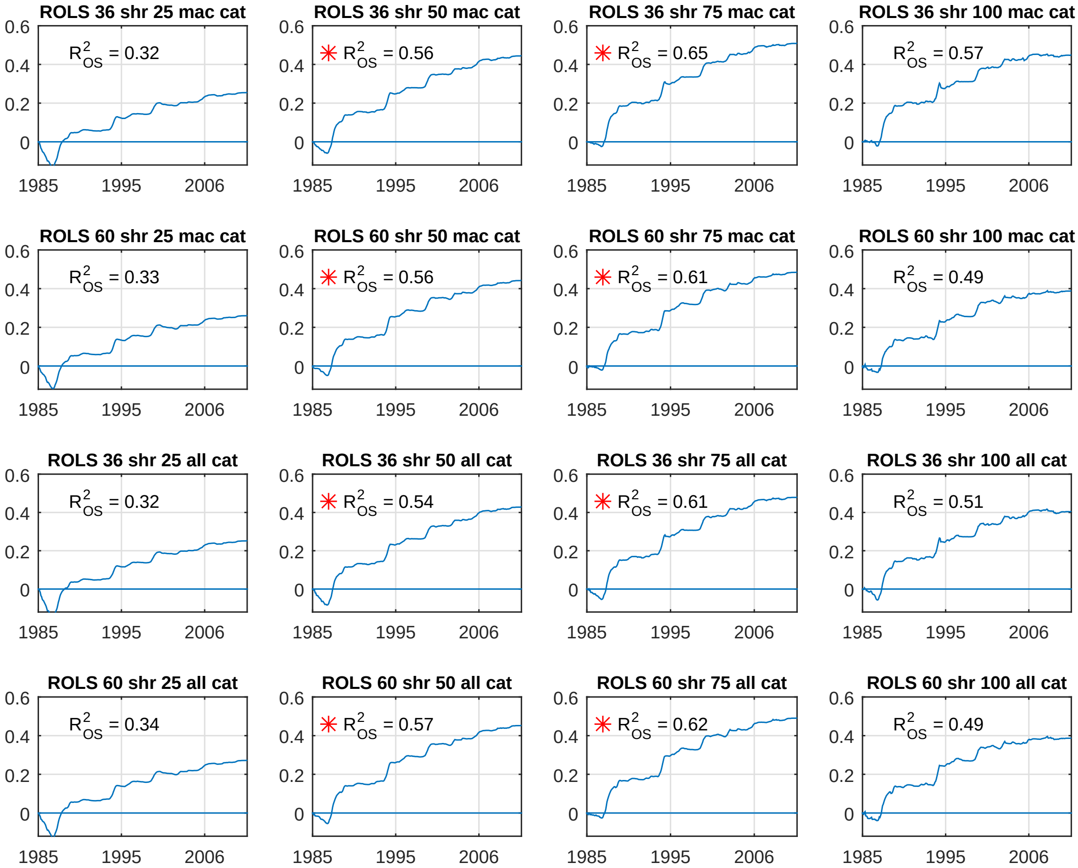


Figure 5-4: The cumulative squared forecast error of the historical average benchmark minus the cumulative squared forecast error of the selected forecast sets and their R^2_{OS}. The red asterisk indicates the forecast sets that are also included in the final model confidence set. The abbreviation '*model* shr x' corresponds to $1-x$ shrinkage towards equal weights. In the case of the recursive OLS weighting schemes 36 or 60 corresponds to the size of the rolling window in months. In the case of the complete subset regression model, the number after k corresponds to the amount of predictor variables that were used to construct the combined forecast.

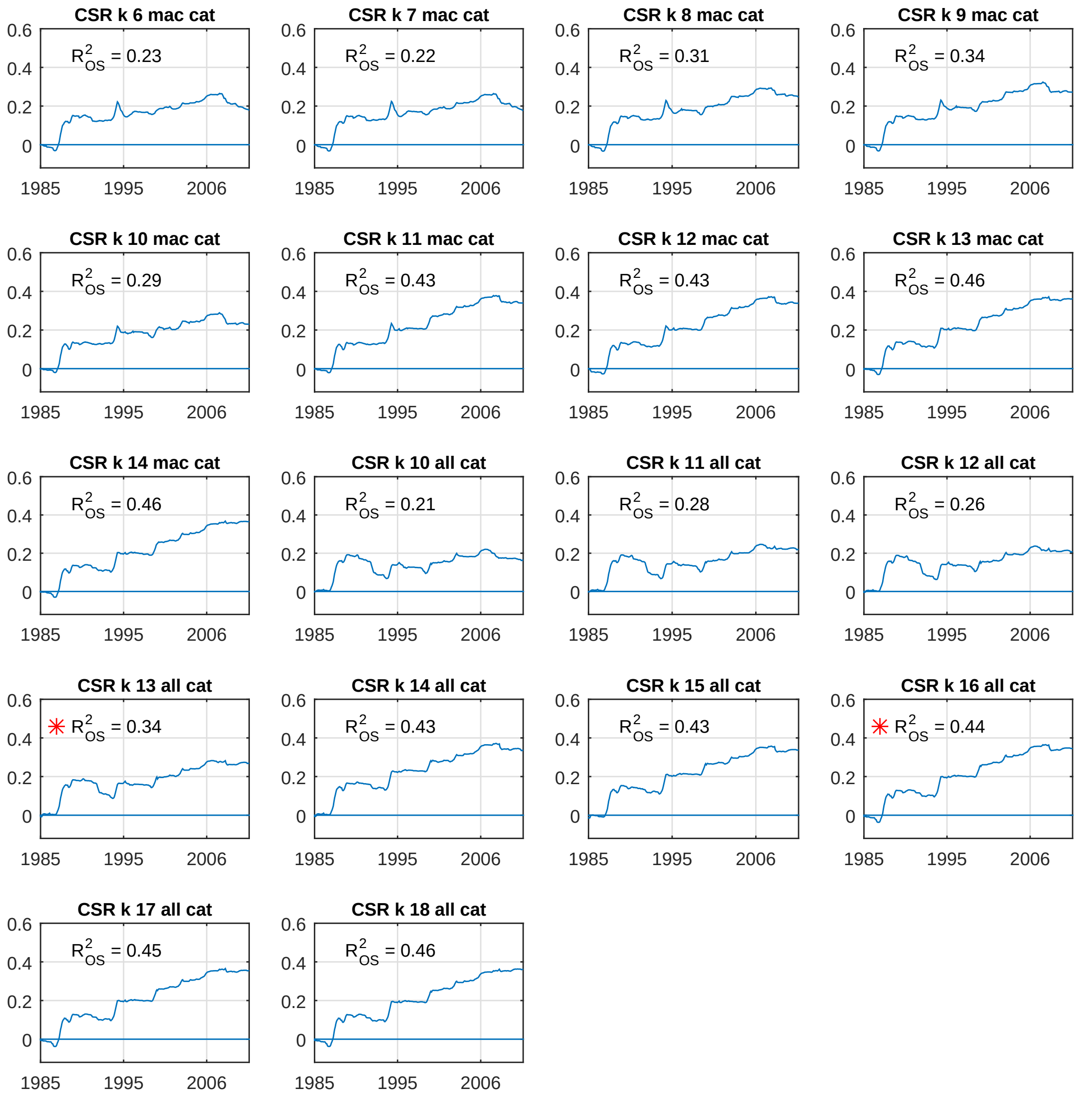


Figure 5-4: Continued.

Examining the recursive OLS forecast sets first shows that the four forecast sets that are constructed with 75% shrinkage indeed perform worse than the rest of the forecast sets. When the four forecast sets to which no shrinkage is applied are examined, it can be observed that, with one exception, they perform a lot better than the ones with 75% shrinkage and only slightly worse than those with 25% and 50% shrinkage. The forecast set that is constructed with no shrinkage, a rolling window of 36 and the macroeconomic predictors, even seems to perform as well as, or even better than those with 50% shrinkage, but it is not included in the model confidence set. On the whole the difference between the forecast sets that are constructed with the macroeconomic predictors, and their counterparts which are constructed with both the macroeconomic and financial predictors, is very small; in the cases where the difference is more than 0.01, the macroeconomic forecast set performs better.

If the complete subset regression forecast sets are examined, is it apparent that increasing the number of regressors improves the performance. In the case when only macroeconomic predictors are used, the forecast sets where $k = 11, ..., 14$, out of 14, perform the best. The same is the case when looking at the forecast sets that were constructed with both the macroeconomic and financial predictors. In this case, out of 18 possible predictors, the forecast sets that perform the best are the ones where the amount of regressors are 14 through 18. To examine the difference between the corresponding forecast sets, the macroeconomic $k = 14$ is compared to the all $k = 18$, as this corresponds to adding the four financial predictor variable sets. As was the case with the recursive OLS forecast sets, the difference is practically none, and the same happen when the other corresponding forecast sets are compared. The model confidence set includes only two forecast sets that were constructed using the complete subset regression method, both constructed with both the macroeconomic and financial set of predictors and 13 and 16 regressors respectively. This is remarkable as the performance of the $k = 16$ forecast set seems very similar to eight forecast sets which have a R^2_{OS} of mre than 0.40. And the forecast set which is included and was constructed with 13 regressors seems to perform significantly worse throughtout the whole period than those forecast sets. Comparing the results of all 24 forecast sets and the model confidence set results it is even more surprising that the two complete subset regression forecast sets are included when the no shrinkage recursive OLS forecast sets are excluded, as those forecast sets seem to outperform the benchmark more consistently.

5-2 Comparison with other maturities

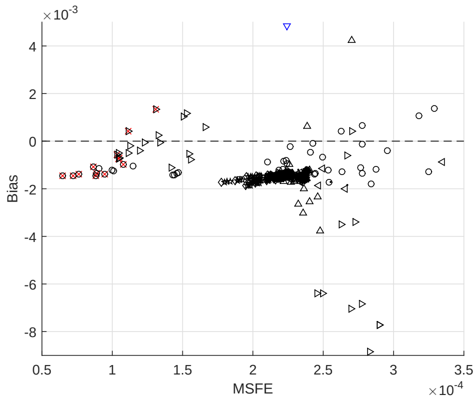
In this report the same predictor variables are used to construct excess bond returns with four different maturities, namely 2, 3, 4 and 5 years. In the previous section the results of the forecast sets for the 5 year maturity are described. Here the results of the three other maturities are briefly described and these are compared with those of the 5 year maturity.

Figure 5-5 presents the results of the forecast sets for the 2, 3 and 4 year maturity and the results of their respective model confidence sets. It is apparent straightaway that the results are very similar to each other and to those of the 5 year maturity. Like with the 5 year maturity, for each of these maturities a large part of the forecast set results is clustered in one point, which in terms of MSFE is almost the same as that of the benchmark. For each maturity it is also the case that there are a number of forecast sets that have a slighty different bias or a slightly worse MSFE. But, as with the results for the 5 year maturity, for each of these maturities there is a group of exactly 34 forecast sets that outperform both the benchmark and the other 719 forecast set in terms of MSFE. And in each case these are again only recursive OLS and complete subset regression forecast sets. Moreover these 34 are exactly the same forecast sets for each maturity. In each case it is still clear that the recursive OLS forecast sets have a smaller MSFE than the complete subset regression forecast sets, with the exception of the recursive OLS forecast sets to which 75% shrinkage was applied.

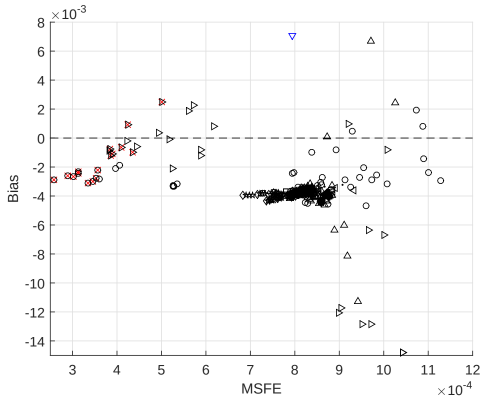
Table 5-3: A breakdown of the model confidence sets for the four different year maturities

	SIM	ROLS	BG1	BG2	BG3	BG4	BG5	CSR	Total
2 years	0	8	0	0	0	0	0	5	13
3 years	0	9	0	0	0	0	0	7	16
4 years	1	10	0	0	3	0	0	5	19
5 years	0	8	0	0	0	0	0	2	10

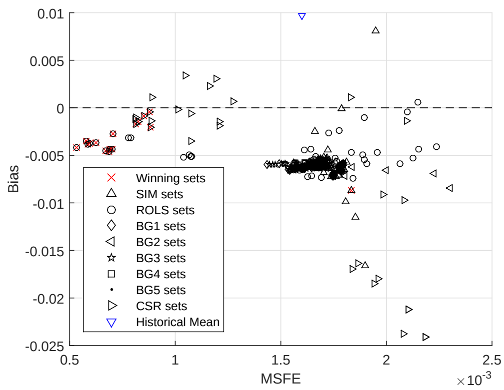
Figure 5-5 also shows which forecast sets are included in the model confidence set for each maturity, and Table 5-3 presents the breakdown by forecast combination method of these



(a) 2 year maturity



(b) 3 year maturity



(c) 4 year maturity

Figure 5-5: MSFE and bias of all the forecast sets of the 2, 3 and 4 year maturity bonds. The model confidence sets, $\widehat{\mathcal{M}}^{*,(2)}_{95\%}$, $\widehat{\mathcal{M}}^{*,(3)}_{95\%}$ and $\widehat{\mathcal{M}}^{*,(4)}_{95\%}$ are represented by the red marks

model confidence sets. The cumulative squared error of the benchmark minus the cumulative squared of these forecast sets are presented in Appendix C.
The model confidence set results are also similar for each maturity. They exclude the recursive OLS forecast sets with 75% shrinkage, and they each keep between eight and ten of the twelve remaining recursive OLS forecast set. With the complete subset regression forecast sets, the 2, 3 and 4 year maturities include a few more forecast sets than the 5 year maturity did. But they all exclude some of these sets that seem to do at least as well as some of the forecast sets that are included.
Finally in the case of the 4 year maturity the model confidence set includes four forecast sets from two completely different weighting schemes. One constructed with the simple best performing weighting scheme and three with the BG3 weighting scheme, and on top of that all four are constructed with only the financial set of predictor variables. However if Figure 5-5c is examined more closely, it is apparent that the results of the simple best performing forecast set lie just under the cluster of results, and the results of the three BG3 forecast sets are within the cluster. In addition the R^2_{OS} of each of these four forecast sets is negative, indicating that they all perform worse than the benchmark.

5-3 Robustness Tests

To test the robustness of the results in this report, this section will first examine the effect of applying different signficance levels to the model confidence set procedure and examining the differences in the resulting model confidence sets. And second the robustness of the forecast combination methods will be tested. This is done by dividing the entire period into smaller subperiods and re-applying each forecast combination method to the subperiods.

5-3-1 Significance Level

The model confidence set procedure is re-applied twice to the combined forecast sets, the first time α is set to 0.10 and the second time to 0.025, corresponding to a 90% and 97.5%

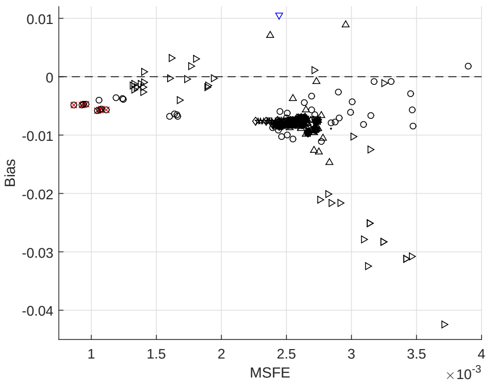
significance level respectively. Figure 5-6 presents the results of these model confidence set procedures. Table 5-4 gives a breakdown of these model confidence sets by forecast combination method.

The 90% model confidence set excludes two extra forecast sets in addition to the 95% model confidence set. These are the two complete subset regression forecast sets that remained, and the result is that the 90% model confidence set contains eight forecast sets, which are all constructed with the recursive OLS weighting scheme. Examining Figure 5-6a these are also the forecast sets which have the lowest MSFE, with the exception of a single other forecast set. These eight forecast sets are the forecast sets that also consistently have the lowest MSFEs in each of the other maturities, so it might be beneficial to apply the model confidence set procedure with a significance level of 90% instead of 95% when applied to such large sets in the future.

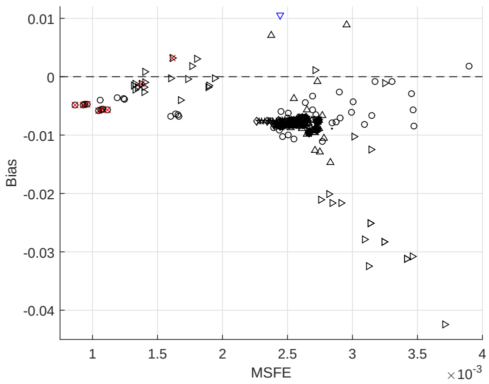
As can be expected from the 97.5% model confidence set, more forecast set are included than in the other two. In fact, forecast sets constructed with each forecast combination method are included. However if Figure 5-6c is examined, it is clear that many of the forecast sets that are included do not perform better than the other forecast sets. In fact many of the recursive OLS and complete subset regression forecast sets that perform a lot better are still not included in the 97.5% model confidence set.

Table 5-4: A breakdown of the model confidence sets for the 5 year maturity at three different levels of α

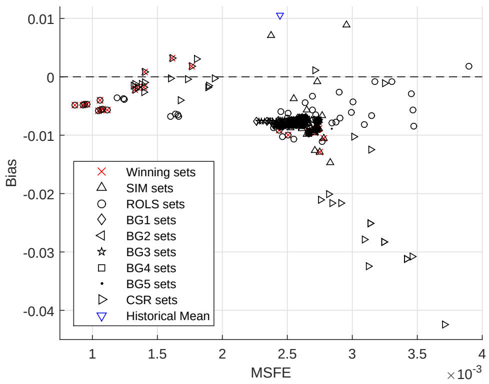
Significance level	SIM	ROLS	BG1	BG2	BG3	BG4	BG5	CSR	Total
90%	0	8	0	0	0	0	0	0	8
95%	0	8	0	0	0	0	0	2	10
97.5%	9	11	14	6	35	11	4	6	96



(a) MCS results at 90% significance level



(b) MCS results at 95% significance level

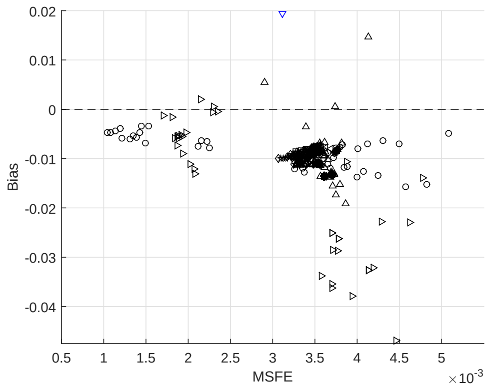


(c) MCS results at 97.5% significance level

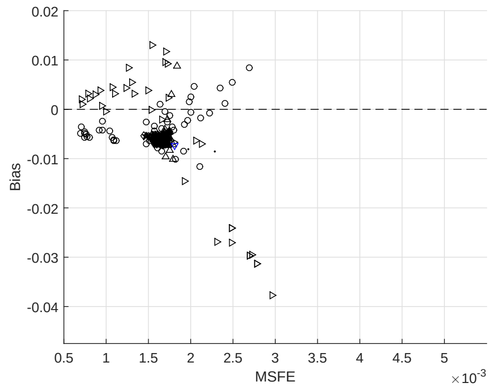
Figure 5-6: MSFE and bias of all the forecast sets of the 5 year maturity bonds. The model confidence sets, $\widehat{\mathcal{M}}^{*,(5)}_{90\%}$, $\widehat{\mathcal{M}}^{*,(5)}_{95\%}$ and $\widehat{\mathcal{M}}^{*,(5)}_{97.5\%}$ are represented by the red marks

5-3-2 Time period

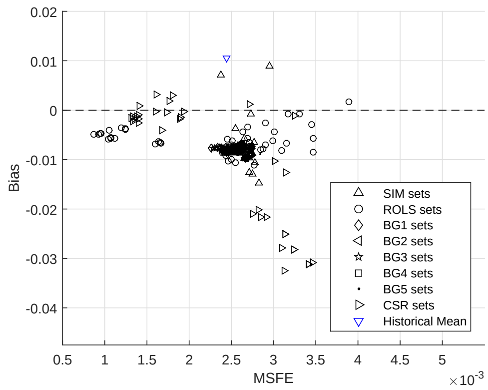
The forecast combination methods in this report are all applied and verified with a single period. To verify whether the results are consistent when the methods are applied to different datasets, the period is now divided into subperiods. The forecast sets in this report are all constructed for the period from Febuary 1985 till December 2011, this is referred to as the *Complete* period. To test the robustness of the results, the period is divided into two parts, Febuary 1985 till July 1998, and August 1998 till December 2011, these are referred to as the *First* period and *Second* period respectively. The forecast sets are reconstructed for each of these periods. The results are presented in Figure 5-7.



(a) *First* period



(b) *Second* period



(c) *Full* period

Figure 5-7: MSFE and bias of all the forecast sets of the 5 year maturity bonds for the two subperiods and the complete period

When the *First* period results are examined in Figure 5-7a, it observed that as in the *Full* period, the recursive OLS and complete subset regression forecast sets outperform the benchmark and the other forecast sets. Another similarity to the *Full* period is that the other forecast sets all at best have the same MSFE as the benchmark, but most of them have a higher MSFE. Some of the complete subset regression forecasts do however have a larger bias than that of the recursive OLS forecast sets; this is in contrast to the *Full* period, as the complete subset regression forecast sets each have a smaller bias in this case.

In the *Second* period all the forecast sets have smaller MSFEs than in the other two periods. This could indicate that the excess bond returns were less volatile in this subperiod. But, as was the case in the *First* and the *Full* period, the recursive OLS and complete subset regression forecast sets outperform the benchmark as well as the other forecast sets. In this period the best performing complete subset regression methods have about the same MSFE as the best performing recursive OLS forecast sets, and the bias of these sets is even slightly smaller.

6. Conclusions

The objective of this report is to see whether forecast combinations can be a solution to using large sets of predictors to forecast excess bond returns. The results in this report show that there are two methods of forecast combinations that are able to construct accurate and robust forecasts of one-year excess U.S. government bond returns and are able to do this consistently for different bond maturities. This conclusion is the result of the analysis of four questions: First, do macroeconomic and financial variables contain useful and complementary information which can be used by forecast combination methods to construct accurate excess bond return forecasts? Second, when applying forecast combination methods to forecasts, which are constructed from predictor variables, is it beneficial to first reduce the number of forecasts by constructing clustered forecasts with either a category or performance based cluster method? Third, if a large number of forecast sets of the excess bond returns are constructed will some of them outperform both the other forecast sets and the benchmark, and is the model confidence set procedure able to identify these forecast sets? And finally, if some method(s) construct forecast sets that outperform the other forecast set(s), will this method or these methods be able to construct forecast set(s) that consistently outperform both the other forecast sets and the benchmark for different maturities and time periods? The analysis of these questions will each be discussed individually.

Two types of predictor variables were available in this report, macroeconomic and financial. The combined forecast sets were constructed using both of these separately, as well as simultaneously. The conclusion that can be drawn from this, is that the macroeconomic predictor variables provide useful information when constructing excess bond return forecasts, whereas this is not the case for the financial predictor variables. The performance of each of the forecast sets that are constructed using solely the financial predictors for a given maturity are almost identical in terms of MSFE and bias. In addition the financial predictor variables add no complementary information to the macroeconomic variables. This is a result of the fact that for each forecast set that is constructed with the macroeconomic variables, its counterpart, which is constructed with the same cluster method, forecast combination method and other parameter settings, but uses both the macroeconomic and financial predictors, has either a similar or a worse performance. This suggests that the financial predictors do not contain useful or complementary information for the prediction of U.S. government bonds. As other papers that only used a small amount of financial variables to predict bond returns reported positive results, a recommendation for future research is that the effects of adding a single or a couple of financial variables, such as Cochrane and Piazzesi [2005]'s single factor model, to the macroeconomic variables should be studied.

The forecast combination methods are applied to three sets of forecast sets: the 279 forecast sets that are constructed straight from the predictor variables, the 18 category clustered forecast sets and the 20 forecast sets which are clustered according to past performance. One conclusion that can be drawn from the results is that clustering the forecast sets according to their past performance does not significantly improve the performance of the resulting combined forecast sets. For most of the forecast combination methods, applying the category cluster method does not improve the performance either; however there are a number of combined forecast sets that are constructed using the category cluster method whose performance is significantly better than that of the other combined forecast sets. At each maturity the final model confidence set consists solely of these combined forecast sets that were clustered by category. It is possible that predictor variables that fall into the same category contain the same information and by combining them, the amount of overlapping information is reduced.

Twelve methods of forecast combinations are used to construct the forecast sets. Of these twelve, there are two which constructed forecast sets that outperform both the other forecast sets and the benchmark are constructed. These are the recursive OLS weighting scheme and the complete subset regression method. The forecast sets are constructed when these two methods are applied to the category clustered forecast sets. These are the forecast sets that are included in the final model confidence sets of each maturity. When the 'best performing' forecast sets are selected based on their MSFE, at each maturity the same 34 forecast sets are selected, 16 recursive OLS and 18 complete subset regression sets. None of the other forecast sets perform significantly better than either each other or the benchmark.

Within the different recursive OLS forecast sets, eight parameter settings produce the most consistent results. These are those which are constructed with a rolling window of 36 or 60 months which use the macroeconomic predictor variables or both the macroeconomic and financial predictors, and with either 25% or 50% shrinkage towards equal weights. In the case of the 5 year maturity, the R_{OS}^2 is between 0.54 and 0.65 for these forecast sets. The use of shrinkage towards equal weights improves the performance of these forecast sets and it can therefore be concluded that the use of shrinkage can be a valuable addition when using forecast combinations to forecast excess bond returns.

The forecast sets that were constructed with the complete subset regression perform slightly worse in most cases than the recursive OLS forecast sets, but they still outperform all the other forecast sets and the benchmark. Additionally the bias of these forecast sets is on the whole lower than those of the recursive OLS forecast sets. The complete subset regression forecast sets, constructed with either the macroeconomic predictor or both the macroeconomic and financial predictors sets, are included in the model confidence sets. In both cases, increasing the number of predictor variables when performing the complete subset regression method improves the performance. The most accurate forecast sets are constructed when the maximum number of regressors is used and substracting regressors slowly diminishes the performance. The complete subset regression method constructs different forecasts with each set of k predictors and then applies equal weights to combine them. Further research could be done to examine whether applying other weighting schemes, such as the

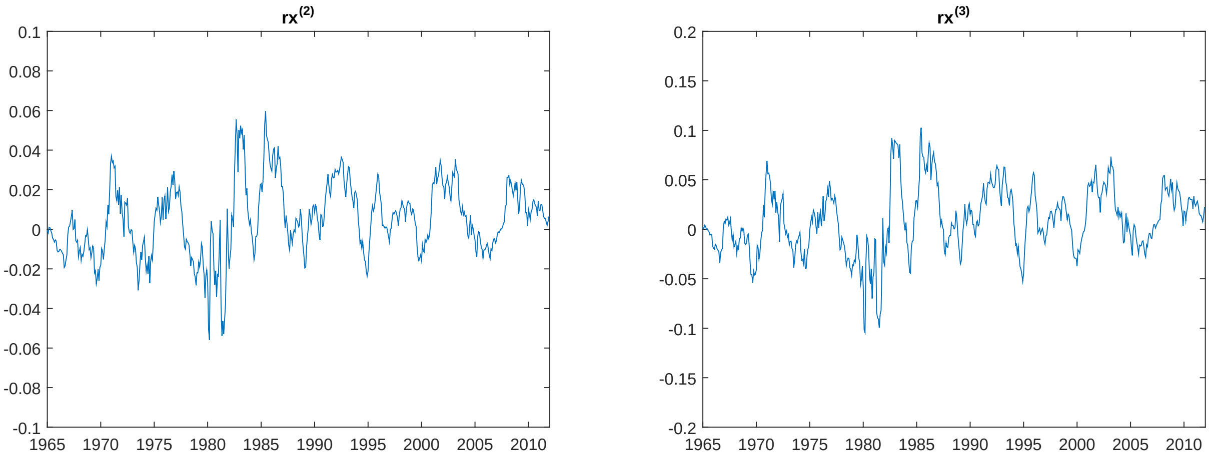
recursive OLS weighting scheme, could improve the performance of these forecasts.
A conclusion of using these two methods is that the recursive OLS weighting scheme provides more accurate forecasts, but there is certainly also merit in applying the complete subset regression method to forecast the excess bond returns. The smaller bias that the complete subset regression achieves could be useful in some cases.

From the results it is clear that the recursive OLS weighting scheme and the complete subset regression method are able to construct forecast sets that consistently outperform the other forecast sets for different maturities. The model confidence set procedure is also able to identify these forecast sets. With the exception of the 4 year maturity, the model confidence set procedure only included forecast sets constructed with one of these two forecast combination methods. However the model confidence set procedure seems to be inconsistent in which forecast sets it includes. Some of the forecast sets is excludes have very similar results to forecast sets which it includes. Sometimes forecast sets are even excluded, which in terms of MSFE, seem to significantly outperform other forecast sets, which the model confidence set procedure does include. In some cases, such as the 4 year maturity, four forecast sets are included, which clearly do not significantly outperform the other forecast sets. Likewise, when the significance level is increased from 95% to 97.5%, the number of forecast sets included increases from ten to 96, but these 96 forecast sets are certainly not the 96 forecast sets which perform the best in terms of MSFE. When the significance level is set to 90% however, eight forecast sets are included. These eight forecast sets are the recursive OLS forecast sets that consistently outperform the other forecast sets, including the other recursive OLS and complete subset regression forecast set, at each maturity. This indicates that the model confidence set procedure has trouble accurately selecting the 'best performing' forecast sets at a 95% significance level and does a better job of this when applied with a 90% significance level.

The same conclusion can be drawn about the robustness of the recursive OLS and complete subset regression combination method, when used to construct forecast sets for subsets of the time period. The two forecast set methods provide the forecast sets that outperform the benchmark as well as the other forecast sets for each subperiod. Therefore, the two methods can also provide accurate forecasts for other datasets. In the *Second* period the complete subset regression forecast sets with the best results have a MSFE which is the same as that of the best performing recursive OLS forecast sets.

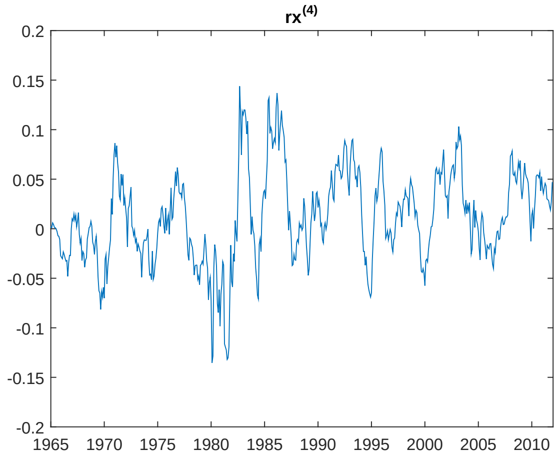
A. Data

A-1 Excess Bond Return Properties



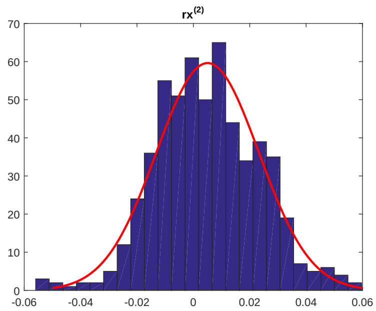
(a) 2 year maturity **(b)** 3 year maturity

Figure A-1: One-year excess bond returns, for the different maturities

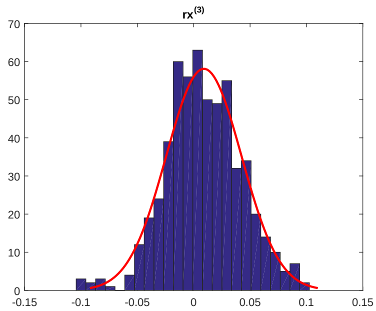


(c) 4 year maturity

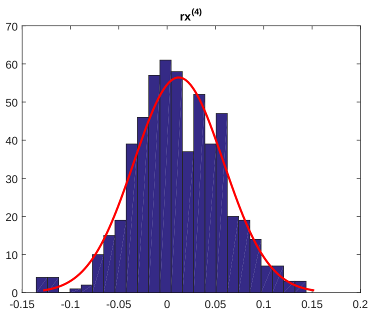
Figure A-1: Continued.



(a) 2 year maturity

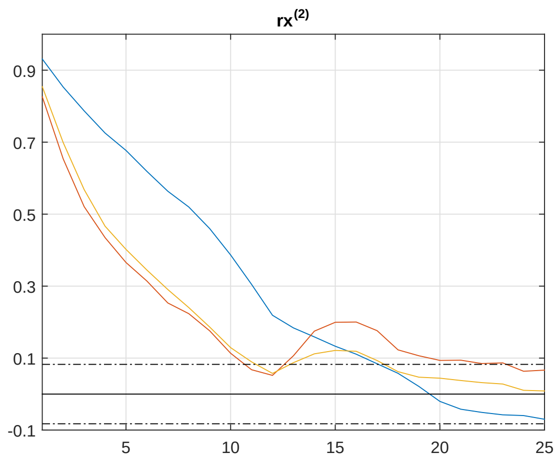


(b) 3 year maturity

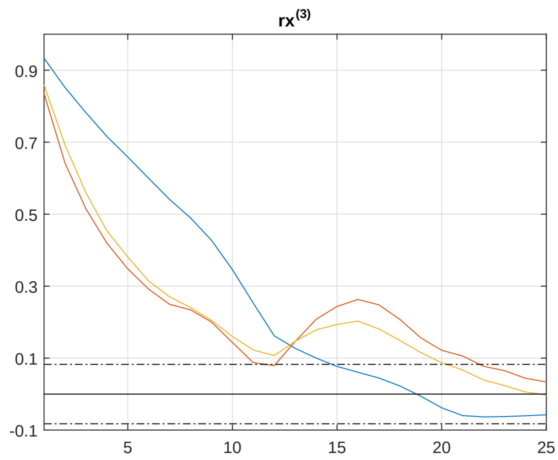


(c) 4 year maturity

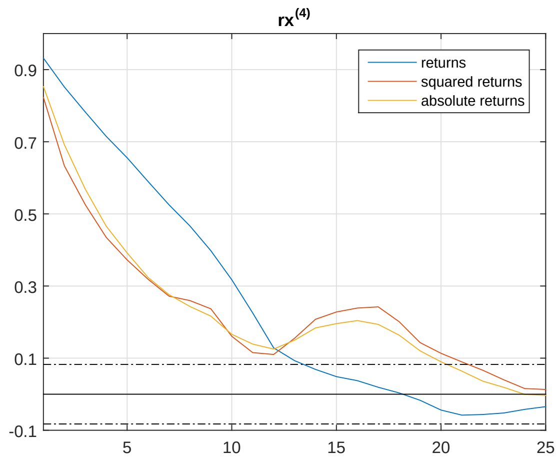
Figure A-2: Histograms of the one-year excess bond returns for the different maturities



(a) 2 year maturity



(b) 3 year maturity



(c) 4 year maturity

Figure A-3: Autocorrelations of the sets of returns, squared returns and absolute returns for the different maturity

A-2 Predictor Variables

Table A-1: Macroenconomic predictor variables

Category	Series Number	Short name	Trans	Description
Out	1	PI	Δln	Personal Income (AR, Bil. Chain 2000 $) (TCB)
Out	2	PI less transfers	Δln	Personal Income Less Transfer Payments (AR, Bil. Chain 2000 $) (TCB)
PCE	3	Real Consumption	Δln	Real Consumption (AC) a0m224/gmdc (a0m224 is from TCB)
Mon	4	M&T sales	Δln	Manufacturing and Trade Sales (Mil. Chain 1996 $) (TCB)
RTS	5	Retail sales	Δln	Sales of Retail Stores (Mil. Chain 2000 $) (TCB)
Out	6	IP: total	Δln	Industrial Production Index - Total Index
Out	7	IP: products	Δln	Industrial Production Index - Products, Total
Out	8	IP: final prod	Δln	Industrial Production Index - Final Products
Out	9	IP: cons gds	Δln	Industrial Production Index - Consumer Goods
Out	10	IP: cons dble	Δln	Industrial Production Index - Durable Consumer Goods
Out	11	IP: cons nondble	Δln	Industrial Production Index - Nondurable Consumer Goods
Out	12	IP: bus eqpt	Δln	Industrial Production Index - Business Equipment
Out	13	IP: matls	Δln	Industrial Production Index - Materials
Out	14	IP: dble matls	Δln	Industrial Production Index - Durable Goods Materials
Out	15	IP: nondble matls	Δln	Industrial Production Index - Nondurable Goods Materials
Out	16	IP: mfg	Δln	Industrial Production Index - Manufacturing (Sic)
Out	17	IP: res util	Δln	Industrial Production Index - Residential Utilities
Out	18	IP: fuels	Δln	Industrial Production Index - Fuels
Out	19	NAPM prodn	lv	Napm Production Index (Percent)
Out	20	Cap util	Δlv	Capacity Utilization (Mfg.) (TCB)
EMP	21	Help wanted indx	Δlv	Index of Help-Wanted Advertising in Newspapers (1967=100;Sa)
EMP	22	Help wanted/unemp	Δlv	Employment: Ratio; Help-Wanted Ads:No. Unemployed Clf
EMP	23	Emp CPS total	Δln	Civilian Labor Force: Employed, Total (Thous.,Sa)
EMP	24	Emp CPS nonag	Δln	Civilian Labor Force: Employed, Nonagric. Industries (Thous.,Sa)
EMP	25	U: all	Δlv	Unemployment Rate: All Workers, 16 Years & Over (%,Sa)
EMP	26	U: mean duration	Δlv	Unemploy. By Duration: Average (Mean) Duration in Weeks (Sa)
EMP	27	U < 5 wks	Δln	Unemploy. By Duration: Persons Unempl.Less than 5 Wks (Thous.,Sa)
EMP	28	U 5-14 wks	Δln	Unemploy. By Duration: Persons Unempl. 5 to 14 Wks (Thous.,Sa)
EMP	29	U 15+ wks	Δln	Unemploy. By Duration: Persons Unempl. 15 Wks + (Thous.,Sa)
EMP	30	U 15-26 wks	Δln	Unemploy. By Duration: Persons Unempl. 15 to 26 Wks (Thous.,Sa)
EMP	31	U 27+ wks	Δln	Unemploy. By Duration: Persons Unempl. 27 Wks + (Thous,Sa)
EMP	32	UI claims	Δln	Average Weekly Initial Claims, Unemploy. Insurance (Thous.) (TCB)
EMP	33	Emp: total	Δln	Employees on Nonfarm Payrolls: Total Private
EMP	34	Emp: gds prod	Δln	Employees on Nonfarm Payrolls - Goods-Producing
EMP	35	Emp: mining	Δln	Employees on Nonfarm Payrolls âĂŞ Mining
EMP	36	Emp: const	Δln	Employees on Nonfarm Payrolls - Construction
EMP	37	Emp: mfg	Δln	Employees on Nonfarm Payrolls - Manufacturing

Table A-1: continued

Category	Series Number	Short name	Trans	Description
EMP	35	Emp: mining	Δln	Employees on Nonfarm Payrolls - Mining
EMP	36	Emp: const	Δln	Employees on Nonfarm Payrolls - Construction
EMP	37	Emp: mfg	Δln	Employees on Nonfarm Payrolls - Manufacturing
EMP	38	Emp: dble gds	Δln	Employees on Nonfarm Payrolls - Durable Goods
EMP	39	Emp: nondbles	Δln	Employees on Nonfarm Payrolls - Nondurable Goods
EMP	40	Emp: services	Δln	Employees on Nonfarm Payrolls - Service-Providing
EMP	41	Emp: TTU	Δln	Employees on Nonfarm Payrolls - Trade, Transportation, and Utilities
EMP	42	Emp: wholesale	Δln	Employees on Nonfarm Payrolls - Wholesale Trade
EMP	43	Emp: retail	Δln	Employees on Nonfarm Payrolls - Retail Trade
EMP	44	Emp: FIRE	Δln	Employees on Nonfarm Payrolls - Financial Activities
EMP	45	Emp: Govt	Δln	Employees on Nonfarm Payrolls - Government
EMP	46	Agg wkly hours	Δln	Employee Hours in Nonag. Establishments (AR, Bil. Hours) (TCB)
EMP	47	Avg hrs	lv	Avg Weekly Hrs of Prod or Nonsup Workers on Private Nonfarm Payrolls
EMP	48	Overtime: mfg	Δlv	AvgWeekly Hrs of Prod or Nonsup Workers on Private Nonfarm Payrolls
EMP	49	Avg hrs: mfg	lv	Average Weekly Hours, Mfg. (Hours) (TCB)
EMP	50	NAPM empl	lv	Napm Employment Index (Percent)
HSS	51	Starts: nonfarm	ln	Housing Starts:Nonfarm(1947-58);Total Farm&Nonfarm(1959-) (Thous.,Saar)
HSS	52	Starts: NE	ln	Housing Starts:Northeast (Thous.U.)S.A.
HSS	53	Starts: MW	ln	Housing Starts:Midwest(Thous.U.)S.A.
HSS	54	Starts: South	ln	Housing Starts:South (Thous.U.)S.A.
HSS	55	Starts: West	ln	Housing Starts:West (Thous.U.)S.A.
HSS	56	BP: total	ln	Housing Authorized: Total New Priv Housing Units (Thous.,Saar)
HSS	57	BP: NE	ln	Houses Authorized by Build. Permits:Northeast (Thou.U.)S.A
HSS	58	BP: MW	ln	Houses Authorized by Build. Permits:Midwest (Thou.U.)S.A.
HSS	59	BP: South	ln	Houses Authorized by Build. Permits:South (Thou.U.)S.A.
HSS	60	BP: West	ln	Houses Authorized by Build. Permits:West (Thou.U.)S.A.
Out	61	PMI	lv	Purchasing Managers Index (Sa)
Ord	62	NAPM new ordrs	lv	Napm New Orders Index (Percent)
Ord	63	NAPM vendor del	lv	Napm Vendor Deliveries Index (Percent)
RTS	64	NAPM Invent	lv	Napm Inventories Index (Percent)
Ord	65	Orders: cons gds	Δln	Mfrs New Orders, Consumer Goods and Materials (Bil. Chain 1982 $) (TCB)
Ord	66	Orders: dble gds	Δln	Mfrs New Orders, Durable Goods Industries (Bil. Chain 2000 $) (TCB)
Ord	67	Orders: cap gds	Δln	Mfrs New Orders, Nondefense Capital Goods (Mil. Chain 1982 $) (TCB)
Ord	68	Unf orders: dble	Δln	Mfrs Unfilled Orders, Durable Goods Indus. (Bil. Chain 2000 $) (TCB)
Inv	69	M&T invent	Δln	Manufacturing and Trade Inventories (Bil. Chain 2000 $) (TCB)
Inv	70	M&T invent/sales	Δlv	Ratio, Mfg. and Trade Inventories to Sales (Based on Chain 2000 $) (TCB)
Mon	71	M1	$\Delta^2 ln$	Money Stock: M1(Curr,Trav.Cks,Dem Dep,Other Ckable Dep) (Bil$,Sa)
Mon	72	M2	$\Delta^2 ln$	Money Stock:M2(M1+Onite Rps,Euro$,G/P&B/D Mmmfs&Sav &Sm Time Dep(Bil$,Sa)
Mon	73	Currency	$\Delta^2 ln$	Money Stock: M3(M2+Lg Time Dep,Term Rps&Inst Only Mmmfs) (Bil$,Sa)
Mon	74	M2 (real)	Δln	Money Supply - M2 in 1996 Dollars (Bci)
Mon	75	MB	$\Delta^2 ln$	Monetary Base, Adj. tor Reserve Requirement Changes (Mil$,Sa)
Mon	76	Reserves tot	$\Delta^2 ln$	Depository Inst Reserves:Total, Adj. tor Reserve Req Chgs (Mil$,Sa)
Mon	77	Reserves nonbor	$\Delta^2 ln$	Depository Inst Reserves:Nonborrowed,Adj. Res Req Chgs (Mil$,Sa)
Mon	78	C&I loan plus	lv	Commercial & Industrial Loans Oustanding in 1996 Dollars (Bci)
Mon	79	Δ C&I loans	$\Delta^2 ln$	Wkly Rp Lg Coml Banks:Net Change Coml & Indus Loans (Bil$,Saar)
Mon	80	Cons credit	Δlv	Consumer Credit Outstanding - Nonrevolving (G19)

Table A-1: continued

Category	Series Number	Short name	Trans	Description
Mon	81	Inst cred/PI	Δln	Ratio, Consumer Installment Credit to Personal Income (Pct.) (TCB)
SPr	82	S&P 500	Δln	S&Ps Common Stock Price Index: Composite (1941-43=10)
SPr	83	S&P: indust	Δln	S&Ps Common Stock Price Index: Industrials (1941-43=10)
SPr	84	S&P div yield	Δlv	S&Ps Composite Common Stock: Dividend Yield (% per Annum)
SPr	85	S&P PE ratio	Δln	S&Ps Composite Common Stock: Price-Earnings Ratio (%,Nsa)
Int	86	Fed Funds	Δlv	Interest Rate: Federal Funds (Effective) (% per Annum,Nsa)
Int	87	Comm paper	Δlv	Cmmercial Paper Rate (AC)
Int	88	3 mo T-bill	Δlv	Interest Rate: U.S.Treasury Bills, Sec Mkt, 3-Mo. (% per Ann,Nsa)
Int	89	6 mo T-bill	Δlv	Interest Rate: U.S.Treasury Bills, Sec Mkt, 6-Mo. (% per Ann,Nsa)
Int	90	1 yr T-bond	Δlv	Interest Rate: U.S.Treasury Const Maturities, 1-Yr. (% per Ann,Nsa)
Int	91	5 yr T-bond	Δlv	Interest Rate: U.S.Treasury Const Maturities, 5-Yr. (% per Ann,Nsa)
Int	92	10 yr T-bond	Δlv	Interest Rate: U.S.Treasury Const Maturities, 10-Yr. (% per Ann,Nsa)
Int	93	Aaa bond	Δlv	Bond Yield: Moodys Aaa Corporate (% per Annum)
Int	94	Baa bond	Δlv	Bond Yield: Moodys Baa Corporate (% per Annum)
Int	95	CP-FF spread	lv	cp90-fyff (AC)
Int	96	3 mo-FF spread	lv	fygm3-fyff (AC)
Int	97	6 mo-FF spread	lv	fygm6-fyff (AC)
Int	98	1 yr-FF spread	lv	fygt1-fyff (AC)
Int	99	5 yr-FF spread	lv	fygt5-fyff (AC)
Int	100	10 yr-FF spread	lv	fygt10-fyff (AC)
Int	101	Aaa-FF spread	lv	fyaaac-fyff (AC)
Int	102	Baa-FF spread	lv	fybaac-fyff (AC)
FX	103	Ex rate: avg	Δln	United States;Effective Exchange Rate (Merm) (Index No.)
FX	104	Ex rate: Switz	Δln	Foreign Exchange Rate: Switzerland (Swiss Franc per U.S.$)
FX	105	Ex rate: Japan	Δln	Foreign Exchange Rate: Japan (Yen per U.S.$)
FX	106	Ex rate: UK	Δln	Foreign Exchange Rate: United Kingdom (Cents per Pound)
FX	107	EX rate: Canada	Δln	Foreign Exchange Rate: Canada (Canadian $ per U.S.$)
Pri	108	PPI: fin gds	$\Delta^2 ln$	Producer Price Index: Finished Goods (82=100,Sa)
Pri	109	PPI: cons gds	$\Delta^2 ln$	Producer Price Index: Finished Consumer Goods (82=100,Sa)
Pri	110	PPI: int matâĂŹls	$\Delta^2 ln$	Producer Price Index: Intermed Mat.Supplies & Components (82=100,Sa)
Pri	111	PPI: crude matâĂŹls	$\Delta^2 ln$	Producer Price Index: Crude Materials (82=100,Sa)
Pri	112	Spot market price	$\Delta^2 ln$	Spot market price index: bls & crb: all commodities (1967=100)
Pri	113	PPI: nonferrous	$\Delta^2 ln$	Index Of Sensitive Materials Prices (1990=100) (Bci-99a)
Pri	114	NAPM com price	lv	Napm Commodity Prices Index (Percent)
Pri	115	CPI-U: all	$\Delta^2 ln$	Cpi-U: All Items (82-84=100,Sa)
Pri	116	CPI-U: apparel	$\Delta^2 ln$	Cpi-U: Apparel & Upkeep (82-84=100,Sa)
Pri	117	CPI-U: transp	$\Delta^2 ln$	Cpi-U: Transportation (82-84=100,Sa)
Pri	118	CPI-U: medical	$\Delta^2 ln$	Cpi-U: Medical Care (82-84=100,Sa)
Pri	119	CPI-U: comm.	$\Delta^2 ln$	Cpi-U: Commodities (82-84=100,Sa)
Pri	120	CPI-U: dbles	$\Delta^2 ln$	Cpi-U: Durables (82-84=100,Sa)
Pri	121	CPI-U: services	$\Delta^2 ln$	Cpi-U: Services (82-84=100,Sa)
Pri	122	CPI-U: ex food	$\Delta^2 ln$	Cpi-U: All Items Less Food (82-84=100,Sa)
Pri	123	CPI-U: ex shelter	$\Delta^2 ln$	Cpi-U: All Items Less Shelter (82-84=100,Sa)
Pri	124	CPI-U: ex med	$\Delta^2 ln$	Cpi-U: All Items Less Medical Care (82-84=100,Sa)
Pri	125	PCE defl	$\Delta^2 ln$	Pce, Impl Pr Defl:Pce (1987=100)
Pri	126	PCE defl: dlbes	$\Delta^2 ln$	Pce, Impl Pr Defl:Pce; Durables (1987=100)
Pri	127	PCE defl: nondble	$\Delta^2 ln$	Pce, Impl Pr Defl:Pce; Nondurables (1996=100)

Table A-1: continued

Category	Category	Category	Category	Category
Pri	128	PCE defl: service	$\Delta^2 ln$	Pce, Impl Pr Defl:Pce; Services (1987=100)
AHE	129	AHE: goods	$\Delta^2 ln$	Avg Hourly Earnings of Prod or Nonsup Workers on Private Nonfarm Payrolls - Goods-Producing
AHE	130	AHE: const	$\Delta^2 ln$	Avg Hourly Earnings of Prod or Nonsup Workers on Private Nonfarm Payrolls - Construction
AHE	131	AHE: mfg	$\Delta^2 ln$	Avg Hourly Earnings of Prod or Nonsup Workers on Private Nonfarm Payrolls Manufacturing
Oth	132	Consumer expect	Δlv	U. of Mich. Index of Consumer Excpectations (Bcd-83)

Table A-2: Financial predictor variables

Category	Series Number	Short name	Source	Description
PYD	1	D_log(DIV)	CRSP	Log difference of the sum of the dividends in the last 4 quarters (divs are not reivested)
PYD	2	D_log(P)	CRSP	Log difference of the CRSP portfolio price when dividends are not reivested
PYD	3	D_DIVreinvested	CRSP	Log difference of the sum of the dividends in the last 4 quarters (divs are reinvested)
PYD	4	D_Preinvested	CRSP	Log difference of the CRSP portfolio price when dividends are reinvested
PYD	5	d-p	CRSP	DIVreinveste - Preinveste = log(DIV) - log(P)
RF	6	R15-R11	French	Small stock value spread constructed from French database
RF	7	CP factor	CP	Piazzesi-Cochrane risk factor, quarterly average (Cochrane and Piazzesi, 2005)
RF	8	Mkt-RF	French	Fama-French market risk factor (Fama and French, 1993)
RF	9	SMB	French	Fama-French risk factor (Fama and French, 1993)
RF	10	HML	French	Fama-French risk factor (Fama and French, 1993) French
RF	11	UMD	French	Momentum risk factor, French data set

Table A-2: continued

Category	Series Number	Short name	Source	Category	Series Number	Short name	Source
I	12	Agric	French	I	44	Paper	French
I	13	Food	French	I	45	Boxes	French
I	14	Beer	French	I	46	Trans	French
I	15	Smoke	French	I	47	Whlsl	French
I	16	Toys	French	I	48	Rtail	French
I	17	Fun	French	I	49	Meals	French
I	18	Books	French	I	50	Banks	French
I	19	Hshld	French	I	51	Insur	French
I	20	Clths	French	I	52	RlEst	French
I	21	MedEq	French	I	53	Fin	French
I	22	Drugs	French	I	54	Other	French
I	23	Chems	French	S&BM	55	ports_2	French
I	24	Rubbr	French	S&BM	56	ports_4	French
I	25	Txtls	French	S&BM	57	ports_5	French
I	26	BldMt	French	S&BM	58	ports_6	French
I	27	Cnstr	French	S&BM	59	ports_7	French
I	28	Steel	French	S&BM	60	ports_8	French
I	29	Mach	French	S&BM	61	ports_9	French
I	30	ElcEq	French	S&BM	62	ports_high	French
I	31	Autos	French	S&BM	63	ports_low	French
I	32	Aero	French	S&BM	64	port2_2	French
I	33	Ships	French	S&BM	65	port2_3	French
I	34	Mines	French	S&BM	66	port2_4	French
I	35	Coal	French	S&BM	67	port2_5	French
I	36	Oil	French	S&BM	68	port2_6	French
I	37	Util	French	S&BM	69	port2_7	French
I	38	Telcm	French	S&BM	70	port2_8	French
I	39	PerSv	French	S&BM	71	port2_9	French
I	40	BusSv	French	S&BM	72	port2_high	French
I	41	Hardw	French	S&BM	73	port2_low	French
I	42	Chips	French	S&BM	74	port3_2	French
I	43	LabEq	French	S&BM	75	port3_3	French

Table A-2: continued

Category	Series Number	Short name	Source	Category	Series Number	Short name	Source
S&BM	76	port3_4	French	S&BM	112	port6_high	French
S&BM	77	port3_5	French	S&BM	113	port6_low	French
S&BM	78	port3_6	French	S&BM	114	port7_2	French
S&BM	79	port3_7	French	S&BM	115	port7_3	French
S&BM	80	port3_8	French	S&BM	116	port7_4	French
S&BM	81	port3_9	French	S&BM	117	port7_5	French
S&BM	82	port3_high	French	S&BM	118	port7_6	French
S&BM	83	port3_low	French	S&BM	119	port7_7	French
S&BM	84	port4_2	French	S&BM	120	port7_8	French
S&BM	85	port4_3	French	S&BM	121	port7_9	French
S&BM	86	port4_4	French	S&BM	122	port7_low	French
S&BM	87	port4_5	French	S&BM	123	port8_2	French
S&BM	88	port4_6	French	S&BM	124	port8_3	French
S&BM	89	port4_7	French	S&BM	125	port8_4	French
S&BM	90	port4_8	French	S&BM	126	port8_5	French
S&BM	91	port4_9	French	S&BM	127	port8_6	French
S&BM	92	port4_high	French	S&BM	128	port8_7	French
S&BM	93	port4_low	French	S&BM	129	port8_8	French
S&BM	94	port5_2	French	S&BM	130	port8_9	French
S&BM	95	port5_3	French	S&BM	131	port8_high	French
S&BM	96	port5_4	French	S&BM	132	port8_low	French
S&BM	97	port5_5	French	S&BM	133	port9_2	French
S&BM	98	port5_6	French	S&BM	134	port9_3	French
S&BM	99	port5_7	French	S&BM	135	port9_4	French
S&BM	100	port5_8	French	S&BM	136	port9_5	French
S&BM	101	port5_9	French	S&BM	137	port9_6	French
S&BM	102	port5_high	French	S&BM	138	port9_7	French
S&BM	103	port5_low	French	S&BM	139	port9_8	French
S&BM	104	port6_2	French	S&BM	140	port9_high	French
S&BM	105	port6_3	French	S&BM	141	port9_low	French
S&BM	106	port6_4	French	S&BM	142	portl_2	French
S&BM	107	port6_5	French	S&BM	143	portl_3	French
S&BM	108	port6_6	French	S&BM	144	portl_4	French
S&BM	109	port6_7	French	S&BM	145	portl_5	French
S&BM	110	port6_8	French	S&BM	146	portl_6	French
S&BM	111	port6_9	French	S&BM	147	portl_7	French

B. Methods

B-1 k-means algorithm

1. Collect $\text{MSFE}_{i,t}^{(n)}$ for $i = 1, ..., N$ in $\mathcal{V}_t^{(n)}$ and let v_i be the i-th element of $\mathcal{V}_t^{(n)}$.
2. Set the amount of desired clusters to k.
3. Determine k initial cluster centers (centroids). This is done with the following sub-algorithm:
   (a) Select a random element from $\mathcal{V}_t^{(n)}$, define this as first centroid and denote it by c_1.
   (b) Compute the squared distances from each other element to c_1. For element v_m denote this as:
   $$d(v_m, c_j) \equiv (v_m - c_1)^2 \tag{B-1}$$
   (c) Select the next centroid, c_2, at random from $\mathcal{V}_t^{(n)}$ with probability
   $$\frac{d^2(v_m, c_1)}{\sum_{j=1}^{n} d^2(v_j, c_1)} \tag{B-2}$$
   (d) Until k centroids are chosen repeat:
      i. Compute the distances from each observation to each centroid and assign each observation to its closest centroid.
      ii. For $m = 1, ..., n$ and $p = 1, ..., j-1$, select centroid j at random from $\mathcal{V}_t^{(n)}$ with probability
      $$\frac{d^2(v_m, c_p)}{\sum_{\{h; v_h \in C_p\}} d^2(v_h, c_p)} \tag{B-3}$$
      where C_p is the set of all observations closest to centroid c_p and v_m belongs to C_p. That is, select each subsequent center with a probability proporitonal to the distance from itself to the closest center that is already chosen.

4. Assigning the elements to the clusters is done in two steps, first with a batch update and second with a online update.
   (a) The batch update algorithm:
      i. Compute the squared distance of each element in $\mathcal{V}_t^{(n)}$ to each centroid.
      ii. All the elements are assigned to the cluster with the closest centroid.
      iii. Compute the average of the observations in each cluster to obtain k new centroid locations.
      iv. Repeat steps (a) through (c) until the assignment of the clusters does not change any more.
   (b) The online update algorithm:
      i. Compute the squared distance of each element in $\mathcal{V}_t^{(n)}$ to each centroid.
      ii. Each element is reassigned individually to a different centroid if the reassignment decreases the sum of distances.
      iii. Compute the average of the observations in each cluster to obtain k new centroid locations.
      iv. Repeat steps (a) through (c) until the assignment of the clusters does not change any more.

B-2 Model Confidence Set

Determining sample and bootstrap statistics:

1. A circular bootstrap scheme with block length $l = 18$ is used to create $B = 10000$ bootstrap resamples.
2. For each forecast set the variables $L_{i,t}^{(n)}$ for $i = 1, ..., m_0$ and $t = 1, ..., T$ are determined. These variables are used to calculate the sample averages for each forecast set.

$$\bar{L}_i^{(n)} \equiv \frac{1}{T}\sum_{t=1}^{T} L_{i,t}^{(n)} \quad \text{for } i = 1, ..., m_0 \tag{B-4}$$

3. The corresponding bootstrap variables are given by

$$L_{b,i,t}^{*,(n)} = L_{i,\tau_b,t}^{(n)} \quad \text{for } b = 1, ...B,\ i = 1, ..., m_0, \text{ and } t = 1, ..., T \tag{B-5}$$

   and these are used to calculate the bootstrap sample averages, $\bar{L}_{b,i}^{*,(n)} \equiv \frac{1}{T}\sum_{t=1}^{T} L_{b,i,t}^{*,(n)}$.
4. Define the relative sample loss statistics of both the forecast sets and their bootstraps as:

$$\bar{d}_{ij}^{(n)} = \bar{L}_i^{(n)} - \bar{L}_j^{(n)} \qquad \bar{d}_{b,ij}^{*,(n)} = \bar{L}_{b,i}^{*,(n)} - \bar{L}_{b,j}^{*,(n)} \tag{B-6}$$

5. Calculate the estimate of $\text{var}(\bar{d}_{ij}^{(n)})$

$$\widehat{\text{var}}\left(\bar{d}_{ij}^{(n)}\right) \equiv \frac{1}{B}\sum_{b=1}^{B}\left(\bar{d}_{b,ij}^{*,(n)} - \bar{d}_{ij}^{(n)}\right)^2 \tag{B-7}$$

Sequential testing:

1. Initialize by setting $\mathcal{M}^{(n)} = \mathcal{M}^{0,(n)}$
2. Define
$$t_{ij}^{(n)} = \frac{\bar{d}_{ij}^{(n)}}{\sqrt{\widehat{\text{var}}\left(\bar{d}_{ij}^{(n)}\right)}} \quad \text{for } i,j \in \mathcal{M}^{(n)} \tag{B-8}$$
and determine the test statistic
$$T_{R,\mathcal{M}}^{(n)} \equiv \max_{i,j\in\mathcal{M}^{(n)}} |t_{ij}^{(n)}| \tag{B-9}$$
3. Define
$$t_{b,ij}^{*,(n)} = \frac{\bar{d}_{b,ij}^{*,(n)} - \bar{d}_{ij}^{(n)}}{\sqrt{\widehat{\text{var}}\left(\bar{d}_{ij}^{(n)}\right)}} \quad \text{for } i,j \in \mathcal{M}^{(n)} \tag{B-10}$$
and determine the bootstrap estimate of $T_{R,\mathcal{M}}^{(n)}$ distribution as
$$T_{b,\mathcal{M}}^{*,(n)} \equiv \max_{i,j\in\mathcal{M}^{(n)}} |t_{b,ij}^{*,(n)}| \quad \text{for } b = 1,\dots,B \tag{B-11}$$
4. The p-value of $H_{0,\mathcal{M}^{(n)}}$ is then given by
$$P_{H_{0,\mathcal{M}^{(n)}}} \equiv \frac{1}{B}\sum_{b=1}^{B} I_{\left\{T_{R,\mathcal{M}}^{(n)} > T_{b,\mathcal{M}}^{*,(n)}\right\}} \tag{B-12}$$
where $I_{\{\bullet\}}$ is an indicatior function.
5. $H_{0,\mathcal{M}^{(n)}}$ is rejected if $P_{H_{0,\mathcal{M}^{(n)}}} < \alpha$. Then $e_{R\mathcal{M}}^{(n)} = \arg\max_{i\in\mathcal{M}^{(n)}} \sup_{j\in\mathcal{M}^{(n)}} t_{ij}^{(n)}$ is eliminated from $\mathcal{M}^{(n)}$.
6. Steps 2 through 5 are repeated until $H_{0,\mathcal{M}^{(n)}}$ is accepted. The resulting set of forecast sets is denoted $\widehat{\mathcal{M}}_{1-\alpha}^{*,(n)}$ and referred to as the $(1-\alpha)$ MCS.

C. Results

Table C-1: A breakdown of the model confidence sets for 2 year maturity

		SIM	ROLS	BG1	BG2	BG3	BG4	BG5	CSR	Total
Mac	Ind	0/5	-	0/12	-	0/36	0/8	-	0/1	0/62
	Cat	0/5	5/8	0/12	0/12	0/36	0/8	0/8	1/13	6/102
	Per	0/5	0/8	0/12	0/12	0/36	0/8	0/8	-	0/89
Fin	Ind	0/5	-	0/12	-	0/36	0/8	-	0/1	0/62
	Cat	0/5	0/8	0/12	0/12	0/36	0/8	0/8	0/3	0/92
	Per	0/5	0/8	0/12	0/12	0/36	0/8	0/8	-	0/89
All	Ind	0/5	-	0/12	-	0/36	0/8	-	0/1	0/62
	Cat	0/5	3/8	0/12	0/12	0/36	0/8	0/8	4/17	7/106
	Per	0/5	0/8	0/12	0/12	0/36	0/8	0/8	-	0/89
	Total	0/45	8/48	0/108	0/72	0/324	0/72	0/48	5/36	13/753

The final row represents the breakdown by the forecast combination method of forecast sets which are included in the final model confidence set for the 2 year maturity.

Table C-2: A breakdown of the model confidence sets for 3 year maturity

		SIM	ROLS	BG1	BG2	BG3	BG4	BG5	CSR	Total
Mac	Ind	0/5	-	0/12	-	0/36	0/8	-	0/1	0/62
	Cat	0/5	4/8	0/12	0/12	0/36	0/8	0/8	2/13	6/102
	Per	0/5	0/8	0/12	0/12	0/36	0/8	0/8	-	0/89
Fin	Ind	0/5	-	0/12	-	0/36	0/8	-	0/1	0/62
	Cat	0/5	0/8	0/12	0/12	0/36	0/8	0/8	0/3	0/92
	Per	0/5	0/8	0/12	0/12	0/36	0/8	0/8	-	0/89
All	Ind	0/5	-	0/12	-	0/36	0/8	-	0/1	0/62
	Cat	0/5	5/8	0/12	0/12	0/36	0/8	0/8	5/17	10/106
	Per	0/5	0/8	0/12	0/12	0/36	0/8	0/8	-	0/89
Total		0/45	9/48	0/108	0/72	0/324	0/72	0/48	7/36	16/753

The final row represents the breakdown by the forecast combination method of forecast sets which are included in the final model confidence set for the 3 year maturity.

Table C-3: A breakdown of the model confidence sets for 4 year maturity

		SIM	ROLS	BG1	BG2	BG3	BG4	BG5	CSR	Total
Mac	Ind	0/5	-	0/12	-	0/36	0/8	-	0/1	0/62
	Cat	0/5	5/8	0/12	0/12	0/36	0/8	0/8	1/13	6/102
	Per	0/5	0/8	0/12	0/12	0/36	0/8	0/8	-	0/89
Fin	Ind	1/5	-	0/12	-	3/36	0/8	-	0/1	4/62
	Cat	0/5	0/8	0/12	0/12	0/36	0/8	0/8	0/3	0/92
	Per	0/5	0/8	0/12	0/12	0/36	0/8	0/8	-	0/89
All	Ind	0/5	-	0/12	-	0/36	0/8	-	0/1	0/62
	Cat	0/5	5/8	0/12	0/12	0/36	0/8	0/8	4/17	9/106
	Per	0/5	0/8	0/12	0/12	0/36	0/8	0/8	-	0/89
Total		1/45	10/48	0/108	0/72	3/324	0/72	0/48	5/36	22/753

The final row represents the breakdown by the forecast combination method of forecast sets which are included in the final model confidence set for the 4 year maturity.

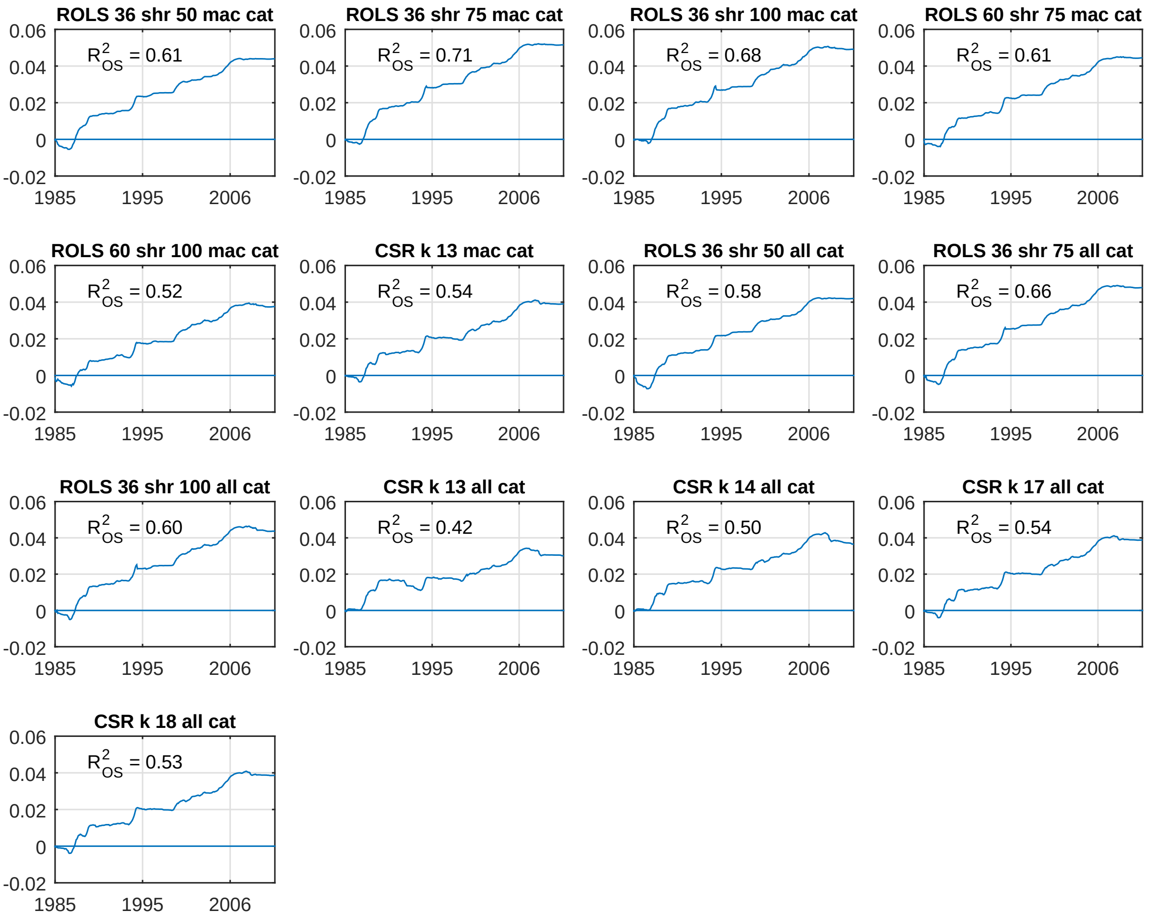


Figure C-1: The cumulative squared forecast error of the historical average benchmark minus the cumulative squared forecast error of the forecast sets that are included in the model confidence set for the 2 year maturity, their R^2_{OS}

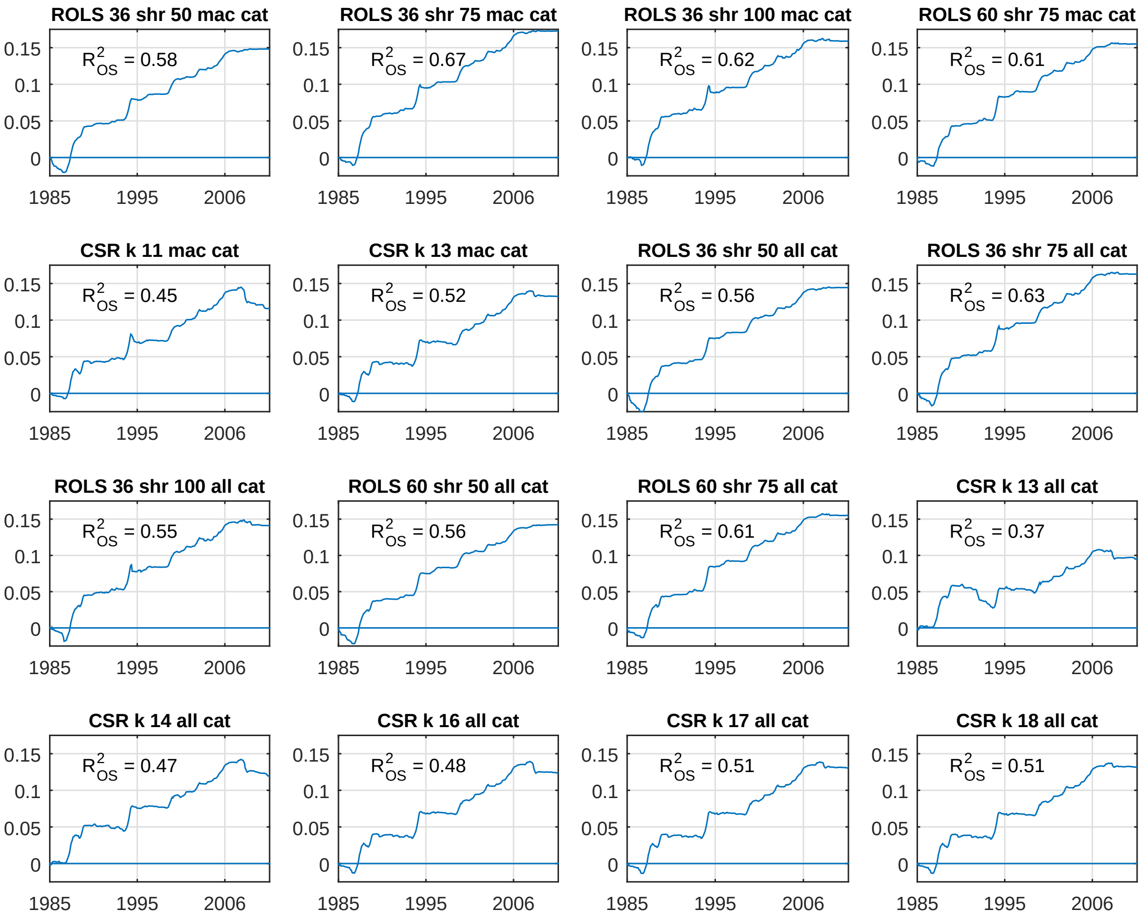


Figure C-2: The cumulative squared forecast error of the historical average benchmark minus the cumulative squared forecast error of the forecast sets that are included in the model confidence set for the 3 year maturity, their R^2_{OS}

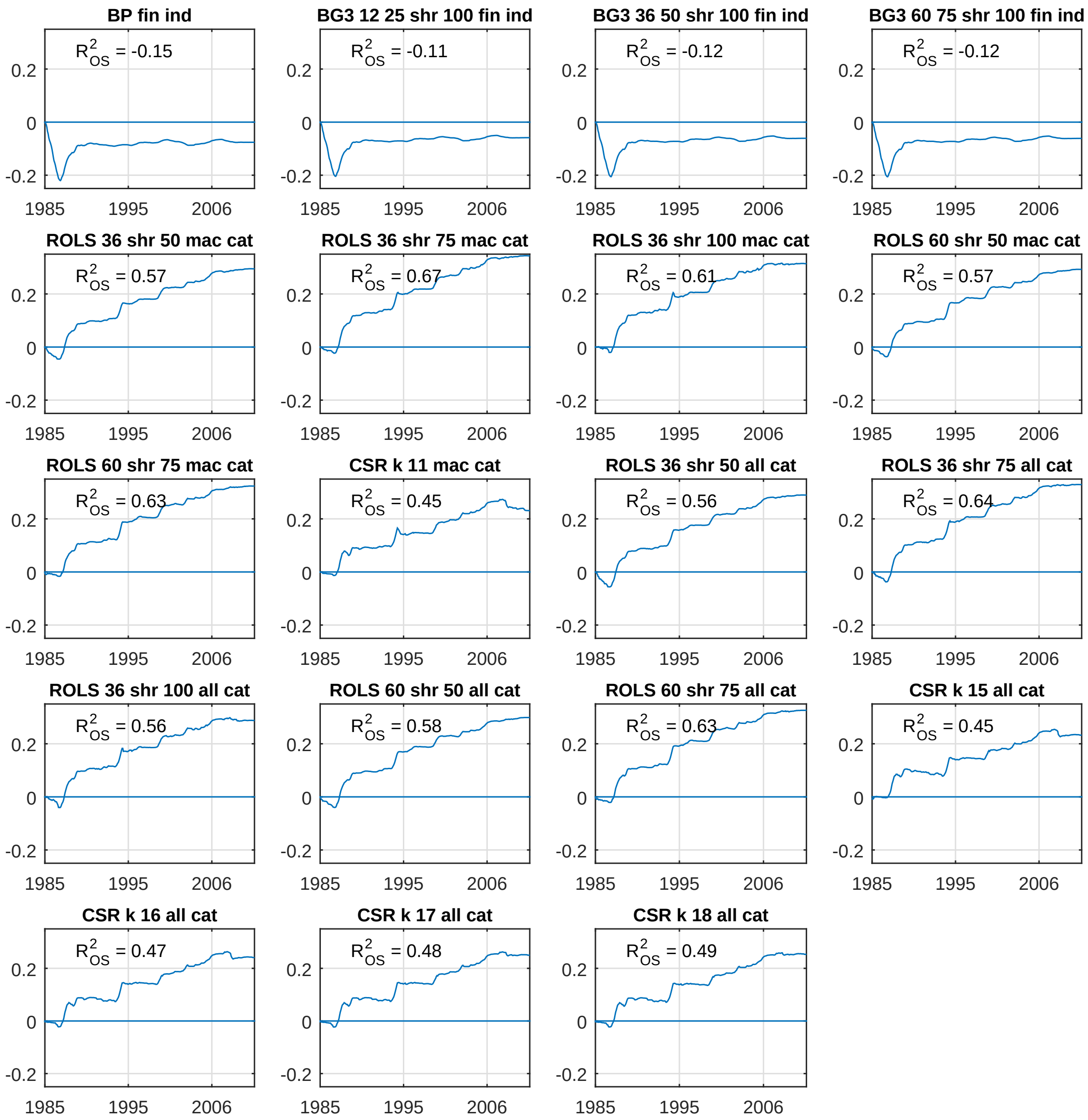


Figure C-3: The cumulative squared forecast error of the historical average benchmark minus the cumulative squared forecast error of the forecast sets that are included in the model confidence set for the 4 year maturity, their R^2_{OS}

Bibliography

Aiolfi and Timmermann. Persistence in forecasting performance and conditional combination strategies. *Journal of Econometrics*, 135(1-2):31–53, 2006.

Bates and Granger. The combination of forecasts. *Or*, 20(4):451–468, 1969.

Campbell and Thompson. Predicting excess stock returns out of sample: Can anything beat the historical average? *Review of Financial Studies*, 21(4):1509–1531, 2008.

Cieslak and Povala. Understanding bond risk premia. 2011.

Cochrane and Piazzesi. Bond risk premia. *American Economic Review*, (95):138–160, 2005.

Cooper and Priestley. Time-varying risk premiums and the output gap. *Review of Financial Studies*, 22(7):2601–2633, 2009.

Diebold and Pauly. The use of prior information in forecast combination. *International Journal of Forecasting*, 6(4):503–508, 1990.

Elliott, Gargano, and Timmermann. Complete subset regressions. *Journal of Econometrics*, 177 (2):357–373, December 2013.

Elliott, Gargano, and Timmermann. Complete subset regressions with large-dimensional sets of predictors. 2015.

Fama and French. The cross-section of expected stock returns. *the Journal of Finance*, 47(2): 427–465, 1992.

Fama and French. Common risk factors in the returns on stocks and bonds. *Journal of financial economics*, 1993.

Hansen. The Model Confidence Set. *Econometrica*, 79(2):453–497, 2011.

Huang and Shi. Determinants of Bond Risk Premia. *Risk Management*, (April), 2010.

Lettau and Ludvigson. Resurrecting the (C) CAPM: A cross-sectional test when risk premia are time-varying. *Journal of Political Economy*, 109(6):1238–1287, 2001.

Ludvigson and Ng. The empirical risk-return relation: A factor analysis approach. *Journal of Financial Economics*, 83(1):171–222, January 2007.

Ludvigson and Ng. Macro Factors in Bond Risk Premia. *Review of Financial Studies*, 22(12): 5027–5067, October 2009.

Christopher M. Miller, Robert T. Clemen, and Robert L. Winkler. The effect of nonstationarity on combined forecasts. *International Journal of Forecasting*, 7(4):515–529, 1992.

Stock and Watson. Combination forecasts of output growth in a seven-country data set. *Journal of Forecasting*, 23(6):405–430, September 2004.

Stock and Watson. Forecasting with many predictors. *Handbook of economic forecasting*, 1(05), 2006.

Allan Timmermann. Chapter 4 Forecast Combinations. In *Handbook of economic forecasting*, volume 1, pages 135–196. 2006.