Macroeconomic factors and equity premium predictability ☆

Daniel Buncic [a,*], Martin Tischhauser [b]

[a] Research & Modelling Division, Sveriges Riksbank, Sweden
[b] Center of Economic Research, ETH Zurich, Switzerland

A R T I C L E   I N F O

A B S T R A C T

Neely et al. (2014) have recently demonstrated how to *efficiently* combine information from a set of popular technical indicators together with the standard Goyal and Welch (2008) predictor variables widely used in the equity premium forecasting literature to improve out-of-sample forecasts of the equity premium using a small number of principal components. We show that forecasts of the equity premium can be further improved by, first, incorporating broader macroeconomic data into the information set, second, improving the selection of the most *relevant* factors and combining the most relevant factors by means of a forecast combination regression, and third, imposing theoretically motivated positivity constraints on the forecasts of the equity premium. We find that in particular our proposed forecast combination approach, which combines forecasts of the most relevant Neely et al. (2014) and macroeconomic factors and further imposes positivity constraints on the equity premium forecasts, generates statistically significant and economically sizeable improvements over the best performing model of Neely et al. (2014).

1. Introduction

Returns in excess of the risk free rate on any asset are financial compensation for bearing systemic risk. This compensation is generally measured by risk premia. The perceived aggregate level of risk in an economy is a function of financial and macroeconomic variables, that is, the state-variables of the real economy. Therefore, any financial and/or macroeconomic variables that help to predict the state of the economy should in theory also help to predict the equity premium.[1] Incorporation information contained in macro-economic variables thus appears to be paramount when construction forecasts of the equity premium. Yet, with the exception of Ludvigson and Ng (2007), there exist very few studies that examine the predictive gains from using a broad set of macroeconomic indicators on forecasts of the equity premium.

The objective of this study is to examine more closely the value of macroeconomic predictors for equity premium forecasts. More specifically, we take the set-up and data of the recent study by Neely, Rapach, Tu, and Zhou (2014) as our benchmark model and then add the *most relevant macroeconomic information* to the predictors set to provide a measure of the *state of the real economy*. Neely et al.

(2014) have recently demonstrated that *efficiently* combining the information from 14 popular technical indicators added to the 14 traditional Goyal and Welch (2008) predictor variables by means of *diffusion indices* or *factors* provides to date *"the best monthly U.S. equity premium forecast"* (Rapach & Zhou, 2013; p. 344). It is thus natural to take the best performing "ALL Factors" model in Neely et al. (2014) as our benchmark and then assess whether macroeconomic information can be used in an *efficient* way to improve forecasts of the equity premium.

   We follow two different strategies to improve forecasts of the equity premium when adding information from the Stock and Watson (2002a, b) macroeconomic panel to the predictor set. First, we extract factors from the joint dataset of the technical indicators, the Goyal and Welch (2008) predictor variables and the macroeconomic panel of Stock and Watson (2002a, b). As it is important to maintain a parsimonious model when constructing forecasts of the equity premium, we use the adaptive Lasso of Zou (2006) as a model selection device to find the most *relevant* predictors for the equity premium.[2] That is, after we have determined the baseline dimension of factors that explain the variation in the joint dataset, we select the most *relevant* factors for the equity premium by means of an adaptive Lasso penalised regression. These selected factors are then used in a predictive regression model to construct forecasts of the equity premium. This approach differs from Neely et al. (2014), who incrementally add factors to the predictor set and determine the appropriate number of factors for the equity premium based on maximizing the adjusted R^2. The adaptive Lasso approach that we employ provides more flexibility in the selection of the factors, as it does not rely upon an incremental addition of factors determined by the magnitude of their eigenvalues to the predictive regression model, thereby allowing for the possibility of, for instance, only the 1st and the 4th factors (instead of all factors from 1 to 4) to be included.

   A disadvantage of our first approach is that it confounds, to some extent at least, what contribution the information contained in the macroeconomic panel has with regards to improving forecast accuracy of the equity premium, since it becomes more difficult to precisely disentangle how the factors are made up. As a second strategy, we therefore also consider a forecast combination approach, where forecasts of the equity premium are constructed *separately* from the most relevant ALL factors of Neely et al. (2014) and the most relevant factors extracted from the macroeconomic panel of Stock and Watson (2002a, b), where the relevant factors are again selected by the adaptive Lasso. These separate forecasts are then optimally combined with their weights determined by the recent performance of the two individual models.

   One important additional aspect that we incorporate into our equity premium forecasting approach is the inclusion of theoretically motivated restrictions. It is well known since at least Campbell and Thompson (2008) that imposing theoretically coherent constraints on the parameter estimates of the predictive regression models and/or the equity premium forecast itself *"almost always"* improve out-of-sample forecast performance. We thus find it important to also include positivity constraints on the predictions of the equity premium in our modelling and forecasting approach.

   The use of *i*) theoretically motivated restrictions, *ii*) forecast combination and *iii*), factors (or diffusion indices) are three of the four strategies emphasized in Rapach and Zhou (2013, see page 332) that are needed for a successful equity premium forecasting implementation. The fourth strategy noted in Rapach and Zhou (2013), regime shifts, is implicitly allowed for in our prediction implementation by the use of a recursive forecasting scheme, where the factors as well as the parameters of the factor predictive regression model are re-estimated at each point in time. Although in a simplified form, such an approach is akin to a time varying parameter regression model, which, as emphasized by West and Harrison (1997) is capable of approximating (at least locally) neglected nonlinearities in the modelling framework. Our proposed forecasting approach thus addresses all four of the listed strategies in Rapach and Zhou (2013) to improve out-of-sample forecasts of the equity premium.

   It is important to stress here that, although related, our study differs from Ludvigson and Ng (2007) in a number of important ways. First, despite including factors extracted from the macroeconomic database of Stock and Watson (2002a, b) in their in-sample predictive regressions (see Table 2 and Section 5.1 for the evaluation results), Ludvigson and Ng (2007) only include 2 factors extracted from their *financial* dataset, which are labelled $\widehat{G}_1$ and $\widehat{G}_3$ (with the first entering in squared form), in their out-of-sample evaluation (see Table 6 in Section 5.3). Their out-of-sample evaluation, therefore, does not assess the benefit of adding macroeconomic information to the predictor set when constructing forecasts of the equity premium. Second, Ludvigson and Ng (2007) select the factors relevant for the equity premium from the two sets of financial and macroeconomic factors in a seemingly ad hoc way, stating simply that a number of specifications and combinations of the different macroeconomic and financial factors were considered, including also various polynomial bases (see page 185 in Ludvigson and Ng (2007)). We follow a clear statistical approach and use the adaptive Lasso as a model selection devise to find the most relevant factors for the equity premium. Third, we use the *'to date best performing'* ALL factors forecasting model of Neely et al. (2014) for monthly data as our benchmark model.

   We find that macroeconomic factors play an important role in improving out-of-sample forecasts of the equity premium. In particular, our proposed forecast combination approach, which combines forecasts of the most relevant Neely et al. (2014) ALL factors and our added macroeconomic factors, and which further imposes positivity constraints on the forecasts of the equity premium generates statistically significant and economically sizable improvements over the best performing ALL factors model of Neely et al. (2014). Monthly out-of-sample R^2 values can be as high as 1.75% and forecast improvements are statistically significant. Moreover, we show that these statistical gains translate into meaningful economic gains to a risk-averse investor. The annualised gain in certainty equivalent return over the ALL factor model of Neely et al. (2014) is in the order of 3.35%. Finally, a secondary finding of our study is that all predictability results, including those of the ALL factors model of Neely et al. (2014), and also our proposed forecast combination

---

[2] Neely et al. (2014), among many others, have forcefully demonstrated that a key aspect of creating a successful prediction environment when including many potential predictor variables is the *efficient* inclusion of the most important information for the equity premium by means of diffusion indices or factors extracted from the combined set of technical indicators and Goyal and Welch (2008) predictor variables.

approach with macroeconomic factors weaken considerably from the beginning of 1992 onwards when compared to the simple historic average forecast. This highlights the need to further study equity premium predictability over this time frame and to asses which alterative predictor variables may be fruitful in explaining asset price movements from 1992 onwards. A novel step in the right direction seems to be taken by Rapach, Ringgenberg, and Zhou (2015), who use information on short interest as a predictor variable and show that it consistently produces sizeable forecast improvements over the historic average benchmark.

The rest of the paper is structured as follows: Section 2 describes the data used in the analysis. In Section 3, the set up of the predictive regressions model is explained, together with the additions that we consider to improve in-sample as well as out-of-sample performance. Section 4 discusses the empirical results of the models presented in the previous section, providing a statistical and an economic evaluation of the proposed forecasting approaches. In Section 5 we conclude the study.

2. Data

We work with the well known and widely studied equity premium modelling dataset of Goyal and Welch (2008), which was also recently employed by Neely et al. (2014).[3] To conserve space here, we relegate all data description to the Appendix. The transformations that we employ to construct the target variable of interest, the equity premium (EQ henceforth), and the Goyal and Welch (2008) predictor variables (simply GW predictors henceforth) are described in detail in Table A.1. From here on, we will denote these predictors by x^{GW} (or ECON as in to Neely et al. (2014)) in our notation. We follow Neely et al. (2014) and add 14 technical indicators based on three popular trend-following trading strategies to the GW predictor variables. These three trading strategies are: *i*) Moving average rules, *ii*) Momentum rules and, *iii*) On-balance Volume. The construction of the technical indicators, together with their summary statistics are provided in Section A.2.1 in the Appendix. We use the notation x^{TECH} to denote the vector of 14 technical indicators of Neely et al. (2014) that we construct.

In addition to the GW predictors and the TECH indicators constructed by Neely et al. (2014), we add data from the macroeconomic panel of Stock and Watson (2002a, b) (simply SW henceforth) to the information set.[4] The SW database includes a broad range of macroeconomic variables relating to real output, employment and hours, housing starts, manufacturing orders, retail sales data, inventories, price indexes, consumer loans and expectations, interest rates and interest rate spreads, stock market indicators, and foreign exchange rates and commodity prices.[5] It is thus fairly comprehensive as it contains high-level aggregates, as well as disaggregated components of the various macroeconomic and financial series that are included. To avoid "*double counting issues*" due to the inclusion of high-level aggregates as well as disaggregated series, we exclude all high-level aggregates from the SW database. Further, as the Goyal and Welch (2008) dataset already includes various interest rates, interest rate spreads and various other variables as predictor variables, we further drop the 10-year treasury constant maturity rate, the 3−month treasury bill rate, and the term spread from the SW database. Lastly, there are a number of variables for which the sample starts much later than January 1959. To avoid shortening the available sample period any further, we dropped those variables as well. From the full set of monthly SW predictors, we retain 66 and collect these in the vector x^{SW} to be used in our empirical analysis.

Analogous to the data handling in Stock and Watson (2012), all macroeconomic variables are subject to preliminary transformations to induce stationarity. The transformation depends on the category of the series.[6] Briefly, real activity variables were transformed to monthly growth rates by taking the first differences of logs, wages and prices were transformed to monthly changes of monthly inflation by taking the second differences of the logs, interest rates were transformed by taking first differences, and interest rate spreads are used in levels. There is one last important '*transformation*' that we apply to the SW data. To be consistent with the definition and timing of the inflation variable INFL$_t$ in the Goyal and Welch (2008) predictor set, we apply one extra lag to all variables related to real output, employment and hours, housing starts, manufacturing orders, retail sales data, inventories, price indexes, consumer loans, and expectations variables to account for the publication delay in the releases of these variables.[7] All 66 SW variables that we use are listed in Table A.5 in the in the Appendix, together with the transformation codes that are applied, as well as some summary statistics.

Our study covers the effective sample period from April 1959 to June 2011. All data used in this study are available from: http://www.danielbuncic.com/data/eqt6edata.zip.

3. Modelling approach

This section describes the modelling approach that we adopt throughout the study. We initially give a background on the traditional framework for assessing equity premium predictability and then proceed to show how macroeconomic factors, factor selection and combination, as well as sign restrictions can be used to improve the in-sample as well as out-of-sample forecast performance of the equity

---

[3] We obtain the predictor data from the program replication files of Neely et al. (2014) available at: http://sites.slu.edu/rapachde/DataProgramFilesMS2014.zip (last retrieved, September 2016). We would like to thank Dave Rapach for making these replication files publicly available, so that an exact replication of their results is possible.

[4] The SW database is available from: http://www.princeton.edu/~mwatson/publi.html (retrieved, May 2014). The monthly data spans the period from January 1959 to June 2011 and is available in updated form from http://www.princeton.edu/~mwatson/ddisk/Stock_Watson_Disentangling_ReplicationFiles_July05_2013.zip.

[5] The motivation for including broad macroeconomic diffusion indexes for forecasting asset prices is discussed in greater detail in Ludvigson and Ng (2009, 2007), Rapach and Zhou (2013) and also Stock and Watson (2012).

[6] We have translated their Gauss code to Matlab and use, for reasons of consistency with their analysis, exactly the same data transformations. To avoid unnecessary repetition of why these transformations were applied, we refer the interested reader to their paper.

[7] See also Buncic and Müller (2017) for more details on updates and revisions to US inflation data.

premium.

3.1. Background

The conventional framework for analyzing equity premium predictability is a *'linear predictive regression'* model taking the form:

$$\text{EQ}_{t+1} = b_0 + \boldsymbol{b}'\boldsymbol{x}_t + \varepsilon_{t+1}, \tag{1}$$

where EQ_{t+1} is the equity premium at time $t+1$, b_0 is an intercept term, $\boldsymbol{b}$ is a $p \times 1$ dimensional vector of slope parameters, $\boldsymbol{x}_t$ is a $p \times 1$ dimensional vector of predictor variables observed at time t with $p \geq 1$, and ε_{t+1} is an unobserved disturbance term with mean zero. Two of the simplest and most commonly followed approaches to investigate equity premium predictability is by means of individual bivariate regressions using one variable at a time from the GW predictor set, or the '*kitchen sink*' regression, which includes the full set of GW predictors in (1) (see for example Goyal and Welch (2008), Campbell and Thompson (2008), Neely et al. (2014) and many others).

The weaknesses of both of these two approaches are well understood in the literature. First, specifying regression models with only one explanatory variable is deemed to yield an under or mis-specified model, resulting in heavily biased parameter estimates and therefore poor out-of-sample forecasts. Second, the '*kitchen sink*' regression approach, which includes the full set of potential predictors often suffers from the well known in-sample overfitting problem. Moreover, in the context of the GW predictor set, a main concern with using the full set of predictors is that there is high degree of cross-correlation in the variables. For instance, the two smallest eigenvalues of the cross-product moment matrix of $\boldsymbol{x}^{\text{GW}}$ are -1.4540×10^{-15} and 3.0247×10^{-15}, resulting in a huge condition number and thus a '*near*' singular design matrix. The consequence of this is that any parameter estimates based on the kitchen sink regression using the full GW predictor set will be highly variable or imprecise, thereby, despite the good in-sample performance, providing extremely poor out-of-sample forecasts of the equity premium. This poor performance of the kitchen sink regression was highlighted by Goyal and Welch (2008) and Rapach, Strauss, and Zhou (2010), among others, however, never openly attributed to the fact that the regressors in the kitchen sink model are highly (near perfectly) collinear.

3.2. Technical indicators and factor regressions

Neely et al. (2014) have recently shown that extending the standard set of GW predictor variables with various technical indicators can lead to improved forecasts of the equity premium. To avoid problems related to using one variable at a time in bivariate regressions or using the entire universe of Goyal and Welch (2008) and technical predictors, Neely et al. (2014) follow a latent factor approach to incorporate all relevant information from the numerous predictor variables, while at the same time trying to retain a parsimonious model to avoid overfitting concerns. The underlying premise of the latent factor based approach is that a few *factors* are driving the co-movements of the large number of potential equity premium predictors.

Let $\mathscr{X}_{it}$ be an element of a general $T \times N$ dimensional *panel* of data $\mathscr{X}$, where $i = 1, \ldots, N$ is the cross-sectional dimension of the panel, ie., the number of predictor variables in our case, and $t = 1, \ldots, T$ is a standard time index. The latent factor approach supposes that $\mathscr{X}_{it}$ has a factor structure taking the form:

$$\mathscr{X}_{it} = \boldsymbol{\lambda}_i'\boldsymbol{F}_t + \varepsilon_{it} \tag{2}$$

where $\boldsymbol{F}_t$ is an $n \times 1$ vector of latent common factors with $n \ll N$, λ_i is a corresponding $n \times 1$ vector of factor loadings and ε_{it} is an idiosyncratic error term. The error term ε_{it} can be specified to be contemporaneously and serially uncorrelated, which yields a strict factor model, or it can be allowed to have mild forms of serial and/or cross correlation resulting in an approximate factor model. Under either scenario, the factors can be estimated consistently using principal components (see, Bai (2003), Bai and Ng (2006) and Stock and Watson (2006, 2011)).

The defining feature of the latent factor model in (2) is that the dimension of the latent factors $\boldsymbol{F}_t$ is substantially smaller than the (cross-sectional) dimension of the panel $\mathscr{X}$, thereby providing a parsimonious representation of the variation in $\mathscr{X}$. Let $\boldsymbol{f}_t \subseteq \boldsymbol{F}_t$. The subset $\boldsymbol{f}_t$ of the latent factors $\boldsymbol{F}_t$ can then be utilised in a *'factor predictive regression'* of the form:

$$\text{EQ}_{t+1} = \beta_0 + \boldsymbol{\beta}'\boldsymbol{f}_t + \varepsilon_{t+1}, \tag{3}$$

where the dimension of $\boldsymbol{f}_t$ is $r \times 1$, with $r \leq n$, β_0 is an intercept term, $\boldsymbol{\beta}$ is an $r \times 1$ dimensional vector of slope parameters, and ε_{t+1} is again an unobserved disturbance term with mean zero.

It is important here to distinguish between the two factors $\boldsymbol{f}_t$ and $\boldsymbol{F}_t$ that we define, and to understand their differences.[8] Recall that $\boldsymbol{F}_t$ is defined as the vector of factors that is driving the variation in the panel $\mathscr{X}$, which contains the full set of variables that are assumed to be important for equity premium prediction. Nevertheless, the equity premium itself does not need to be dependent on the full set of $\boldsymbol{F}_t$, but could be driven by only a subset $\boldsymbol{f}_t$ of $\boldsymbol{F}_t$ instead. This distinction is of significance even when employing a parsimonious factor model setting, as it is likely that there will be factors which are uninformative about the equity premium and therefore redundant in a

---

[8] The importance of this distinction was forcefully emphasized in Ludvigson and Ng (2007, page 177) and also Ng (2013, p. 770) in her handbook chapter on variable selection for predictive regressions. Thus finding an appropriate approach to select those factors most important for equity premium forecasts will be a challenge and we describe later on how we go about implementing this empirically.

forecasting application.

To provide an example of this, consider the in-sample predictive regression results in Table 2 in Neely et al. (2014), which are based on the GW predictors, and where the underlying factors are extract (or estimated) using principal components. Neely et al. (2014) determine the first 3 factors (corresponding to the three largest eigenvalues) to be the most relevant factors, and label these $\widehat{F}_1^{\text{ECON}}$, $\widehat{F}_2^{\text{ECON}}$ and $\widehat{F}_3^{\text{ECON}}$, respectively.[9] From the predictive regression results in Table 2 in Neely et al. (2014), we see, nevertheless, that only the third factor $\widehat{F}_3^{\text{ECON}}$ is significantly different from zero, and also sizable in magnitude. Factors one and two ($\widehat{F}_1^{\text{ECON}}$ and $\widehat{F}_2^{\text{ECON}}$), on the other hand, are not only statistically insignificant in the equity premium factor regressions, but also rather small in size.[10] This example illustrates the important difference between f_t and F_t in the notation above. Once the optimal number of factors F_t has been determined, it will be beneficial to find those factors f_t in F_t that are most informative or predictive for the equity premium EQ_{t+1}, and to discard any other factors that are not deemed important, in order to improve out-of-sample forecast performance.

3.3. Macroeconomic factors, factor selection and combination, and sign restrictions

Ludvigson and Ng (2007, 2009) have recently shown that including information about the state of the macroeconomy by means of macroeconomic 'diffusion indices' or factors extracted from the Stock and Watson (2002a, b) database can improve forecasts of risk premia in bond markets, as well as forecasts of the conditional mean and variance of the equity premium. To assess how useful the information contained in the SW macroeconomic dataset is in the context of the setting adopted in Neely et al. (2014), we add the macroeconomic information in two distinct ways.

First, we extract factors from the joint panel of GW predictors, technical indicators of Neely et al. (2014) and SW macroeconomic variables. Since we are increasing the predictor dimension when adding the SW macroeconomic variables to the GW data and technical indicators of Neely et al. (2014), it is important to maintain a parsimonious model structure to avoid poor out-of-sample forecast performance due to overfitting. We achieve this by using the adaptive Lasso ('aLasso' for short in the subscript notation) penalised regression estimator of Zou (2006) as a model selection device, to determine the most important factors for predicting the equity premium. Second, we use the forecast combination approach of Stock and Watson (2004), where the individual forecasts of the equity premium that are combined are obtained from the most relevant ALL factors of Neely et al. (2014) and the most relevant SW factors, with most relevant in this context meaning that the factors were selected by the adaptive Lasso procedure. The combination weights are determined by the past performance of the individual model's forecasts. Lastly, we follow Campbell and Thompson (2008) and impose positivity restrictions on the forecasts of the equity premium from the adaptive Lasso as well as the model combined forecasts.

3.3.1. Implementation of the adaptive Lasso

Let us define here the factor notation that we use and also provide a few more details on how the adaptive Lasso is implemented. Denote by $\widehat{F}_{t,1:\kappa_a}^{\text{ALL}}$ the $\kappa_a \times 1$ vector of ALL factors at time t extracted by PCA from $\mathscr{X}(\text{ALL}) := [x^{\text{GW}}\ x^{\text{TECH}}]$, ie., from the combined panel of GW predictors and Neely et al. (2014) technical indicators, where the dimension κ_a is determined following the same procedure as in Neely et al. (2014, see footnote 6 in their paper) by the adjusted R^2. Also, let $\widehat{F}_{t,1:\kappa_s}^{\text{SW}}$ be defined as the $\kappa_s \times 1$ dimensional vector of SW factors extracted by PCA from $\mathscr{X}(\text{SW}) := x^{\text{SW}}$, that is, from the Stock and Watson (2002a, b) panel of macroeconomic variables that are described in Section 2. Lastly, let us define the $\kappa_j \times 1$ dimensional vector of JOINT factors extracted from $\mathscr{X}(\text{JOINT}) := [x^{\text{SW}}\ x^{\text{GW}}\ x^{\text{TECH}}]$ by $\widehat{F}_{t,1:\kappa_j}^{\text{JOINT}}$, that is, the factors extracted from the *joint* panel of SW macroeconomic variables, the GW predictors and technical indicators.[11]

Once the vector of factors $\widehat{F}_{t,1:\kappa}$ is determined, instead of including all factors in a predictive regression model such as (3), we use the adaptive Lasso of Zou (2006) as a variable selection device to find those factors that are most important for the equity premium, to avoid issues related to in-sample overfitting. This is achieved by means of a penalised regression that solves the following generic optimisation problem:

$$\widehat{\boldsymbol{\theta}}_{\text{aLasso}} = \operatorname{argmin}_{\{\theta_i\}_{i=1}^{\kappa}} \left\{ \sum_{t=1}^{T-1} \left[EQ_{t+1} - \boldsymbol{\theta}' \widehat{F}_t \right]^2 + \mu \sum_{i=1}^{\kappa} \widehat{w}_i |\theta_i| \right\}, \tag{4}$$

where the μ parameter in (4) is a 'complexity' parameter which controls the amount of 'shrinkage' (or penalty) coming from the $\sum_{i=1}^{\kappa} \widehat{w}_i |\theta_i|$

---

[9] See footnote 6 in Neely et al. (2014) on how the number of factors was determined. Rather than using the information criteria of Bai and Ng (2002) to determine the factor structure in $\mathscr{X}$, Neely et al. (2014) follow a strategy of adding (in increasing order) factors in a predictive regression up to a predetermined maximum value of $\bar{\kappa}$, and then proceed to select the number of factors by choosing the set that gives the highest adjusted R^2 ($\bar{R}_2$). Recall that when PCA is used to extract the factors, the eigenvalues and factors are commonly sorted in descending order, yielding as the first factor the one corresponding to the largest eigenvalue, the second factor corresponding to the second largest eigenvalue and so forth.

[10] The eigenvalues corresponding to these first three factors are 4.45, 2.24 and 1.98, respectively, suggesting that the last factor explains less than half of the variation in $\mathscr{X}$ than the first factor. Yet, with regards to equity premium prediction, it is the third factor that is most important.

[11] For the SW factors $\widehat{F}_{t,1:\kappa_s}^{\text{SW}}$ and the JOINT factors $\widehat{F}_{t,1:\kappa_j}^{\text{JOINT}}$, we follow the mainstream literature on factor models and use the Bai and Ng (2002) IC_{p1} information criterion to determine the appropriate number of factors κ_s. We set the maximum number of factors to 5 and 7 for the SW and JOINT panels.

term, and $\widehat{\boldsymbol{w}}$ is a $\kappa \times 1$ weight vector that is equal to $1/|\widehat{\boldsymbol{\theta}}_{\text{OLS}}|$, with $\widehat{\boldsymbol{\theta}}_{\text{OLS}}$ being the OLS estimate of $\boldsymbol{\theta}$ in (4), ie., when $\mu = 0$.[12]

If the complexity parameter μ is known, the relation in (4) is straightforward to solve using a standard pathwise coordinate descent algorithm (see Hastie, Tibshirani, and Friedman (2009)). In the *i. i. d.* setting, it is common to use cross-validation methods to choose the complexity parameter μ. In our setting this is not feasible, since the $\widehat{\boldsymbol{F}}_{t,1:\kappa_a}^{\text{ALL}}$ factors extracted from the GW predictors and the technical indicators are highly persistent. To avoid having to shorten the out-of-sample evaluation period further by having to specify a tuning or holdout period to determine the complexity parameter μ, we follow Zou, Hastie, and Tibshirani (2007) and use the Akaike Information Criterion (AIC) as defined on page 9 in their paper. We prefer to use AIC over other information criteria, as the AIC is known to yield models with the smallest one-step-ahead mean squared prediction errors.

Note here again that we use the adaptive Lasso in (4) as a model selection device and *not* to form the predictions of the equity premium itself.[13] Thus, once adaptive Lasso estimates of $\boldsymbol{\theta}$ in (4) are available, we follow the suggestions of Belloni and Chernozhukov (2013) and use OLS post Lasso to run predictive regressions of the form:

$$\text{EQ}_{t+1} = \beta_0 + \boldsymbol{\beta}'\widehat{\boldsymbol{f}}_t + \varepsilon_{t+1}, \tag{5}$$

where $\widehat{\boldsymbol{f}}_t$ is the $k \times 1$ ($k \leq \kappa$) vector of *relevant* factors selected from the adaptive Lasso estimates that have not been shrunk to 0. That is, let $1(\widehat{\boldsymbol{\theta}}_{\text{aLasso}}\neq 0)$ denote the $\kappa \times 1$ dimensional vector of indicators that are equal to 1 in position i whenever the statement in the brackets is true, ie., the adaptive Lasso estimate of θ_i is non-zero. Then, $1(\cdot)$ selects those factors in $\widehat{\boldsymbol{F}}_t$ with non-zero $\widehat{\boldsymbol{\theta}}_{\text{aLasso}}$, such that, $\widehat{\boldsymbol{f}}_t = \widehat{\boldsymbol{F}}_t[1(\widehat{\boldsymbol{\theta}}_{\text{aLasso}}\neq 0)]$. In our approach, we will be particularly interested in working with a subset of JOINT factors selected by the adaptive Lasso from the JOINT factor set as $\widehat{\boldsymbol{f}}_t^{\text{JOINT}} = \widehat{\boldsymbol{F}}_{t,1:\kappa_j}^{\text{JOINT}}[1(\widehat{\boldsymbol{\theta}}_{\text{aLasso}}^{\text{JOINT}}\neq 0)]$. Forecasts of the equity premium from the subset of selected JOINT factors are then constructed as:

$$\widehat{\text{EQ}}_{t+1|t}^{\text{JOINT}} = \widehat{\beta}_0 + \widehat{\boldsymbol{\beta}}'\widehat{\boldsymbol{f}}_t^{\text{JOINT}}, \tag{6}$$

where $\widehat{\beta}_0$ and $\widehat{\boldsymbol{\beta}}$ denote the OLS (post Lasso) estimates of a regression of the form in (5) with $\widehat{\boldsymbol{f}}_t^{\text{JOINT}}$ as the vector of predictor variables.

One issue with extracting factors from the JOINT factor set $\mathscr{X}(\text{JOINT})$ above is that we will be mixing highly persistent data, ie., the $\mathscr{X}(\text{ALL})$ data from the GW predictors and Neely et al. (2014) technical indicators, and data that has been transformed to stationarity and thus low persistence, the $\mathscr{X}(\text{SW})$ data, together when extracting the factors. The extracted factors will thus be dominated by the persistent variables, whose past values have more power in predicting the variation in the persistent variables (see also the discussion on pages $220-222$ in Elliott and Timmermann (2016)). For that reason, we also consider a forecast combination strategy, which separates not only the extraction of the factors, but also the selection of the factors, and then proceeds by combining the selected subset of factors in a combination regression. This forecasting strategy is explained next.

3.3.2. Model combination

The second approach that we consider to improve the out-of-sample forecast performance is a forecast combination approach as employed, for instance, in Stock and Watson (2004). To illustrate how this approach is implemented, let us define $\widehat{\text{EQ}}_{t+1|t}^{\text{ssALL}}$ to be the time t forecast of the equity premium at time $t+1$, where only the adaptive Lasso selected subset $\widehat{\boldsymbol{f}}_t^{\text{ALL}} = \widehat{\boldsymbol{F}}_{t,1:\kappa_a}^{\text{ALL}}[1(\widehat{\boldsymbol{\theta}}_{\text{aLasso}}^{\text{ALL}}\neq 0)]$ of extracted ALL factors from Neely et al. (2014) are used as regressors in the OLS post Lasso regression model in (5).[14] Let the resulting one-step-ahead prediction errors be denoted by:

$$\widehat{e}_{t+1|1}^{\text{ssALL}} = \text{EQ}_{t+1} - \widehat{\text{EQ}}_{t+1|t}^{\text{ssALL}}, \tag{7}$$

where

$$\text{MSFE}_{t+1}^{\text{ssALL}} = \frac{1}{t - T_{is}} \sum_{\tau=T_{is}}^{t} \left[\widehat{e}_{\tau+1|\tau}^{\text{ssALL}}\right]^2 \tag{8}$$

is the corresponding mean squared forecast error (MSFE) at time $t+1$, with $t = T_{is}+1, \ldots, T$.[15] Also, let the terms $\widehat{\text{EQ}}_{t+1|t}^{\text{SW}}$, $\widehat{e}_{t+1|1}^{\text{SW}}$ and

---

[12] It is common to demean and standardize the dependent and regressor variables when implementing penalised regressions such as the Lasso, so that the intercept term θ_0 is by construction 0. Also, the general specification of w in Zou (2006) is $1/|\theta_{\text{OLS}}|^\gamma$, with $\gamma > 0$. However, to avoid searching over γ as well, we simply set γ to 1, which is one of the values considered and recommended in Zou (2006).

[13] Belloni and Chernozhukov (2013) in fact highlight that the "*Lasso is often used as an estimator, and most often only as a model selection device*" (page 6). We very much follow the model selection usage of the Lasso and utilize it as a computational tool to determine the most relevant factors for the equity premium. In a somewhat different context, Buncic and Melecky (2014) have also used the Lasso as a model selection device.

[14] Note here that we are reserving the ALL superscript in the $\widehat{\text{EQ}}_{t+1|t}^{\text{ALL}}$ notation to denote the forecast of the equity premium using exactly the same approach as used by Neely et al. (2014) with their ALL factors. That is, we set the maximum number of admissible ALL factors to 4 and then use the adjusted R^2 criterion to determine the best set of factors, which could be between 1 and 4. The ssALL superscript is used here to denote the adaptive Lasso selected subset of ALL factors from Neely et al. (2014).

[15] T_{os} is the number of out-of-sample observations which is equal to $T- T_{is}- h+ 1$, where T is the total number of observations, T_{is} is the in-sample number of observations and h is the forecast horizon (which is equal to 1 here).

Table 1
In-sample factor selected/combined predictive regression results: full sample and two sub-samples.

(1) Variable	(2) Slope	(3) t−stat	(4) R^2	(5) R^2_{Exp}	(6) R^2_{Rec}	(1) Variable	(2) Slope	(3) t−stat	(4) R^2	(5) R^2_{Exp}	(6) R^2_{Rec}
					Full sample: April 1959 to June 2011						
Selected JOINT factors $\widehat{\boldsymbol{f}}^{\text{JOINT}}$						Combination of selected ALL and SW factors					
$\widehat{F}_2^{\text{JOINT}}$	−0.16	−3.09***	1.37	0.24	3.80	$\widehat{F}_1^{\text{ALL}}$	0.11	1.78*	1.80	0.33	4.96
						$\widehat{F}_3^{\text{ALL}}$	0.26	1.87*			
						$\widehat{F}_2^{\text{SW}}$	0.06	0.53			
						Selected SW factors					
						$\widehat{F}_2^{\text{SW}}$	0.28	1.72*	0.47	0.36	0.71
					Sub-sample 1: April 1959 to May 1985						
Selected JOINT factors $\widehat{\boldsymbol{f}}^{\text{JOINT}}$						Combination of selected ALL and SW factors					
$\widehat{F}_2^{\text{JOINT}}$	0.23	3.56***	3.44	−0.30	8.78	$\widehat{F}_1^{\text{ALL}}$	0.05	0.79	7.93	3.25	14.64
						$\widehat{F}_2^{\text{ALL}}$	−0.16	−1.63*			
						$\widehat{F}_3^{\text{ALL}}$	0.49	3.03***			
						$\widehat{F}_4^{\text{ALL}}$	0.37	2.38***			
						$\widehat{F}_1^{\text{SW}}$	−0.07	−0.91			
						$\widehat{F}_3^{\text{SW}}$	−0.06	−0.46			
						$\widehat{F}_4^{\text{SW}}$	−0.22	−1.46			
Neely et al. (2014) ALL Factors						Selected SW factors					
$\widehat{F}_1^{\text{ALL}}$	0.08	1.26	6.96	2.03	14.03	$\widehat{F}_1^{\text{SW}}$	−0.21	−1.09	2.14	2.46	1.08
$\widehat{F}_2^{\text{ALL}}$	−0.14	−1.82*				$\widehat{F}_3^{\text{SW}}$	−0.17	−1.36			
$\widehat{F}_3^{\text{ALL}}$	0.63	4.82***				$\widehat{F}_4^{\text{SW}}$	−0.48	−2.30***			
$\widehat{F}_4^{\text{ALL}}$	0.44	3.25***									
					Sub-sample 2: May 1985 to June 2011						
Selected JOINT factors $\widehat{\boldsymbol{f}}^{\text{JOINT}}$						Combination of selected ALL and SW factors					
$\widehat{F}_1^{\text{JOINT}}$	−0.11	−1.46	2.97	−0.11	12.49	$\widehat{F}_1^{\text{ALL}}$	0.08	0.78	2.76	0.09	11.04
$\widehat{F}_2^{\text{JOINT}}$	0.14	1.59				$\widehat{F}_1^{\text{SW}}$	0.07	0.71			
$\widehat{F}_4^{\text{JOINT}}$	0.24	1.54				$\widehat{F}_2^{\text{SW}}$	0.29	1.70*			
Neely et al. (2014) ALL Factors						Selected SW factors					
$\widehat{F}_1^{\text{ALL}}$	0.15	1.38	1.47	0.16	5.51	$\widehat{F}_2^{\text{SW}}$	0.23	1.98**	0.50	0.23	1.04
$\widehat{F}_2^{\text{ALL}}$	0.14	1.63*									

Notes: This table reports in-sample predictive regression results from the proposed modelling approach which includes macroeconomic factors from the SW panel dataset and which then selects (or combines) the factors using the adaptive Lasso. The results are shown over the full sample period from April 1959 to June 2011, as well as two sub-periods: sub-sample 1 is from April 1959 to May 1985 and sub-sample 2 is from May 1985 to June 2011. The table is arranged in three parts, top, middle and bottom, which correspond to the three sub-periods. Columns 1 to 6 show, respectively, the predictor variable, the estimated slope coefficient, the corresponding t−statistic, as well as goodness of fit measures R^2 (in percent), which are further measured separately over expansion (Exp) and recession (Rec) periods as defined by the NBER. The rows show the subset of factors which were selected with the adaptive Lasso procedure that we outline in Section 3.3.1. For instance, the vector of selected factors from the joint data set is defined as: $\widehat{\boldsymbol{f}}_t^{\text{JOINT}} = \widehat{\boldsymbol{F}}_{t,1:\kappa_j}^{\text{JOINT}} [\mathbf{1}(\widehat{\boldsymbol{\theta}}_{\text{aLasso}}^{\text{JOINT}} \neq 0)]$, where $\widehat{\boldsymbol{\theta}}_{\text{aLasso}}^{\text{JOINT}}$ are the parameter estimates of the adaptive Lasso penalised regression in (4), and κ_j is determined using the IC_{p1} information criterion of Bai and Ng (2002). The ALL and SW factors shown under the "Combination of selected ALL and SW factors" heading, were also selected by using first the adaptive Lasso on the ALL and SW factors independently, and then combined in an OLS regression. t−statistics and corresponding levels of significance are based on the wild bootstrap as before. Significance at the 1, 5, and 10 percent levels are marked, respectively, by ***, ** and *.

$\text{MSFE}_{t+1}^{\text{SW}}$ be defined analogously for the predictions from an OLS post Lasso regression model that only uses the adaptive Lasso selected factors $\widehat{\boldsymbol{f}}_t^{\text{SW}}$ from the full set of SW factors $\widehat{\boldsymbol{F}}_{t,1:\kappa_s}^{\text{SW}}$. Then, the *combined* forecast is defined as:

$$\widehat{\text{EQ}}_{t+1|t}^{\text{COMB}} = \omega_t^{\text{SW}} \widehat{\text{EQ}}_{t+1|t}^{\text{SW}} + \left(1 - \omega_t^{\text{SW}}\right)\widehat{\text{EQ}}_{t+1|t}^{\text{ssALL}}. \tag{9}$$

We follow Stock and Watson (2004) and choose the weights ω_t^{SW} in (9) according to the historical performance of the respective prediction models. We measure past performance of each model by its MSFE, with the weights being proportional to the inverse of the MSFE defined as:

$$\omega_t^{\text{SW}} = \frac{1/\text{MSFE}_t^{\text{SW}}}{1/\text{MSFE}_t^{\text{ssALL}} + 1/\text{MSFE}_t^{\text{SW}}}. \tag{10}$$

Note here that ω_t^{SW} only uses time t information and does not require another calibration period for the ω_t^{SW} weights to be determined. Also, the formulation of the weights in Stock and Watson (2004) allows for *discounting* in the construction of the MSFE at time t.

To abstract from further specification needs and to keep the model combination forecasts as simple as possible, we do not consider such discounting techniques here.[16]

3.3.3. Sign restrictions

Campbell and Thompson (2008) have pointed out that one of the weaknesses in standard out-of-sample evaluations of equity premia predictability such as in Goyal and Welch (2008) is that the analysis is based on *unrestricted* predictive regressions. Campbell and Thompson (2008, page 1516) argue that: "*[a] regression estimated over a short sample period can easily generate perverse results, such as a negative coefficient when theory suggests that the coefficient should be positive*." Moreover, "*[i]n practice, an investor would not use a perverse coefficient but would likely conclude that the coefficient is zero, in effect imposing prior knowledge on the output of the regression.*" Campbell and Thompson (2008) go on to show that restricting regression coefficients (ie., parameter estimates) to zero whenever the estimate has the wrong sign "*almost always*" improve the out-of-sample forecast performance of the predictive regressions.

Following the suggestions of Campbell and Thompson (2008), we impose sign restrictions in our prediction models. Note that Campbell and Thompson (2008) use two different types of sign restrictions. One imposes a 0 restriction on any coefficient that does not provide the theoretically correct sign. The second restriction sets the forecast of the equity premium itself, ie., $\widehat{EQ}_{t+1|t}$, to 0 whenever it is negative. Since the extracted factors have no direct economic meaning, it is not possible to impose restrictions based on theoretical considerations on the coefficients that are obtained. We therefore only impose a positivity restriction on the forecasts of the equity premium, so that negative forecasts are set to 0.

4. Empirical results

We evaluate the empirical performance of our equity premium modelling approach by first assessing the in-sample fit of the model, and then proceeding to an out-of-sample forecast evaluation. Since the focus of our analysis is to show how macroeconomic information, together with factor selection, combination and positivity constraints can be used to improve out-of-sample forecasts of the equity premium, we provide a direct comparison to the results in Neely et al. (2014), which serve as a benchmark. Note that we consider a somewhat shorter sample period from April 1959 to June 2011 due to the availability of the SW macroeconomic data. For that reason we include a full replication of the relevant findings in Neely et al. (2014) over our shorter sample period to facilitate the comparison. These results are reported in detail in Table A.6 in Appendix A.2. to conserve space and avoid repetition.[17] To summarize our findings here, the explanatory power of the GW predictor set is somewhat weaker over the April 1959 to June 2011 sample period, while that of the TECH technical indicators remains largely the same.

4.1. In-sample results of the macroeconomic factors, selection and combination results

In Table 1 we show the in-sample factor predictive regression results from our proposed modelling approach, which adds the SW macroeconomic panel to the data set of Neely et al. (2014) and then selects and combines the set of *relevant* factors determined by our adaptive Lasso procedure. We also report results for the adaptive Lasso selected factors obtained from the SW macroeconomic panel alone.[18] Further, to provide a broader overview of the in-sample performance of our proposed approach, we show fitting results for three different periods in Table 1. These are from April 1959 to June 2011 (full sample), as well as two sub-periods. Sub-sample 1 is from April 1959 to May 1985. Sub-sample 2 is from May 1959 to June 2011. The two sub-periods are constructed simply by splitting the full sample into two equally sized sub-samples of 313 observations. Our motivation for doing this is to show that the selected factors change non-negligibly, so that it may not be overly informative to take a given time period and to try to interpret these factors by relating them to the underlying variables that make up the panel of predictors.

The results in Table 1 are arranged symmetrically, showing the predictor variable, the estimated slope coefficient, the corresponding t−statistic, as well as goodness of fit measures R^2 (in percent), which are further measured separately over expansion (Exp) and recession (Rec) periods as defined by the NBER in columns 1 to 6, respectively.[19] The top third of Table 1 shows the results for the entire sample

---

[16] As an alternative to the simple forecast combination approach that we consider here, it would also be possible to construct dynamically averaged forecasts from all the factors that are extracted at each point in time. Such a dynamic model averaging approach is considered in Buncic and Piras (2016) and Buncic and Moretto (2015) in the context of exchange rate and copper forecasting. Due to the computational burden that this approach brings, and to avoid introducing another econometric model, we focus on a simple recursive regression representation.

[17] We should re-emphasise here again that a substantial component of the equity premium is unpredictable. The predictability results which we discuss in this section should thus be put into the perspective of other studies in the equity premium forecasting literature. Quantitatively, this means that a monthly R^2 of 0.5%, despite being seemingly small in magnitude, indicates an economically significant degree of predictability, in-sample as well as out-of-sample. From the perspective of an asset pricing model, one should indeed only expect a limited degree of predictability. Nevertheless, this *'small'* magnitude is still likely to have a considerable effect on the utility of a risk-averse investor (see Campbell and Thompson (2008), Zhou (2010), Rapach et al. (2013), Neely et al. (2014) and others for further discussion).

[18] We thank an anonymous referee for suggesting this to us. These results are provided as additional information and are not further discussed in the text.

[19] More concretely, the R^2 across expansion and recessions states are computed as.

$$R_S^2 = 1 - \frac{\sum_{t=1}^{T} 1_t^S \widehat{\varepsilon}_{i,t}^2}{\sum_{t=1}^{T} 1_t^S \left(EQ_t - \overline{EQ}\right)^2}, \forall\, S = \{Rec, Exp\}. \tag{11}$$

The variable 1_t^S is an economic state indicator, defined so that 1_t^{Rec} (1_t^{Exp}) is equal to unity when month t is a recession (expansion) state and zero otherwise. The fitted residuals $\widehat{\varepsilon}_{i,t}$ are from the full sample estimates of the bivariate equity premium regressions in (1). The term $\overline{EQ}$ in (A.8) is the sample mean of EQ_t.

period, the middle for Sub-sample 1 and the bottom for Sub-sample 2. Over each of the three considered sample periods we show results for the selected JOINT factors, the combined relevant factors and for the selected SW factors. In the two sub-samples, we also report results from the Neely et al. (2014) ALL Factors approach analogous to the full sample results given in Table A.6. As is well known, standard asymptotic $t-$ statistics for the estimates in Table 1 are likely to be inflated, with highly persistent regressors further distorting test sizes (see Stambaugh (1999)). We therefore follow Neely et al. (2014) and many others in this literature and perform a wild bootstrap to compute critical values of the $t-$statistics. By using a wild bootstrap, we account for the persistence in the regressors, the correlations between the equity premium and the predictor innovations, and general forms of heteroskedasticity.[20]

From the full sample results in the top of Table 1 we can see that both of the approaches that we propose to reduce the set of factors by either *selecting* the *relevant* factors from the joint panel of data $\mathscr{X}(\text{JOINT}) := [\boldsymbol{x}^{\text{SW}}\ \boldsymbol{x}^{\text{GW}}\ \boldsymbol{x}^{\text{TECH}}]$ or by *combining* the *selected relevant* factors from $\mathscr{X}(\text{ALL}) := [\boldsymbol{x}^{\text{GW}}\ \boldsymbol{x}^{\text{TECH}}]$ and $\mathscr{X}(\text{SW}) := [\boldsymbol{x}^{\text{SW}}]$ lead to marginally worse in-sample fits when compared to the Neely et al. (2014) in-sample results reported in Table A.6, when measured by the regression R^2. We should re-emphasise here that this is expected, since the sole purpose of employing a penalised regression estimator such as the adaptive Lasso as a selection device is to avoid overfitting and loss of out-of-sample forecast performance due to the inclusion of too many 'unimportant' predictors. What is interesting to see here is that the shrinkage that the adaptive Lasso imposes on the factors extracted from $\mathscr{X}(\text{JOINT})$ is quite strong, where only the second PCA factor $\widehat{F}_2^{\text{JOINT}}$ is selected to be included in the OLS post Lasso factor regression. This factor is highly significant and accounts on its own for about 1.37% of the variation in the equity premium. From the "*combination of selected ALL and SW factors*" regression results in the right part of Table 1, it can be seen that the adaptive Lasso shrinks the second ALL factor $\widehat{F}_2^{\text{ALL}}$ to 0, so that only the two significant factors remain from the three ALL factors of Neely et al. (2014).[21] From the vector of SW factors $\widehat{F}^{\text{SW}}$, only the second SW factor $\widehat{F}_2^{\text{SW}}$ is selected by the adaptive Lasso procedure, which enters, nevertheless, '*insignificantly*' into the factor regression. The regression R^2 of 1.80% is only marginally lower than the 1.87% from the ALL factors regression of Neely et al. (2014).

The results from the two sub-samples are rather different. Over the first sub-sample from April 1959 to May 1985, four ALL factors are selected by the Neely et al. (2014) approach, with the overall fit from this model being quite good, yielding an R^2 of nearly 7%. From the vector of $\widehat{F}^{\text{JOINT}}$ factors, the adaptive Lasso selects only one relevant factor, $\widehat{F}_2^{\text{JOINT}}$, thereby again imposing fairly strong shrinkage. This factor is also highly significant and explains nearly 3.5% of the variation in the equity premium. From the combination approach, the selected relevant factors include all four of the ALL factors of Neely et al. (2014) and further three factors from the vector of SW factors $\widehat{F}^{\text{SW}}$. Interestingly, these three SW factors are deemed to be statistically insignificant. Nevertheless, the overall fit of the combination of selected ALL and SW factors approach is rather strong, yielding an R^2 of nearly 8%. As a whole, the in-sample explanatory power of the three different approaches that are presented in the middle of Table 1 is rather high in the first sub-sample.

For the second sub-sample from May 1985 to June 2011, the results shown in the bottom panel of Table 1 are indicative of an overall weakening in the explanatory power of the Neely et al. (2014) ALL factors. This is evident not only from the overall drop in the R^2 to 1.47%, but also from the fact that only two of the possible maximum of 4 Neely et al. (2014) ALL factors are selected, i.e., $\widehat{F}_1^{\text{ALL}}$ and $\widehat{F}_2^{\text{ALL}}$. From the selected JOINT factors results, it is interesting to see that three JOINT factors from $\widehat{F}^{\text{JOINT}}$ are now selected, rather than only one as in the full and first sub-sample periods. This leads to a noticeable increase in the fit, with an R^2 of nearly 3%. The combination approach reported on the right side in the bottom of Table 1 selects the first ALL factor and the first two SW factors, providing an in-sample R^2 of 2.76%. Overall, it can be seen that the statistical significance of the selected models is weaker in the second sub-sample than in the first.

4.2. Out-of-sample forecast evaluation

We use data from April 1959 to April 1974 as our in-sample fitting period, and then assess the out-of-sample forecast performance of our proposed modelling approach over the evaluation period from May 1974 to June 2011. The length of the in-sample fitting period is the same as in Neely et al. (2014), that is, 181 observations are used initially to start the forecast recursions, leaving 446 data points for out-of-sample evaluation. We would like to emphasise here that we *intentionally* choose the number of in-sample observations to be exactly the same as in Neely et al. (2014) to avoid concerns related to data mining, that is, searching over the best '*fitting*' out-of-sample period. As can be seen from the actual and fitted plots in Fig. 1, this choice of the number of in-sample observations implies an out-of-sample period that starts during the recession of the first oil price shock in 1974. This period is marked by a severe downturn in the equity premium which, evidently, has an adverse effect on the forecast performance of all considered models, but most notably so on the ones that we propose in this paper.

We also follow Neely et al. (2014) and employ an expanding estimation window in our out-of-sample analysis. Such an approach mimics the real time behaviour of investors or financial market participants, which typically update their prediction models as new data

---

[20] We use standard '*' notation to denote significance at the 1, 5, and 10 percent levels, with significance here being defined as the probability of the wild bootstrap draws being larger than the estimated slope coefficient.

[21] It may seem a bit unusual to see here that although their magnitudes are nearly unchanged, their statistical significance is somewhat weaker, so that both factors are significant only at the 10% level, rather then at the 1% and 5%. The three ALL factors are orthogonal by construction, so that the exclusion of one of these three factors should not have any effect on the size of the coefficients or their standard errors. So the change in the level of significance must be coming in from the second SW factor that is included in the regression. Indeed, computing the correlation matrix of the regressors for this regression shows that the correlation between $\widehat{F}_1^{\text{ALL}}$ and $\widehat{F}_2^{\text{SW}}$ is around 0.3.
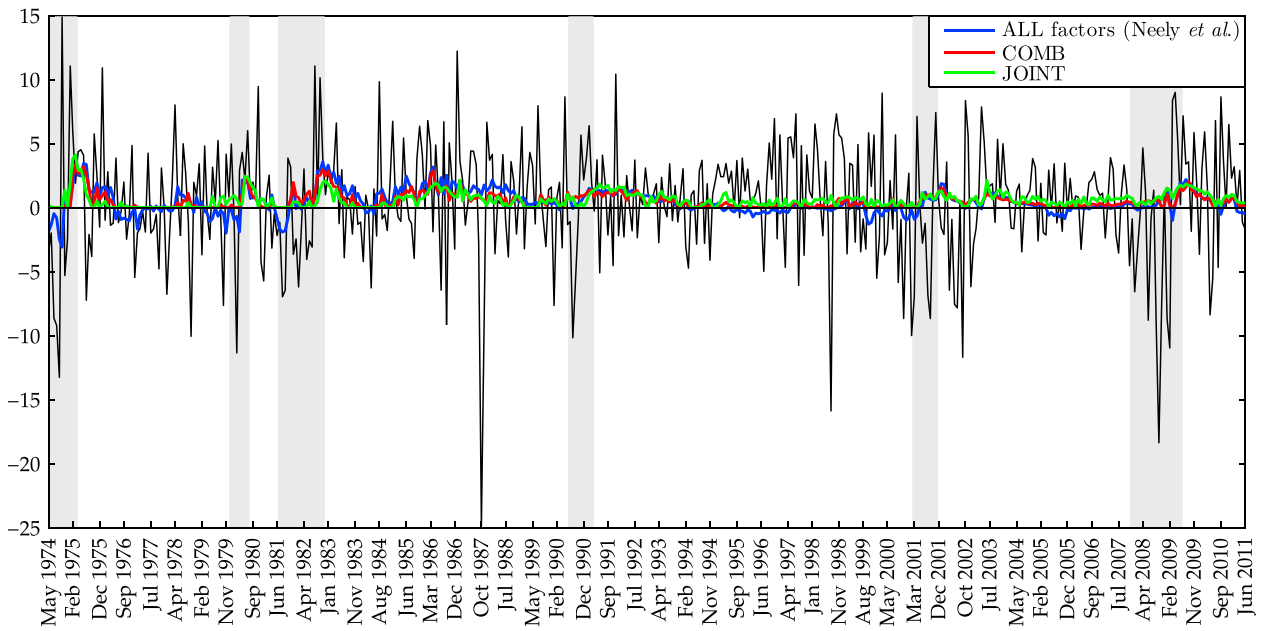
Fig. 1. Comparison of actual and predicted values over the out-of-sample period from May 1974 to June 2011. The actual equity premium is drawn by the thin solid black line. Forecasted values from the ALL factors model of Neely et al. (2014) as well as our proposed COMB and JOINT approaches which include the positivity restrictions are drawn by the solid blue, red and green lines, respectively.

become available. Further, using an expanding estimation window can accommodate, at least to some extent, parameter and/or model instabilities which may have occurred over time. One last point that we would like to stress here is that we extract the factors recursively trough time as new information becomes available, and not based on the full sample. This mimics the real time flow of information accessible to the forecasting agent and avoids the well known 'look ahead bias' when constructing forecasts that rely upon measures extracted from the data.[22]

4.2.1. Evaluation criteria

We assess the out-of-sample forecast performance of our proposed approach by means of the Clark and West (2007) MSFE adjusted $t-$statistic (denoted by CW$-$ statistic henceforth) and also the Campbell and Thompson (2008) out-of-sample R^2 (denoted by R_{os}^2). We follow the suggestion in Clark and West (2007, page 294) and compute the MSFE adjusted $t-$statistic as:

$$\text{CW} - \text{statistic} = \frac{\overline{cw}}{\sqrt{\mathbb{V}ar(\overline{cw})}}, \tag{12}$$

where $\overline{cw} = T_{os}^{-1} \sum_{t=T_{is}}^{T} cw_{t+1}$ and $\mathbb{v}(\overline{cw})$ is the variance of the (out-of-sample) mean of cw_{t+1}, which can simply be obtained as the heteroskedasticity and autocorrelation (HAC) robust $t-$statistic on the intercept term from a regression of cw_{t+1} on a constant.[23] The cw_{t+1} sequence is defined as:

$$cw_{t+1} = \left[\widehat{e}_{t+1|t}^{\text{HA}}\right]^2 - \left[\widehat{e}_{t+1|t}^{\mathcal{M}}\right]^2 + \left[\widehat{\text{EQ}}_{t+1|t}^{\text{HA}} - \widehat{\text{EQ}}_{t+1|t}^{\mathcal{M}}\right]^2, \tag{13}$$

where $\widehat{\text{EQ}}_{t+1|t}^{\text{HA}} = \frac{1}{t} \sum_{\tau=1}^{t} \text{EQ}_\tau$ and $\widehat{e}_{t+1|t}^{\text{HA}} = \text{EQ}_{t+1} - \widehat{\text{EQ}}_{t+1|t}^{\text{HA}}$ are historic average forecasts and corresponding one-step ahead forecasts errors, respectively.[24] The superscript $\mathcal{M}$ in (13) is used as an index for the different forecasting approaches that are considered in the out-of-sample comparison. That is, $\mathcal{M} = \{\text{ALL}, \text{JOINT}, \text{COMB}\}$, where ALL denotes the Neely et al. (2014) ALL factor modelling approach, and JOINT and COMB the prediction approaches defined in (6) and (9), respectively.

---

[22] Similar issues arise when using output gaps as predictor variables, as these require the 'estimation' of the permanent component of GDP and this is frequently done with a HP filter. To avoid look ahead bias, it is thus common to roll through the out-of-sample period and construct an estimate of the permanent component at each point in time.

[23] See also the discussion in Section 2.1 in Diebold (2015) for more background on this in the context of the traditional Diebold-Mariano (DM) statistic.

[24] Note here that the $[\widehat{e}_{t+1|t}^{\text{HA}}]^2 - [\widehat{e}_{t+1|t}^{\mathcal{M}}]^2$ term is the standard Diebold and Mariano (1995) sequence that is computed to test for (unconditional) superior predictive ability. The second term $[\widehat{\text{EQ}}_{t+1|t}^{\text{HA}} - \widehat{\text{EQ}}_{t+1|t}^{\mathcal{M}}]^2$ is an adjustment term that arises due to the nested nature of the models being compared and performs a bias correction (see Clark and West (2007) for more details).

Note here that the CW−statistic assesses a test of the null hypothesis that the MSFE of the HA benchmark model is equal to the MSFE of the forecasts from model $\mathcal{M}$, against the one sided alternative hypothesis that the historic average's MSFE is greater than that of model $\mathcal{M}$. A rejection of the null hypothesis hence suggests that forecasts from model $\mathcal{M}$ are (on average and unconditionally) *significantly* better than those from HA. It should be highlighted here that the CW−statistic is particularly suitable in the given context, as it is designed for a comparison of *nested* (forecasting) models. The benchmark model is the HA model, which can be obtained from the predictive factor regression models by restricting the slope coefficient to zero.

The Campbell and Thompson (2008) R_{os}^2 is defined as:

$$R_{os}^2 = 1 - \frac{\text{MSFE}^{\mathcal{M}}}{\text{MSFE}^{\text{HA}}}, \tag{14}$$

where MSFE^{ℓ} for all $\ell = \{\text{HA}, \mathcal{M}\}$ is the mean squared forecast error computed over the entire out-of-sample period T_{os}, that is, $\text{MSFE}^{\ell} = T_{os}^{-1} \sum_{t=T_{is}}^{T} [\widehat{e}_{t+1|t}^{\ell}]^2$. Intuitively, the R_{os}^2 statistic in (14) measures the reduction in the MSFE of the proposed model $\mathcal{M}$ relative to the benchmark HA model. When $R_{os}^2 > 0$, then this is an indication that the proposed model performs better than the benchmark model in terms of MSFE, while $R_{os}^2 < 0$ suggests that the benchmark model performs better.

In addition to the CW− statistic of Clark and West (2007) and the out-of-sample R^2 of Campbell and Thompson (2008), we also compute and plot the cumulative difference of the squared forecast errors (cumSFE) of the HA and $\mathcal{M}$ model forecasts over the out-of-sample period. This cumulative difference (denoted by cumSFE) is commonly used in the forecasting literature as a visual tool to highlight the predictive performance of a proposed model relative to the benchmark over time (see Goyal and Welch (2008) and Rapach, Strauss, and Zhou (2013), among many others). In our setting, this difference is computed as:

$$\text{cumSFE}_{t+1} = \sum_{\tau=T_{is}}^{t} \left(\left[\widehat{e}_{\tau+1|\tau}^{\text{HA}} \right]^2 - \left[\widehat{e}_{\tau+1|\tau}^{\mathcal{M}} \right]^2 \right), \forall t = T_{is}, \dots, T. \tag{15}$$

A value of cumSFE$_{t+1}$ above zero indicates that the cumulative sum of the squared forecast errors from the historical average benchmark are larger than those of the proposed model $\mathcal{M}$, suggesting that model $\mathcal{M}$ produces more accurate forecasts. In general, a rising value in cumSFE$_{t+1}$ means that model $\mathcal{M}$ forecasts produce better predictions than the HA.

4.2.2. Forecast evaluation results

Before we turn our attention to a formal evaluation of the out-of-sample forecast performance, it will be instructive here to perform a fairly simple test to determine what the value added of the information from the SW macroeconomic panel is. We investigate this by means of an encompassing test developed by Harvey, Leybourne, and Newbold (1998). This test assess whether the out-of-sample forecasts of the equity premium using only the ALL factors of Neely et al. (2014) can be improved upon when the selected SW factors are added as regressors.

We implement the test within the following regression setup (see equation (2) in Harvey et al. (1998)):

$$\widehat{e}_{t+1|t}^{\text{ALL}} = \eta \left(\widehat{e}_{t+1|t}^{\text{ALL}} - \widehat{e}_{t+1|t}^{\text{SW}} \right) + \nu_{t+1}, \forall t = T_{is}, \dots, T, \tag{16}$$

where $\widehat{e}_{t+1|t}^{\text{ALL}}$ and $\widehat{e}_{t+1|t}^{\text{SW}}$ are, respectively, one-step-ahead forecast errors from the ALL factor model of Neely et al. (2014) and the adaptive Lasso selected factors $\widehat{f}_t^{\text{SW}}$. Under the null hypothesis of $\mathscr{H}_0 : \eta = 0$, we have that the forecasts $\widehat{\text{EQ}}_{t+1|t}^{\text{ALL}}$ from the ALL factor model of Neely et al. (2014) '*encompass*' those from $\widehat{\text{EQ}}_{t+1|t}^{\text{SW}}$, so that there is no relevant excess information in $\widehat{\text{EQ}}_{t+1|t}^{\text{SW}}$ once conditioning on $\widehat{\text{EQ}}_{t+1|t}^{\text{ALL}}$. If the above null hypothesis is rejected, than $\widehat{\text{EQ}}_{t+1|t}^{\text{SW}}$ contains information beyond that in $\widehat{\text{EQ}}_{t+1|t}^{\text{ALL}}$, which can be exploited to improve forecasts of the equity premium. We obtain a point estimate of 0.73 for η, with a HAC robust t−statistic of 2.71, resulting in a one-sided p−value of 0.0034. Thus, the null hypothesis that the ALL factors forecasts $\widehat{\text{EQ}}_{t+1|t}^{\text{ALL}}$ encompass those from $\widehat{\text{EQ}}_{t+1|t}^{\text{SW}}$ is strongly rejected.[25] From this we can conclude that the subset of selected SW factors contain valuable information that can be used to improve out-of-sample forecasts of the equity premium.

Formal out-of-sample forecast evaluation results are presented in Table 2 below. Column 1 lists the different modelling approaches that are being evaluated. Columns 2 to 6 show, respectively, the MSFE, the Campbell and Thompson (2008) out-of-sample R^2 (in percent), the Clark and West (2007) MSFE adjusted t−statistic (CW− statistic), and the out-of-sample R^2 computed over expansion and recession periods separately (also in percent). In Table 2 we show the historic average (HA) MSFE in the top left and then three different blocks of models that are evaluated. In the top block we show the replicated results for the three different ECON, TECH and ALL factor groups of Neely et al. (2014), following exactly their factor selection approach that is based on the adjusted R^2. Here we also report the results from the selected SW factor regressions. The last two blocks show the results based on our proposed JOINT and COMB approaches as described in Section 3.3 (see equations (6) and (9) in particular), where the results in the bottom block further impose the

---

[25] Results based on the d_t sequence defined in equation (16) in Harvey et al. (1998), which in our setup is defined as $d_t = \widehat{e}_{t+1|t}^{\text{ALL}}(\widehat{e}_{t+1|t}^{\text{ALL}} - \widehat{e}_{t+1|t}^{\text{SW}})$ leads qualitatively to the same conclusion, yielding a HAC t−statistic of 2.5907 with a one-sided p−value of 0.0096.

Table 2
Out-of-sample forecast evaluation results: May 1974 to June 2011.

(1) Modelling approach	(2) MSFE	(3) R^2_{os}	(4) CW–statistic	(5) $R^2_{os,Exp}$	(6) $R^2_{os,Rec}$
Historic average (HA)	21.0473				
Neely et al. (2014) ECON factors	21.6309	−2.7729	0.9739	−3.8160	−0.5881
Neely et al. (2014) TECH factor	20.9167	0.6202	1.2601	−0.0249	1.9714
Neely et al. (2014) ALL factors	21.2795	−1.1033	1.5032*	−3.6682	4.2690
Selected SW factors	20.9962	0.3378	1.4837	0.7429	−0.5107
JOINT, without $\widehat{EQ}_{t+1|t} > 0$ restriction	21.2997	−1.1993	1.0062	−2.2425	0.9858
COMB, without $\widehat{EQ}_{t+1|t} > 0$ restriction	20.9951	0.2477	1.6315*	−1.0149	2.8923
JOINT, with $\widehat{EQ}_{t+1|t} > 0$ restriction	21.0830	−0.1699	1.4142*	−0.8241	1.2002
COMB, with $\widehat{EQ}_{t+1|t} > 0$ restriction	20.9066	0.6682	2.1609**	−0.0881	2.2525

Notes: This table reports out-of-sample forecast evaluation results over the May 1974 to June 2011 of the various forecasting approaches that we consider. In column one, the considered models are listed. In columns 2 to 6, the mean squared forecast error (MSFE), the out-of-sample R^2 of Campbell and Thompson (2008), the CW– statistic of Clark and West (2007) and expansion and recession R^2_{os} are listed. Significance at the 1, 5, and 10 percent levels are marked by ***, ** and *, respectively. The top third of the table shows the replication results of the Neely et al. (2014) approach over the May 1974 to June 2011 period. The middle third of the table shows the results of our proposed JOINT and COMB modelling approach without the positivity constraint on the equity premium forecast being imposed. At the bottom of the table we show the JOINT and COMB results with the positivity constraint being imposed.

positivity restriction discussed in Section 3.3.3 on the forecasts of the equity premium from the JOINT and COMB approaches.

From the results in Table 2 we can see that, as with the in-sample evaluation, overall out-of-sample predictability using the ECON, TECH, or ALL factors of Neely et al. (2014) is noticeably weaker over our evaluation period from May 1974 to June 2011 than reported in their paper.[26] Interestingly, only the out-of-sample forecast results based on the TECH factor produce a positive and sizeable R^2_{os} of 0.6202%, which, nonetheless, yields a fairly low (and statistically insignificant) CW– statistic. Conversely, while the ALL factors based forecasts produce a significant CW– statistic at the 10% level, the R^2_{os} is negative, with a value of − 1.1033%.[27] The ECON factors based forecasts of the equity premium produce a negative R^2_{os} of − 2.7729% and also a small and insignificant CW– statistic.

Looking over the JOINT and COMB results which do not impose the positivity restriction on the forecasts of the equity premium in the middle of Table 2, we see that in particular the combination of the separately (adaptive Lasso) selected factors $\widehat{f}_t^{ALL}$ and $\widehat{f}_t^{SW}$ (COMB approach) seems to noticeably improve the out-of-sample forecast performance. The out-of-sample R^2 is 0.2477%, with a CW– statistic of 1.4142, which is significant at the 10% level. The performance of the JOINT approach is somewhat less satisfactory, with the out-of-sample forecast evaluation results being similar to those of the Neely et al. (2014) ALL factors.[28] That is, the R^2_{os} is negative at − 1.1993% and the CW– statistic is around 1. A possible explanation of why the JOINT approach does not yield as good out-of-sample forecast results as the COMB approach could be due to the difficulty of extracting factors from a mix of highly persistent data (ie., the ECON and TECH predictors) and the much less persistent SW data. The COMB approach is not affected by this difference in persistence, as the forecasts of the equity premium are combined rather than the variables themselves, which will be only weakly autocorrelated.

From the results of the JOINT and COMB approaches which do incorporate the positivity restriction on the forecasts of the equity premium shown down the bottom of Table 2, it is evident that statistically significant improvements over the historic average forecast are attained. More specifically, for the JOINT approach, the CW– statistic is significant at the 10% level, while for the COMB approach, the CW– statistic is significant at the 5% level, being as high as 2.1609. The out-of-sample R^2 is sizeable at 0.6682% for the COMB forecasts, being nonetheless, negative for the JOINT forecasts. Overall the statistical results suggest that the inclusion of macroeconomic information in the form of factors or diffusion indices, together with factor selection and combination methods, as well as positivity restrictions on the forecasts of the equity premium can be successfully used to improve out-of-sample forecast performance. To get a feel for how well the different models fit the equity premium data, to conserves space without any discussion, we show a plot of actual and predicted values in Fig. 1.

To provide a visual assessment of the forecast performance of the models over time, we show plots of the cumulative sum of squared forecast errors (cumSFE) defined in (15) for the May 1974 to June 2011 period in Fig. 2. The cumSFE of the COMB and JOINT approaches are drawn as red and green solid lines for the better performing models that impose the positivity restriction on the equity premium forecasts. Red and green dashed lines show the cumSFE for COMB and JOINT that do not impose this positivity restriction. The

---

[26] We will come back to that when we look at the full comparison later.

[27] Note here that, as discussed in Neely et al. (2014), as well as Clark and West (2007) and McCracken (2007), this seemingly unusual result of having a negative R^2_{os} and yet a significant CW– statistic is due to the fact that a test comparing nested models is performed. The predictive (factor) regression model will produce noisier forecasts than the historical average when the true data generating process is a simple constant mean model, because of the need to extract the factors and estimate the slope coefficients. The sampling variation from the parameter estimation then translates into noisier forecasts.

[28] We will see later in Fig. 2, that not only the R^2_{os} are similar in magnitude, but also the actual forecasts track each other reasonably closely.
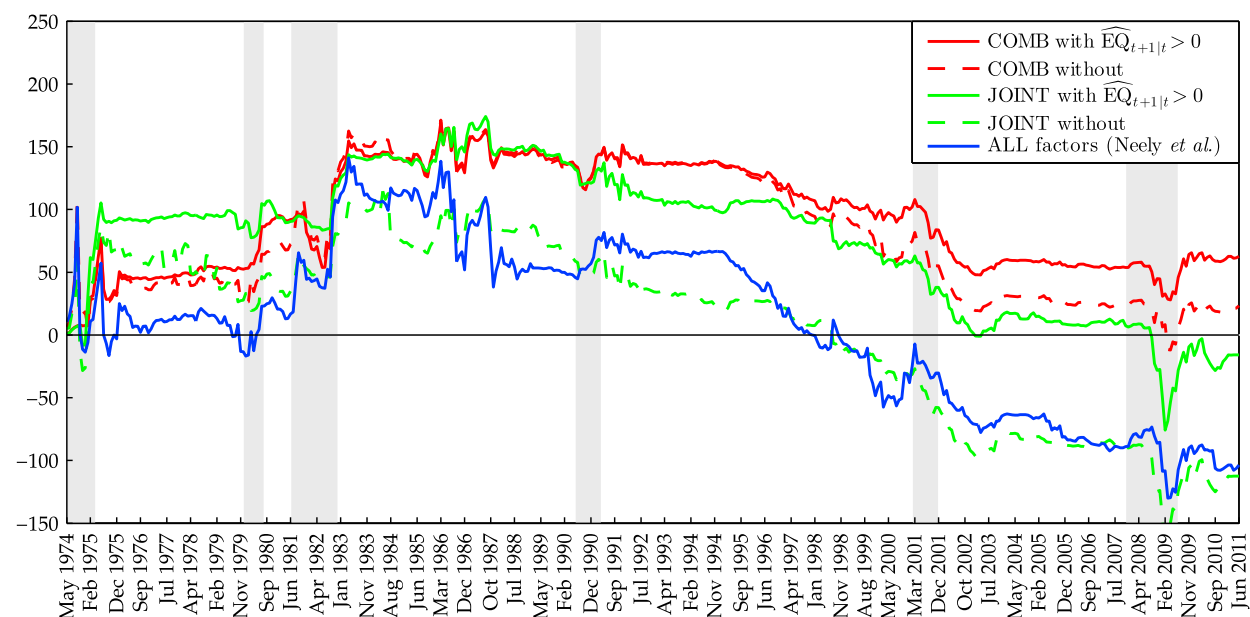
Fig. 2. Time series plots of the cumulative sum of squared forecast errors (cumSFE) for the out-of-sample period from May 1974 to June 2011. The red, green and blue solid lines are cumSFEs from the COMB, JOINT and ALL factors approach of Neely et al. (2014), respectively, where the COMB and JOINT forecasts also impose the positivity restriction that the equity premium forecast $\widehat{EQ}_{t+1|t}$ has to be positive (denoted by "with $\widehat{EQ}_{t+1|t} > 0$" in the legend). The black solid line shows the results for the selected SW factors. The dashed green and red lines show forecasts from the same COMB and JOINT approaches, but where the positivity constraint is not enforced. The benchmark model in the computation of the cumSFE is the historic average (HA).

blue solid line provides the benchmark cumSFE of the ALL factors of Neely et al. (2014).[29] The black solid line shows the performance of the selected SW factor forecasts. Vertical shadings in Fig. 2 highlight U.S. recession periods as classified by the NBER.

The most interesting features of the cumSFE sequence plotted in Fig. 2 can be summarised as follows. First, the cumSFE of the COMB and JOINT equity premium forecasts which include the positivity constraint are consistently above zero, with the only exception being the late 2008 period onwards, where the forecasts from the JOINT modelling approach produce cumSFEs that fall below 0. Most of the predictive gains seem to be realised during the early part of the out-of-sample evaluation period, with the recession in the early 1980s providing a boost in predictive performance to all models that we consider. From the early 1990s onwards, the performance starts to worsen consistently for all models (relative to the HA benchmark forecast), with, nonetheless, the combination approach still providing the best model based forecasts. Second, the benefits from imposing the positivity constraint on the forecasts from the COMB and JOINT approaches is clearly visible by the solid green and red lines being consistently above their dashed counterparts. Overall, the imposition of this constraint improves predictions from the JOINT modelling approach considerably more than for the COMB approach. Third, the cumSFE sequence from the forecasts of the ALL factors approach of Neely et al. (2014) is only from 1980 to 1998 consistently above zero, highlighting the overall much weaker performance over our May 1974 to June 2011 out-of-sample period.

Given the statistical as well as visual evidence of an overall weaker performance of the Neely et al. (2014) ALL factors approach over our out-of-sample period, we show in panel (a) of Fig. 3 a comparison to the ALL factors cumSFE computed over the January 1966 to April 1974 out-of-sample period to understand why these differences occur. This ALL factors cumSFE from January 1966 to April 1974 is drawn as the solid black line in panel (a) of Fig. 3. Note that we set the $x-$axis date range to be from January 1966 to December 2011, and then fill the unavailable first 100 and last 6 observations of our COMB, JOINT and ALL cumSFE vectors with 'NANs' to be able to plot these together in one figure. Also, we normalise the COMB, JOINT and ALL factors cumSFEs by adding the May 1974 value of the ALL factors cumSFE from the original Neely et al. (2014) data to them. The two vertical black lines mark our out-of-sample evaluation period from May 1974 to June 2011. The horizontal black line shows the location of the corresponding zero line in Fig. 2.

From panel (a) in Fig. 3 we can see that a substantial part of the forecast improvement over the original Neely et al. (2014) time frame is realised at the very beginning of the out-of-sample period, that is, from January 1966 to mid 1974.[30] This creates a fairly large base in the forecast improvement, leading to an overall out-of-sample R^2 of 1.79% over the full January 1966 to December 2011 period. Recall that our out-of-sample R^2 values were smaller at 0.6682% and -0.1699%. To account for the positive impact of the initial out-of-sample period (which is not available to us) on the forecast evaluation results, we perform a counterfactual analysis to measure what our R_{os}^2 values would have been if one was to *'incorporate'* this initial improvement into our COMB and JOINT predictions. We implement this by

---

[29] We do not show plots of the ECON and/or TECH factor forecasts reported in Table 2 to avoid clutter and to maintain readability. The ALL factors model is considered to be the best performing model in Neely et al. (2014).

[30] We can see from the plots that the predictive improvement as a whole is quite substantial during the (first oil price shock induced) recession from 1973 to 1975.
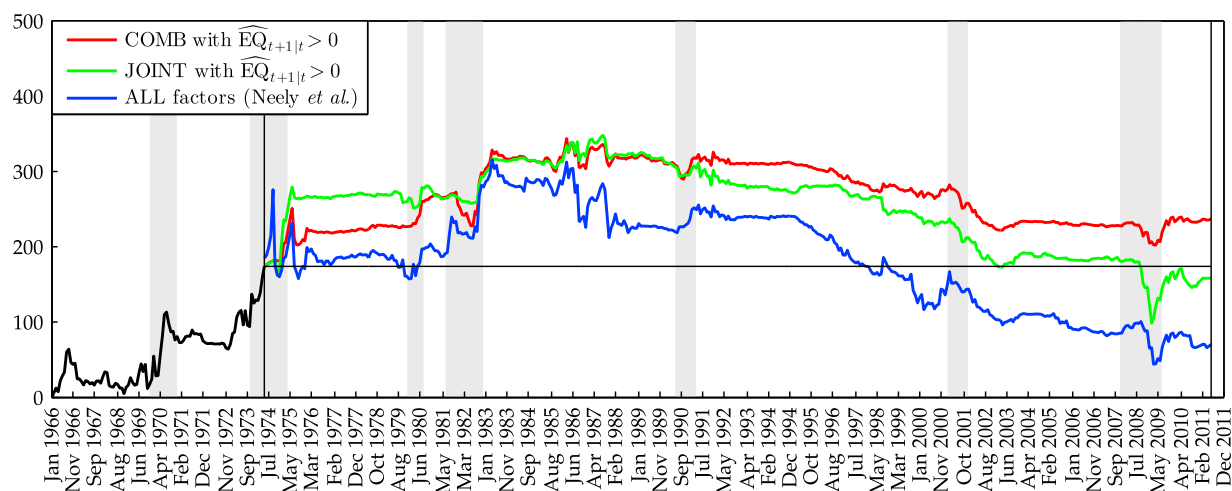
(a) Time series plot of cumSFE over Neely *et al.* out−of−sampel period (January 1966 to December 2011)
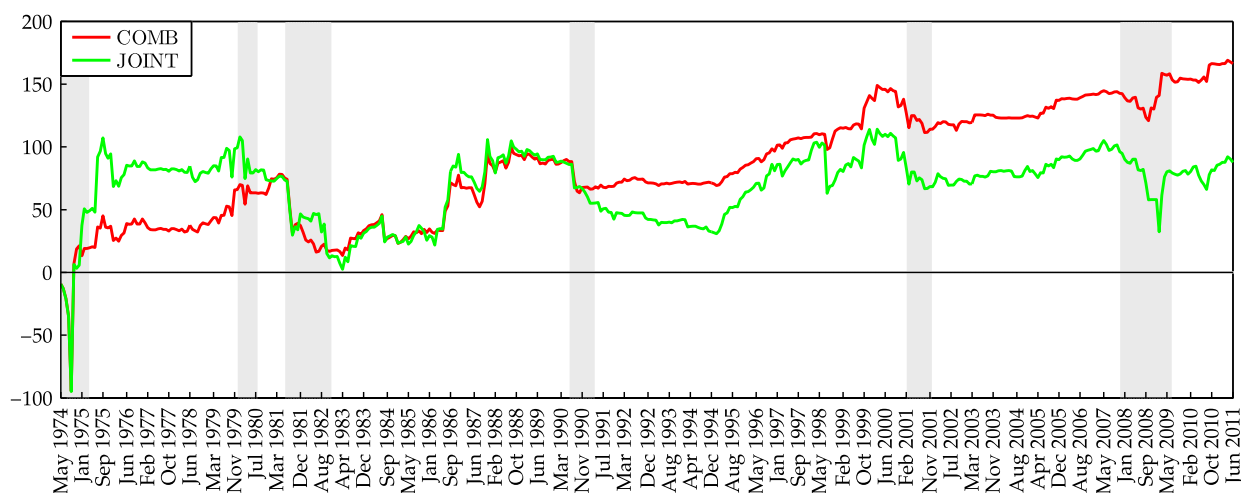
(b) Time series plot of cumSFE with ALL factor model of Neely *et al.* as the benchmark (May 1974 to June 2011)

Fig. 3. Comparison of cumSFEs. In the top panel we show the cumSFE evolution (relative to HA forecast) starting from January 1966, which is the beginning of the out-of-sample period in Neely et al. (2014). From the period January 1966 to April 1974 we compute the squared forecast errors exactly as in Neely et al. (2014) and then concatenate these to the squared forecasts errors that we obtain from our out-of-sample period beginning in May 1974. The cumSFE over the January 1966 to April 1974 that is obtained from the original Neely et al. (2014) files is shown by the black solid line. The COMB, JOINT and ALL factors cumSFEs are marked by red, green and blue lines, respectively. The vertical thin black lines in May 1974 and June 2011 mark the beginning and end of our out-of-sample period. The horizontal line denotes the corresponding 0 line in Fig. 2. The bottom of the panel shows the cumSFEs of the COMB and JOINT forecasting approaches (which include the positivity constraints), but now with the ALL factor model of Neely et al. (2014) serving as the benchmark.

Table 3
Asset allocation results: May 1974 to June 2011.

(1) Model	(2) CER	(3) Δ CER	(4) ΔCER_{Exp}	(5) ΔCER_{Rec}	(6) Sharpe ratio	(7) Turnover	(8) Δ CER(50bp)
HA	3.2025				0.0515	2.1019	
Neely et al. (2014) ALL factors	3.1890				0.0517	11.1152	
JOINT	5.8025	2.6135	0.0910	10.4255	0.0818	8.9540	2.2754
COMB	6.5446	3.3556	−0.7246	11.5705	0.0972	8.3545	3.0174

Notes: This table reports the portfolio performance from a simple asset allocation exercise, following the set-up in Neely et al. (2014). The first column lists the different models that are evaluated. Column 2 shows the certainty equivalent return (CER) for a mean-variance investor, as defined in (19) in annualised percentage terms (ie., multiplied by 1200). Column 3 shows the relative CER denoted by ΔCER using the Neely et al. (2014) ALL factor model as the benchmark. Columns 4 and 5 show the same ΔCER values, but now computed over expansion and recession periods separately. Columns 6 and 7 give the monthly Sharpe ratio and average monthly turnover. Column 8 shows the ΔCER measure when proportional transaction costs of 50bp are included in the calculation of the CER.

simply taking the forecast errors from the original ALL factors of Neely et al. (2014) for the January 1966 to April 1974 period and concatenating these to the forecast errors obtained from our COMB and JOINT approaches from May 1974 to June 2011. These concatenated forecast errors are then used to construct counterfactual MSFEs and out-of-sample $R^{2'}$s. For the COMB and JOINT forecasting approaches which impose the positivity constraint on the equity premium predictions, we obtain counterfactual R_{os}^2 values of 1.9537% and 1.2361%, respectively.

Although the above presented counterfactual analysis gives an interesting indication of how our proposed COMB and JOINT approaches would have performed if macroeconomic data for the full out-of-sample period of Neely et al. (2014) would have been available to us, we now perform a direct comparison of our COMB and JOINT factor forecasts to the ALL factor forecasts of Neely et al. (2014) over the May 1974 to June 2011 period. To provide an initial visual assessment of this comparison, we show the time series evolution of the $cumSFE_{t+1}$ sequence in panel (b) of Fig. 3, where the forecast errors from the ALL factor model now serve as the benchmark, that is, we use $\hat{e}_{t+1|t}^{ALL}$ in place of $\hat{e}_{t+1|t}^{HA}$ in (15). As can be seen from the red and green solid lines, from January 1975 onwards, the COMB and JOINT approaches (with positivity constraint) produce a consistently above zero and, more importantly, upward trending cumSFE sequence. Only the first few observations at the very beginning of the evaluation period are negative (one of these substantially so). Nevertheless, despite this negative influence, out-of-sample R^2 values are 1.7522% and 0.9232% for the COMB and JOINT approaches, producing respective CW− statistics of 1.8913 and 1.8042, which are significant at the 5% level (one sided p−values are 0.0293 and 0.0356). If one was to drop the first 5 out-of-sample observations, the R_{os}^2 values would have been 2.8244% and 1.9757%, with corresponding CW− statistics of 2.5873 and 2.3346, highlighting further the strong improvement over the ALL factors forecasts of Neely et al. (2014).

Overall, we can summarize this section by concluding that significant improvements in out-of-sample forecast performance can be obtained when macroeconomic information in the form of factors are incorporated into statistical prediction models that utilize factor selection and forecast combination methods. Moreover, as emphasized in Campbell and Thompson (2008), imposing positivity restrictions on forecasts of the equity premium further improve out-of-sample performance.

4.3. Assessing the economic value

How much economic value does the statistical forecast improvement bring to an investor in terms of utility? To answer this question, we follow again the set-up in Neely et al. (2014), Campbell and Thompson (2008) and many others, and measure the economic value of our equity premium forecasts for a risk-averse investor. That is, we compute the certainty equivalent return (CER) for a mean-variance investor that allocates her wealth across a portfolio with one risky and one risk-free asset, where the allocation decision is based on forecasts of the equity premium. The share α_t of the portfolio that is allocated to the risky asset (at the end of month t which is to be held until month $t+1$), is determined by

$$\alpha_t = \gamma^{-1} \frac{\widehat{EQ}_{t+1|t}}{\hat{\sigma}_{t+1|t}^2}, \tag{17}$$

where γ is a parameter that specifies the relative risk aversion of the investor, $\widehat{EQ}_{t+1|t}$ is the forecast of the equity premium given information up to time t, and $\hat{\sigma}_{t+1|t}^2$ is the variance of $\widehat{EQ}_{t+1|t}$. The return on the portfolio in month $t+1$ (denoted by r_{t+1}^p) is given by[31]:

$$r_{t+1}^p = r_{t+1}^f + \alpha_t EQ_{t+1}, \tag{18}$$

---

[31] Recall that the equity premium is constructed as the difference between the log return on the S&P 500 index and the risk free bill, that is, as $r_{t+1}^{SP500} - r_{t+1}^f$. When constructing a weighted portfolio with one risky and one risk-free asset, we would thus form the portfolio as $\alpha_t r_{t+1}^{SP500} + (1 - \alpha_t) r_{t+1}^f$, which can of course be rewritten as $r_{t+1}^f + \alpha_t (r_{t+1}^{SP500} - r_{t+1}^f)$, where the latter term is, by definition, the equity premium EQ_{t+1}.

where r_{t+1}^{f} is the return on the risk-free asset and EQ_{t+1} is the realised equity premium at time $t+1$. The CER of the investor's portfolio is then given by

$$CER_p = \widehat{\mu}_p - \frac{1}{2}\gamma\widehat{\sigma}_p^2, \tag{19}$$

where $\widehat{\mu}_p$ and $\widehat{\sigma}_p^2$ are the (sample) mean and variance of the portfolio return r_{t+1}^p, respectively.

Our asset allocation results are reported in Table 3.[32] We follow standard practice in the literature and constrain the portfolio weights α_t to lie between 0 and 1.5. In line with the formulation in Neely et al. (2014), we further set the relative risk aversion coefficient γ to 5 and use a five-year rolling window of past monthly returns to obtain an estimate of the forecast variance $\widehat{\sigma}_{t+1|t}^2$. Column 1 in Table 3 lists the models that are being assessed. Columns 2 and 3 show the CER and relative CER (denoted by ΔCER) both multiplied by 1200 to be interpreted as annual percentage portfolio fees that an investor would be willing to pay to have access to our predictive regression forecasts. The relative CERs for the JOINT and COMB models are computed with the Neely et al. (2014) ALL factor model as the benchmark, that is:

$$\Delta CER^{\zeta} = 1200\big(CER^{\zeta} - CER^{ALL}\big), \tag{20}$$

for $\zeta = \{\text{JOINT}, \text{COMB}\}$. Columns 4 and 5 show the same ΔCER measure, however now computed over expansion and recession periods separately. Columns 6 and 7 show monthly Sharpe ratios and monthly (average) turnover, which is the percentage of wealth traded each month.[33] The last column contains relative CERs (or CER gains) when transaction costs are incorporated into the computation of the utility gain. In line with the evaluation approach in Neely et al. (2014), we consider a proportional transaction cost of 50 basis points (bp) per transaction.[34]

From the results in Table 3 we can initially see that for the historical average (HA) forecast, the CER, Sharpe ratio as well as turnover values over our May 1974 to June 2011 period are similar in magnitude to the values reported in Neely et al. (2014). For instance, the CER is somewhat lower at 3.2 as opposed to the 3.54 reported in Neely et al. (2014), the monthly Sharpe ratio is marginally higher at 0.0515, compared to the 0.05 in Neely et al. (2014), while average monthly turnover in our sample period is at 2.10%, thus very close to the 2.09% value reported in Neely et al. (2014). Looking over the ALL factor model results, we can see that the CER is marginally smaller than for our HA forecasts, while the Sharpe ratio is nearly the same. A noteworthy difference to the results in Neely et al. (2014) is that the (absolute) monthly turnover is around 11 over our sample period, with a relative turnover (relative to HA) of only about 5.29, which is quite a bit lower than the relative monthly turnover of 7.51 found in Neely et al. (2014).

Turning to the JOINT and COMB asset allocation results in the bottom part of Table 3, it is evident that the improvements in CER over the ALL factor model of Neely et al. (2014) are sizable, yielding annual gains of 2.61% and 3.35%, respectively. As can be seen in Table 3 and also in line with the results in Neely et al. (2014), most of these gains are realised over recession periods. Monthly Sharpe ratios that are based on portfolios constructed from the JOINT and COMB forecasting approaches are noticeably smaller at 0.0818 and 0.0972 when compared to the ALL factor model results in Neely et al. (2014) of 0.16. This lower result in terms of the Sharpe ratio is due to the overall weaker out-of-sample forecast performance when compared to the longer period in Neely et al. (2014). What is interesting to see from the JOINT and COMB results is that average monthly turnover is quite low when compared to the ALL factors approach, with absolute values of around 8.95 and 8.35, respectively (or relative values of 4.26 and 3.97). What this indicates is that the positivity restriction imposes more smoothness on the portfolio weights α_t, resulting in a lower number of transactions (on average). Lower turnover helps to keep the ΔCER(50bp) higher when accounting for the impact of trading costs on the certainty equivalent return. As is evident, the drop in ΔCER when proportional trading costs of 50bp are included is only of an order of around 0.3.

5. Conclusion

We show how out-of-sample forecasts of the equity premium can be significantly improved by incorporating the following three aspects into the forecasting design. First, by including macroeconomic data in the form of factors or *diffusion indices* into the information set, second, by improving the selection of the most relevant factors and by combining the most relevant factors by means of combination regressions, and third, by imposing theoretically motivated positivity restrictions on the forecasts of the equity premium. Our findings demonstrate that the information contained in the factors extracted from the macroeconomic database of Stock and Watson (2002a, b) improve upon the recently proposed ALL factors based approach of Neely et al. (2014), which combines the traditional Goyal and Welch (2008) predictors with a set of technical indicators, and then compresses this information using principal components. Applying standard out-of-sample evaluation criteria over an out-of-sample period from April 1974 to June 2011 we show that improvements in the out-of-sample R^2 can be as high as 1.7522% from our best performing forecast combination approach (COMB).

It should be highlighted here that — despite the overall success of our proposed *factor* forecast combination approach — the forecast performance noticeably weakened from the beginning of 1992 onwards, as is evident from the time series evolution of the cumSFE plots

---

[32] Table 3 reads as in Neely et al. (2014) to facilitate a direct comparison to their results. We add one extra column to show the full CER values as well.

[33] For ease of comparison, we show (average) monthly turnover in column 7, rather than relative ones as reported in Neely et al. (2014) to keep the interpretation of entries under the turnover heading constant and to avoid confusion about whether these are absolute or relative measures.

[34] See Neely et al. (2014) for further details on how these are computed.

(relative to the historic average). The information content in these predictors thus seems to have become obsolete over time, suggesting that financial market participants have increasingly shifted their conditioning information to other predictors. Aggregate short interest seems to be one of these new predictors. In fact, Rapach et al. (2015) have recently concluded that "*short interest is the strongest known predictor of aggregate stock returns*" thus far. A challenge for future research is to find complementarities between the short interest predictors and the more traditional GW based fundamental predictors, as it seems unlikely that one would consider the traditional predictor base to be entirely uninformative.

Appendix to. 'Macroeconomic Factors and Equity Premium Predictability'

This appendix provides additional background information on the sources of the data and the construction of the technical indicators. We show some summary statistics in tables as well as various time series plots. We also provide some additional in-sample test results in this section.

A.1 Description of the data

A.1.1 Goyal and Welch Predictors

We follow the mainstream literature on equity premium predictability and construct the equity premium as well as the economic and financial predictor variables following the approach of Goyal and Welch (2008), Neely et al. (2014) and many others.[35] More specifically, the transformations that we employ to construct the equity premium (EQ for short) and the Goyal and Welch (2008) predictor variables (simply GW predictors henceforth) are described in detail in Table A.1 below.

Table A.1
Goyal and Welch (2008) predictors as used in Neely et al. (2014).

EQ	Equity premium: log returns of S&P 500 including dividends minus log return of a risk-free bill.
DP	Log dividend-price ratio: log of 12-month moving sum of dividends paid on the S&P 500 minus the log of the price of the S&P 500.
DY	Log dividend yield: log of 12-month moving sum of dividends paid on the S&P 500 Index minus the log of lagged S&P 500 prices.
EP	Log earnings-price ratio: log of 12-month moving sum of earnings on the S&P 500 Index minus the log of S&P 500 prices.
DE	Log dividend-payout ratio: log of 12-month moving sum of dividends paid on the S&P 500 minus log of 12-month moving sum of earnings on the S&P 500.
RVOL	Equity premium volatility: 12-month moving standard deviation estimator (based on Mele (2007)).
BM	Book-to-market ratio: book-to-market value ratio for the Dow Jones Industrial Average.
NTIS	Net equity expansion: ratio of 12-month moving sum of net equity issues by New York Stock Exchange (NYSE) listed stocks to total end-of-year market capitalization of NYSE stocks.
TBL	Treasury bill rate: interest rate on three-month Treasury bill (secondary market).
LTY	Long-term yield: long-term government bond yield.
LTR	Long-term return: return on long-term government bonds.
TMS	Term spread: long-term yield minus treasury bill rate.
DFY	Default yield spread: Moody's BAA corporate bond yield minus Moody's AAA corporate bond yield.
DFR	Default return spread: long-term corporate bond return minus the long-term government bond return.
INFL	Inflation: calculated from the CPI for all urban consumers (same as in Neely et al. (2014), using two periods lagged inflation to account for the delay in CPI releases).

To provide a baseline comparison of the properties of the data covering our sample period from April 1959 to June 2011 to that used in Neely et al. (2014) (ie., from December 1950 to December 2011), we show summary statistics in Table A.2 and time series plots in Fig. A.1 of all the GW predictor variables that are constructed in Table A.1.

---

[35] We obtain the predictor data from the program replication files of Neely et al. (2014) available at: http://sites.slu.edu/rapachde/DataProgramFilesMS2014.zip (last retrieved, September 2016). We would like to thank Dave Rapach for making these replication files publicly available, so that an exact replication of their results is possible.
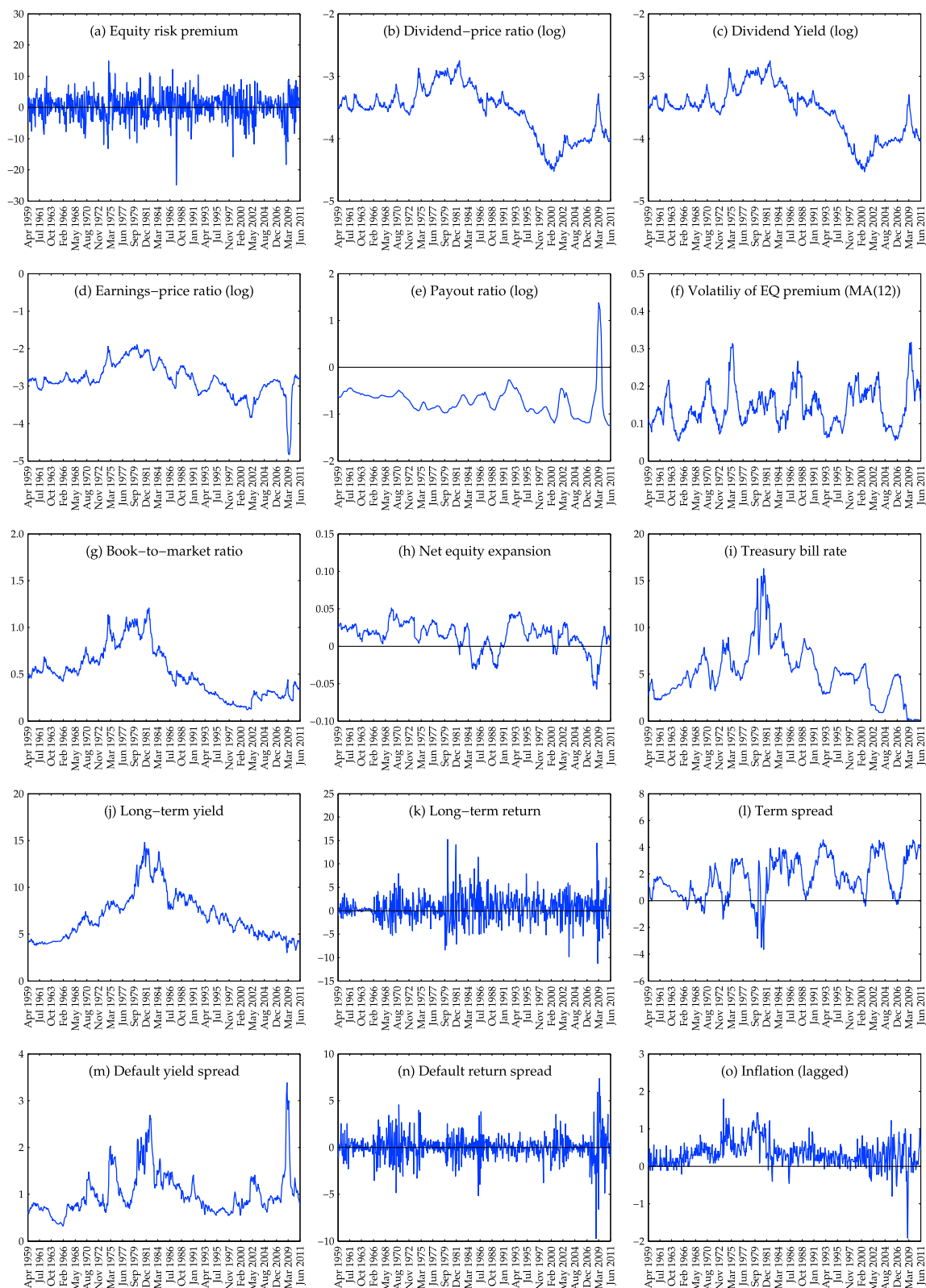
Fig. A.1. Time series plots of the 14 economic and financial predictor variables of Goyal and Welch (2008) and the equity risk premium.

Table A.2

Summary statistics of Goyal and Welch (2008) predictors from 1959:4 to 2011:6.

(1) Variable	(2) Description	(3) Mean	(4) Median	(5) Std.Dev	(6) Skew	(7) Kurt	(8) Min	(9) Max	(10) ACF(1− 3)			(11) PACF(1− 3)		
EQ	Equity risk premium	0.3461	0.7621	4.3376	−0.68	5.49	−24.84	14.87	0.06	−0.03	0.04	0.06	−0.04	0.04
DP	Dividend-price ratio (log)	−3.5545	−3.4921	0.3983	−0.45	2.49	−4.52	−2.75	0.99	0.98	0.98	0.99	−0.08	0.02
DY	Dividend Yield (log)	−3.5495	−3.4894	0.3984	−0.46	2.52	−4.53	−2.75	0.99	0.98	0.97	0.99	−0.08	0.03
EP	Earnings-price ratio (log)	−2.8295	−2.8591	0.4426	−0.74	5.85	−4.84	−1.90	0.99	0.97	0.95	0.99	−0.39	−0.15
DE	Payout ratio (log)	−0.7250	−0.7166	0.3131	2.78	18.26	−1.24	1.38	0.98	0.94	0.89	0.98	−0.76	−0.16
RVOL	Volatility of EQ (MA(12))	0.1443	0.1358	0.0524	0.71	3.35	0.05	0.32	0.96	0.93	0.89	0.96	−0.05	−0.07
BM	Book-to-market ratio	0.5198	0.4937	0.2623	0.60	2.54	0.12	1.21	0.99	0.99	0.98	0.99	−0.02	0.01
NTIS	Net equity expansion	0.0134	0.0157	0.0189	−1.00	4.27	−0.06	0.05	0.98	0.95	0.92	0.98	−0.07	−0.11
TBL	Treasury bill rate	5.1543	4.9600	2.9146	0.86	4.49	0.03	16.30	0.99	0.96	0.94	0.99	−0.34	0.20
LTY	Long-term yield	6.8766	6.4000	2.4840	0.93	3.36	3.03	14.82	0.99	0.98	0.97	0.99	−0.07	0.09
LTR	Long-term return	0.6083	0.4000	2.8836	0.45	6.15	−11.24	15.23	0.03	−0.08	−0.03	0.03	−0.08	−0.03
TMS	Term spread	1.7223	1.6500	1.4990	−0.17	2.68	−3.65	4.55	0.95	0.90	0.85	0.95	−0.12	0.07
DFY	Default yield spread	1.0141	0.8900	0.4628	1.74	7.02	0.32	3.38	0.97	0.92	0.87	0.97	−0.31	0.14
DFR	Default return spread	0.0099	0.0700	1.3933	−0.29	10.62	−9.75	7.37	−0.07	−0.06	−0.03	−0.07	−0.06	−0.04
INFL	Inflation (lagged)	0.3280	0.2954	0.3459	−0.03	6.96	−1.92	1.79	0.61	0.42	0.33	0.61	0.07	0.08

Notes: This table reports standard summary statistics of the equity risk premium and the 14 Goyal and Welch (2008) predictor variables as used in Neely et al. (2014). In the first two columns of the table, the short variable name and the variable description are shown. In columns three to nine, we show common sample statistics such as the mean, median (Med), standard deviation (Std.dev), skewness (Skew), kurtosis (Kurt), as well as the minimum (Min) and maximum (Max) of the series. In the last six columns (grouped into blocks of three), the first to third order autocorrelation function (ACF) and partial ACF (PACF) are given. The sample period is from April 1959 to June 2011.

Briefly, the summary statistics in Table A.2 show that the overall properties of the variables over our shorter sample period are very similar to those in Neely et al. (2014). Some of the interest rates and interest rate spreads, namely TBL, LTY, LTR, TMS, and DFY, exhibit marginally higher means compared to the Neely et al. (2014) time period. Nevertheless, in particular the autocorrelation structure is almost identical to that reported in Neely et al. (2014), showing, as expected, rather strong autocorrelations of the valuation ratios, the interest rates and interest rate spreads. The most notable difference is the apparently lower mean value in the (log) equity risk premium EQ, that is, we obtain a value of 0.3461 in Table A.2 while the value reported in Neely et al. (2014, see Table 1) is 0.47.

A.1.2 Neely et al. Technical Indicators

As in Neely et al. (2014), we construct 14 technical indicators based on three popular trend-following trading strategies. These three trading strategies are: *i*) Moving average rules, *ii*) Momentum rules and, *iii*) On-balance Volume.

The moving average (MA) rule, which gives a buy signal, MA(s, l) = 1, or a sell signal, MA(s, l) = 0, respectively, is computed by comparing two moving averages of different lengths as:

$$\text{MA}_t(s, l) = \begin{cases} 1, & \text{if } \text{MA}_t(s) \geq \text{MA}_t(l), \\ 0, & \text{otherwise,} \end{cases} \tag{A.1}$$

where

$$\text{MA}_t(n) = n^{-1} \sum_{i=0}^{n-1} P_{t-i}, \tag{A.2}$$

$\forall n = s$, l and P_t being the level of the S&P 500 index. The terms s and l determine the *length* of the short (s) and long (l) MA filters that are applied to the asset price. We use again the same specification as in Neely et al. (2014) and consider short MA filters of 1–3 months horizons (ie., $s = \{1, 2, 3\}$) and long MA filters of 9 and 12 months horizons ($l = \{9, 12\}$). As the short MA is more sensitive to changes in the price level P_t, it will thus be more responsive to signal changes in price trends.

The momentum (MOM) rule is defined as:

$$\text{MOM}_t(m) = \begin{cases} 1, & \text{if } P_t \geq P_{t-m}, \\ 0, & \text{otherwise.} \end{cases} \tag{A.3}$$

It generates a buy signal whenever the current period's price level is higher than that from m periods ago. The rational behind this strategy is that when $P_t \geq P_{t-m}$, we have positive momentum and thus expect higher excess returns in the future. We use $m = \{9, 12\}$, which translates into 9 and 12 months momentum signals.

The on-balance volume (VOL) rule is computed as:

$$\text{VOL}_t(s, l) = \begin{cases} 1, & \text{if } \text{MA}_t^{(\text{OBV})}(s) \geq \text{MA}_t^{(\text{OBV})}(l), \\ 0, & \text{otherwise,} \end{cases} \tag{A.4}$$

where

$$\text{MA}_t^{(\text{OBV})}(n) = n^{-1} \sum_{i=0}^{n-1} \text{OBV}_{t-i}, \tag{A.5}$$

with $n = s$, l and $s = \{1, 2, 3\}$ and $l = \{9, 12\}$ being the length of the moving average period. The on-balance volume variable OBV_t in (A.5) is defined as:

$$\text{OBV}_t = \sum_{k=1}^{t} \text{VOL}_k D_k, \tag{A.6}$$

where

$$D_k = \begin{cases} 1, & \text{if } P_k \geq P_{k-1}, \\ -1, & \text{otherwise}, \end{cases} \tag{A.7}$$

and VOL_k is trading volume in period k.[36] The on-balance volume indicator is meant to capture the notion that relatively high recent trading volume in combination with recent price increases are taken as signs of a strong positive market trend, and thus lead to a buy signal in the technical indicator. The construction and abbreviations of all technical indicators is summarised in Table A.3 below. All technical indicators are collected in the vector $\boldsymbol{x}^{\text{Tech}}$.

Table A.3
Neely et al. (2014) technical indicators.

MA(1,9)	Buy signal when MA of prices of length 1 lies above MA of length 9.
MA(1,12)	Buy signal when MA of prices of length 1 lies above MA of length 12.
MA(2,9)	Buy signal when MA of prices of length 2 lies above MA of length 9.
MA(2,12)	Buy signal when MA of prices of length 2 lies above MA of length 12.
MA(3,9)	Buy signal when MA of prices of length 3 lies above MA of length 9.
MA(3,12)	Buy signal when MA of prices of length 3 lies above MA of length 12.
MOM(9)	Buy signal when price at time t lies above price in $(t-9)$.
MOM(12)	Buy signal when price at time t lies above price in $(t-12)$.
VOL(1,9)	Buy signal when MA of on-balance volume of length 1 lies above MA of length 9.
VOL(1,12)	Buy signal when MA of on-balance volume of length 1 lies above MA of length 12.
VOL(2,9)	Buy signal when MA of on-balance volume of length 2 lies above MA of length 9.
VOL(2,12)	Buy signal when MA of on-balance volume of length 2 lies above MA of length 12.
VOL(3,9)	Buy signal when MA of on-balance volume of length 3 lies above MA of length 9.
VOL(3,12)	Buy signal when MA of on-balance volume of length 3 lies above MA of length 12.

In total, the relations in (A.1), (A.3), and (A.4) define 14 technical indicator variables. In Table A.4, we show summary statistics analogous to the GW predictors to provide a broad overview of the properties of the technical indicators that are constructed. The 14 technical indicators generate buy signals between 66% and 71% of the time over our sample period from April 1959 to June 2011. It can also be seen that, as is common for moving average based constructed technical variables, they are fairly persistent, with first order autocorrelations ranging between 0.61 and 0.82. These technical indicators are, nonetheless, not as persistent as some of the GW predictors summarised in Table A.2.

Table A.4
Summary statistics of Neely et al. (2014) technical indicators from 1959:4 to 2011:6.

(1) Variable	(2) Description	(3) Mean	(4) Median	(5) Std.Dev	(6) Skew	(7) Kurt	(8) Min	(9) Max	(11) ACF(1 − 3)			(12) PACF(1 − 3)		
MA(1,9)	MA(1) > MA(9)	0.6709	1	0.4703	−0.73	1.53	0	1	0.69	0.54	0.43	0.69	0.13	0.02
MA(1,12)	MA(1) > MA(12)	0.6979	1	0.4595	−0.86	1.74	0	1	0.78	0.65	0.55	0.78	0.11	0.02
MA(2,9)	MA(2) > MA(9)	0.6804	1	0.4703	−0.73	1.53	0	1	0.75	0.59	0.46	0.75	0.04	0.02
MA(2,12)	MA(2) > MA(12)	0.6963	1	0.4602	−0.85	1.73	0	1	0.83	0.69	0.58	0.83	0.02	0.00
MA(3,9)	MA(3) > MA(9)	0.6773	1	0.4679	−0.76	1.58	0	1	0.79	0.61	0.47	0.79	−0.02	−0.01
MA(3,12)	MA(3) > MA(12)	0.7011	1	0.4581	−0.88	1.77	0	1	0.82	0.69	0.58	0.83	0.02	0.02
MOM(9)	Momentum 9 Months	0.6900	1	0.4629	−0.82	1.67	0	1	0.76	0.70	0.59	0.77	0.28	−0.02
MOM(12)	Momentum 12 Months	0.7075	1	0.4553	−0.91	1.83	0	1	0.83	0.75	0.65	0.84	0.19	−0.04
VOL(1,9)	VOL MA(1) > MA(9)	0.6677	1	0.4714	−0.71	1.51	0	1	0.61	0.55	0.44	0.61	0.29	0.05
VOL(1,12)	VOL MA(1) > MA(12)	0.6979	1	0.4595	−0.86	1.74	0	1	0.72	0.66	0.52	0.72	0.32	−0.06
VOL(2,9)	VOL MA(2) > MA(9)	0.6645	1	0.4725	−0.70	1.49	0	1	0.77	0.58	0.47	0.77	−0.01	0.07
VOL(2,12)	VOL MA(2) > MA(12)	0.6979	1	0.4595	−0.86	1.74	0	1	0.84	0.68	0.58	0.84	−0.06	0.11
VOL(3,9)	VOL MA(3) > MA(9)	0.6836	1	0.4654	−0.79	1.62	0	1	0.77	0.61	0.47	0.78	0.02	−0.01
VOL(3,12)	VOL MA(3) > MA(12)	0.6932	1	0.4615	−0.84	1.70	0	1	0.84	0.72	0.60	0.85	0.02	−0.04

---

[36] We use the trading volume data that is provided by Neely et al. (2014) in their data files for reasons of data consistency. The source of the volume data is Google Finance.

Notes: This table reports standard summary statistics of the 14 technical indictors as used by Neely et al. (2014). In the first two columns of the table, the short variable name and the variable description are shown. In columns three to nine, we show common sample statistics such as the mean, median (Med), standard deviation (Std.dev), skewness (Skew), kurtosis (Kurt), as well as the minimum (Min) and maximum (Max) of the series. In the last six columns (grouped into blocks of three), the first to third order autocorrelation function (ACF) and partial ACF (PACF) are given. The sample period is from April 1959 to June 2011.

A.1.3 Stock and Watson macroeconomic dataset

Table A.5 below shows summary statistics, together with variable definitions and transformation codes of the Stock and Watson (2012) dataset.

Table A.5
Overview, description and summary statistics of Stock and Watson dataset.

(1) Entry	(2) Code	(3) Variable Descriptions	(4) Mean	(5) Median	(6) Std.Dev	(7) Skew	(8) Kurt	(9) Min	(10) Max
1	5	Industrial Production: Durable Materials	0.0031	0.0043	0.0177	−0.0239	22.1595	−0.1285	0.1684
2	5	Industrial Production: Non-Durable Materials	0.0019	0.0018	0.0108	−1.3585	11.7764	−0.0765	0.0407
3	5	Industrial Production: Durable Consumer Goods	0.0025	0.0029	0.0233	0.1069	10.3298	−0.1316	0.1281
4	5	Industrial Production: Automotive products	0.0025	0.0034	0.0525	0.2800	19.3041	−0.3792	0.3498
5	5	Industrial Production: Non-Durable Consumer Goods	0.0016	0.0017	0.0076	−0.0971	3.2198	−0.0241	0.0247
6	5	Industrial Production: Business Equipment	0.0038	0.0044	0.0123	−0.9219	7.9717	−0.0825	0.0521
7	5	Industrial Production: Consumer Energy Products	0.0024	0.0026	0.0213	−0.1637	4.0212	−0.0676	0.0740
8	5	All Employees: Durable Goods Manufacturing	−0.0003	0.0003	0.0068	−1.3636	12.5005	−0.0486	0.0295
9	5	All Employees: Construction	0.0010	0.0017	0.0096	0.0086	8.9546	−0.0580	0.0563
10	5	All Employees: Education & Health Services	0.0031	0.0030	0.0019	0.2312	4.7847	−0.0045	0.0110
11	5	All Employees: Financial Activities	0.0018	0.0020	0.0020	−0.6687	4.2172	−0.0067	0.0069
12	5	All Employees: Information Services	0.0007	0.0012	0.0038	−0.5378	4.4921	−0.0157	0.0156
13	5	All Employees: Professional & Business Services	0.0025	0.0030	0.0028	−1.1898	5.7811	−0.0108	0.0103
14	5	All Employees: Leisure & Hospitality	0.0022	0.0025	0.0032	0.0364	7.9423	−0.0143	0.0210
15	5	All Employees: Other Services	0.0026	0.0027	0.0023	−0.2492	3.5170	−0.0067	0.0105
16	5	All Employees: Natural Resources & Mining	0.0001	0.0000	0.0087	−0.1992	6.5047	−0.0343	0.0417
17	5	All Employees: Trade, Transportation & Utilities	0.0013	0.0015	0.0025	−0.3172	5.3284	−0.0109	0.0144
18	5	All Employees: Retail Trade	0.0016	0.0014	0.0030	0.8629	13.1387	−0.0124	0.0236
19	5	All Employees: Wholesale Trade	0.0016	0.0016	0.0031	−0.3025	6.8719	−0.0142	0.0172
20	5	All Employees: Federal	0.0012	0.0015	0.0026	−0.3304	4.5083	−0.0086	0.0147
21	5	All Employees: State government	0.0000	0.0000	0.0045	−2.5183	21.7849	−0.0382	0.0118
22	5	All Employees: Local government	0.0020	0.0016	0.0031	0.2817	4.5843	−0.0112	0.0148
23	2	Unemployment Rate - 16–19 yrs	0.0019	0.0018	0.0027	−0.1346	5.5928	−0.0107	0.0129
24	2	Unemployment Rate - 20 yrs & over, Men	0.0169	0.0000	0.8434	0.0714	3.4772	−2.7000	3.1000
25	2	Unemployment Rate - 20 yrs & over, Women	0.0059	0.0000	0.2037	0.1905	4.6030	−1.0000	0.8000
26	5	Number Unemployed for Less than 5 Weeks	0.0037	0.0000	0.2156	0.3370	4.5945	−0.8000	1.0000
27	5	Number Unemployed for 5–14 Weeks	0.0011	0.0012	0.0540	0.2231	4.5371	−0.1828	0.2741
28	5	Civilians Unemployed for 15–26 Weeks	0.0015	−0.0005	0.0558	0.2470	3.9288	−0.2204	0.2289
29	5	Number Unemployed for 27 Weeks & over	0.0018	0.0010	0.0780	−0.0253	3.9695	−0.3631	0.2881
30	5	Employment Level - Part-Time for Economic Reasons, All Industries	0.0033	0.0010	0.0728	0.2432	3.8012	−0.2144	0.2810
31	2	Average Weekly Hours: Manufacturing	0.0023	−0.0008	0.0428	0.4694	4.1048	−0.1258	0.1627
32	2	Average Weekly Hours: Overtime: Manufacturing	0.0019	0.0000	0.2603	0.5520	21.1492	−1.9000	2.3000
33	5	Housing Starts in Midwest Census Region	0.0024	0.0000	0.1412	0.1377	9.5492	−0.9000	0.8000
34	5	Housing Starts in Northeast Census Region	−0.0020	0.0029	0.1657	−0.0298	5.0535	−0.6202	0.6562
35	5	Housing Starts in South Census Region	−0.0026	0.0063	0.1911	−0.2694	4.1809	−0.7395	0.6355
36	5	Housing Starts in West Census Region	−0.0009	0.0000	0.0997	0.0593	4.0751	−0.3470	0.4305
37	5	Sales of retail stores (mil. Chain 2000 $)	−0.0015	−0.0093	0.1225	0.0410	3.5487	−0.4415	0.3826
38	5	Manufacturers' new orders durable goods industries (bil. chain 2000 $)	0.0026	0.0030	0.0126	−0.1079	8.3314	−0.0771	0.0750
39	5	Manufacturers' new orders, consumer goods & materials (mil. 1982 $)	0.0020	0.0038	0.0360	−0.1890	4.3856	−0.1516	0.1552
40	5	Manufacturers' unfilled orders durable goods indus. (bil. chain 2000 $)	0.0014	0.0015	0.0206	0.0706	4.0032	−0.0768	0.0787
41	5	Manufacturers' new orders, non-defence capital goods (mil. 1982 $)	0.0026	0.0021	0.0107	0.2454	3.9915	−0.0307	0.0541
42	5	Index of supplier deliveries – vendor performance (pct.)	0.0027	0.0051	0.0742	−0.0591	4.0372	−0.2584	0.2897
43	5	Manufacturing & trade inventories (bil. Chain 2005 $)	−0.0001	0.0039	0.0672	−0.8953	9.4678	−0.3822	0.3220
44	6	Producer Price Index: Finished Consumer Goods	0.0027	0.0031	0.0042	−0.5114	4.2703	−0.0136	0.0154
45	6	Producer Price Index: Finished Consumer Foods	0.0000	0.0000	0.0068	−0.2021	8.1826	−0.0368	0.0368
46	6	Producer Price Index: Industrial Commodities	0.0000	0.0000	0.0100	−0.9529	16.1065	−0.0733	0.0696
47	6	Producer Price Index: Intermediate Materials: Supplies & Components	0.0000	0.0000	0.0086	−0.2662	8.5073	−0.0470	0.0472
48	5	Producer Price Index: Crude Petroleum Deflated by PCEPILFE	0.0000	−0.0001	0.0137	−0.0533	13.5516	−0.1024	0.0998
49	2	Effective Federal Funds Rate	0.0000	0.0000	0.0092	−1.5458	15.4882	−0.0687	0.0383
50	2	Moody's Seasoned Baa Corporate Bond Yield - Government Security 10y Spread	0.0000	0.0000	0.0075	−1.1723	15.5047	−0.0443	0.0444
51	1	Treasury Bill 6m - Treasury Bill 3m	0.0029	0.0022	0.0995	0.0331	9.3489	−0.4929	0.5054
52	1	Government Security 1y - Treasury Bill 3m	0.0019	−0.0016	0.0751	−0.3825	9.5518	−0.3676	0.4810
53	1	CP3FM-TB3MS	−0.0038	0.0100	0.5497	−2.1784	44.1578	−6.6300	3.0600
54	5	M2 Money Stock Deflated by PCEPILFE	0.1515	0.1300	0.1977	−0.0723	11.5950	−1.0100	1.4500
55	5	Commercial and Industrial Loans at all Commercial Banks Deflated by PCEPILFE	0.5650	0.4900	0.4194	1.0691	6.1591	−0.9400	2.9300
56	5	Total Non-revolving Credit Outstanding Deflated by PCEPILFE	0.5381	0.3740	0.4947	2.3640	11.8036	−0.0360	4.2540
57	5	Real Estate Loans at All Commercial Banks Deflated by PCEPILFE	0.0028	0.0032	0.0085	−0.0740	4.9575	−0.0322	0.0450
58	5	Total Revolving Credit Outstanding Deflated by PCEPILFE	0.0028	0.0033	0.0075	−0.0552	5.5316	−0.0357	0.0450

(*continued on next page*)

Table A.5 (*continued*)

(1) Entry	(2) Code	(3) Variable Descriptions	(4) Mean	(5) Median	(6) Std.Dev	(7) Skew	(8) Kurt	(9) Min	(10) Max
59	5	Total Consumer Credit Outstanding Deflated by PCEPILFE	0.0027	0.0030	0.0065	−2.9209	36.8709	−0.0763	0.0234
60	5	Common Stock Prices: Dow Jones Industrial Average	0.0050	0.0052	0.0060	0.5901	7.3930	−0.0167	0.0442
61	5	FRB Nominal Major Currencies Dollar Index (Linked to EXRUS in 1973:1)	−0.0008	0.0000	0.0150	−0.1084	4.4392	−0.0539	0.0647
62	5	Foreign Exchange Rate: Switzerland (Swiss Franc Per U.S.$)	−0.0026	−0.0003	0.0254	−0.2748	4.3443	−0.0875	0.0928
63	5	Foreign Exchange Rate: Japan (Yen Per U.S.$)	−0.0024	0.0000	0.0237	−0.6918	5.1664	−0.1052	0.0807
64	5	Foreign Exchange Rate: United Kingdom (Cents Per Pound)	−0.0009	−0.0001	0.0218	−0.4324	6.2669	−0.1108	0.0952
65	5	Foreign Exchange Rate: Canada (Canadian $ Per U.S.$)	0.0000	−0.0002	0.0125	0.7375	15.1655	−0.0601	0.1129
66	5	Consumer expectations NSA (University of Michigan)	−0.0006	−0.0016	0.0556	−0.0092	6.2868	−0.2415	0.3096

This table shows the variables from the Stock and Watson database that we use, together with the transformations that we apply. The transformation codes in column (2) are as follows: 1 = no transformation, 2 = first difference, 4 = logarithm, 5 = first difference of logarithms, 6 = second difference of logarithms. PCEPILFE is the Personal Consumption Expenditures Excluding Food and Energy (Chain-Type Price Index) used to deflate the series.

A.2 April 1959 to June 2011 comparison to Neely et al. (2014) results

To provide a benchmark comparison of the bivariate regressions of the equity premium on each of the Goyal and Welch (2008) predictors, as well as the technical indicators of Neely et al. (2014) we show replicate results of Neely et al. (2014) over our shorter time period. These are shown in the upper part of Table A.6. The results in Table A.6 are arranged symmetrically, showing the predictor variable, the estimated slope coefficient, the corresponding t−statistic, as well as goodness of fit measures R^2 (in percent), which are further measured separately over expansion (Exp) and recession (Rec) periods as defined by the NBER in columns 1 to 6, respectively.[37] The top half of Table A.6 shows the bivariate regression results, while the factor based regressions are shown down the bottom. Since standard asymptotic t−statistics for the estimates in Table A.6 are likely to be inflated, with highly persistent regressors further distorting test sizes (see Stambaugh (1999)), we perform a wild bootstrap to compute critical values of the t−statistics. By using a wild bootstrap, we account for the persistence in the regressors, the correlations between the equity premium and the predictor innovations, and general forms of heteroskedasticity.[38]

Comparing the bivariate regression results in Table A.6 to those provided in Table 2 in Neely et al. (2014), we see that only three (as opposed to 6) of the Goyal and Welch (2008) financial and economic predictors are significantly different from zero at the 5% level (based on one-sided upper-tail bootstrapped p−values), which also exceed the R^2 'threshold' of 0.5%. These are the RVOL, LTR, and TMS predictors. Interestingly, other popular predictors such as the dividend yield (DY), the treasury bill rate (TBL) and the long term yield (LTY) are not statistically significant, with the parameter estimates on TBL and TLY being also particularly small in magnitude. All three significant predictors have substantially higher R^2 values during recessions, suggesting that the biggest gains from using these predictors are realised during economic downturns. This contrast in performance is particularly stark for the long-term government bonds rate (LTR) predictor variable.

Looking over our results for the technical indicators in the right upper block in Table A.6, we can see that these are more consistent with the results obtained in Neely et al. (2014). All the slope coefficients from the bivariate technical indicator regressions are significant at the 10% level, and are further sizable in magnitude. Even the VOL(3,9) indicator, which is not significant in Neely et al. (2014), is significant at the 10% level over our sample period. The technical indicators generally have also similarly sized R^2 values, again showing a rather strong predictive performance during recession periods.[39]

The results from the factor predictive regressions in the bottom panel of Table A.6 show that running in-sample OLS using the extracted factors as predictors considerably increases the fit of the model. We follow the mainstream literature on factor models and treat the '*estimated*' factors $\widehat{\boldsymbol{F}}_t$ in all in-sample tests as if they were observed, ie., pre-estimation of the factors does not affect the consistency or precision of the estimates of the second stage factor predictive regression parameters (see Bai and Ng (2006) for the technical details and also Ludvigson and Ng (2007, 2009) who follow the same strategy empirically). Alternatively, one could try to account for the first step estimation error by working out the standard errors analytically as shown by Pagan (1986) and implemented in Brand, Buncic, and Turunen (2010) in a constructed regressor context. Bootstrap methods provide also an alternative, as illustrated in Ludvigson and Ng (2007) and Hatemi-J, Roca, and Buncic (2006).

---

[37] More concretely, the R^2 across expansion and recessions states are computed as.

$$R_S^2 = 1 - \frac{\sum_{t=1}^{T} 1_t^S \widehat{\varepsilon}_{i,t}^2}{\sum_{t=1}^{T} 1_t^S (\text{EQ}_t - \overline{\text{EQ}})^2}, \forall\, S = \{\text{Rec}, \text{Exp}\}. \tag{A.8}$$

The variable 1_t^S is an economic state indicator, defined so that 1_t^{Rec} (1_t^{Exp}) is equal to unity when month t is a recession (expansion) state and zero otherwise. The fitted residuals $\widehat{\varepsilon}_{i,t}$ are from the full sample estimates of the bivariate equity premium regressions in (1). The term $\overline{\text{EQ}}$ in (A.8) is the sample mean of EQ$_t$.

[38] We use standard '*' notation to denote significance at the 1, 5, and 10 percent levels, with significance here being defined as the probability of the wild bootstrap draws being larger than the estimated slope coefficient.

[39] Here, the coefficients are interpreted as follows: given a buy signal, say from the MA(1,12) indicator, next month's equity premium is expected to increase by 84 basis points.

For consistency reasons, we again follow Neely et al. (2014) and select the appropriate number of factors using their adjusted R^2 approach, setting a maximum value for the admissible number of factors to 3, 1, and 4 for the Goyal and Welch (2008) (ECON) predictors, the technical indicators (TECH), and the ALL factor group, respectively. The number of factors selected from the GW predictor set (denoted by ECON in Table A.6) and the technical indicators (denoted by TECH) are the same as over their period. The fit of the ECON factors seems to be somewhat weaker, with an R^2 value of only 0.76%, as opposed to 1.18% reported in Neely et al. (2014), however, the fit of the TECH factor is overall in line with the results in Neely et al. (2014). The ALL factors regression results down the bottom of Table A.6 are also weaker than in Neely et al. (2014), with an R^2 of 1.87% compared to 2.02%. Further, only 3 ALL factors are selected instead of 4, where the magnitude as well as the significance of these 3 correspond to the F_1^{ALL}, F_2^{ALL} and F_4^{ALL} ALL factors in Neely et al. (2014).

Table A.6
In-sample predictive regression results of Neely et al. (2014), April 1959 to June 2011.

(1) Variable	(2) Slope	(3) t−stat	(4) R^2	(5) R^2_{Exp}	(6) R^2_{Rec}	(1) Variable	(2) Slope	(3) t−stat	(4) R^2	(5) R^2_{Exp}	(6) R^2_{Rec}
Bivariate regressions											
Goyal and Welch (2008) predictors						Neely et al. (2014) technical indicators					
DP	0.50	1.07	0.21	0.10	0.45	MA(1,9)	0.56	1.36*	0.37	−0.44	2.11
DY	0.57	1.23	0.28	0.04	0.79	MA(1,12)	0.84	1.96**	0.80	−0.21	2.98
EP	0.22	0.46	0.05	0.07	0.01	MA(2,9)	0.66	1.61*	0.51	−0.39	2.42
DE	0.36	0.54	0.07	−0.01	0.23	MA(2,12)	0.96	2.23***	1.03	−0.22	3.71
RVOL	7.79	2.53***	0.88	0.55	1.61	MA(3,9)	0.73	1.77**	0.61	−0.09	2.13
BM	0.20	0.26	0.01	−0.02	0.10	MA(3,12)	0.54	1.27*	0.33	−0.14	1.34
NTIS	3.15	0.29	0.02	0.21	−0.40	MOM(9)	0.65	1.54*	0.49	−0.13	1.81
TBL	0.07	1.12	0.23	0.17	0.36	MOM(12)	0.67	1.56*	0.49	−0.20	1.99
LTY	0.02	0.26	0.01	0.02	0.00	VOL(1,9)	0.73	1.79**	0.62	−0.58	3.19
LTR	0.14	2.12**	0.89	−0.44	3.73	VOL(1,12)	0.97	2.27***	1.06	−0.28	3.94
TMS	0.22	1.86**	0.56	0.18	1.39	VOL(2,9)	0.86	2.12**	0.87	−0.24	3.25
DFY	0.36	0.78	0.15	0.16	0.13	VOL(2,12)	0.95	2.21**	1.01	0.10	2.97
DFR	0.17	0.88	0.31	0.07	0.84	VOL(3,9)	0.55	1.32*	0.35	−0.24	1.62
INFL	0.13	0.20	0.01	−0.10	0.25	VOL(3,12)	0.90	2.13**	0.91	0.08	2.68
Neely et al. (2014) factor regressions											
ECON Factors						TECH Factor					
$\widehat{F}_1^{ECON}$	0.02	0.30	0.76	0.76	0.77	$\widehat{F}_1^{TECH}$	0.13	2.37***	0.89	−0.27	3.38
$\widehat{F}_2^{ECON}$	−0.24	−2.12**									
$\widehat{F}_3^{ECON}$	−0.07	−0.46									
ALL Factors											
$\widehat{F}_1^{ALL}$	0.12	1.85**	1.87	0.08	5.72						
$\widehat{F}_2^{ALL}$	0.07	0.78									
$\widehat{F}_3^{ALL}$	0.28	2.27***									

Notes: This table reports in-sample predictive regression results corresponding to those in Table 2 of Neely et al. (2014), nevertheless over the sample period from April 1959 to June 2011. Predictor variables are as defined in Neely et al. (2014). The factors $\widehat{F}$ are selected as in Neely et al. (2014). t−statistics and corresponding levels of significance are based on a wild bootstrap, which accounts for the persistence in the regressors, the correlations between the equity premium and the predictor innovations, as well as general forms of heteroskedasticity. Significance at the 1, 5, and 10 percent levels are marked, respectively, by ***, ** and *.

References

Bai, Jushan (2003). Inferential theory for factor models of large dimensions. *Econometrica, 71*(1), 135–171.

Bai, Jushan, & Ng, Serena (2006). Confidence intervals for diffusion index forecasts and inference for factor-augmented regressions. *Econometrica, 74*(4), 1133–1150.

Bai, Jushan, & Ng, Serena (2002). Determining the number of factors in approximate factor models. *Econometrica, 70*(1), 191–221.

Belloni, Alexandre, & Chernozhukov, Victor (2013). Least squares after model selection in high-dimensional sparse model. *Bernoulli, 19*, 521–547.

Brand, Claus, Buncic, Daniel, & Turunen, Jarkko (2010). The impact of ECB monetary policy decisions and communication on the yield curve. *Journal of the European Economic Association, 8*(6), 1266–1298.

Buncic, Daniel & Melecky, Martin (2014). Equilibrium credit: The reference point for macroprudential supervisors. *Journal of Banking and Finance, 41*, 135–154 (April).

Buncic, Daniel, & Moretto, Carlo (2015). Forecasting copper prices with dynamic averaging and selection models. *North American Journal of Economics and Finance, 33*(1), 1–38.

Buncic, Daniel, & Müller, Oliver (2017). Measuring the output gap in Switzerland with linear opinion pools. *Economic Modelling, 74*, 153–177 (August).

Buncic, Daniel, & Piras, Gion Donat (2016). Heterogenous agents, the financial crisis and exchange rate predictability. *Journal of International Money and Finance, 60*, 313–359 (February).

Campbell, John Y., & Thompson, Samuel B. (2008). Predicting excess stock returns out of Sample: Can anything beat the historical average? *Review of Financial Studies, 21*(4), 1509–1531.

Clark, Todd E., & West, Kenneth D. (2007). Approximately normal tests for equal predictive accuracy in nested models. *Journal of Econometrics, 138*(1), 291–311.

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

Diebold, Francis X., & Mariano, Roberto S. (1995). Comparing predictive accuracy. *Journal of Business and Economic Statistics, 13*(1), 253–263.

Elliott, Graham, & Timmermann, Allan (2016). *Economic forecasting*. Princeton, New Jersey, USA: Princeton University Press.

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

Harvey, David I., Leybourne, Stephen J., & Newbold, Paul (1998). Tests for forecast encompassing. *Journal of Business and Economic Statistics, 16*(2), 254–259.

Hastie, Trevor, Tibshirani, Robert, & Friedman, Jerome (2009). *The elements of statistical learning: Data mining, inference, and prediction* (2nd ed.). Springer New York.

Hatemi-J, Abdulnasser, Roca, Eduardo, & Buncic, Daniel (2006). Bootstrap causality tests of the relationship between the equity markets of the US and other developed countries: Pre- and Post-September 11. *Journal of Applied Business Research, 22*(3), 65–74.

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

Ludvigson, Sydney C., & Ng, Serena (2009). Macro factors in bond risk premia. *The Review of Financial Studies, 22*(12), 5027–5067.

McCracken, Michael W. (2007). Asymptotics for out of sample tests of Granger causality. *Journal of Econometrics, 140*(2), 719–752.

Mele, A. (2007). Asymmetric stock market volatility and the cyclical behavior of expected returns. *Journal of Financial Economics, 86*, 446–478.

Neely, Christopher J., Rapach, David E., Tu, Jun, & Zhou, Guofu (2014). Forecasting the equity risk premium: The role of technical indicators. *Management Science, 60*(7), 1772–1791.

Ng, Serena (2013). Variable selection in predictive regressions. In Graham Elliott, & Allan Timmermann (Eds.), *Handbook of economic forecasting, 2A* pp. 753–788. Elsevier.

Pagan, Adrian R. (1986). Two stage and related estimators and their applications. *Review of Economic Studies, 53*(4), 517–538.

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

Rapach, David E., Strauss, Jack K., & Zhou, Guofu (2013). International stock return predictability: What is the role of the United States? *The Journal of Finance, 68*(4), 1633–1662.

Rapach, David E., Ringgenberg, Matthew C., & Zhou, Guofu (2015). *Short interest and aggregate stock returns*. Working Paper. Available from: http://sites.slu.edu/rapachde/Returns_short_interest_09-10-2015.pdf?attredirects=0&d=1.

Rapach, David E., & Zhou, Guofu (2013). Forecasting stock returns. In Graham Elliott, & Allan Timmermann (Eds.), *Handbook of economic forecasting, 2A* pp. 328–383. Elsevier.

Stambaugh, Robert F. (1999). Predictive regressions. *Journal of Financial Economics, 54*, 375–421.

Stock, James H., & Watson, Mark W. (2002a). Forecasting using principal components from a large number of predictors. *Journal of the American Statistical Association, 97*, 1167–1179.

Stock, James H., & Watson, Mark W. (2002b). Macroeconomic forecasting using diffusion indexes. *Journal of Business and Economic Statistics, 20*(2), 147–162.

Stock, James H., & Watson, Mark W. (2004). Combination forecasts of output growth in a seven-country data set. *Journal of Forecasting, 23*, 405–430.

Stock, James H., & Watson, Mark W. (2006). Forecasting with many predictors. In Graham Elliott, Clive W. J. Granger, & Allan Timmermann (Eds.), *Handbook of economic forecasting, 1* pp. 515–554. Elsevier. chap. 10.

Stock, James H., & Watson, Mark W. (2011). Dynamic factor models. In Michael P. Clements, & David F. Hendry (Eds.), *Oxford handbook of forecasting*. Oxford University Press.

Stock, James H., & Watson, Mark W. (2012). *Disentangling the channels of the 2007–2009 recession*. NBER Working Paper No. 18094. National Bureau of Economic Research.

West, M., & Harrison, J. (1997). *Bayesian forecasting and dynamic models*. Springer Verlag.

Zhou, Guofu (2010). How much stock return predictability can we expect from an asset pricing model? *Economics Letters, 108*(2), 184–186.

Zou, Hui (2006). The adaptive lasso and its oracle properties. *Journal of the American Statistical Association, 101*(476), 1418–1429.

Zou, Hui, Hastie, Trevor, & Tibshirani, Robert (2007). On the degrees of freedom of the lasso. *The Annals of Statistics, 35*(5), 2173–2192.