UvA ECONOMETRICS

Discussion Paper: 2011/10

Getting the Most Out of Macroeconomic Information for Predicting Stock Returns and Volatility

Cem Çakmakli and Dick van Dijk

UvA UNIVERSITEIT VAN AMSTERDAM

Getting the Most Out of Macroeconomic Information for Predicting Stock Returns and Volatility [*]

Cem Çakmaklı [†,1] and Dick van Dijk [‡,2]

[1] *Department of Quantitative Economics, University of Amsterdam*
[2] *Econometric Institute, Erasmus University Rotterdam*

December 2011

Abstract

This paper documents that factors extracted from a large set of macroeconomic variables bear useful information for predicting monthly US excess stock returns and volatility over the period 1980-2005. Factor-augmented predictive regression models improve upon both benchmark models that only include valuation ratios and interest rate related variables, and possibly individual macro variables, as well as the historical average excess return. The improvements in out-of-sample forecast accuracy are both statistically and economically significant. The factor-augmented predictive regressions have superior market timing ability and volatility timing ability, while a mean-variance investor would be willing to pay an annual performance fee of several hundreds of basis points to switch from the predictions offered by the benchmark models to those of the factor-augmented models. An important reason for the superior performance of the factor-augmented predictive regressions is the stability of their forecast accuracy, whereas the benchmark models suffer from a forecast breakdown during the 1990s.

Keywords: return predictability, model uncertainty, dynamic factor models, variable selection
JEL Classification: C22, C53, G11, G12

1 Introduction

Stock return predictability remains an issue of intense debate. On the one hand, a substantial number of papers reports positive forecasting results, with the most successful predictor variables being valuation ratios such as the dividend yield, the price-earnings ratio, and the book-to-market ratio, or other financial variables such as the short-term interest rate, the yield spread, or the credit spread, see Keim and Stambaugh (1986); Campbell and Shiller (1988); Fama and French (1988); Ferson and Harvey (1991); Ang and Bekaert (2007); Campbell and Thompson (2008); Wachter and Warusawitharana (2009), among many others. On the other hand, studies such as Welch and Goyal (2008) contest this view, arguing that over long time spans, no single predictor variable can outperform the historical mean return in terms of forecast accuracy, both in-sample and out-of-sample.

Interestingly, although stock returns are assumed to be linked to business conditions, macroeconomic variables such as output growth and inflation do not seem to add any predictive power for stock returns beyond the above-mentioned valuation ratios and interest rate related variables. The few exceptions to this rule include the consumption-wealth ratio (Lettau and Ludvigson, 2001), the expenditure share on housing Piazzesi *et al.* (2007), expenditure on durables as a fraction of its stock (Gomes *et al.*, 2009), and survey-based measures of expected business conditions (Campbell and Diebold, 2009).

Several complicating issues plague research on the predictive content of macroeconomic information for stock returns. First and foremost, numerous macroeconomic variables are available, but typically only a small number of variables is considered as possible predictors in a return regression. Presumably, the information embedded in the selected variables is more limited than the information that is actually available to an investor, who might consider a much larger set of macro variables.

Obviously, there are sound statistical reasons for limiting the number of (macro)

variables in a predictive regression model. Expanding the set of predictors would exacerbate the issue of parameter estimation uncertainty, which translates into additional uncertainty in the resulting return forecasts. At the same time, with the number of predictors being restricted it becomes crucial to select the most informative variables for consideration. This essentially leads to the issue of model uncertainty, as it is not clear *a priori* which macroeconomic variables bear the most relevant information for stock returns.

Second, the relations between excess stock returns and individual predictors appear to be highly unstable, see Bossaerts and Hillion (1999); Pesaran and Timmermann (2002); Paye and Timmermann (2006); Rapach and Wohar (2006); Lettau and Van Nieuwerburgh (2008); Ravazzolo *et al.* (2008); Pettenuzzo and Timmermann (forthcoming), among others. Hence, if a given macroeconomic variable is found to be useful for forecasting stock returns over a certain period, this might very well break down at some point. Put differently, the predictive ability of individual variables strongly fluctuates over time. Note that this aggravates the model uncertainty problem in the sense that the set of variables to include in a predictive regression model is unstable.

In this paper, we apply dynamic factor models to jointly handle the issues of model uncertainty, parameter estimation uncertainty, and structural instability. Specifically, we use principal component analysis to construct a small number of factors from a large set of macro variables. These factors are then included in a predictive regression model for excess stock returns. The motivation for considering this approach is derived from macroeconomic forecasting, where various successful applications of factor models have appeared in recent years for prediction variables such as output growth and inflation, see Stock and Watson (2002b); Forni *et al.* (2003); Mönch (2008), among others. The same approach has been adopted by Ludvigson and Ng (2007) for examining the relationship between expected stock returns

and volatility and by Ludvigson and Ng (2009) for analyzing the effects of macroeconomic conditions on bond risk premia. Independent, concurrent research by Bai (2010) also demonstrates that macro factors can be exploited for improving excess stock return predictions[1]

While the issue of return predictability is still disputed, there appears to be more consensus that the level of volatility is predictable to some extent. Several approaches have been developed for this purpose, ranging from GARCH and stochastic volatility models to realized volatility measures based on high-frequency returns, see Andersen *et al.* (2006) for a recent review. In addition, several studies document a relation between stock market volatility and macroeconomic conditions, with volatility being substantially higher during business cycle recessions, see Schwert (1989) and Hamilton and Lin (1996), among others. In this paper we also consider the use of macro factors extracted for predicting the volatility of monthly stock returns.

When using the factor-based approach for forecasting excess stock returns and volatility, several choices need to be made. Among others, we have to decide upon which individual variables are used in the factor construction and how many factors are included in the predictive regression model. While we adopt a standard approach for the latter issue (following previous studies in the macro forecasting literature), we examine the former issue in detail. Specifically, we consider the added value of several approaches for pre-selection of the macro variables that enter the factor construction stage, relative to the standard approach of simply including all available variables. This relates to the issue that in the standard approach the factors are taken to be the first few principal components of the macro variables, which are constructed without taking into account the purpose for which the factors are going to be used.

---

[1]We only became aware of this project while completing the current work. Several features distinguish our paper from Bai (2010). First, Bai (2010) only considers predictive regressions for excess return, while we also use macro factors for volatility. Second, we examine the possibility of pre-selection of macro variables that are used in the factor construction, while Bai (2010) does not. (She instead examines the possibility to select the most informative factors after their construction.)

In case the factors are used in predictive regressions as we do in the present paper, it need not necessarily be the case that these principal components contain the most relevant information for the variable that we aim to forecast, as shown by Boivin and Ng (2006) and Bai and Ng (2008), among others. The variable pre-selection procedures that we consider here explicitly take into account the predictive ability of the individual macro variables for the variables of interest, i.e. the excess stock return and volatility.

We conduct an empirical analysis of the usefulness of the factor-augmented regression models for out-of-sample prediction of monthly US excess stock returns and volatility over the period January 1980 - December 2005. In addition to the most commonly applied financial variables such as the dividend yield and the short-term interest rate, we construct factors based on a comprehensive set of 116 individual macroeconomic variables. The added value of the macro factors in the predictive regressions is assessed both in statistical as well as in economic terms. On the one hand, we consider the directional accuracy of the forecasts. For excess returns, we examine the ability of the factor-augmented predictive regressions to correctly predict its sign (market timing). For volatility, we consider the sign relative to the historical median level (volatility timing). Hit ratios, defined as the proportion of forecasts for which the actual and predicted signs match, are used to evaluate the directional accuracy of the forecasts. On the other hand, we assess the economic value of the return and volatility forecasts by using them in active mean-variance investment strategies. In addition to standard performance measures such as the Sharpe Ratio, we use a utility based metric to evaluate how much an investor would be willing to pay to use the predictions from the factor-augmented models rather than those from a benchmark model.

Our results provide convincing evidence that the factor-based approach is successful in predicting excess returns and volatility along both dimensions. First, the

factor-based predictive regressions have superior market timing and volatility timing ability compared to several benchmark models that include only valuation ratios and interest rate related variables, and possibly individual macro variables. In addition, the historical average return forecast is beaten as well. Second, using the excess return and volatility forecasts in active investment strategies, we find considerable economic value of the factor-based predictive regression models. A mean-variance investor would be willing to pay an annual performance fee of several hundreds of basis points to switch from the predictions offered by the benchmark models to those of the factor-based models. The active investment strategies also outperform passive buy-and-hold benchmarks. An important finding of our analysis is that the performance of the factor-based predictive regression models is stable over time. Whereas the benchmark models, especially the model including financial variables only, suffer from a substantial deterioration in performance during the 1990s, the predictive accuracy of the factor-based models remains stable or even improves.

Several alternative approaches are available to handle the issue of model uncertainty. First, we may consider variable selection, trying to identify the most relevant individual predictors at each point in time, as in Pesaran and Timmermann (1995). This strategy has several drawbacks though. For example, possibly valuable information in the variables that are not selected for inclusion in the predictive regression is discarded completely. Additionally, selection is necessarily based on past performance, which is no guarantee for future performance, especially in light of the well-documented instability of predictive relations between returns and financial and macro variables. Second, we may apply (Bayesian) model averaging (Avramov, 2002; Cremers, 2002), or forecast combination (Aiolfi and Favero, 2005; Timmermann, 2008; Huang and Lee, 2010; Rapach *et al.*, 2010). In principle this is an attractive way to account for model uncertainty, but still only a limited number of variables can be considered in this approach.

The paper is structured as follows. Section 2 develops the specification of the factor-augmented predictive regression models. Two procedures for preselecting the macro variables that are included in the factors are also discussed. Section 3 provides details on the data set used in the empirical analysis, various implementation issues, and the benchmark models that are used to judge the performance of the factor-augmented predictive regressions. Sections 4 and 5 present the empirical results, focusing on market timing and volatility timing ability of the predictive regressions, and on their economic value in investment strategies, respectively. Section 6 draws conclusions.

2 Factor-augmented predictive regressions

We consider one-period ahead forecasts of excess stock returns and volatility based on linear regression models. In this section, the model specification is described in detail in terms of predictive regressions for excess returns. Exactly the same set-up is used for (log) volatility.

In the conventional approach, the predictive regression model for excess returns takes the form

$$r_{e,t+1} = \beta_0 + \beta_x' x_t + \beta_z' z_t + \varepsilon_{t+1}, \tag{1}$$

where $r_{e,t+1}$ is the excess stock return in period $t+1$, $x_t = (x_{1t}, \ldots, x_{kt})'$ is a $(k \times 1)$ vector of macroeconomic predictor variables, and $z_t = (z_{1t}, \ldots, z_{lt})'$ is an $(l \times 1)$ vector of financial predictor variables. Throughout it is assumed that the values of x_t and z_t are available at the end of period t.

The key issue in the empirical specification of (1) is the content of the vectors x_t and z_t. Multiple candidates typically are available, especially for the macroeconomic variables. Including all of them is not recommended in that case, as the resulting parameter estimation uncertainty would adversely affect forecast accuracy. As dis-

6

cussed in the introduction, one possibility to avoid this issue would be to apply a variable selection procedure. This typically boils down to selecting those variables that have had the strongest predictive ability for excess returns during the recent past. Alternatively, model averaging or forecast combination may be employed, constructing a weighted average of forecasts from a range of models of the form (1) with different variables in x_t and z_t. In both approaches, the number of variables that can be considered is still fairly limited. Furthermore, variable selection procedures eventually completely discard the information embedded in those variables not selected for inclusion in the model.

The approach we consider in this paper rests on the assumption that the k available macroeconomic variables obey a factor structure of the form

$$x_t = \Lambda f_t + e_t, \tag{2}$$

where f_t is an $(r \times 1)$ vector of common factors, with r being much smaller than k. The factors f_t are assumed to be mutually orthogonal and in decreasing order in terms of their ability to explain the variance of the macroeconomic variables x_t. The factors are unobserved but can be estimated consistently by means of principal component analysis, as discussed in Stock and Watson (2002a, 2002b), among others. Given such estimates of the factors, we consider a factor-augmented predictive regression for excess returns of the form

$$r_{e,t+1} = \beta_0 + \beta_f' f_t + \beta_z' z_t + \varepsilon_{t+1}. \tag{3}$$

The main advantage of this factor-based approach is that the number of macro variables k can be rather large, while still the information in all variables is exploited.

A possible drawback of the above procedure is that the factors f_t are constructed regardless of the relation between the individual predictors x_{it}, $i = 1, \ldots, k$ and the

target variable $r_{e,t+1}$. Intuitively, while the first few principal components capture the bulk of the variation in the macroeconomic variables x_t, it need not necessarily be the case that this is the information that is most relevant for forecasting excess returns. More accurate forecasts may be obtained by taking into account the specific target when constructing the factors. Recent studies such as Bair *et al.* (2006); Boivin and Ng (2006) and Bai and Ng (2008) show that when the variables used for factor construction are pre-selected based on the predictive ability of x_{it} for the target variable, the forecasting performance of the factor-augmented models can indeed be improved in comparison with the models that simply use all the available macro variables.

Existing procedures for constructing factors that take into account the target variable typically boil down to reducing the initial set of macroeconomic variables $\{x_{it}\}_{i=1}^{k}$ by applying some sort of variable selection procedure. Here we consider the 'thresholding' rules advocated by Bai and Ng (2008). These are described intuitively in the remainder of this section, with full details being provided in Appendix A.

2.1 Pre-selecting macroeconomic variables

The first variable selection procedure we consider, called 'hard' thresholding, is based on the individual predictive content of x_{it} for $r_{e,t+1}$. Specifically, for each predictor x_{it}, $i = 1, 2, \ldots, k$, we regress $r_{e,t+1}$ onto x_{it} and z_t, that is, we estimate the model

$$r_{e,t+1} = \beta_0 + \beta_{x,i}x_{it} + \beta'_{z,i}z_t + \varepsilon_{i,t+1}. \tag{4}$$

We then construct the factors f_t by principal components as before, but only from those predictors for which the coefficient $\beta_{x,i}$ in this regression is significantly different from zero at a given significance level α.

The hard thresholding procedure evaluates each predictor in isolation, ignor-

ing the (possibly correlated) information in other predictors. This may lead to the selection of a subset of x_t consisting of very similar macroeconomic predictors. To overcome this problem, Bai and Ng (2008) suggest a 'soft' thresholding rule, which does take other candidate variables into account when selecting the predictors. Specifically, the soft thresholding procedure makes use of the Least Angle Regression (LARS) methodology, originating from Efron et al. (2004), to select a limited number of informative variables out of a large group. The main idea is to add predictors one at a time, starting with the predictor that is most correlated with the target variable. This predictor's coefficient is increased from its starting value of zero, up to the point where the residual is equally correlated with the predictor chosen initially and a second predictor. This second predictor is then added to the so-called 'active set', and the coefficients on both predictors in this set are now simultaneously increased, in such a way that the residual is equally correlated with these two predictors and with a third predictor, which then enters the active set. This process can be repeated until the set of candidate predictors is exhausted (provided the number of candidates k is less than the number of observations T). In this way, a complete ranking of the candidate predictors is obtained. Alternatively, LARS can be used as a variable selection method by stopping the algorithm after a prespecified number of predictors has been selected. This procedure closely approximates the more well-known Lasso method proposed by Tibshirani (1996); see Efron et al. (2004) and Bai and Ng (2008) for a discussion of this similarity. It should be noted that the soft thresholding procedure as descirbed above is not computationally intensive due to the availability of an efficient algorithm for LARS proposed in Efron *et al.* (2004). In fact, the soft thresholding procedure has a computational cost of the same order as ordinary least squares.

Intuitively, soft thresholding is similar to a 'specific-to-general' model selection procedure or 'forward stepwise regression', where predictors are added to the model

one at a time. The predictor that is added in a given iteration is the variable that has the maximum correlation with the residual vector from a regression of the target variable to the predictors already selected previously. The main difference is that soft thresholding performs variable selection and shrinkage of regression coefficients simultaneously. This follows from the fact that LARS effectively optimizes an objective function that can be written in terms of a penalized regression. This shrinkage property of the LARS has the potential of providing an additional advantage when the objective is to forecast excess stock returns. As shown in Campbell and Thompson (2008), shrinkage of the regression coefficients (either using theoretical considerations or econometric techniques) improves the accuracy of predictive regressions for excess returns.

3   Data, implementation details, and benchmark models

3.1   Data

We aim to predict monthly excess returns and volatility on the Standard & Poor's (S&P) 500 index. The sample period runs from January 1970 until December 2005. As discussed below, we use a ten-year moving window for model specification and estimation, such that one-month ahead forecasts are made for January 1980 onwards. The risk-free rate is proxied by the 1-month T-bill rate, for which we use the Fama-Bliss rate from CRSP. Unfortunately, this series was discontinued at the end of 2001. For the remaining years in our sample period we use the 4-week T-Bill rate measured at close on the last trading day of each month. For constructing a measure of monthly volatility, we follow the recent literature that indicates that highly accurate measures of volatility for a certain period can be obtained from higher frequency returns (see Fleming *et al.*, 2003; Andersen *et al.*, 2006, among others). This motivates us to

10

opt for the 'realized volatility' based on daily returns. Specifically, following French *et al.* (1987) and Akgiray (1989), we measure the monthly realized volatility by

$$\sigma_t^2 = \sum_{i=1}^{n_t} (r_{i,t} - \bar{r}_t)^2 \left[1 + \frac{2}{n_t} \sum_{j=1}^{n_t} (n_t - j) \hat{\phi}_t^j \right], \tag{5}$$

where n_t denotes the number of trading days in month t, and $\bar{r}$ and $\hat{\phi}_t$ denote the mean and the first-order autocorrelation, respectively, of the daily returns in month t. The correction to the sum of squared daily returns in (5) aims to account for the non-zero first-order autocorrelation in daily index returns due to non-synchronous trading of securities and other market microstructure effects (see French *et al.*, 1987; Lo and MacKinlay, 1990, for example). In the predictive regression models, we use the natural logarithm of the volatility measure, $\log(\sigma_t^2)$, as dependent variable to assure positive predictions and compute the one-period ahead expectation of volatility assuming a lognormal distribution for the realized volatility measure, see Andersen *et al.* (2006).

As discussed in the previous section, we use two types of predictors, namely macroeconomic and financial variables, in the predictive regression models for excess returns and its volatility. Our selection of financial predictors z_t consists of commonly accepted variables, namely the dividend yield, the price earnings ratio, the risk free rate and its first lag and the default spread (defined as difference between the yields on Moody's Baa and Aaa rates). Valuation ratios that combine accounting measures of companies with their market valuation have a long tradition in stock return predictability and have been found to be positively correlated with future excess returns, see Fama and French (1988) and Campbell and Thompson (2008), among others. We include the risk free rate as it is also used in the construction of the excess returns and we use the default spread as a proxy for market risk. In the predictive regression model for volatility, we also include the first lag of the natural

11

logarithm of the volatility measure to account for the strong persistence in stock return volatility.

For the macroeconomic variables x_t that are used for extracting the factors f_t, we use an updated version of the dataset of Stock and Watson (2005). The dataset consists of 127 macroeconomic variables.[2] We exclude four variables related to stock prices and corresponding financial ratios as well as all interest rate related variables, as we intend to separate the information in financial indicators and macroeconomic conditions. Put differently, we aim for the financial predictors and the macroeconomic factors to embed genuine, complementary information for predicting the excess returns and volatility. The remaining dataset consists of 106 macro variables, classified in the following twelve categories: output and income (17), employment and hours (29), sales (2), consumption (1), housing starts and sales (10), inventories (3), orders (5), exchange rates (5), money and credit quantity aggregates (10), price indexes (20), average hourly earnings (3) and miscellaneous (1), with the number of series in the category given in parentheses. We transform the series, whenever appropriate, to ensure stationarity by using log-levels, annual differences and annual log differences. Outliers are defined as the observations that deviate from the median value of the sample by more than six interquantile ranges and in each individual series these are replaced by the median value of the previous five observations (see Stock and Watson, 2005, for details). To avoid any look-ahead bias, the macroeconomic variables are lagged by one month. By this we mean that x_t contains the values of the macro variables concerning month $t-1$, which typically are released at some point during month t. Detailed description of the individual macroeconomic variables is provided in Appendix B.

---

[2]The original dataset of Stock and Watson (2005) includes 132 variables but due to the data availability when updating the variables from their original sources we had to exclude 5 variables.

3.2 Implementation details

The empirical application of the factor-augmented predictive regression (3) with (2) requires several decisions to be made.

First, we emphasize that we use a fully recursive forecasting approach, in the sense that for predicting the excess return and volatility in period $t+1$, we only use historically available information up to and including period t for all parts of the model specification, including the pre-selection of macro variables, the construction of the factors with principal component analysis, and the estimation of the coefficients in the predictive regressions. Given the existing evidence for the presence of structural breaks in the relation between excess returns and financial and macro predictors (see the references in the introduction), we use a moving window approach, with the window length fixed at ten years (120 monthly observations).[3] As demonstrated by Pesaran and Timmermann (2004) ignoring structural breaks can be especially costly when predicting the sign of excess returns, which is one of the aspects we focus on in our empirical analysis.

Second, a crucial issue concerns the number of factors r to be included in the predictive regression (3). Following Stock and Watson (2002b), among others, we use the Bayesian Information Criterion (BIC) for this purpose, where at least one and at most six factors are included. Furthermore, we generalize the model in (3) allowing for up to two lags of each factor to be included. To keep the computational burden at a feasible level, we do not perform an exhaustive search across all possible combinations of the first six principal components and lag structures. Instead, we assume that factors are included sequentially in order of importance,[4] while the number of lags is assumed to be the same for all included factors.[5]

---

[3] An alternative approach would be to consider a dynamic 'optimal' window length as in Pesaran and Timmermann (2002, 2007), based on statistical tests for structural breaks.

[4] By this we mean that when the j-th principal component is included in the predictive regression for $j = 2, \ldots, 6$, the first $j-1$ principal components are as well.

[5] We examine the added value of selecting the number of factors (and their lags) by examining the performance of a predictive regression model with a fixed number of factors. For this purpose,

Figure 1 summarizes the results of this model selection procedure. The graphs on the left show that BIC selects parsimonious models for the excess return regressions. For most windows only the first principal component is included, especially when the factors are constructed from all macro variables. Multiple factors are included more often when pre-selection of the macro variables is applied. At first sight this finding may seem puzzling, in the sense that it intuitively less factors should be required to capture the information in a smaller set of variables. However, it should be kept in mind that the thresholding rules aim to select those macro variables that have relevant predictive information for excess returns. This should then also apply to the principal components of the selected subset of variables, such that more factors may be included. The same observations applies to the predictive regressions for volatility, although multiple factors also are included quite frequently when no preselection of the macro variables is done. Interestingly, for both returns and volatility lags of the factors are almost never included, except for the excess return regressions during a short period centered around 1980. This may be interpreted as showing that any relevant information contained in these macro variables is incorporated in stock prices and volatility within the next month.

- insert Figure 1 about here -

Third, we need to specify the significance level in the hard thresholding procedure. Given the limited predictability of excess returns, we employ fairly liberal (one-sided) significance levels of 0.10 and 0.15, corresponding with threshold levels α for the t-statistic of $\beta_{x,i}$ of 1.28 and 1.04 (in absolute value), respectively, in (4). For these significance levels the average number of variables selected into the predictive regression for excess returns is equal to 35 and 45. For the volatility models, the

---

we consider the factor-augmented model that only includes the first principal component (and no lags), as this is supposed to capture the bulk of the variation in the macroeconomic variables. Results for this single-factor model are qualitatively similar to the results reported here, and are available upon request.

corresponding numbers are 48 and 58. Figure 2 shows that the number of selected variables varies substantially over time. For example, for the excess return regressions and a significance level of 0.10, it ranges between 6 (for windows ending in 1982) and 79 (for some windows ending in 2001). It is also interesting to note that for the predictive regressions for volatility the number of selected variables seems to gradually decline over time. There seems to be no relation between the number of individual variables selected for the factor construction and the number of factors included in the predictive regressions, in the sense that the correlation between these two characteristics is very close to zero in all cases.

<center>- insert Figure 2 about here -</center>

Fourth, we need to decide upon a decision rule for the soft thresholding approach for determining the subset of predictors to be included in the construction of the factors f_t. The LARS algorithm provides a complete ordering of the available macroeconomic variables according to their predictive ability. Hence, in principle it can be combined with any decision rule. An obvious approach would be to use information criteria, but in practice this results in the selection of a rather small number of predictors. Considering the limited predictability of excess returns, we select the first 60 variables that are included in the 'active set' in the LARS algorithm.

3.3   Benchmark models

We consider three benchmark models to evaluate the value added of the factor-augmented predictive regression models. The first benchmark model discards the information in the macroeconomic variables altogether and only includes the financial variables z_t that are also used in the factor-augmented regressions. This benchmark enables us to observe if the macro factors indeed bear additional information for predicting the excess stock returns and volatility. The second benchmark model

<center>15</center>

uses individual macroeconomic variables instead of factors. Specifically, we include the three variables with the largest t-values (in absolute value) for their coefficient in the regression (4) in the hard thresholding procedure.[6] The aim of this benchmark is to examine the benefits of using factors that contain information from a large number of macroeconomic variables relative to using the most promising individual macroeconomic predictors. The third benchmark model revisits the issue of predictive (in)ability of financial and macroeconomic predictors by using the historical average as the prediction of the excess return, see Welch and Goyal (2008). Given its strong persistence, for volatility we use a random walk forecast in this case.

4 Market timing and volatility timing

We evaluate the performance of the factor-augmented predictive regressions in two different ways. In the next section, we assess their economic value when the excess return and volatility forecasts are used in active investment strategies. In this section, we consider the quality of the forecasts in statistical terms by evaluating their directional accuracy. For excess returns, we examine the ability of the factor-augmented predictive regressions to correctly predict its sign, which we refer to as market timing. For volatility, it seems crucial to be able to predict high and low volatility regimes and therefore, following Marquering and Verbeek (2004a), we consider the sign relative to the median over the relevant 10-year moving window that is used for factor construction and model estimation, which we label volatility timing. We focus on the directional accuracy of the return and volatility predictions, as this generally is quite closely related to their economic value in terms of the performance of investment strategies based on the forecasts, see Leitch and Tanner (1991).

---

[6]In an earlier version of the paper, we used annual inflation, annual growth of industrial production and the annual growth rate of money supply, as in Pesaran and Timmermann (1995). This approach is obviously not as flexible as the current approach as it uses same variables each period. Results for this alternative specification are available upon request.

4.1 Methodology

We assess the market (volatility) timing ability by means of the hit ratio, defined as the proportion of months for which the sign of the excess return (volatility) is predicted correctly. We formally test the null hypothesis of no market timing ability by means of the statistic proposed in Pesaran and Timmermann (1992). This generalization of the Henriksson-Merton (HM) statistic tests whether the empirical hit ratio is significantly higher than the expected hit ratio under independence of the signs of the actual value and the forecast. For excess returns, the statistic is defined as

$$\text{PT} = \frac{\hat{P} - \hat{P}_*}{\sqrt{\hat{V}(\hat{P}) - \hat{V}(\hat{P}_*)}} \tag{6}$$

where $\hat{P}$ is the actual hit ratio $\hat{P} = \sum_{t=0}^{n-1} I_{\{r_{e,t+1}\hat{r}_{e,t+1}\}}/n$ with $I_{\{A\}}$ being an indicator function that takes the value 1 if the argument A (in this case the product of the actual excess return $r_{e,t+1}$ and the forecast $\hat{r}_{e,t+1}$) is positive and 0 otherwise, and n is the number of months in the forecast period. Similarly $\hat{P}_*$ is the expected hit ratio under independence, computed as $P_* = \hat{P}_r\hat{P}_{\hat{r}} + (1 - \hat{P}_r)(1 - \hat{P}_{\hat{r}})$ where $\hat{P}_r = \sum_{t=1}^{n} I_{\{r_{e,t+1}\}}/n$ and $\hat{P}_{\hat{r}} = \sum_{t=1}^{n} I_{\{\hat{r}_{e,t+1}\}}/n$ are the proportion of months for which the actual returns and the predicted returns, respectively, are positive. Finally, $V(\hat{P}_*)$ and $V(\hat{P})$ are estimates of the variances of $\hat{P}_*$ and $\hat{P}$. Under the null hypothesis of no market timing ability, the PT statistic is standard normally distributed asymptotically.

The PT test relies upon the assumption that the indicator functions $I_{\{r_{e,t+1}\}}$ and $I_{\{\hat{r}_{e,t+1}\}}$ do not exhibit serial dependence. For the monthly excess returns this assumption may hold, but given the extensive evidence on volatility clustering it is likely to be violated for volatility. We assess this issue in the following way. The PT statistic as defined in (6) is asymptotically equivalent to a (one-sided) test of

17

the significance of the slope coefficient in the regression

$$I_{\{r_{e,t+1}\}} = a + bI_{\{\hat{r}_{e,t+1}\}} + \eta_t, \tag{7}$$

see Breen *et al.* (1989). We examine the significance of the slope coefficient b by means of its t-statistic based on heteroscedasticity and autocorrelation consistent (HAC) standard errors to address the relevance of serial correlation in the indicator functions. The resulting test is denoted as BGJ-HAC.

In addition to the separate evaluation of the directional accuracy of the excess return and volatility forecasts, we also consider the market and volatility timing ability of the predictive regression models jointly. For this purpose, we employ a generalization of the nonparametric HM test proposed by Marquering and Verbeek (2004b) allowing for four categories, namely combinations of positive or negative excess return with high or low volatility. More explicitly, the test is based on the following contingency table, where ξ_t denotes the median level of volatility during the 10-year moving window ending in month t:

	$r_{e,t+1} \leq 0$ $\sigma^2_{t+1} \leq \xi_t$	$r_{e,t+1} \leq 0$ $\sigma^2_{t+1} > \xi_t$	$r_{e,t+1} > 0$ $\sigma^2_{t+1} \leq \xi_t$	$r_{e,t+1} > 0$ $\sigma^2_{t+1} > \xi_t$	Total
$\hat{r}_{e,t+1} \leq 0, \hat{\sigma}^2_{t+1} \leq \xi_t$	$n_{1,1}$	$n_{1,2}$	$n_{1,3}$	$n_{1,4}$	$n_{1,0}$
$\hat{r}_{e,t+1} \leq 0, \hat{\sigma}^2_{t+1} > \xi_t$	$n_{2,1}$	$n_{2,2}$	$n_{2,3}$	$n_{2,4}$	$n_{2,0}$
$\hat{r}_{e,t+1} > 0, \hat{\sigma}^2_{t+1} \leq \xi_t$	$n_{3,1}$	$n_{3,2}$	$n_{3,3}$	$n_{3,4}$	$n_{3,0}$
$\hat{r}_{e,t+1} > 0, \hat{\sigma}^2_{t+1} > \xi_t$	$n_{4,1}$	$n_{4,2}$	$n_{4,3}$	$n_{4,4}$	$n_{4,0}$
Total	$n_{0,1}$	$n_{0,2}$	$n_{0,3}$	$n_{0,4}$	n

Given the above contingency table the HM test statistic takes the form

$$\text{HM} = \sum_{i,j=1}^{4} \left[\frac{\left(n_{i,j} - \frac{n_{i,0}n_{0,j}}{n} \right)^2}{\frac{n_{i,0}n_{0,j}}{n}} \right], \tag{8}$$

Under the null hypothesis of no joint market timing ability, the test statistic follows a χ^2 distribution with nine degrees of freedom.

18

Like the PT test in (6), the HM statistic in (8) assumes the absence of serial correlation in the signs of the actual and predicted excess returns and volatility. As discussed before, this is a rather unrealistic assumption for volatility in particular, invalidating the HM test. We address this issue by means of the testing framework recently proposed by Pesaran and Timmermann (2009). Pesaran and Timmermann (2009) develop tests of independence for multicategorical variables based on canonical correlations from dynamically augmented reduced rank regressions, which are robust in the presence of serial dependencies. The crucial ingredient of this approach is to define (3×1) vectors $y_{t+1} = (y_{1,t+1}, y_{2,t+1}, y_{3,t+1})'$ and $\hat{y}_{t+1} = (\hat{y}_{1,t+1}, \hat{y}_{2,t+1}, \hat{y}_{3,t+1})'$ representing the occurrence of the different categories. Specifically, in line with the contingency table given above, define $y_{i,t+1}$, $i = 1, 2, 3$ as

$$y_{1,t+1} = 1 \text{ if } r_{e,t+1} \leq 0 \text{ and } \sigma^2_{t+1} \leq \xi_t, \text{ and } =0 \text{ otherwise,}$$

$$y_{2,t+1} = 1 \text{ if } r_{e,t+1} \leq 0 \text{ and } \sigma^2_{t+1} > \xi_t, \text{ and } =0 \text{ otherwise,}$$

$$y_{3,t+1} = 1 \text{ if } r_{e,t+1} > 0 \text{ and } \sigma^2_{t+1} \leq \xi_t, \text{ and } =0 \text{ otherwise,}$$

and $\hat{y}_{i,t+1}$, $i = 1, 2, 3$, similarly but based on the predictions $\hat{r}_{e,t+1}$ and $\hat{\sigma}^2_{t+1}$. As shown by Pesaran and Timmermann (2009), in case the multicategorical variables are serially independent, a test of independence can be based on the canonical correlation coefficients between y_{t+1} and $\hat{y}_{t+1}$. We may then consider both a maximum canonical correlation test and a trace test based on the average canonical correlation. The latter in fact is identical to the standard Pearson chi-square contingency table test of independence, as given in (8). In the case of serially dependent outcomes, a valid test of independence can be constructed using canonical correlations of suitably filtered versions of y_{t+1} and $\hat{y}_{t+1}$.

The tests of Pesaran and Timmermann (2009) can also be obtained within a regression framework. Specifically, in the absence of serial dependence a test of

19

independence between the actual and the predicted signs of the excess returns and volatility corresponds with testing $\gamma = 0$ in the regression

$$\theta' y_{t+1} = c + \gamma' \hat{y}_{t+1} + \eta_t, \tag{9}$$

where θ is a vector of weights for the different categories of the actual signs. Note that in case of two categories, this reduces to the regression in (7). In the multicategory case, the null hypothesis $\gamma = 0$ may be tested straightforwardly by means of an F-test, but the result obviously depends on the values of the 'nuisance parameters' θ. A standard approach in this case is to take the maximum F-test over all possible choices of θ, which is identical to testing the null hypothesis that the largest canonical correlation between y_{t+1} and $\hat{y}_{t+1}$ is equal to zero. Note that the largest canonical correlation between y_{t+1} and $\hat{y}_{t+1}$ being equal to zero when the actual and predicted signs of the excess returns and volatility are independent implies that in fact all three canonical correlations between y_{t+1} and $\hat{y}_{t+1}$ are zero in this case. This leads to the idea of the trace statistic, which exactly tests this hypothesis. The trace statistic also corresponds with testing the null hypothesis that the rank of the matrix Π is equal to 0 in the reduced rank regression

$$y_{t+1} = a + \Pi \hat{y}_{t+1} + \eta_t. \tag{10}$$

As shown by Pesaran and Timmermann (2009), in the case of serial dependence in the actual and predicted signs, robust tests of independence may be obtained by augmenting the regressions (9) and (10) by an appropriate number of lagged values of y and $\hat{y}$. Alternatively, the serial correlation may be account for by using a HAC covariance matrix In our empirical analysis, we employ the trace statistic based on (10) with a HAC covariance matrix with Bartlett weights, using the Andrews (1991) procedure for automatic lag length selection. As discussed in Pesaran and Timmer-

mann (2009) is asymptotically chi-squared distributed with nine degrees of freedom, but this may be a poor approximation to the finite-sample distribution given the discrete nature of y_{t+1} and $\hat{y}_{t+1}$ in (10). For this reason, we employ simulation to obtain the appropriate critical values for the sample size in our empirical application.

4.2   Empirical results

Table 1 displays hit ratios for the excess return and volatility forecasts for all models considered. We show results for the complete out-of-sample period from January 1980 until December 2005, as well as for subperiods spanning the first and second half. Panels A and B of the table show hit ratios for excess returns and volatility individually. The bottom panel displays hit ratios for joint market timing and volatility timing, i.e. the numbers shown are the proportion of months for which the signs of excess returns and volatility both are predicted correctly. The corresponding test statistics for evaluating the significance of the market and volatility timing are shown in Table 2.

<center>- insert Table 1 about here -</center>

Panel A of Table 1 reveals several interesting results concerning the market timing ability of the factor-augmented predictive regressions. First, over the complete out-of-sample period the factor-augmented models achieve hit ratios between 57.4% and 62.5%, which all are significantly higher than the expected hit ratio under independence according to the PT test and the HAC t-statistic of the slope coefficient in (7), see Table 2. It is also interesting to note that in most cases the values of the PT test and the BGJ-HAC statistic are close, which corresponds to the lack of serial dependence in the signs of the actual and predicted excess returns.

<center>- insert Table 2 about here -</center>

<center>21</center>

The hit ratios achieved by the factor-augmented models exceed those of benchmark models I and II, with the single exception that the model using the strictest hard thresholding rule (TRS II) has a slightly lower hit ratio than the first benchmark model. This finding suggests that macroeconomic variables do bear relevant additional information for market timing, relative to the valuation ratios and interest rate related variables considered here. Moreover, it is better to extract the information from a large panel of macroeconomic variables by means of factors than to include a small number of specific macro variables as in the second benchmark model. In fact, the low hit ratio of this model (also compared to the first benchmark model) suggests that including individual macroeconomic variable actually deteriorates market timing ability. Relative to the third benchmark model, the performance of the factor-augmented predictive regressions seems less convincing, with only the standard factor model achieving a clearly higher hit ratio. However, to a large extent the market timing ability of the third benchmark is spurious, in the sense that only four percent of the excess return predictions from this model are negative. Hence, an investor using this approach would almost always invest in stocks without trying to time the market. This is confirmed by the test results in Table 2, which shows that the third benchmark model does not have significant market timing ability. In fact, for the second sub-period the test statistics cannot be computed due to the fact that after 1992 all excess return predictions from this model are positive.

Second, comparing the predictions from the different factor-augmented models among each other, it is seen that the 'standard' approach of including all available macroeconomic variables for extracting the factors at each point in time performs best. Apparently, market timing ability is not enhanced by pre-selecting the (presumably) most informative predictors which the other three methods aim to achieve.

Third, the market timing ability of the factor-augmented predictive regressions seems to be fairly stable over time, in the sense that the hit ratios for the first and

second half of the out-of-sample period do not differ dramatically. In fact, the hit ratios are higher for the period 1993-2005 for all four approaches. The PT and BGJ-HAC results in Table 2 also show that the significance of the market timing ability does not differ across the two sub-periods. This finding is in sharp contrast to existing evidence, which suggests that the market timing ability of typical predictive regressions has reduced considerably since the early 1990s. This is indeed confirmed by our results for the first benchmark model, which achieves the highest hit ratio of almost 63% during the period from 1980-1992, but then suffers a dramatic decline by ten percentage points to only 53% in the period 1993-2005. The corresponding PT and BGJ-HAC statistics in Table 2 decline from highly significant values of 2.73 and 2.42 for the first sub-period to insignificant values close to zero for the second subperiod.

We examine the (in)stability of the market timing ability of the different models in more detail by computing hit ratios for rolling windows of 60 months. These are displayed in Figure 3 together with the expected hit ratios under independence of the signs of the excess return forecasts and realizations. The fifth panel of Figure 3 clearly reveals the large decline of the market timing ability of predictive regressions that include only financial variables between 1995-1997. After 1995 the hit ratios for the second benchmark model mostly are even below the expected hit ratios under independence, confirming the limited ability of this model with individual macro variables to predict the sign of the excess return correctly. In contrast, the performance of the 'standard' factor-augmented predictive regression model is much more stable and consistently above the expected hit ratio under independence. The hit ratios of the factor models with pre-selection of macro variables show considerably more fluctuation, although they do exceed the hit ratios expected under independence most of the time. For example, for the first hard-thresholding model, the hit ratio varies between 41% for windows around 1990 and approximately 65% for

windows centered around 1985, 1995 and after the year 2000. Finally, the bottom panel of Figure 3 clearly demonstrates the lack of market timing ability of the historical mean return, in the sense that the actual hit ratio generally coincides with the expected hit ratio under independence.

- insert Figure 3 about here -

The market timing results suggest that the factors extracted from macroeconomic variables may have useful information for predicting the sign of the excess stock returns. Ultimately, accurately predicting the magnitude of the excess returns may be at least as important as correct sign predictions. We assess this aspect here in two different ways. First, we examine the significance of the slope coefficient in a regression of the actual excess returns on the corresponding forecasts,

$$r_{e,t+1} = a + b\hat{r}_{e,t+1} + \eta_t, \tag{11}$$

where again we use HAC standard errors to account for serial correlation. This is a robust version of the test used in Bossaerts and Hillion (1999). The results in Table 2 show that the performance of all models declines in the second half of the out-of-sample period. While the slope coefficient is significantly different from zero (and positive) at the 5% level or better for the subperiod 1980-1992, it becomes insignificant for the period 1993-2005. For the first benchmark model, the estimate of b in (11) even turns negative. The only exception is that the factor-augmented model based on soft thresholding appears to retain some of its predictive ability, with the HAC t-test being significant at the 10% level for the latter period.

Second, we evaluate the mean squared prediction error (MSE) of the factor-augmented predictive regressions relative to the MSE of the historical mean prediction used in the third benchmark model. Specifically, following Welch and Goyal (2008) and Campbell and Thompson (2008), among others, we consider the out-of-

24

sample R^2 defined as

$$R^2_{\text{OoS}} = 1 - \frac{\sum_{t=0}^{n-1}(r_{e,t+1} - \hat{r}_{e,t+1})^2}{\sum_{t=0}^{n-1}(r_{e,t+1} - \bar{r}_{e,t+1})^2}, \tag{12}$$

where $\hat{r}_{e,t+1}$ is the return prediction obtained from one of the factor-augmented regression models and $\bar{r}_{e,t+1}$ is the historical average return computed using the ten-year moving window ending in period t, that is, the excess return prediction used in the third benchmark model. Positive (negative) values of R^2_{OoS} indicate that the factor-augmented model achieves a smaller (larger) MSE than the historical average return. By inspecting the path of R^2_{OoS} for increasing values of n, we can also assess the stability of the relative predictive accuracy of the two competing models. When the R^2_{OoS} increases, it means that the factor-augmented model outperforms the historical average, and vice versa when the R^2_{OoS} decreases. Figure 4 displays the path of the R^2_{OoS} during the complete out-of-sample period for the four factor-augmented predictive regressions. Over the complete out-of-sample period, the R^2_{OoS} values are negative at around -0.10, such that the historical average return achieves a smaller MSE than the elaborate factor-augmented predictive regressions. Interestingly, this appears to be due completely to a small number of erratic forecasts obtained from the factor-augmented models at the beginning of the sample period. This results in large fluctuations in the R^2_{OoS} values, such that they drop to levels below -1 in July 1980. Following this unfortunate start the R^2_{OoS} generally shows an increasing pattern, indicating that over the remainder of the forecast period the factor-augmented regressions actually produce considerably more accurate excess return predictions than the historical average return.

- insert Figure 4 about here -

Next, we turn to the volatility timing results as shown in panel B of Table 1. As volatility exhibits much more persistence than excess returns, predicting the state

25

of volatility one month ahead may be expected to be less difficult than predicting the sign of excess returns. This is indeed reflected by the high values of the hit ratios for the complete out-of-sample period as well as both subperiods. All models have hit ratios above 60%, with a substantial increase in the second half of the out-of-sample period. For the complete out-of-sample period 1980-2005 the factor-augmented models outperform the first and second benchmark only marginally. For the period after 1992, we find larger differences in hit ratios up to 5% between the factor-augmented model based on soft thresholding and the second benchmark model that includes individual macro variables. The statistics in Table 2 show that all models have significant volatility timing ability, both for the complete out-of-sample period as well as the two subperiods. Note that in this case the value of the BGJ-HAC statistic is considerably smaller than the PT statistic, due to the presence of positive serial correlation in the signs of the actual and predicted volatility.

Figure 5 shows the hit ratios for volatility forecasts over 5-year rolling windows, together with the expected hit ratios under independence of the sign forecasts and the actual signs. We observe that all models exhibit a similar pattern in their hit ratio, showing a gradual increase after 1995. Note that this does not necessarily imply that the volatility timing ability of the predictive regressions has improved over time. For example, the actual hit ratios are at the highest level for moving windows ending between 2000 and 2003, close to 90% for most models. This also is the case, however, for the expected hit ratios under independence, suggesting that the impressive values of the hit ratios are in fact not due to superior volatility timing ability but due to volatility clustering. Apart from this exceptional period, though, we do observe that the actual hit ratios are well above their expected values under independence, suggesting that in general volatility timing is possible with predictive regression models.

- insert Figure 5 about here -

Finally, we consider the joint market and volatility timing ability, which is crucial for a mean-variance investor who would like to make investment decisions based on predictions of both excess returns and volatility. From Table 1, we observe that the hit ratios of joint timing take lower values compared to those of market and volatility timing individually. This shows that predicting the true state of the excess return and volatility jointly is more problematic than predicting them separately. The difference in the performance of the factor-augmented models and benchmark models in market timing ability is reflected in the joint timing results as this difference becomes larger during the second subsample. As in the market timing case many of the test statistics cannot be computed for the third benchmark model. Again this is due to the fact that some of the categories are never realized for the third benchmark due to the lack of market timing ability.

5 Economic value

In the previous section we analyzed the market timing and volatility timing ability of the factor-augmented predictive regression models. As mentioned before, this generally provides some indication of the performance of active trading strategies based on the excess return and volatility predictions. In this section, we assess this economic value of the factor-augmented predictive regression models explicitly. For this purpose, we consider a mean-variance investor with a one-month horizon, who decides upon the fraction of her wealth to be allocated to stocks using forecasts of excess returns and volatility. We use a utility based metric to evaluate how much the investor would be willing to pay to use the predictions from the factor-augmented models rather than those from the benchmarks. Specifically, we employ the procedure proposed in West *et al.* (1993), which is frequently used for measuring the economic value of dynamic investment strategies, see Fleming *et al.* (2001, 2003), Marquering and Verbeek (2004a), De Pooter *et al.* (2008), and Della Corte *et al.*

(2009, 2010), among others.

5.1 Methodology

We consider a risk-averse investor with mean-variance preferences, managing a portfolio consisting of stocks and risk-free T-bills. Each month the investor decides upon the fraction of wealth to be invested in stocks using the objective function

$$\max_{w_{t+1}} \quad \mathsf{E}_t(r_{p,t+1}) - \frac{1}{2}\gamma\mathsf{Var}_t(r_{p,t+1}), \tag{13}$$

where $\mathsf{E}_t(r_{p,t+1})$ and $\mathsf{Var}_t(r_{p,t+1})$ are the expected value and the variance of the portfolio return in period $t+1$ conditional on the information in period t, and γ represents the degree of relative risk aversion (RRA). The portfolio return is given by

$$r_{p,t+1} = r_{f,t+1} + w_{t+1}r_{e,t+1},$$

where $r_{f,t+1}$ denotes the risk-free return, and w_{t+1} is the fraction of wealth allocated to stocks. We consider a myopic investor with a horizon of one month, such that the investor solves the maximization problem in (13) each time period in order to determine the portfolio weights for the next month. The optimal portfolio weight for stocks is given by

$$w_{t+1}^* = \frac{\mathsf{E}_t(r_{e,t+1})}{\delta\mathsf{Var}_t(r_{p,t+1})}. \tag{14}$$

As (14) indicates, the implementation of the dynamic mean-variance strategy requires predictions of the excess return and volatility for computing the optimal portfolio weight for stocks. We assume that our investor employs the recursive predictions of the expected excess returns and volatility obtained from the factor-augmented regression models and the benchmark models for this purpose.

In order to mitigate the effects of forecast uncertainty the investor imposes restrictions on the portfolio weight. Specifically, we consider the following two scenar-

ios. First, the investor uses the restriction $w^*_{t+1} \in [-1, 2]$, thus allowing for limited short-sales and leveraging of the portfolio. In the second scenario, short-sales and leveraging are not allowed at all, thus $w^*_{t+1} \in [0, 1]$.

Using the new return and volatility predictions that become available each month, the mean-variance investor adjusts her portfolio weights if necessary. Hence, transaction costs are an inevitable factor that must be taken into account. At the start of month $t + 1$, the investor changes the allocation to stocks from w_t to w_{t+1}. We assume that transaction costs are a fixed proportion c of the wealth invested, such that the portfolio return is reduced by

$$c_{t+1} = 2c|w_{t+1} - w_t|, \tag{15}$$

where the multiplication by 2 follows from the fact that the investor rebalances her positions in both stocks and the riskfree T-bills. We consider no, medium and high transaction costs by setting c equal to 0%, 0.1% and 0.3%, respectively.

Assuming quadratic utility, we can use the average realized utility to consistently estimate the expected utility obtained from a given level of initial wealth W, see West *et al.* (1993), that is,

$$\overline{U} = \frac{W}{n} \sum_{t=0}^{n-1} \left(R_{p,t+1} - \frac{1}{2} \frac{\gamma}{(1+\gamma)} R^2_{p,t+1} \right), \tag{16}$$

where $R_{p,t+1} = 1 + r_{p,t+1} - c_{t+1}$ is the gross return on the portfolio net of transaction costs.

We assess the economic value of the factor-augmented predictive regression models by computing a 'performance fee', denoted as Δ, that an investor should be willing to pay for switching from one investment strategy to another. For example, suppose that holding a dynamic portfolio constructed using one of the benchmark models yields the same average utility as holding a dynamic portfolio constructed using a

29

factor-augmented predictive model that is subject to annual expenses, Δ, expressed as a fraction of wealth invested. Because the investor would be indifferent between the two strategies we can interpret Δ as the maximum fee an investor should be willing to pay to switch from the benchmark strategy to the strategy based on the factor-augmented models. The magnitude of this fee is given by the value of Δ that satisfies

$$\sum_{t=0}^{n-1} \left\{ \left(R_{p,t+1}^a - \Delta \right) - \frac{1}{2} \frac{\gamma}{(1+\gamma)} \left(R_{p,t+1}^a - \Delta \right)^2 \right\} = \sum_{t=0}^{n-1} \left\{ R_{p,t+1}^b - \frac{1}{2} \frac{\gamma}{(1+\gamma)} (R_{p,t+1}^b)^2 \right\} \tag{17}$$

where the superscripts a and b denote the alternative strategies.

We compare the economic value of the factor-augmented predictive models relative to the three benchmark models. Additionally, we also consider three buy-and-hold strategies. These strategies assume that the investor invests a constant fraction of 100% (0%), 50% (50%), or 0% (100%) of her wealth in stocks (risk-free T-bills), respectively. Throughout the analysis, we set the relative risk aversion γ equal to 6.[7] Finally, for conventional comparison of the portfolio performances we also compute average returns, standard deviations, and Sharpe Ratios for each portfolio.

5.2 Empirical results

The results in Table 3 indicate that over the complete out-of-sample period from January 1980 until December 2005, investment strategies based on the factor-augmented predictive regressions outperform those based on the benchmark models, with a few exceptions. This holds irrespective of whether we consider Sharpe Ratios, as shown in the third column, or the performance fees for switching from the benchmark models to the factor-augmented predictive models in the three rightmost columns. While the magnitude of the economic value of the macro factors is affected by the level of

---

[7]In order to examine the sensitivity of the results to the level of risk aversion, we also consider risk aversion levels of 2 and 10. Results are shown in Appendix C.

transaction costs and the restrictions imposed on the portfolio weights (as discussed in more detail below), it remains positive in all scenarios.

As an example, consider the performance of the factor-augmented models relative to the first benchmark model in the scenario where limited short sales and leverage are allowed ($w_{t+1}^* \in [-1, 2]$) and transaction costs are moderate (0.1%). In that case the portfolios based on the factor-augmented regressions achieve annualized Sharpe Ratios in the range 0.45-0.55, compared to 0.35 for the first benchmark model. This difference is mostly due to a higher average return achieved by the factor-augmented models ranging between 12% and 13.5% compared to 10.5% for the benchmark. The volatilities of the portfolio returns are comparable, at around 15%. The mean-variance investor is willing to pay a performance fee ranging from 131 to 326 basis points annually for switching from the first benchmark model to one of the factor-augmented models. This confirms the results of the market timing analysis in the previous section, in that the macroeconomic factors add relevant information concerning the expected returns and volatility to the information embedded in the financial variables in this first benchmark model. The possibilities to exploit this additional information become somewhat more limited when short sales and leverage are abandoned completely. The performance fees become smaller in this case, although they remain substantial ranging between 109 and 142 basis points.

The outperformance of the factor-augmented predictive regression models relative to the second benchmark model is particularly impressive. Again considering the scenario with medium transaction costs and limited short sales and leverage, the Sharpe Ratio of this benchmark is a disappointing 0.10, mostly due to a low average return of 7.2%. The performance fees in this scenario range between 622 and 816 basis points. Put differently, the investor is willing to pay between 6-8% on

an annual basis to replace individual macroeconomic variables by factors based on a large database in the predictive regression models. Again, the performance fees are reduced, to about 3-3.5%, when the portfolio weights are restricted to be between 0 and 1.

The third benchmark model, which uses the historical average excess return and the previous month's volatility level as predictions, appears more difficult to beat. When limited short sales and leverage are allowed the third benchmark model actually performs better than most of the factor-augmented models, with performance fees of around 100 basis points. Interestingly, all of the factor-augmented models yield higher average returns than this benchmark model. The third benchmark model achieves a considerably lower standard deviation of returns, however. Still, the factor-augmented model with soft thresholding performs slightly better than the third benchmark model by an annual performance fee of 50 basis points. Moreover, when short sales and leverage are completely restricted, all factor-augmented models have a positive performance fee as well as higher Sharpe ratios. In this case, an investor is willing to pay between 137 and 170 basis points annually to switch from the third benchmark model to the factor-augmented models. The main consequence of abandoning short sales and leverage is a substantial reduction in the standard deviation of portfolio returns based on the factor-augmented models from 15% to 8% on average, see the second column of Table 3. While the portfolio standard deviation is also reduced for the third benchmark model, the decline is much less pronounced (from 13.3% to 10%).

The performance fees of the factor-augmented models relative to the three benchmark models are relatively stable across the three different levels of transaction costs considered. By contrast, when comparing the dynamic strategies with static buy and hold strategies, the level of transaction costs becomes much more important in evaluating the relative portfolio performance, especially when limited short sales and

leveraging are allowed. In that case, the factor-augmented models still render positive performance fees when transaction costs are not included. Medium and high transaction cost levels, however, lead to negative performance fees. Especially the mixed portfolio with a constant weight of 50% in stocks performs well, with Δ being approximately equal to -500 basis points in the scenario with high transaction costs. This picture changes dramatically when short sales and leveraging are not allowed. In this case, under medium transaction costs all factor-augmented models have a positive performance fee ranging from 86 to 119 basis points annually against the static mixed portfolio. The performance fees against the other two static portfolios are considerably higher. The investor is willing to pay around 250 basis points annually to switch from the buy-and-hold strategy in stocks to one of the factor augmented models, and more than 300 basis points to switch from the static strategy that always invests in T-bills. Interestingly, even if the transaction costs are high, factor-augmented models still perform better than or as good as the static strategies. An important finding is that the performance fees are highest when the factor models with thresholding rules are used for portfolio construction. This indicates that a sensible pre-selection of the set of macroeconomic variables from which factors are constructed adds even more economic value than simply using all available macro variables as in the standard factor-augmented model.

The performance fees for switching from the static portfolios to the factor-augmented predictive regressions indicate the importance of imposing restrictions on the portfolio weights, to avoid excessive exposure to stocks but also to limit transaction costs. This is reflected in Table 4 where we display the mean and the standard deviation of the portfolio weights. For all factor-augmented models (as well as for the benchmark model), we observe that the standard deviation of the portfolio weights is reduced by 50-60% (from around 1.00 to 0.45, on average) when short sales and leverage are abandoned.

- insert Table 4 about here -

A crucial finding from the market timing analysis in the previous section was the stability of the performance of the factor-augmented predictive regressions, whereas the performance of the benchmark models deteriorates substantially during the second half of the out-of-sample period. This is also reflected in their (relative) economic value during the subperiods 1980-1992 and 1993-2005, as shown in Tables 5 and 6. The portfolios based on the predictions obtained from the benchmark models experience a pronounced decline in their Sharpe Ratio. For example, the first benchmark model (including financial variables only) achieves a Sharpe ratio of 0.59 over the period 1980-1992, when limited short sales and leverage are allowed and transaction costs are medium. This drops to 0.00 for the period 1993-2005. The reduction in Sharpe ratio is less dramatic when the portfolio weights are more restricted, but still substantial from 0.66 to 0.28. For the factor-augmented models, the Sharpe ratio is approximately the same during the two sub-periods.

- insert Tables 5 and 6 about here -

The performance fees reveal a similar pattern. While the dynamic strategies based on the benchmark models mostly achieve positive performance fees against the static buy-and-hold strategies during the first half of the out-of-sample period, these turn negative for the period after 1992. Strikingly, the factor-augmented models generally have higher performance fees during the second sub-period. Most notably this holds for those factor models where thresholding rules are employed for factor extraction. Specifically, models with hard thresholding with restricted portfolio weights perform better than all passive strategies even when the transaction costs are at the highest level. Allowing for short sales and leverage of the portfolio increases the performance fee for switching from the benchmark models to the factor-augmented models. However, at the same time it reduces the performance

34

fees for switching from the static strategies, as more volatile returns cause a loss of utility. For the first half of the sample period only the first benchmark model performs marginally better than the factor-augmented models when the portfolio weights are completely restricted. When limited short sales and leverage are allowed the performance fee for switching to a factor-augmented model ranges from −18 to −534 basis points under medium transaction costs.

A more detailed picture of the evolution of the performance of the factor-augmented models relative to the benchmark models over the out-of-sample period is provided by Figure 6, showing the fees for switching from each of the three benchmark models to factor-augmented regressions for moving windows of 60 months (with medium transaction costs). An interesting observation from these graphs is that for the first benchmark model, the superior performance of the factor-augmented models already starts around the beginning of the 1990's, whereas for the third benchmark model it only starts around the year 2000. In fact, the third benchmark model is seen to perform particularly well during most of the 1990s, with negative performance fees for all four factor-augmented models. This is due to the low standard deviation of portfolio returns obtained from the third benchmark model combined with the prolonged bull market with low volatility during this period. As discussed before, the third benchmark model always predicts moderate positive excess returns after the mid 1980's but does not have any market timing ability. This strategy seems to work for the bull market of the mid 1990's when volatility was also very low. However, with the advent of more volatile periods towards the end of this decade, the market timing ability of the predictive models gains more importance. This explains the dramatic improvement in the performance of the factor-augmented models starting from the end of the 1990's, with the fees relative to the third benchmark model increasing to almost 1200 basis points at the end of 2003.

- insert Figure 6 about here -

An important result from Table 6 is that the factor-augmented models that use thresholding rules for predictor selection perform better than the factor model that simply uses all available variables, especially when limited short sales and leverage are allowed for. This confirms the findings of Bai and Ng (2008) in the context of forecasting inflation. When the aim is to predict a variable with limited predictive content such as stock returns, our findings show that a reasonable set of thresholding rules is even more crucial also economically. For switching from the passive strategy of investing 100% in stocks to the first hard thresholding rule with completely restricted portfolio weights, for example, an investor would pay 284 basis points per annum compared to a fee of only 168 for the standard factor model. This difference increases when limited short sales are allowed, for which the precision of return predictions becomes more important rather than the prediction of the direction of the market. When this scenario is considered the factor model with the soft thresholding rule, which is relatively more sophisticated than others, outperforms the other factor models.

The majority of previous studies focuses on the market timing ability of predictive models for (excess) returns only. This is either because the objective of the econometric methods is directed only at return prediction or because the investment strategy is 'discrete', involving switching between stocks and risk-free investments based only on the predicted sign of the excess return (see Pesaran and Timmermann, 1995; Aiolfi and Favero, 2005, among others). From this perspective, it is also interesting to evaluate the performance of the factor-augmented models using only the information in the excess return forecasts. We implement this in our mean-variance framework by only using the return forecast in (14) while setting the conditional volatility equal to the median volatility value during the estimation sample.[8] Table 7 exhibits the results of these strategies for the complete out-of-sample period

---

[8]Recall that we also used this as benchmark value for determining "high" and "low" volatile periods.

and the two subperiods. For the sake of brevity, we only display results for medium transaction costs (0.1%). Results for other transactions cost levels are in line with these and available upon request.

- insert Table 7 about here -

A comparison of the results in Table 7 with those in Panel B of Tables 3-6 reveals that ignoring the volatility predictions increases both the mean and volatility of the portfolio returns. The overall effect on the Sharpe ratio is slightly negative. The performance fees for switching from the benchmark models to factor-augmented models increase substantially for almost all cases. Especially for the third benchmark model, which uses the historical average excess return as prediction, the performance fees almost double. This clearly shows how this benchmark model benefits from volatility timing rather than market timing in returns, confirming the results of Fleming *et al.* (2001, 2003). The benefits of volatility timing do not seem so pronounced for factor-augmented models as the increase of the performance fee for switching from static portfolios is very limited. When portfolio weights are completely restricted, in almost all cases the factor-augmented models have positive performance fees, regardless whether the competing strategies are static or dynamic but based on benchmark models.

6 Conclusions

While the basic question whether stock returns are predictable at all has not been settled conclusively, the voluminous literature on this topic seems to agree that individual macroeconomic variables do not have predictive ability for returns, at least not for prolonged periods of time. The relation between stock returns and individual macroeconomic predictors appears to be subject to relatively frequent structural breaks. Hence, when a given macroeconomic variable may appear to be

37

useful for forecasting stock returns over a certain period, this might very well break down at some point. Put differently, studying return predictability with a limited number of individual macroeconomic variables over a long time span is unlikely to find positive results.

While the issue of return predictability is still disputed, there appears to be more consensus that the level of volatility is predictable to some extent. In addition, several studies document a relation between stock market volatility and macroeconomic conditions, with volatility being substantially higher during business cycle recessions. Nevertheless, it is unclear a priori which specific macroeconomic variables have the most relevant information for future volatility.

The point of departure for this paper is that in reality we may consider numerous macroeconomic variables, and their combined information may be useful for predicting future stock returns and volatility to some extent. Based upon an empirical analysis we document that factors extracted from a large set of macroeconomic variables do indeed have predictive ability for monthly US excess stock returns and volatility for the period 1980-2005. Factor-augmented predictive regression models improve upon both benchmark models that only include valuation ratios and interest rate related variables and possibly individual macro variables, as well as naive predictors such as the historical average excess return and previous month's volatility. The improvements in out-of-sample forecast accuracy are both statistically and economically significant. The factor-augmented predictive regressions have superior market timing ability and volatility timing ability, while a mean-variance investor would be willing to pay an annual performance fee of several hundreds of basis points to switch from the predictions offered by the benchmark models to those of the factor-augmented models. An important reason for the superior performance of the factor-augmented predictive regressions is the stability of their forecast accuracy, whereas the benchmark models suffer from a forecast breakdown during the 1990s.

Appendix A   Thresholding rules

This appendix provides descriptions of the two procedures for preselection of a subset of macro variables that are used for constructing the factors that enter the predictive regressions. Further details of the thresholding rules can be found in Bai and Ng (2008).

A.1   Hard thresholding

The 'hard' thresholding procedure selects macro variables x_{it} based on their individual predictive content for the excess returns $r_{e,t+1}$. The algorithm for hard thresholding is as follows:

1. For each macroeconomic predictor x_{it}, $i = 1, 2, \cdots, k$, estimate the regression

$$r_{e,t+1} = \beta_0 + \beta_{x,i} x_{it} + \beta'_{z,i} z_t + \varepsilon_{i,t+1},$$

   by least squares, where $r_{e,t+1}$ denotes the excess stock return in period $t + 1$ and z_t is the $(l \times 1)$ vector of valuation ratios and interest rate related variables that are always included.

2. Sort the macro variables in decreasing order of the absolute value of the t-statistic of $\hat{\beta}_{x,i}$, denoted t_i.

3. Choose a (one-sided) significance level α and select the subset of predictors for which $|t_i|$ exceeds the corresponding threshold.

4. Construct the factors f_t by means of the principal components for the selected subset of predictors.

A.2   Soft thresholding

The 'soft' thresholding procedure takes other candidate macro variables into account when selecting the predictors, based on the Least Angle Regression (LARS) methodology. Predictors are added to the model one at a time, starting with the variable that is most correlated with the excess return. This predictor's coefficient is increased from its starting value of zero, up to the point where the residual is equally correlated with the predictor chosen initially and a second predictor. This second predictor is then added to the so-called 'active set', and the coefficients on both predictors in this set are now simultaneously increased, in such a way that the residual is equally correlated with these two predictors and with a third predictor, which then enters the active set. This process can be repeated until the set of candidate predictors is exhausted (provided the number of candidates k is less than the number of observations T). In this way, a complete ranking of the candidate predictors is obtained.

Technically, the j-th iteration of the LARS algorithm proceeds as follows.

1. Compute the correlations $\hat{c}_{(j)}$ between the candidate predictor variables and the residual excess return after the $(j-1)$-st iteration (thus, with $j-1$ candidate predictors already included in the active set). Collecting the predictors (which are assumed to be standardized to have mean zero and variance one) in the $(T \times k)$ matrix X and denoting the estimate of the expected excess return obtained in the $(j-1)$-st iteration as $\hat{\mu}_{(j-1)}$, we have $\hat{c}_{(j)} = X'(r_e - \hat{\mu}_{(j-1)})$.

2. Identify the candidate predictors with the largest absolute correlations and define $\mathcal{S}_{(j)}$ as the set of the corresponding indices, that is

$$\hat{c}_{(j),\max} = \max_i\{|\hat{c}_{(j),i}|\} \quad \text{and} \quad \mathcal{S}_{(j)} = \{i : |\hat{c}_{(j),i}| = \hat{c}_{(j),\max}\},$$

where $\hat{c}_{(j),i}$ denotes the i-th element of $\hat{c}_{(j)}$.

3. Define the active matrix $X_{(j)}$ corresponding to $\mathcal{S}_{(j)}$ by 'signing' the predictors, that is

$$X_{(j)} = (s_{(j),i} X_i)_{i \in \mathcal{J}} \quad \text{with} \quad s_{(j),i} = \text{sign}(\hat{c}_{(j),i}).$$

4. Let $G_{(j)} = X'_{(j)} X_{(j)}$ and $A_{(j)} = (\iota'_{(j)} G'_{(j)} \iota_{(j)})^{-1/2}$ where $\iota_{(j)}$ is a vector of ones corresponding to the size of $\mathcal{S}_{(j)}$. Then define $u_{(j)} = X_{(j)} w_{(j)}$, where $w_{(j)} = A_{(j)} G^{-1}_{(j)} \iota_{(j)}$, and $a_{(j)} \equiv X' u_{(j)}$.

5. Update the estimate of the expected excess return as $\hat{\mu}_{(j)} = \hat{\mu}_{(j-1)} + \hat{\gamma} u_{(j)}$, where

$$\hat{\gamma} = \min_{i \in \mathcal{S}^c_{(j)}}{}^+ \left\{ \frac{\hat{c}_{(j)} - \hat{c}_{(j),i}}{\hat{A}_{(j)} - \hat{a}_{(j),i}} , \ \frac{\hat{c}_{(j)} + \hat{c}_{(j),i}}{\hat{A}_{(j)} + \hat{a}_{(j),i}} \right\}$$

with $\min^+$ indicating that the minimum is taken over only positive components. The algorithm is initialized with $\hat{\mu}_{(0)} = 0$.

Appendix B Macroeconomic variables

Table B.1: Base Set of Predictors used for Extracting Factors

Short Name	Transf.	Description
Output and Income		
PI	Δln	Personal Income (AR, Bil. Chain 2000$) (TCB)
PI less transfers	Δln	Personal Income Less Transfer Payments (AR, Bil. Chain 2000$) (TCB)
IP: total	Δln	Industrial Production Index - Total Index
IP: products	Δln	Industrial Production Index - Products, Total
IP: final prod	Δln	Industrial Production Index - Final Products
IP: cons gds	Δln	Industrial Production Index - Consumer Goods
IP: cons dble	Δln	Industrial Production Index - Durable Consumer Good
IP: cons nondble	Δln	Industrial Production Index - Nondurable Consumer Good
IP: bus eqpt	Δln	Industrial Production Index - Business Equipment
IP: matls	Δln	Industrial Production Index - Materials
IP: dble mats	Δln	Industrial Production Index - Durable Goods Materials
IP: nondble mats	Δln	Industrial Production Index - Nondurable Goods Materials
IP: mfg	Δln	Industrial Production Index - Manufacturing (Sic)
IP: res util	Δln	Industrial Production Index - Residential Utilities es
IP: fuels	Δln	Industrial Production Index - Fuels
NAPM prodn	lv	NAPM Production Index (Percent)
Cap util	Δlv	Capacity Utilization (Mfg) (TCB)
Employment and hours		
Help wanted index	lv	Index Of Help-Wanted Advertising In Newspapers (1967=100;Sa)
Help wan/emp	lv	Employment: Ratio; Help-Wanted Ads: No. Unemployed Clf
Emp CPS total	Δln	Civilian Labor Force: Employed, Total (Thous. Sa)
Emp CPS nonag	Δln	Civilian Labor Force: Employed, Nonagric. Industries (Thous.,Sa)
U: all	lv	Unemployment Rate: All Workers, 16 Years & Over (%,Sa)
U: mean duration	Δlv	Unemploy.By Duration: Average(Mean)Duration In Weeks (Sa)
U < 5 wks	Δln	Unemploy.By Duration: Persons Unempl.Less Than 5 Wks (Thous.,Sa)
U 5-14 wks	Δln	Unemploy.By Duration: Persons Unempl.5 To 14 Wks (Thous.,Sa)
U 15+ wks	Δln	Unemploy.By Duration: Persons Unempl.15 Wks + (Thous.,Sa)
U 15-26 wks	Δln	Unemploy.By Duration: Persons Unempl.15 To 26 Wks (Thous.,Sa)
U 27+ wks	Δln	Unemploy.By Duration: Persons Unempl.27 Wks + (Thous,Sa)
UI claims	Δln	Average Weekly Initial Claims, Unemploy. Insurance (Thous.) (TCB)
Emp: total	Δln	Employees On Nonfarm Payrolls: Total Private
Emp: gds prod	Δln	Employees On Nonfarm Payrolls-Goods-Producing
Emp: mining	Δln	Employees On Nonfarm PayrollsMining
Emp: const	Δln	Employees On Nonfarm PayrollsConstruction
Emp: mfg	Δln	Employees On Nonfarm PayrollsManufacturing
Emp: dble gds	Δln	Employees On Nonfarm PayrollsDurable Goods
Emp: nondbles	Δln	Employees On Nonfarm Payrolls-Nondurable Goods
Emp: services	Δln	Employees On Nonfarm Payrolls-Service-Providing
Emp: TTU	Δln	Employees On Nonfarm Payrolls-Trade, Transportation, And Utilities
Emp: wholesale	Δln	Employees On Nonfarm Payrolls-Wholesale Trade
Emp: retail	Δln	Employees On Nonfarm Payrolls-Retail Trade
Emp: FIRE	Δln	Employees On Nonfarm Payrolls-Financial Activities
Emp: Govt	Δln	Employees On Nonfarm PayrollsGovernment
Avg hrs	lv	Avg Weekly Hrs of Workers On Private Nonfarm Payrolls - Goods-Producing
Overtime: mfg	Δlv	Avg Weekly Hrs of Workers On Private Nonfarm Payrolls - Mfg Overtime
Avg hrs: mfg	lv	Average Weekly Hours, Mfg. (Hours) (TCB)
NAPM empl	lv	Napm Employment Index (Percent)
Sales		
Retail sales	Δln	Sales Of Retail Stores (Mil. Chain 2000$) (TCB)
M&T sales	Δln	Manufacturing And Trade Sales (Mil. Chain 1996$) (TCB)
Consumption		
Consumption	Δln	Real Consumption (AC) a0m224/gmdc (a0m224 is from TCB)
Housing starts and sales		
HStarts: Total	ln	Housing Starts:Nonfarm(1947-58);Total Farm&Nonfarm(1959-) (Thous.,Saar)
HStarts: NE	ln	Housing Starts:Northeast (Thous.U.)S.A.
HStarts: MW	ln	Housing Starts:Midwest(Thous.U.)S.A.
HStarts: South	ln	Housing Starts:South (Thous.U.)S.A.

Short Name	Transf.	Description
HStarts: West	ln	Housing Starts:West (Thous.U.)S.A.
BP: Total	ln	Housing Authorized: Total New Priv Housing Units (Thous.,Saar)
BP: NE	ln	Houses Authorized By Build. Permits:Northeast(Thou.U.)S.A
BP: MW	ln	Houses Authorized By Build. Permits:Midwest(Thou.U.)S.A.
BP: South	ln	Houses Authorized By Build. Permits:South(Thou.U.)S.A.
BP: West	ln	Houses Authorized By Build. Permits:West(Thou.U.)S.A.
Inventories		
NAPM invent	lv	Napm Inventories Index (Percent)
M&T invent	Δln	Manufacturing And Trade Inventories (Bil. Chain 2000$) (TCB)
M&T invent/sales	Δlv	Ratio, Mfg. And Trade Inventories To Sales (Based On Chain 2000$) (TCB)
Orders		
PMI	lv	Purchasing Managers Index (Sa)
NAPM new ordrs	lv	Napm New Orders Index (Percent)
NAPM vendor del	lv	Napm Vendor Deliveries Index (Percent)
Orders: cons gds	Δln	Mfrs New Orders, Consumer Goods And Materials (Bil. Chain 1982$) (TCB)
Orders: cap gds	Δln	Mfrs New Orders, Nondefense Capital Goods (Mil. Chain 1982$) (TCB)
Exchange rates		
Ex rate: avg	Δln	United States;Effective Exchange Rate(Merm)(Index No.)
Ex rate: Switz	Δln	Foreign Exchange Rate: Switzerland (Swiss Franc Per U.S.$)
Ex rate: Japan	Δln	Foreign Exchange Rate: Japan (Yen Per U.S.$)
Ex rate: UK	Δln	Foreign Exchange Rate: United Kingdom (Cents Per Pound)
Ex rate: Canada	Δln	Foreign Exchange Rate: Canada (Canadian$ Per U.S.$)
Money and credit quantity aggregates		
M1	Δln	Money Stock:M1(Curr,Trav.Cks,Dem Dep,Other Ckable Dep)(Bil$,Sa)
M2	Δln	Money Stock:M2M1+Onite Rps, Euro$,G/P&B/D Mmmfs&Sav&Sm Time Dep(Bil$,Sa)
M2 (real)	Δln	Money Supply:M2 In 1996 Dollars (Bci)
MB	Δln	Monetary Base, Adj For Reserve Requirement Changes(Mil$,Sa)
Reserves tot	Δln	Depository Inst Reserves:Total, Adj For Reserve Req Chgs(Mil$,Sa)
Reserves nonbor	Δln	Depository Inst Reserves:Nonborrowed,Adj Res Req Chgs(Mil$,Sa)
C&I loans	Δln	Commercial&Industrial Loans Oustanding In 1996 Dollars (Bci)
C&I loans	lv	Wkly Rp Lg Coml Banks:Net Change Coml&Indus Loans(Bil$,Saar)
Cons credit	Δln	Consumer Credit Outstanding-Nonrevolving(G19)
Inst cred/PI	Δlv	Ratio, Consumer Installment Credit To Personal Income (Pct.) (TCB)
Price indexes		
PPI: fin gds	Δln	Producer Price Index: Finished Goods (82=100,Sa)
PPI: cons gds	Δln	Producer Price Index: Finished Consumer Goods (82=100,Sa)
PPI: int matls	Δln	Producer Price Index:I ntermed Mat.Supplies & Components(82=100,Sa)
PPI: crude matls	Δln	Producer Price Index: Crude Materials (82=100,Sa)
Com: spot price	Δln	Spot market price index: bls&crb: all commodities(1967=100)
NAPM com price	lv	Napm Commodity Prices Index (Percent)
CPI-U: all	Δln	Cpi-U: All Items (82-84=100,Sa)
CPI-U: apparel	Δln	Cpi-U: Apparel&Upkeep (82-84=100,Sa)
CPI-U: transp	Δln	Cpi-U: Transportation (82-84=100,Sa)
CPI-U: medical	Δln	Cpi-U: Medical Care (82-84=100,Sa)
CPI-U: comm.	Δln	Cpi-U: Commodities (82-84=100,Sa)
CPI-U: dbles	Δln	Cpi-U: Durables (82-84=100,Sa)
CPI-U: services	Δln	Cpi-U: Services (82-84=100,Sa)
CPI-U: ex food	Δln	Cpi-U: All Items Less Food (82-84=100,Sa)
CPI-U: ex shelter	Δln	Cpi-U: All Items Less Shelter (82-84=100,Sa)
CPI-U: ex med	Δln	Cpi-U: All Items Less Midical Care (82-84=100,Sa)
PCE defl	Δln	Pce, Impl Pr Defl:Pce (1987=100)
PCE defl: dlbes	Δln	Pce, Impl Pr Defl:Pce; Durables (1987=100)
PCE defl: nondble	Δln	Pce, Impl Pr Defl:Pce; Nondurables (1996=100)
PCE defl: services	Δln	Pce, Impl Pr Defl:Pce; Services (1987=100)
Average hourly earnings		
AHE: goods	Δln	Avg Hourly Earnings of Workers On Private Nonfarm Payrolls - Goods-Producing
AHE: const	Δln	Avg Hourly Earnings of Workers On Private Nonfarm Payrolls Construction
AHE: mfg	Δln	Avg Hourly Earnings of Workers On Private Nonfarm Payrolls - Manufacturing
Miscellaneous		
Consumer expect	Δlv	U. Of Mich. Index Of Consumer Expectations(Bcd-83)

Note: This table lists the individual macro series used in factor extraction. The column headed 'Transf.' indicates the transformation applied to original series: lv = no transformation (levels are used), ln = logarithm of the level, Δlv = first differences of the level, and Δln = first differences of the log level.

Appendix C Alternative levels of risk aversion

Table C.1: Performance of active trading strategies for RRA=2

	μ	σ	SR	Δ_{100}	Δ_{50}	Δ_0	$\Delta_{\text{BM I}}$	$\Delta_{\text{BM II}}$	$\Delta_{\text{BM III}}$
Panel A: January 1980 - December 2005									
$w^*_{t+1} \in [-1, 2]$									
SFM	17.40	21.41	0.55	147	371	714	447	1146	228
TRS I	16.80	19.19	0.59	187	405	747	480	1178	262
TRS II	15.60	19.26	0.52	49	273	615	348	1046	129
TRS III	16.20	20.06	0.52	88	312	655	387	1086	169
BM I	12.96	21.41	0.35	−300	−75	268			
BM II	6.48	22.55	0.03	−1000	−775	−432			
BM III	15.84	23.04	0.45	−81	144	487			
$w^*_{t+1} \in [0, 1]$									
SFM	12.36	11.22	0.59	−16	207	549	151	435	100
TRS I	12.48	8.80	0.80	46	269	610	213	497	162
TRS II	11.76	9.01	0.69	−37	186	527	130	414	79
TRS III	11.52	9.84	0.59	−78	145	486	89	373	38
BM I	10.92	11.15	0.48	−167	56	398			
BM II	8.04	11.33	0.21	−451	−228	113			
BM III	11.88	13.16	0.48	−116	107	449			
Panel B: January 1980 - December 1992									
$w^*_{t+1} \in [-1, 2]$									
SFM	21.48	23.69	0.59	210	441	805	−124	868	364
TRS I	19.56	21.03	0.55	144	374	738	−190	801	298
TRS II	19.08	20.89	0.55	101	332	695	−232	758	255
TRS III	21.12	22.38	0.59	243	473	838	−91	901	397
BM I	22.56	23.52	0.62	334	565	929			
BM II	12.96	24.04	0.21	−659	−428	−63			
BM III	17.28	22.66	0.42	−154	76	441			
$w^*_{t+1} \in [0, 1]$									
SFM	15.12	12.75	0.59	−11	218	580	10	259	154
TRS I	14.88	9.25	0.80	48	277	638	68	317	213
TRS II	14.40	9.35	0.73	−2	227	588	18	267	163
TRS III	15.12	11.99	0.62	2	231	593	23	272	167
BM I	15.00	12.82	0.59	−20	209	571			
BM II	12.24	11.78	0.38	−270	−41	321			
BM III	13.68	13.27	0.45	−165	64	426			
Panel C: January 1993 - December 2005									
$w^*_{t+1} \in [-1, 2]$									
SFM	13.44	18.78	0.52	82	300	622	1004	1418	93
TRS I	14.04	17.11	0.62	214	432	754	1136	1549	226
TRS II	12.00	17.39	0.48	−7	211	533	915	1329	5
TRS III	12.72	18.46	0.48	24	242	564	947	1360	36
BM I	3.24	18.67	0.00	−923	−704	−383			
BM II	0.00	20.78	−	−1338	−1120	−797			
BM III	14.40	23.38	0.45	−12	207	530			
$w^*_{t+1} \in [0, 1]$									
SFM	9.60	9.39	0.66	−22	195	516	291	610	46
TRS I	10.08	8.24	0.80	43	260	581	356	675	112
TRS II	9.00	8.56	0.62	−72	145	465	240	559	4
TRS III	9.00	9.53	0.55	−95	122	442	217	536	−27
BM I	6.72	9.04	0.35	−313	−96	225			
BM II	3.84	10.77	0.03	−632	−415	−94			
BM III	10.08	13.09	0.48	−69	149	470			

Note: The table displays performance measures for the active mean-variance investment strategies based on monthly excess return and volatility predictions for the period January 1980 - December 1992 with relative risk aversion γ in (16) set equal to 2. Results are shown for medium transaction costs (0.1%). See Table 3 for further details.

Table C.2: Performance of active trading strategies for RRA=10

	μ	σ	SR	Δ_{100}	Δ_{50}	Δ_0	$\Delta_{\text{BM I}}$	$\Delta_{\text{BM II}}$	$\Delta_{\text{BM III}}$
Panel A: January 1980 - December 2005									
$w_{t+1}^* \in [-1, 2]$									
SFM	10.44	11.85	0.42	168	−382	−280	68	458	−282
TRS I	11.40	13.20	0.45	79	−481	−377	−23	373	−379
TRS II	10.56	12.78	0.38	61	−495	−392	−40	354	−394
TRS III	11.76	11.99	0.52	287	−264	−162	187	577	−164
BM I	9.24	11.43	0.31	99	−448	−346			
BM II	7.56	13.13	0.14	−294	−853	−750			
BM III	10.08	9.08	0.48	437	−97	2			
$w_{t+1}^* \in [0, 1]$									
SFM	9.84	7.10	0.59	582	56	153	105	243	79
TRS I	9.72	6.65	0.62	603	79	176	127	266	101
TRS II	9.48	6.58	0.59	588	64	160	112	251	86
TRS III	9.96	6.86	0.62	612	87	184	136	274	110
BM I	9.00	7.41	0.45	478	−49	48			
BM II	7.80	7.59	0.28	339	−188	−91			
BM III	9.36	7.52	0.48	505	−23	75			
Panel B: January 1980 - December 1992									
$w_{t+1}^* \in [-1, 2]$									
SFM	14.28	14.27	0.45	92	−581	−487	−154	498	−586
TRS I	13.92	16.28	0.38	−293	−990	−892	−547	126	−995
TRS II	13.44	15.76	0.38	−243	−934	−837	−495	172	−939
TRS III	14.52	13.99	0.48	163	−507	−412	−81	568	−511
BM I	15.12	13.89	0.55	244	−425	−331			
BM II	11.16	15.45	0.24	−414	−1101	−1004			
BM III	11.40	7.27	0.52	627	5	92			
$w_{t+1}^* \in [0, 1]$									
SFM	12.60	8.28	0.59	668	41	129	−18	170	−4
TRS I	11.64	7.69	0.52	617	−7	81	−66	121	−51
TRS II	11.52	7.59	0.52	611	−13	74	−72	115	−58
TRS III	12.12	7.79	0.59	658	33	121	−26	162	−11
BM I	12.96	8.42	0.62	687	59	147			
BM II	11.04	8.31	0.38	498	−129	−41			
BM III	11.40	6.69	0.55	664	44	131			
Panel C: January 1993 - December 2005									
$w_{t+1}^* \in [-1, 2]$									
SFM	6.60	8.63	0.35	238	−200	−92	272	420	−7
TRS I	8.76	9.08	0.59	420	−19	89	454	603	174
TRS II	7.80	8.80	0.48	340	−98	11	375	524	95
TRS III	9.00	9.49	0.55	400	−41	68	434	584	153
BM I	3.24	7.93	−	−34	−469	−362			
BM II	3.84	10.18	0.03	−185	−629	−519			
BM III	8.64	10.57	0.48	−185	−629	−519			
$w_{t+1}^* \in [0, 1]$									
SFM	6.96	5.54	0.59	498	69	175	220	313	156
TRS I	7.80	5.40	0.76	588	159	265	311	403	247
TRS II	7.44	5.27	0.73	564	136	242	287	380	223
TRS III	7.68	5.72	0.73	566	137	243	289	381	224
BM I	5.04	6.03	0.24	278	−152	−46			
BM II	4.44	6.62	0.14	186	−246	−139			
BM III	7.32	8.24	0.45	347	−89	19			

Note: The table displays performance measures for the active mean-variance investment strategies based on monthly excess return and volatility predictions for the period January 1980 - December 1992 with relative risk aversion γ in (16) set equal to 10. Results are shown for medium transaction costs (0.1%). See Table 3 for further details.

45

References

Aiolfi, M. and C. A. Favero (2005), Model Uncertainty, Thick Modeling and The Predictability of Stock Returns, *Journal of Forecasting*, 24, 233–254.

Akgiray, V. (1989), Conditional Heteroscedasticity in Time Series of Stock Returns: Evidence and Forecasts, *Journal of Business*, 62, 55–80.

Andersen, T. G., T. Bollerslev, P. F. Christoffersen, and F. X. Diebold (2006), Volatility and Correlation Forecasting, in G. Elliot, C. W. J. Granger, and A. Timmermann (eds.), *Handbook of Economic Forecasting*, Amsterdam: North-Holland, pp. 777–878.

Andrews, D. W. K. (1991), Heteroskedasticity and Autocorrelation Consistent Covariance Matrix Estimator, *Econometrica*, 60, 953–966.

Ang, A. and G. Bekaert (2007), Stock Return Predictability: Is it There?, *The Review of Financial Studies*, 20, 651–707.

Avramov, D. (2002), Stock Return Predictability and Model Uncertainty, *Journal of Financial Economics*, 64, 423–458.

Bai, J. (2010), Equity Premium Predictions with Adaptive Macro Indexes, Federal Reserve Bank of New York Staff Reports No. 475.

Bai, J. and S. Ng (2008), Forecasting Economic Time Series Using Targeted Predictors, *Journal of Econometrics*, 146, 304–317.

Bair, E., T. Hastie, D. Paul, and R. Tibshirani (2006), Prediction by Supervised Principal Components, *Journal of the American Statistical Association*, 101, 119–137.

Boivin, J. and S. Ng (2006), Are More Data Always Better for Factor Analysis?, *Journal of Econometrics*, 132, 169–194.

Bossaerts, P. and P. Hillion (1999), Implementing Statistical Criteria to Select Return Forecasting Models: What Do We Learn?, *Review of Financial Studies*, 12, 405–428.

Breen, W., L. R. Glosten, and R. Jagannathan (1989), Economic Significance of Predictable Variations in Stock Returns, *Journal of Finance*, 44, 1177–1189.

Campbell, J. Y. and R. J. Shiller (1988), The Dividend-Price Ratio and Expectations of Future Dividends and Discount Factors, *Review of Financial Studies*, 1, 195–228.

Campbell, J. Y. and S. B. Thompson (2008), Predicting Excess Stock Returns Out of Sample: Can Anything Beat the Historical Average?, *Review of Financial Studies*, 21, 1509–1531.

Campbell, S. D. and F. X. Diebold (2009), Stock Returns and Expected Business Conditions: Half a Century of Direct Evidence, *Journal of Business & Economic Statistics*, 27, 266–278.

Cremers, K. J. M. (2002), Stock Return Predictability: A Bayesian Model Selection Perspective, *Review of Financial Studies*, 15, 1223–1249.

De Pooter, M., M. Martens, and D. van Dijk (2008), Predicting the Daily Covariance Matrix for S&P 100 Stocks Using Intraday Data - But Which Frequency to Use?, *Econometric Reviews*, 27, 199–229.

Della Corte, P., L. Sarno, and I. Tsiakas (2009), An Economic Evaluation of Empirical Exchange Rate Models, *Review of Financial Studies*, 22, 3491–3530.

Della Corte, P., L. Sarno, and I. Tsiakas (2010), A Century of Equity Premium Predictability and the ConsumptionWealth Ratio: An International Perspective, *Journal of Empirical Finance*, 17, 313–331.

Efron, B., T. Hastie, L. Johnstone, and R. Tibshirani (2004), Least Angle Regression, *Annals of Statistics*, 32, 407–499.

Fama, E. F. and K. R. French (1988), Dividend Yields and Expected Stock Returns, *Journal of Financial Economics*, 22, 3–25.

Ferson, W. and C. R. Harvey (1991), The Variation of Economic Risk Premiums, *Journal of Political Economy*, 99, 385–415.

Fleming, J., C. Kirby, and B. Ostdiek (2001), The Economic Value of Volatility Timing, *Journal of Finance*, 56, 329–352.

Fleming, J., C. Kirby, and B. Ostdiek (2003), The Economic Value of Volatility Timing Using "Realized" Volatility, *Journal of Financial Economics*, 67, 473–509.

Forni, M., M. Hallin, M. Lippi, and L. Reichlin (2003), Do Financial Variables Help Forecasting Inflation and Real Activity in the Euro Area?, *Journal of Monetary Economics*, 50, 1243–1255.

French, K. R., G. W. Schwert, and R. F. Stambaugh (1987), Expected Stock Returns and Volatility, *Journal of Financial Economics*, 19, 3–29.

Gomes, J. F., L. Kogan, and M. Yogo (2009), Durability of Output and Expected Stock Returns, *Journal of Political Economy*, 117, 941–986.

Hamilton, J. D. and G. Lin (1996), Stock Market Volatility and the Business Cycle, *Journal of Applied Econometrics*, 11, 573–593.

Huang, H. and T.-H. Lee (2010), To Combine Forecasts or to Combine Information?, *Econometric Reviews*, 29, 534–570.

47

Keim, D. B. and R. F. Stambaugh (1986), Predicting Returns in the Stock and Bond Markets, *Journal of Financial Economics*, 17, 357–390.

Leitch, G. and J. E. Tanner (1991), Economic Forecast Evaluation: Profits versus the Conventional Error Measures, *American Economic Review*, 81, 580–590.

Lettau, M. and S. C. Ludvigson (2001), Consumption, Aggregate Wealth, and Expected Stock Returns, *Journal of Finance*, 56, 815–849.

Lettau, M. and S. Van Nieuwerburgh (2008), Reconciling the Return Predictability Evidence, *Review of Financial Studies*, 21, 1607–1652.

Lo, A. W. and A. C. MacKinlay (1990), Data-Snooping Biases in Tests of Financial Asset Pricing Models, *Review of Financial Studies*, 3, 431–467.

Ludvigson, S. C. and S. Ng (2007), The Empirical Risk-Return Relation: A Factor Analysis Approach, *Journal of Financial Economics*, 83, 171–222.

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

Marquering, W. and M. Verbeek (2004a), The Economic Value of Predicting Stock Index Returns and Volatility, *Journal of Financial and Quantitative Analysis*, 39, 407–429.

Marquering, W. and M. Verbeek (2004b), A Multivariate Nonparametric Test for Return and Volatility Timing, *Finance Research Letters*, 1, 250–260.

Mönch, E. (2008), Forecasting the Yield Curve in a Data-Rich Environment: A No-arbitrage Factor-augmented VAR Approach, *Journal of Econometrics*, 146, 26–43.

Paye, B. S. and A. Timmermann (2006), Instability of Return Prediction Models, *Journal of Empirical Finance*, 13, 274–315.

Pesaran, M. H. and A. Timmermann (1992), A Simple Nonparametric Test of Predictive Performance, *Journal of Business & Economic Statistics*, 10, 561–565.

Pesaran, M. H. and A. Timmermann (1995), Predictability of Stock Returns: Robustness and Economic Significance, *Journal of Finance*, 50, 1201–1228.

Pesaran, M. H. and A. Timmermann (2002), Market Timing and Return Prediction under Model Instability, *Journal of Empirical Finance*, 9, 495–510.

Pesaran, M. H. and A. Timmermann (2004), How Costly is it to Ignore Breaks When Forecasting the Direction of a Time Series?, *International Journal of Forecasting*, 20, 411–425.

Pesaran, M. H. and A. Timmermann (2007), Selection of Estimation Window in the Presence of Breaks, *Journal of Econometrics*, 137, 134–161.

Pesaran, M. H. and A. Timmermann (2009), Testing Dependence Among Serially Correlated Multicategory Variables, *Journal of the American Statistical Association*, 104, 325–337.

Pettenuzzo, D. and A. Timmermann (forthcoming), Predictability of Stock Returns and Asset Allocation under Structural Breaks, *Journal of Econometrics*.

Piazzesi, M., M. Schneider, and S. Tuzel (2007), Housing, Consumption and Asset Pricing, *Journal of Financial Economics*, 83, 531–569.

Rapach, D. E., J. K. Strauss, and G. Zhou (2010), Out-of-Sample Equity Premium Prediction: Combination Forecasts and Links to the Real Economy, *Review of Financial Studies*, 23, 821–862.

Rapach, D. E. and M. E. Wohar (2006), Structural Breaks and Predictive Regression Models of Aggregate U.S. Stock Returns, *Journal of Financial Econometrics*, 4, 238–274.

Ravazzolo, F., D. van Dijk, R. Paap, and P. H. Franses (2008), Bayesian Model Averaging in the Presence of Structural Breaks, in M. Wohar and D. E. Rapach (eds.), *Frontiers of Economics and Globalization Series Vol.3*, Emerald Group Publishing, pp. 561–594.

Schwert, G. W. (1989), Why Does Stock Market Volatility Change over Time?, *Journal of Finance*, 44, 1115–1153.

Stock, J. H. and M. W. Watson (2002a), Forecasting Using Principal Components From a Large Number of Predictors, *Journal of the American Statistical Association*, 97, 1167–1179.

Stock, J. H. and M. W. Watson (2002b), Macroeconomic Forecasting Using Diffusion Indexes, *Journal of Business & Economic Statistics*, 20, 147–162.

Stock, J. H. and M. W. Watson (2005), Implications of Dynamic Factor Models for VAR Analysis, NBER Working Paper No. 11467.

Timmermann, A. (2008), Elusive Return Predictability, *International Journal of Forecasting*, 24, 1–18.

Wachter, J. A. and M. Warusawitharana (2009), Predictable Returns and Asset Allocation: Should a Skeptical Investor Time the Market?, *Journal of Econometrics*, 148, 162–178.

Welch, I. and A. Goyal (2008), A Comprehensive Look at The Empirical Performance of Equity Premium Prediction, *Review of Financial Studies*, 21, 1455–1508.

West, K. D., H. J. Edison, and D. Cho (1993), A Utility-Based Comparison of Some Models of Exchange Rate Volatility, *Journal of International Economics*, 35, 23–45.

Tables and figures

Table 1: Market timing and volatility timing

	SFM	TRS I	TRS II	TRS III	BM I	BM II	BM III
Panel A: Market timing							
Jan 1980 - Dec 2005	0.625	0.587	0.574	0.596	0.580	0.526	0.592
Jan 1980 - Dec 1992	0.622	0.558	0.558	0.590	0.628	0.532	0.577
Jan 1993 - Dec 2005	0.628	0.615	0.590	0.603	0.532	0.519	0.609
Panel B: Volatility timing							
Jan 1980 - Dec 2005	0.721	0.724	0.728	0.728	0.712	0.718	0.731
Jan 1980 - Dec 1992	0.622	0.635	0.641	0.615	0.628	0.647	0.628
Jan 1993 - Dec 2005	0.821	0.814	0.814	0.840	0.795	0.788	0.833
Panel C: Joint market timing and volatility timing							
Jan 1980 - Dec 2005	0.455	0.429	0.420	0.446	0.407	0.385	0.442
Jan 1980 - Dec 1992	0.397	0.346	0.359	0.391	0.397	0.346	0.359
Jan 1993 - Dec 2005	0.513	0.513	0.481	0.500	0.417	0.423	0.526

Note: Panel A (market timing) shows hit ratios for the sign of the monthly excess stock return. Panel B (volatility timing) shows hit ratios for the sign of monthly volatility relative to the median volatility level during the 10-year rolling window that is used for model specification and estimation. Panel C (joint market timing and volatility timing) shows hit ratios for the signs of the monthly excess stock return and volatility. SFM denotes the 'standard' factor-augmented predictive regressions where factors are constructed using all available macroeconomic variables. TRS I and TRS II stand for the factor-augmented predictive regressions where factors are constructed after employing hard thresholding rules with thresholds $|t| = 1.28$ and 1.04, respectively. TRS III stands for the factor-augmented predictive regression where the factors are constructed using the first 60 predictors included in the soft thresholding procedure. BM I denotes the benchmark model that uses only financial variables z_t, BM II denotes the model that uses the three individual macroeconomic predictors with the highest t statistics in the regression (4). BM III denotes the benchmark model where the historical average is used as the excess return forecast and a random walk forecast is used for volatility.

Table 2: Tests for market timing and volatility timing

	SFM	TRS I	TRS II	TRS III	BM I	BM II	BM III
Panel A: Market timing							
PT							
Jan 1980 - Dec 2005	3.46***	2.73***	2.14**	2.66***	2.00**	0.63	0.12
Jan 1980 - Dec 1992	2.50***	1.62*	1.49*	1.86**	2.73***	0.91	0.05
Jan 1993 - Dec 2005	2.38***	2.24**	1.50*	1.89**	0.11	−0.08	NA
BGJ-HAC							
Jan 1980 - Dec 2005	3.12***	2.84***	2.08**	2.70***	1.82**	0.62	0.10
Jan 1980 - Dec 1992	2.47***	1.74**	1.55*	1.94**	2.42***	0.88	0.00
Jan 1993 - Dec 2005	2.05**	2.17**	1.30*	1.84**	0.10	0.10	NA
BH-HAC							
Jan 1980 - Dec 2005	2.60***	2.40***	2.24**	3.24***	2.02**	1.70**	−0.81
Jan 1980 - Dec 1992	2.76***	2.27**	2.05**	2.88***	3.07***	1.99**	−0.56
Jan 1993 - Dec 2005	0.60	0.95	1.00	1.47*	−0.76	0.12	NA
Panel B: Volatility timing							
PT							
Jan 1980 - Dec 2005	8.11***	8.16***	8.32***	8.32***	7.80***	7.98***	8.17***
Jan 1980 - Dec 1992	3.40***	3.71***	3.98***	3.25***	3.54***	3.95***	3.19***
Jan 1993 - Dec 2005	8.11***	7.83***	7.91***	8.56***	7.48***	7.29***	8.34***
BGJ-HAC							
Jan 1980 - Dec 2005	4.65***	4.67***	4.70***	4.65***	4.48***	4.67***	4.79***
Jan 1980 - Dec 1992	2.53***	2.72***	2.82***	2.39***	2.87***	2.87***	2.24**
Jan 1993 - Dec 2005	4.17***	4.15***	4.10***	4.33***	3.94***	3.95***	4.61***
BH-HAC							
Jan 1980 - Dec 2005	3.47***	3.22***	3.27***	3.38***	4.13***	3.07***	2.34***
Jan 1980 - Dec 1992	1.69**	1.59*	1.50*	1.56*	1.71**	1.62*	4.80***
Jan 1993 - Dec 2005	5.64***	5.26***	5.48***	5.69***	4.60***	4.62***	11.84***
Panel C: Joint market and volatility timing							
HM							
Jan 1980 - Dec 2005	81.91***	75.75***	76.97***	81.36***	70.46***	66.53***	84.43***
Jan 1980 - Dec 1992	27.08***	20.75**	24.00***	27.78***	37.71***	22.87***	28.20***
Jan 1993 - Dec 2005	75.43***	70.36***	68.19***	77.75***	60.61***	62.45***	NA
Trace CC							
Jan 1980 - Dec 2005	30.28***	26.64***	26.77***	27.63***	25.48***	23.05***	33.77***
Jan 1980 - Dec 1992	18.04**	11.31	14.09	22.74***	21.77***	11.72	NA
Jan 1993 - Dec 2005	21.16**	22.06***	19.16**	19.82**	17.32**	20.26**	NA

Note: PT is the test statistic for market (volatility) timing ability as defined in (6). BGJ-HAC is the HAC *t*-statistic of the slope coefficient in (7). BH-HAC is the HAC *t*-statistic of the slope coefficient in (11). HM is the generalized Henriksson-Merton statistic for joint market and volatility timing in (8). The Trace CC statistic of Pesaran and Timmermann (2009) uses the sum of canonical correlations for testing market and volatility timing jointly. NA indicates that the test could not be computed. *, **, *** indicate significance at the 10%, 5% and 1% level, respectively. Model abbreviations are as in Table 1.

Table 3: Performance of active trading strategies: Jan 1980 - Dec 2005, RRA=6

	μ	σ	SR	Δ_{100}	Δ_{50}	Δ_0	$\Delta_{\text{BM I}}$	$\Delta_{\text{BM II}}$	$\Delta_{\text{BM III}}$
Passive Strategies									
I. 100% Market	13.56	15.10	0.52						
II. 50% Market	9.60	7.55	0.52						
III. 0% Market	5.64	0.90	−						
Active Strategies									
Panel A: No transaction costs									
$w_{t+1}^* \in [-1, 2]$									
SFM	13.68	14.79	0.55	41	−109	120	200	642	−72
TRS I	13.92	15.38	0.55	12	−138	91	171	615	−101
TRS II	13.32	15.00	0.52	−15	−164	65	145	587	−127
TRS III	14.88	14.48	0.62	191	41	270	349	791	78
BM I	11.64	14.65	0.42	−159	−308	−80			
BM II	8.76	16.32	0.21	−606	−757	−526			
BM III	13.08	13.30	0.55	112	−37	190			
$w_{t+1}^* \in [0, 1]$									
SFM	11.52	8.69	0.69	280	134	358	119	314	136
TRS I	11.16	7.38	0.76	311	165	388	149	344	167
TRS II	10.92	7.48	0.69	277	132	354	116	311	133
TRS III	11.40	8.31	0.69	287	141	364	125	320	142
BM I	10.44	8.83	0.55	162	16	239			
BM II	8.64	9.18	0.35	−33	−180	44			
BM III	10.92	9.98	0.52	145	−1	223			
Panel B: Medium transaction costs (0.1%)									
$w_{t+1}^* \in [-1, 2]$									
SFM	12.48	14.79	0.45	−81	−231	−2	182	673	−94
TRS I	12.72	15.38	0.45	−108	−258	−28	156	649	−121
TRS II	12.12	14.96	0.45	−133	−282	−53	131	622	−145
TRS III	13.56	14.45	0.55	63	−86	142	326	816	50
BM I	10.56	14.69	0.35	−263	−413	−184			
BM II	7.20	16.35	0.10	−761	−911	−681			
BM III	12.12	13.30	0.48	13	−136	91			
$w_{t+1}^* \in [0, 1]$									
SFM	11.04	8.69	0.62	232	86	310	109	325	138
TRS I	10.68	7.34	0.69	265	119	342	142	358	170
TRS II	10.44	7.48	0.66	232	86	309	109	324	137
TRS III	10.92	8.31	0.62	238	92	315	114	330	143
BM I	10.08	8.83	0.48	124	−22	201			
BM II	8.04	9.18	0.28	−93	−239	−16			
BM III	10.44	9.98	0.48	95	−51	173			
Panel C: High transaction costs (0.3%)									
$w_{t+1}^* \in [-1, 2]$									
SFM	10.08	14.86	0.31	−327	−477	−248	147	739	−138
Trsy I	10.32	15.42	0.31	−351	−501	−272	124	717	−162
TRS II	9.72	15.00	0.28	−372	−522	−293	102	694	−183
TRS III	9.24	14.72	0.24	−395	−545	−316	79	670	−7
BM I	8.52	14.72	0.21	−474	−623	−395			
BM II	4.20	16.45	−	−1075	−1226	−995			
BM III	10.08	13.34	0.35	−188	−336	−109			
$w_{t+1}^* \in [0, 1]$									
SFM	10.08	8.73	0.52	136	−10	213	89	349	178
TRS I	9.84	7.38	0.55	173	27	250	126	385	141
TRS II	9.48	7.48	0.52	140	−5	217	93	352	145
TRS III	9.96	8.35	0.52	139	−7	216	92	352	144
BM I	9.36	8.87	0.42	47	−99	124			
BM II	6.96	9.25	0.14	−213	−360	−136			
BM III	9.48	9.98	0.38	−5	−152	−73			

Note: The table displays performance measures for the active mean-variance investment strategies based on monthly excess return and volatility predictions for the period January 1980 - December 2005. Columns headed μ and σ contain the annualized mean and standard deviation of the portfolio returns (in percent). SR denotes the Sharpe ratio. Columns headed Δ contain performance fees (in annualized basis points) for switching from the strategy indicated by the subscript to the strategy indicated by the corresponding row. The subscript x with $x =$100, 50, and 0 refer to the static strategies with x% of wealth invested in stocks. Model abbreviations are as in Table 1.

Table 4: Mean and standard deviation of portfolio weights

	SFM	TRS I	TRS II	TRS III	BM I	BM II	BM III
Panel A: $w_{t+1}^* \in [-1, 2]$				Mean			
Jan 1980 - Dec 2005	0.50	0.32	0.34	0.46	0.51	0.33	0.90
Jan 1980 - Dec 1992	0.42	0.08	0.13	0.34	0.46	0.22	0.69
Jan 1993 - Dec 2005	0.58	0.56	0.56	0.59	0.56	0.44	1.11
				Standard Deviation			
Jan 1980 - Dec 2005	0.99	1.12	1.10	1.04	0.99	1.14	0.68
Jan 1980 - Dec 1992	0.98	1.15	1.13	0.99	0.96	1.21	0.62
Jan 1993 - Dec 2005	1.00	1.03	1.02	1.07	1.01	1.05	0.67
Panel B: $w_{t+1}^* \in [0, 1]$				Mean			
Jan 1980 - Dec 2005	0.47	0.39	0.41	0.45	0.47	0.43	0.66
Jan 1980 - Dec 1992	0.45	0.33	0.35	0.41	0.46	0.41	0.41
Jan 1993 - Dec 2005	0.49	0.46	0.48	0.50	0.47	0.46	0.50
				Standard Deviation			
Jan 1980 - Dec 2005	0.44	0.44	0.44	0.44	0.45	0.45	0.35
Jan 1980 - Dec 1992	0.43	0.43	0.43	0.43	0.44	0.46	0.43
Jan 1993 - Dec 2005	0.44	0.44	0.45	0.46	0.45	0.44	0.45

Note: The table displays means and standard deviations of the portfolio weight for stocks in the active mean-variance investment strategies based on monthly predictions of excess returns and volatility. Model abbreviations are as in Table 1.

Table 5: Performance of active trading strategies: Jan 1980 - Dec 1992, RRA=6

	μ	σ	SR	Δ_{100}	Δ_{50}	Δ_0	$\Delta_{\text{BM I}}$	$\Delta_{\text{BM II}}$	$\Delta_{\text{BM III}}$
Passive Strategies									
I. 100% Market	16.20	15.93	0.52						
II. 50% Market	11.88	7.93	0.52						
III. 0% Market	7.68	0.83	–						
Active Strategies									
Panel A: No transaction costs									
$w_{t+1}^* \in [-1, 2]$									
SFM	17.40	17.01	0.59	14	−182	57	−176	602	−239
TRS I	16.32	18.19	0.48	−246	−444	−203	−437	347	−501
TRS II	15.84	17.60	0.45	−215	−412	−172	−405	376	−467
TRS III	17.16	16.28	0.59	65	−130	107	−124	651	−187
BM I	18.84	16.66	0.66	189	−6	232			
BM II	13.20	18.57	0.31	−594	−792	−551			
BM III	14.76	11.54	0.62	246	55	288			
$w_{t+1}^* \in [0, 1]$									
SFM	14.64	9.94	0.69	348	158	389	−22	209	152
TRS I	13.08	8.24	0.66	290	102	332	−77	152	95
TRS II	12.84	8.28	0.62	268	79	309	−100	130	73
TRS III	14.16	9.32	0.69	341	152	382	−28	202	145
BM I	15.00	10.12	0.73	370	180	411			
BM II	12.48	9.77	0.48	139	−50	180			
BM III	12.84	9.42	0.55	196	7	237			
Panel B: Medium transaction costs (0.1%)									
$w_{t+1}^* \in [-1, 2]$									
SFM	16.20	17.01	0.48	−112	−308	−69	−165	614	−271
TRS I	15.00	18.19	0.42	−373	−571	−330	−427	358	−534
TRS II	14.64	17.56	0.38	−332	−528	−289	−385	397	−492
TRS III	15.96	16.32	0.52	−63	−258	−20	−116	660	−221
BM I	17.52	16.70	0.59	53	−142	96			
BM II	11.88	18.57	0.21	−733	−931	−690			
BM III	13.80	11.50	0.52	155	−36	197			
$w_{t+1}^* \in [0, 1]$									
SFM	14.16	9.94	0.66	297	108	339	−18	209	151
TRS I	12.60	8.24	0.59	244	55	285	−70	156	98
TRS II	12.48	8.24	0.59	224	35	265	−90	136	78
TRS III	13.68	9.35	0.66	288	99	330	−27	200	142
BM I	14.52	10.12	0.66	316	126	357			
BM II	12.00	9.77	0.45	88	−101	130			
BM III	12.36	9.42	0.48	146	−43	188			
Panel C: High transaction costs (0.3%)									
$w_{t+1}^* \in [-1, 2]$									
SFM	13.80	17.11	0.35	−367	−563	−324	−145	641	−337
TRS I	12.48	18.26	0.28	−633	−831	−590	−409	382	−603
TRS II	12.24	17.56	0.28	−570	−767	−527	−347	440	−541
TRS III	13.44	16.39	0.35	−321	−516	−278	−100	683	−292
BM I	14.88	16.80	0.42	−222	−417	−179			
BM II	9.12	18.64	0.07	−1017	−1215	−974			
BM III	12.00	11.50	0.38	−29	−219	13			
$w_{t+1}^* \in [0, 1]$									
SFM	13.20	10.01	0.55	196	6	237	−12	209	148
TRS I	11.76	8.24	0.48	150	−38	192	−56	164	103
TRS II	11.52	8.28	0.48	135	−53	176	−71	149	88
TRS III	12.72	9.39	0.52	182	−7	223	−25	196	135
BM I	13.44	10.18	0.55	207	18	249			
BM II	10.92	9.80	0.35	−14	−203	28			
BM III	11.28	9.42	0.38	47	−142	89			

Note: The table displays performance measures for the active mean-variance investment strategies based on monthly excess return and volatility predictions for the period January 1980 - December 1992. See Table 3 for further details.

Table 6: Performance of active trading strategies: Jan 1993 - Dec 2005, RRA=6

|	μ	σ	SR	Δ_{100}	Δ_{50}	Δ_{0}	$\Delta_{\text{BM I}}$	$\Delta_{\text{BM II}}$	$\Delta_{\text{BM III}}$
Passive Strategies									
I. 100% Market	11.04	14.24	0.52						
II. 50% Market	7.32	7.14	0.52						
III. 0% Market	3.60	0.45	−						
Active Strategies									
Panel A: No transaction costs									
$w^*_{t+1} \in [-1, 2]$									
SFM	9.96	12.12	0.52	67	−39	180	554	680	87
TRS I	11.64	11.85	0.69	260	155	374	747	873	281
TRS II	10.80	11.78	0.62	178	73	292	665	791	199
TRS III	12.48	12.26	0.73	313	207	427	800	927	140
BM I	4.32	11.95	0.07	−487	−592	−373			
BM II	4.32	13.61	0.07	−617	−723	−503			
BM III	11.40	14.86	0.52	−21	−127	94			
$w^*_{t+1} \in [0, 1]$									
SFM	8.40	7.14	0.69	215	111	327	253	415	121
TRS I	9.24	6.34	0.90	331	227	443	369	531	237
TRS II	8.88	6.55	0.80	286	183	398	325	486	192
TRS III	8.64	7.07	0.69	234	130	346	272	434	333
BM I	5.88	7.14	0.31	−38	−142	74			
BM II	4.92	8.38	0.14	−201	−305	−88			
BM III	9.12	10.46	0.52	−201	−9	208			
Panel B: Medium transaction costs (0.1%)									
$w^*_{t+1} \in [-1, 2]$									
SFM	8.76	12.09	0.42	−53	−158	61	512	731	76
TRS I	10.56	11.85	0.59	145	39	258	709	928	273
TRS II	9.60	11.78	0.52	56	−49	170	620	840	185
TRS III	11.28	12.26	0.62	182	76	296	746	967	310
BM I	3.48	11.95	0.00	−565	−670	−451			
BM II	2.64	13.65	−	−790	−896	−675			
BM III	10.32	14.86	0.45	−130	−237	−15			
$w^*_{t+1} \in [0, 1]$									
SFM	7.92	7.10	0.62	168	64	280	231	438	125
TRS I	8.76	6.30	0.83	284	181	396	347	553	241
TRS II	8.40	6.55	0.73	238	134	350	301	507	195
TRS III	8.16	7.03	0.62	187	83	299	250	456	144
BM I	5.64	7.14	0.28	−63	−167	49			
BM II	4.20	8.42	0.07	−270	−374	−158			
BM III	8.52	10.46	0.48	43	−61	157			
Panel C: High transaction costs (0.3%)									
$w^*_{t+1} \in [-1, 2]$									
SFM	6.36	12.09	0.21	−295	−400	−181	427	836	52
TRS I	8.16	11.85	0.38	−89	−194	25	632	1040	257
TRS II	7.20	11.81	0.31	−191	−296	−77	530	938	156
TRS III	8.64	12.33	0.42	−84	−189	30	638	1047	263
BM I	1.92	11.99	−	−721	−827	−608			
BM II	0.72	13.79	−	−1138	−1244	−1023			
BM III	8.16	14.93	0.31	−351	−458	−236			
$w^*_{t+1} \in [0, 1]$									
SFM	6.96	7.10	0.48	73	−31	185	186	483	134
TRS I	7.80	6.30	0.66	190	86	301	302	599	250
TRS II	7.44	6.55	0.59	140	37	252	253	550	201
TRS III	7.20	7.03	0.52	92	−12	204	205	502	153
BM I	5.16	7.10	0.21	−113	−217	−1			
BM II	2.88	8.49	−	−411	−515	−299			
BM III	7.56	10.50	0.38	−61	−166	52			

Note: The table displays performance measures for the active mean-variance investment strategies based on monthly excess return and volatility predictions for the period January 1993 - December 2005. See Table 3 for further details.

Table 7: Performance of active trading strategies based on excess return forecasts

|	μ	σ	SR	Δ_{100}	Δ_{50}	Δ_0	$\Delta_{\text{BM I}}$	$\Delta_{\text{BM II}}$	$\Delta_{\text{BM III}}$
Panel A: January 1980 - December 2005									
$w_{t+1}^* \in [-1, 2]$									
SFM	14.28	18.05	0.48	−252	−405	−172	353	1191	−45
TRS I	13.56	17.67	0.45	−271	−423	−191	332	1169	−64
TRS II	13.08	17.60	0.42	−310	−462	−230	293	1129	−103
TRS III	14.28	17.60	0.48	−200	−352	−120	403	1239	7
BM I	10.80	18.15	0.28	−605	−758	−525			
BM II	4.92	20.30	−	−1464	−1619	−1382			
BM III	12.24	15.80	0.42	−205	−355	−125			
$w_{t+1}^* \in [0, 1]$									
SFM	12.00	10.18	0.62	232	86	310	159	516	231
TRS I	11.52	8.49	0.69	287	141	364	213	568	286
TRS II	10.92	8.38	0.62	235	90	313	162	517	235
TRS III	11.52	9.80	0.59	245	98	323	403	1239	207
BM I	10.44	10.25	0.45	74	−73	151			
BM II	7.32	10.95	0.14	−284	−431	−206			
BM III	10.80	11.81	0.45	1	−147	79			
Panel B: January 1980 - December 1992									
$w_{t+1}^* \in [-1, 2]$									
SFM	18.84	19.85	0.55	−204	−404	−160	−131	1044	−198
TRS I	16.68	19.71	0.45	−397	−596	−353	−324	850	−391
TRS II	16.68	19.43	0.45	−355	−554	−311	−282	890	−349
TRS III	17.16	18.71	0.52	−217	−415	−174	−145	1021	−212
BM I	18.96	18.98	0.59	−73	−271	−29			
BM II	10.44	21.44	0.14	−1263	−1465	−1218			
BM III	11.52	10.57	0.38	−5	−195	36			
$w_{t+1}^* \in [0, 1]$									
SFM	14.76	11.71	0.62	239	48	281	−3	291	222
TRS I	13.92	8.90	0.69	341	152	382	101	392	325
TRS II	13.20	8.73	0.62	279	90	320	39	330	263
TRS III	14.40	10.81	0.62	259	69	300	18	310	243
BM I	14.52	11.33	0.59	242	51	283			
BM II	11.64	11.40	0.35	−52	−242	−10			
BM III	11.64	10.46	0.38	16	−174	58			
Panel C: January 1993 - December 2005									
$w_{t+1}^* \in [-1, 2]$									
SFM	9.72	15.97	0.38	−301	−408	−185	803	1331	100
TRS I	10.56	15.35	0.45	−153	−260	−37	948	1474	247
TRS II	9.48	15.42	0.38	−269	−376	−154	831	1358	130
TRS III	11.28	16.39	0.48	−186	−294	−70	920	1449	173
BM I	2.64	16.97	−	−1110	−1218	−993			
BM II	−0.60	18.91	−	−1659	−1768	−1541			
BM III	12.96	19.68	0.48	−410	−519	−291			
$w_{t+1}^* \in [0, 1]$									
SFM	9.12	8.35	0.66	225	121	337	315	732	240
TRS I	9.00	8.00	0.66	232	128	345	322	739	248
TRS II	8.64	7.97	0.62	192	88	304	282	698	207
TRS III	8.64	8.63	0.59	157	53	270	248	665	215
BM I	6.24	8.87	0.31	−90	−194	23			
BM II	2.88	10.29	−	−510	−614	−397			
BM III	9.84	13.03	0.48	−16	−122	99			

Note: The table displays performance measures for the active mean-variance investment strategies based on monthly excess return predictions, with the volatility fixed at the median level during the 10-year rolling window that is used for model specification and estimation. Results are shown for medium transaction costs (0.1%) and with RRA=6. See Table 3 for further details.

Figure 1: Number of factors and their lags used in each period
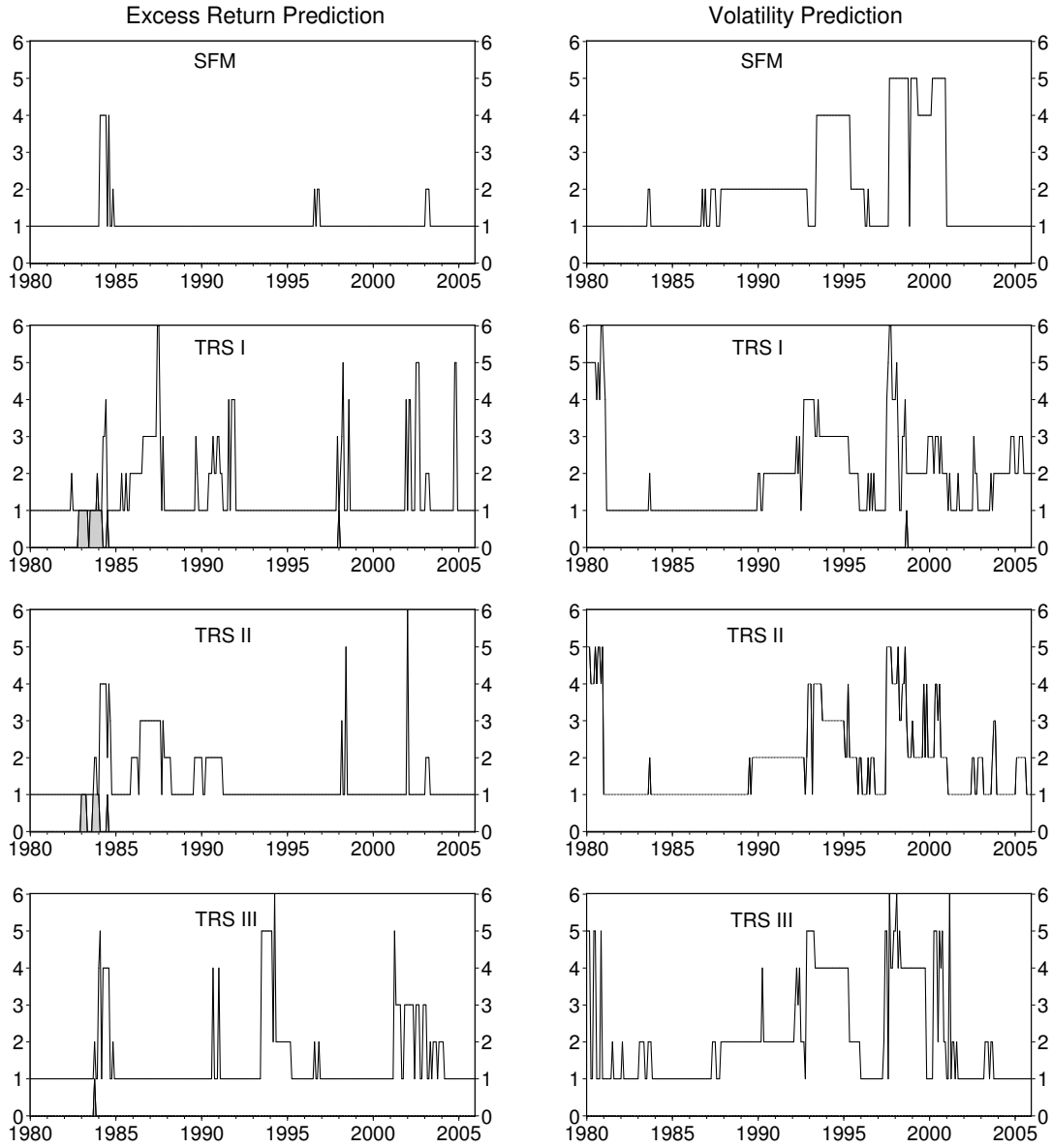
Note: The graphs show the number of factors (solid line) and the number of lags (shaded area) included in the predictive regression models for excess returns (left) and volatility (right), for the moving window ending at the date on the horizontal axis. The selection of the number of factors and their lags is made with BIC, where at least one and at most six factors are included and up to two lags of each factor. Model abbreviations are as in Table 1.

Figure 2: Number of factors and their lags used in each period
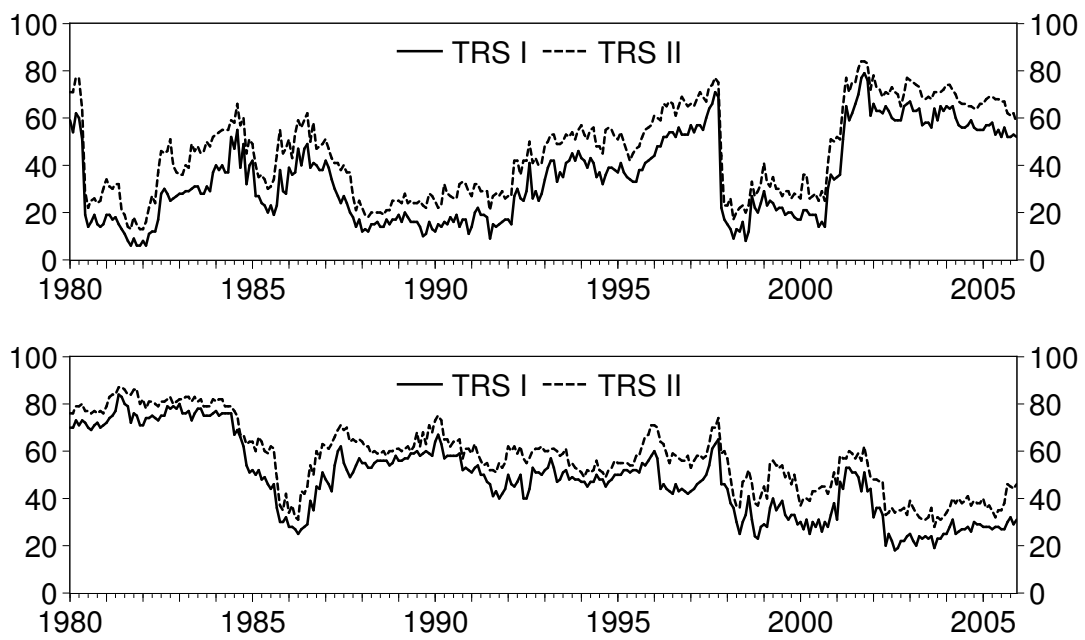
Note: The graphs show the number of macroeconomic variables that are selected using the hard thresholding approach with significance levels 0.10 (TRS I) and 0.15 (TRS II) for constructing factors that enter the predictive regressions for excess returns (top panel) and for volatility (bottom panel).

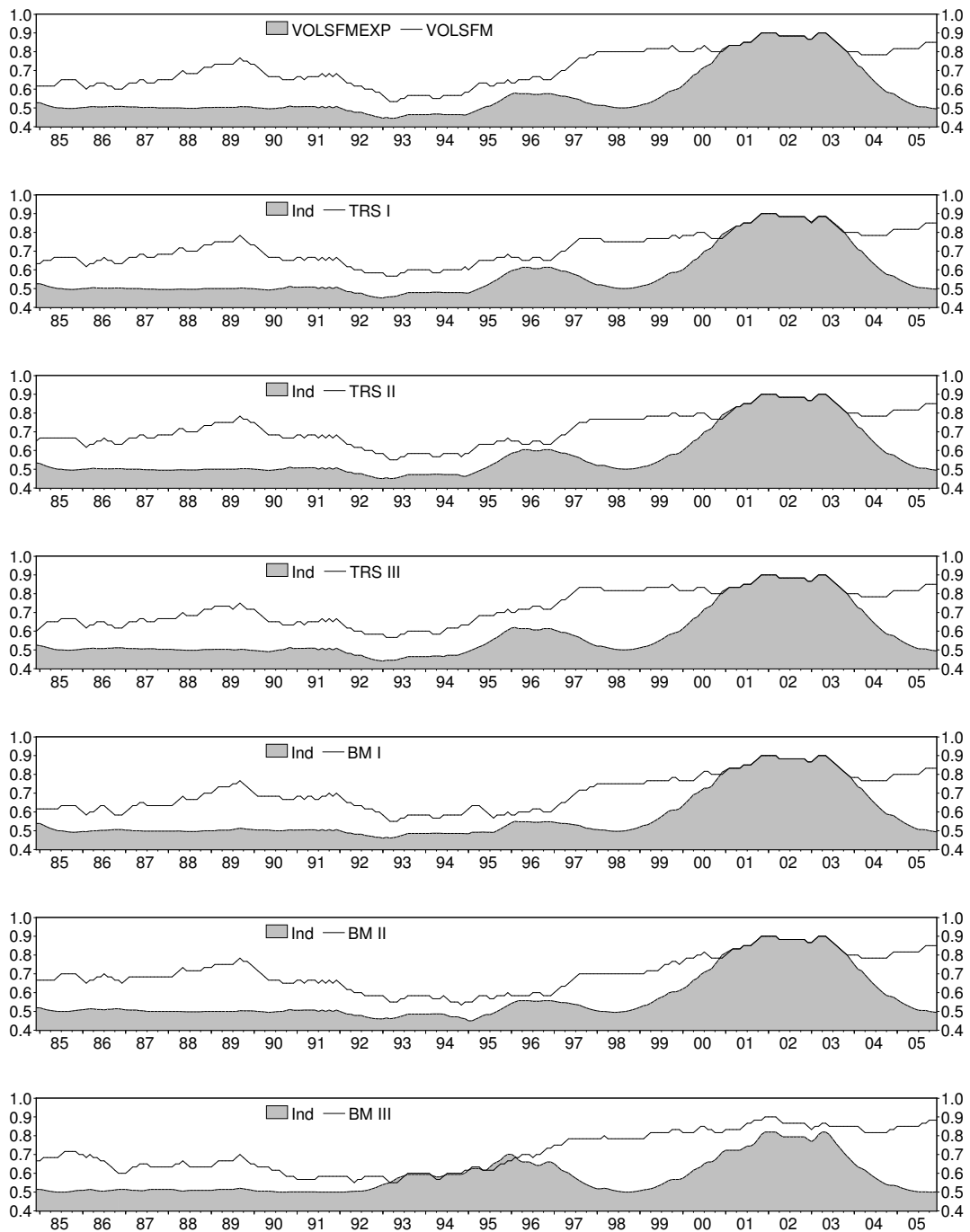Figure 3: Hit ratio for excess returns during five-year moving windows
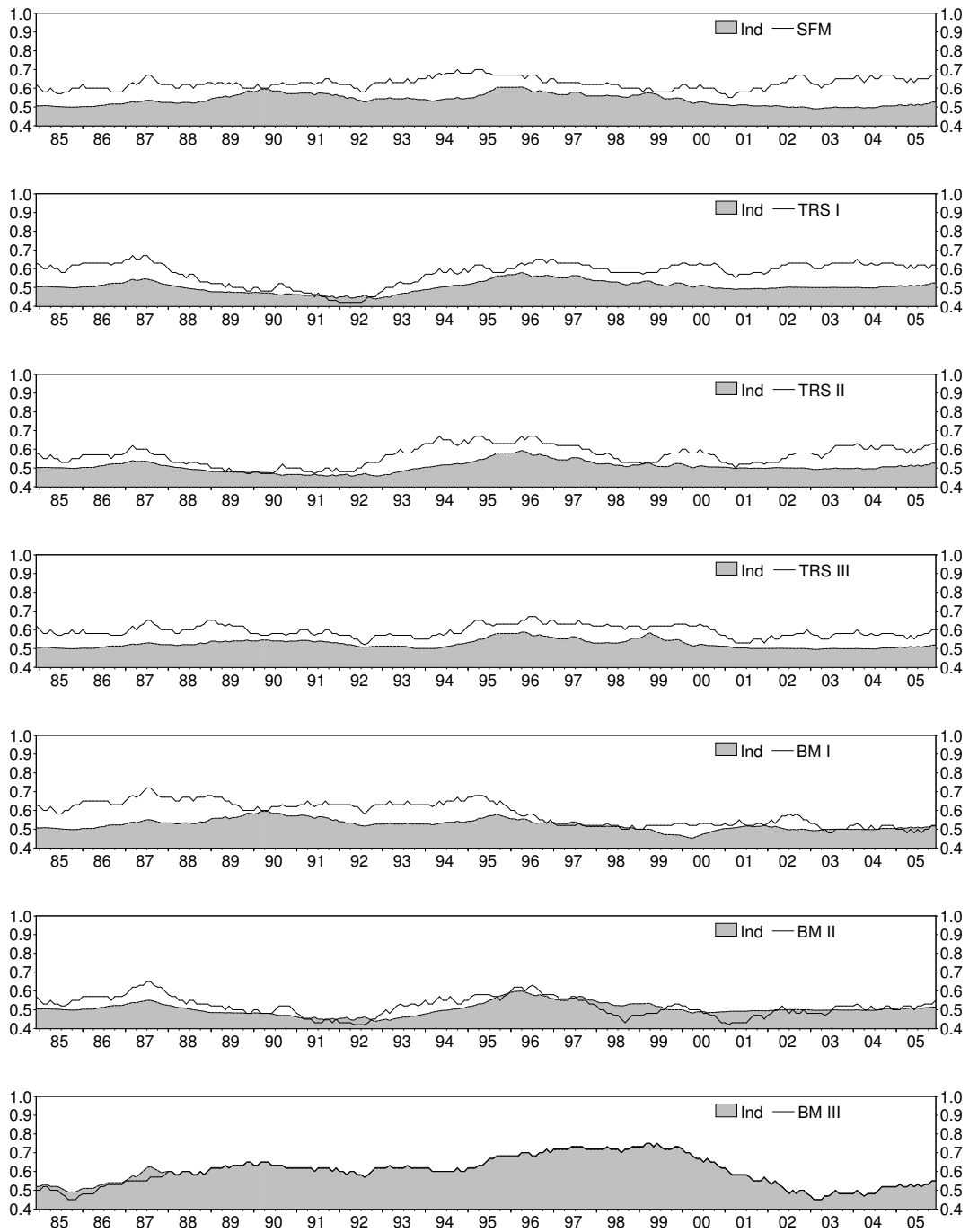
Note: The solid line is the hit ratio for the sign of the monthly excess stock return, during the five-year moving window ending at the date displayed on the horizontal axis. The shaded area indicates the expected hit ratio under the assumption of independence between the actual and predicted signs. Model abbreviations are as in Table 1.

Figure 4: Out-of-sample R^2 of factor-augmented predictive regression models for excess returns
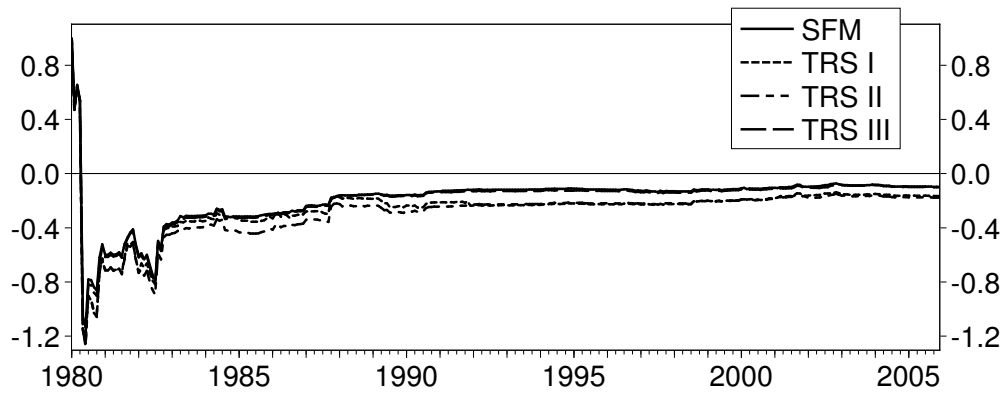

Note: The graph shows the time path of R^2_{OoS}, defined in (12), that is, one minus the ratio of the cumulative squared prediction errors for the factor-augmented regression relative to the cumulative squared prediction errors for the prevailing historical mean excess return. Model abbreviations are as in Table 1.

Figure 5: Hit ratio for volatility during five-year moving windows

Note: The solid line is the hit ratio for the sign of the monthly volatility (relative to the median volatility level during the 10-year rolling window that is used for model specification and estimation), during the five-year moving window ending at the date displayed on the horizontal axis. The shaded area indicates the expected hit ratio under the assumption of independence between the actual and predicted signs. Model abbreviations are as in Table 1.

61

Figure 6: Economic performance of factor-augmented predictive regression models during five-year moving windows
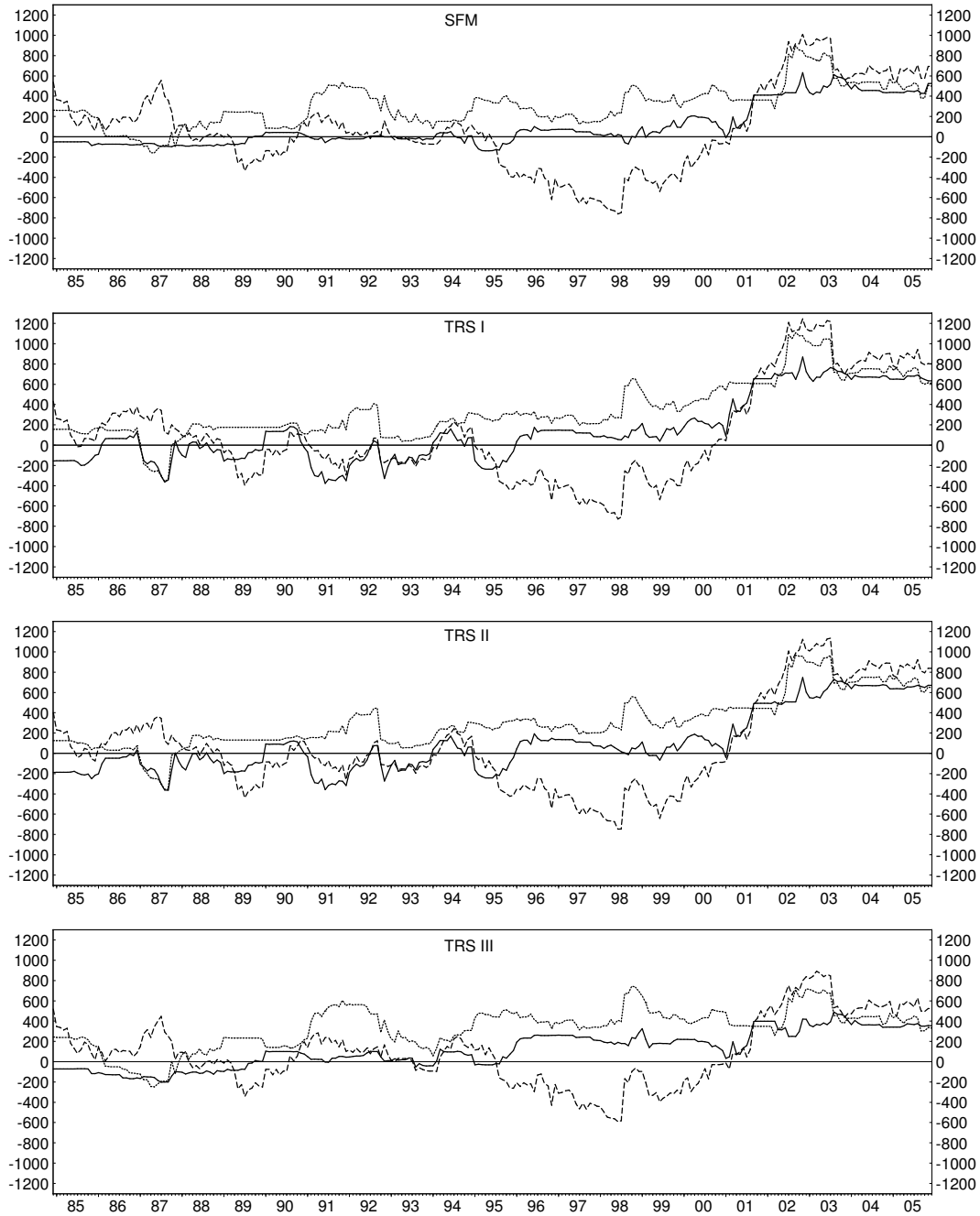
Note: Performance fees Δ (in basis points per annum) computed according to (17) for five-year moving windows ending at the date displayed on the horizontal axis. The solid line, the dotted line and the dashed lines represent the performance fees of the indicated factor-augmented predictive regression relative to the first, second, and third benchmark models. Model abbreviations are as in Table 1.