Equity premium prediction: Are economic and technical indicators instable?

by Fabian Bätje and Lukas Menkhoff

Abstract

We show that technical indicators deliver economic value in predicting the U.S. equity premium. A crucial element of this value stems from the relative stability of returns over the full sample period from 1950 to 2013. Results tentatively improve over time and mostly beat alternatives over subperiods. By contrast, economic indicators perform only until the 1970s, thereafter they basically lose predictive power, even when the last crisis is considered. Translating the predictive power of technical indicators into a standard investment strategy delivers an average Sharpe ratio of 0.6 for investors who had entered the market at any point in time.

Fabian Bätje, Department of Economics, Leibniz University Hannover, Königsworther Platz 1, D-30167 Hannover, Germany; e-mail: baetje@gif.uni-hannover.de.

Lukas Menkhoff, Kiel Institute for the World Economy, 24100 Kiel, Germany, and Kiel University; e-mail: menkhoff@gif.uni-hannover.de; tel. ++49 (0)431 8814 216.

Equity premium prediction: Are economic and technical indicators instable?

1   Introduction

There is a long-standing debate whether the equity premium is predictable or not. Whereas predictability seemed to be largely accepted for some time (e.g., Campbell and Shiller, 1988a,b, Fama and French, 1988, 1989, Cochrane, 2008), Goyal and Welch (2008) present strong evidence challenging the view of predictability. They show that economic indicators used for equity premium prediction perform instable over time. In particular does forecasting performance mainly arise from the periods up to the early 1970s but not from the later decades. Seen from this perspective, many earlier results in favor of predictability may be rather driven by specific samples than showing a systematic relation.

There are two recent developments, however, which motivate a new analysis. First, economic indicators predict the equity premium quite well in crisis times which might lead to much improved forecasting results due to the consideration of the major crisis of 2008/09. Second, Neely et al. (2014) show that their universe of 14 technical trading rules is well able to predict the equity premium out-of-sample. The performance of these indicators is better than that of 14 conventional indicators which are based on economic reasoning, such as the "dividend-price ratio".

Building on these recent developments we thoroughly examine the possible instability of economic and technical indicators for predicting the U.S. equity premium. Not surprisingly, all indicators are instable over time but to very different degrees. Thus it is important for an assessment of economic value to look at both, returns on forecasting and the instability of these forecasts. Our main finding is that technical indicators do have economic value. Transforming this kind of equity premium prediction into a conventional investment strategy generates an average Sharpe ratio of 0.6 for an investor who entered the market at any point in time since the mid 1960s (no transaction costs imposed). Reassuringly this performance is tentatively increasing over time and also quite stable over sub-periods. Therefore, technical indicators perform superior to economic indicators which lose their predictive power after the 1970s, even when we consider the recent crisis period.

Our procedure closely follows main studies in this field, such as Goyal and Welch (2008) or Neely et al. (2014), in order to make our analysis directly comparable to this

benchmark literature. Thus we choose the same selection of 14 technical and 14 economic indicators being used to predict the equity premium over the sample period from 1950 to the end of 2013. Based on this replication of earlier results we implement various tests to uncover potential forecasting instability over time.

We contribute to the literature in three ways: (1) we thoroughly examine the instability of technical indicators, (2) we also apply a new stability test to comprehensively analyze instability and (3) we assess the instability of economic value of forecasting indicators. Regarding the sample period, we consider data until the end of 2013 compared to the last considered years of 2005 in Goyal and Welch (2008) and 2011 in Neely et al. (2014). This sample extension is interesting when repeating the Goyal and Welch prediction approach as the additional years include a deep crisis period but do not improve predictability of economic indicators. Compared to Neely et al. (2014) the extension is just two years which does not qualitatively change their findings.

Regarding the *instability of technical indicators*, we complement the analysis of Neely et al. (2014) and document that reported results for various sub-periods are inferior to observed full sample findings. Thus, also the use of technical indicators produces instable results over time, however, much less than for the economic indicators.

Regarding our *new stability test* we first examine the empirical relationship between the equity premium and the forecasting variables by conventional break tests. Unfortunately, results are thin and inconsistent. Thus we propose another approach which basically mirrors the commonly used recursive procedure (e.g. Goyal and Welch, 2003): instead of using a fixed starting point and enlarging the sample period from there on, we fix the end point and shorten the sample period step by step. This procedure introduces the idea of rolling windows but avoids unreliable results from a standard rolling window approach. Results show that technical indicators – different from economic indicators – produce meaningful forecasts until recently.

Finally, regarding the *economic value* we assess forecasts by utility-based metrics. In detail, we consider a mean-variance investor who optimizes his risk-return profile dependent on the predicted equity premium. Subsequently, portfolio performance is determined by the certainty equivalent return and the Sharpe ratio using various risk-aversion coefficients, transaction costs and constraints regarding portfolio weights. We find that technical indicators are able to beat alternative investment strategies in almost all relevant cases. By contrast, the

declining predictive ability of economic indicators translates into disappointing outcomes of respective investment strategies.

Our research belongs to four strands of literature, reflecting our respective contributions. We first refer to studies predicting the equity premium by economic indicators as surveyed by Rapach and Zhou (2013). This line of research has developed established indicators, such as the dividend-price ratio, other valuation ratios, the inflation rate, the stock market volatility and short term interest rates, as well as credit and term spreads. These studies are reflected and refined in the collection of papers summarized by Spiegel (2008), including Goyal and Welch (2008), Campbell and Thompson (2008), Cochrane (2008) and Lettau and Van Nieuwerburgh (2008). More recently, Rapach et al. (2010) suggested that combination measures of economic indicators may lead to better forecasting results. Second, our study is inspired by earlier works showing the usefulness of certain technical indicators for predictions on equity markets, such as Brock et al. (1992), Brown et al. (1998) and Lo et al. (2000) and Neely et al. (2014). Third, our main concern here is, however, not predictability as such but its potential instability. Thus, we use additional tests beyond the standard recursive approach (e.g. Goyal and Welch, 2003) for examining instability of predictive ability over time. We apply break tests, recently used by Rapach and Wohar (2006) and Paye and Timmermann (2006), however, with limited success compared to previous evidence. Therefore we employ a recursive estimation setting based on rolling initialization periods which should be suited to account for such occasional breaks. Fourth and finally, we evaluate the predictive performance in economic terms which provides a direct linkage to its practical usefulness. In this respect, we follow Cenesizoglu and Timmermann (2012) and Rapach and Zhou (2013), among others.

This paper is organized in five more sections. Section 2 informs about the approach and the data. Section 3 contains our examination of out-of-sample equity premium prediction, including analysis of instability. Results are assessed in Section 4 where we analyze the economic value from the preceding section. Further robustness tests are provided in Section 5 and Section 6 concludes.

2   Approach and data

This section provides background information for the research presented in later sections. It describes the forecasting approach (Section 2.1) and the data being used (Section 2.2).

2.1 Forecasting approach

Our empirical application is based on the typical specification for the equity premium prediction, i.e.

$$r_{t+1} = \alpha_i + \beta_i x_{i,t} + \varepsilon_{i,t+1} \tag{1}$$

where r_{t+1} is the equity premium at time $t+1$ and $x_{i,t}$ is the one-month lagged predictive variable stemming from a broad set of economic variables and technical trading rules, indexed by i. $\varepsilon_{i,t+1}$ denotes the corresponding equity premium innovation. In addition, we also make use of forecasting strategies which should yield superior prediction performance by addressing concerns of in-sample overfitting, model uncertainty and parameter instability (summarized by Rapach and Zhou, 2013). In detail, used forecasting strategies incorporate information from the full set of predictor variables stemming from economic variables, technical indicators or both. We follow Neely et al. (2014) in this respect and estimate latent factor structure models, proposed by Stock and Watson (2002a,b). Regarding the amount of principal components used in the predictive setting, we employ the Schwarz information criterion (SIC), assuming a maximum number of three common components based on the set of 14 economic variables and technical indicators, and four based on the full set of 28 predictors. Results on alternative strategies are presented in the robustness section.

Given the empirical finding that out-of-sample evidence of equity premium prediction falls behind in-sample exercises (see for example, Bossaerts and Hillion, 1999, Goyal and Welch, 2003, 2008), our application is solely based on ex-ante identification (see Campbell, 2008). Therefore, we are interested in whether predictor variables deliver equity premium forecasts in a real-time setting and, more precisely, whether they outperform the historical average commonly used as a benchmark specification. To address the out-of-sample aspect of our analysis, the predictive regression in (1) is converted into a real-time setting, where we split the total sample into an initialization period [1:s-1] and an out-of-sample evaluation period [s:T]. More specifically, one step ahead forecasts are obtained by recursive estimates.

Out-of-sample forecast accuracy is then assessed by the R_{OS}^2 evaluation statistic, suggested by Campbell and Thompson (2008)

$$R_{OS}^2 = 1 - \frac{\sum_{s=1}^{T}(r_s - \hat{r}_s)^2}{\sum_{s=1}^{T}(r_s - \bar{r}_s)^2} \tag{2}$$

where $\hat{r}_s$ represents the out-of-sample forecasts based on the predictive variables and $\bar{r}_s$ is the forecast using the historical average instead. Moreover, to examine whether predictors significantly contain information above and beyond the historical average, we make use of the MSFE-adjusted test statistic proposed by Clark and West (2007). Statistical inference is based on the null hypothesis of equal or lower mean squared forecasting errors under the benchmark specification against the one-sided alternative of lower mean squared forecasting errors using the predictive variable under analysis. During the course of examination, we allow for various specifications to check whether empirical results are stable and economically important.

2.2 Data description

Our sample covers monthly observations from December 1950 through December 2013, for a total of 757 observations which should be reasonably long for our objective of stability screening. The dataset and the sample size are related to Neely et al. (2014) and updated by two additional years. Our application is based on forecasting the monthly U.S. equity premium which is defined as the difference between the continuously compounded log return of the S&P 500 (including dividends) and the log return on a risk-free bill. We make use of 14 economic variables that have been used prevalently in the empirical literature and for comparison purposes we also focus on 14 predictive variables stemming from the category of technical trading rules. A detailed variable description is given in the Data Appendix.

Economic indicators. The set of 14 economic predictor variables is a representative outline of variables commonly used to predict the equity return (see, for example, Goyal and Welch, 2008, Rapach et al., 2010). These variables comprise information about **stock characteristics**: (log) Dividend-price ratio (DP); (log) Dividend yield (DY); (log) Earnings-price ratio (EP); (log) Dividend-payout ratio (DE); Equity risk premium volatility (RVOL); Book-to-market ratio (BM) and Net equity expansion (NTIS) in addition to **interest related information**: Treasury bill rate (TBL); Long-term yield (LTY); Long-term Return (LTR);

Term spread (TMS); Default yield spread (DFY); Default return spread (DFR) and Inflation (INFL).[1]

Technical indicators. Following Neely et al. (2014) the full set of 14 technical indicators is based on 3 kinds of popular technical trading strategies. At the end of each period, i.e. in our setting each month, all of these indicators provide a buy (sell) signal based on recent price movements. We generate six technical trading strategies based on **moving-average rules** which compare short- (1, 2, 3 months) and long-term (9, 12 months) moving averages to detect changes in stock price trends. In addition, we obtain two technical trading strategies by comparing current with past stock prices, i.e. **momentum rules**. If the current price level exceeds the previous level (9, 12 months periods ago) then the trading rule generates a buy signal, i.e. a trend-following perspective. The third category is based on **volume rules**. These six technical trading indicators relate volume to price changes (short-term=1, 2, 3 months; long-term=9, 12 months) to detect strong price trend movements, as proposed by Granville (1963). The importance of volume comes from the interpretation that price movements confirmed by high trading volume generate more serious signals of stock price trends.

Descriptive statistics. Descriptive statistics for the U.S. equity premium and predictor variables are reported in Table I. The equity premium provides on average a return of 0.51% per month with a monthly standard deviation of 4.22% which leads on average to an annualized Sharpe ratio of 0.42. Summary statistics on technical indicators show a sample mean in the range of 0.67 to 0.73 which involves buy signals in at least two-third of the whole sample range. First order autocorrelation coefficients are highly statistically significant and in the range of 0.60 to 0.83. This tentatively supports the underlying assumption of technical analysis that past price trends persist into the future.

TABLE I about here

Economic predictors on the other hand confirm previous findings of highly statistically significant persistency near the unit root for almost all variables. With the exception of the long-term return (LTR), the default return spread (DFR) and the inflation rate (INFL), all economic variables are highly autocorrelated, with 1st order autocorrelation coefficients near 1. Second to third autocorrelation coefficients illustrate that the persistent behavior of economic variables decays comparatively slower over time relative to technical indicators.

---

[1] We follow Neely et al. (2014) by using a slightly different volatility measure proposed by Mele (2007) which

3   Out-of-sample equity premium prediction

This section presents our prediction results in four steps. We start with results for the full sample period, thus mainly replicating earlier exercises for a somewhat longer period (Section 3.1). Then we apply the Goyal and Welch (2003) stability procedure on the economic and technical indicators (Section 3.2) and analyze these time series with conventional break tests (Section 3.3). Finally, we apply a rolling-recursive estimation approach to measure performance stability over time (Section 3.4).

3.1 Out-of-sample prediction results

As a first step of our empirical analysis we document forecasting results of the 14 economic and 14 technical indicators for the full sample period. This allows comparison with earlier studies, in particular with Neely et al. (2014) which cover a somewhat shorter period from January 1966 to December 2011. We show that adding two additional years of observations does not qualitatively change results (Table II): economic indicators perform poorly, technical indicators perform much better and all indicators tend to perform much better during recession periods than during expansion periods.

TABLE II about here

In more detail, Panel A shows that among the economic indicators only two, i.e. RVOL and LTR, outperform the historical average benchmark with statistical significant positive R^2_{OS}. For three further variables we can reject the null hypothesis of $R^2_{OS} \leq 0$ at conventional significance levels of %5 and 10% even though the R^2_{OS} is negative.[2] By contrast all technical indicators have a positive R^2_{OS}, four of them have a comparatively high out-of-sample predictive performance in the range of 0.66% to 0.83%. The MSFE-adjusted test statistic indicates that seven technical indicators exhibit statistical significant forecast accuracy; five positive R^2_{OS} are significant at the 10% level, while only two indicators are significant at the 5% level.

Results reported in Panel B and C provide information from the full set of economic and technical indicators by forming principal components. As expected, the R^2_{OS} for the economic

---

[2] Clark and West (2007) mentioned that the null hypothesis can be rejected even if we observe negative R^2_{OS} due to the adjustment term which accounts for upward bias in the MSFE produced by parameter estimates that are zero under the null.

variables is negative and that for the technical indicators is positive. Nevertheless, p-values for the MSFE-adjusted test statistic are below 0.05 for economic variables, whereas technical indicators outperform the historical average solely at the 10% level. Principal components based on economic as well as technical indicators (Panel C) indicate highly statistical significant outperformance at the 1% level with a R_{OS}^2 of 0.71% for the full sample. Overall, this supports the notion that technical indicators contain information above and beyond economic variables over the business cycle, as shown by Neely et al. (2014).

Finally, forecasting power of nearly all predictor variables is predominantly located in recession periods, which is in line with earlier findings motivated by Fama and French (1989), Cochrane (1999, 2007) and highlighted by Henkel et al. (2011), among others.

3.2 Dynamic out-of-sample prediction performance

As mentioned by Timmermann (2008) "Most of the time the forecasting models perform rather poorly, but there is evidence of relatively short-lived periods with modest return predictability" which might lead to positive R_{OS}^2 over the full sample period. This is in line with findings by Goyal and Welch (2008), who show that the predictive ability of economic variables sharply increases during the Oil price shock recession in the 1970s but that the same models perform poorly if these unusual years are excluded from the sample.

To examine whether the forecast performance over the full sample, as presented in Table II, may benefit from short-lived periods, we follow Goyal and Welch (2003, 2008) in this section. They propose focusing on the cumulative sum of differences in the squared forecast errors under the benchmark specification and the squared forecast errors based on predictive variables (CDSFE).

$$CDSFE(s,i) = \sum_{s=1}^{T}((r_s - \bar{r}_s)^2 - (r_s - \hat{r}_{s,i})^2) \tag{3}$$

FIGURE I about here

To save space Figure I shows the out-of-sample performance of principal component indicators, relative to the benchmark, at each point in time. First, values above zero indicate a positive performance of the predictive model up to the point in time that is considered. Second, an increasing process contributes positively, whereas a declining line implies that predictive performance is negative in the period under consideration. The three panels show the predictive performance for three principal components, representing economic indicators,

9

technical indicators and all indicators (figures for all 28 single indicators are available on request).

Overall, we confirm earlier findings: (i) We show that none prediction model outperforms the historical average consistently over time, i.e. there are no continuously upward sloping curves. (ii) Local predictability is concentrated in recessions rather than expansions. (iii) The indicator PC_{ECON} (see Panel A) provides some outperformance up to the first half of the sample with a sharp increase in the predictive performance during the 1970s recession (as mentioned by Goyal and Welch, 2008) and around the 1980s recessions.[3] (iv) None of the 14 single economic indicators performs considerably better than PC_{ECON}. (v) The performance of principal component predictive regressions based on technical indicators, PC_{Tech} (see Panel B), is never much worse than the benchmark over longer periods, i.e. there are only small negative values, and the long-term trend is rather upwards than downwards. (vi) Looking at the 14 technical indicators individually largely confirms these findings. (vii) Finally, Panel C shows forecasting performance by combining information from economic and technical indicators (PC_{ALL}). The overall pattern follows PC_{ECON} but is moderated by the influence of PC_{Tech}.

Given this strong time-dependent predictive ability, further analysis seems warranted, to analyze whether predictability is solely driven by specific samples or whether predictor variables show a systematic relation. These aspects are analyzed in Sections 3.3 and 3.4.

3.3 Structural stability tests

Early evidence of instability in the prediction performance, using valuation ratios (see Lettau and Ludvigson, 2001, Goyal and Welch, 2003, and Ang and Bekaert, 2007 for example), has recently being linked to the presence of occasional break dates. But the possibility of occasional changes seems not to be restricted to economic variables. Park and Irwin (2007) mention that also technical trading strategies are subject to substantial changes and their profitability tends to vanish after the late 1990's.

In Section 3.2 above, we have related the equity premium to predictor variables in a recursive estimation setting. This results in the most efficient coefficient estimates by incorporating more information as it becomes available. Nevertheless, these out-of-sample forecasts are based on the presumption, that the underlying relationship is constant or sparsely

---

[3] This finding is in line with the strong deterioration in the predictive performance of dividend-price ratios since the mid 90s, shown by Lettau and Ludvigson (2001), Goyal and Welch (2003) and Ang and Bekaert (2007) which results from a sharp increase in their persistency.

time-varying. However, recent literature (e.g., Pesaran and Timmermann, 2002, Lettau and Van Nieuwerburgh, 2008, Rapach et al., 2010, Pettenuzzo and Timmermann, 2011) highlights the effects of model and parameter instability due to occasional structural breaks. Such breaks might also explain weak out-of-sample results compared to its in-sample counterparts (see Clark and McCracken, 2005).[4] Rapach and Wohar (2006) and Paye and Timmermann (2006) provide evidence for the presence of structural breaks in the 1990s and highlight that the relationship between the equity premium and dividend-price ratio substantially decreased after 1990. Interest rate related variables, like the term spread, offer breakpoints in the 1970s. Accordingly, ignoring the presence of possible breaks would lead to biased estimates and thus failure to predict the equity premium out-of-sample.

Postulating at least one breakpoint up to time T, the data generating process exhibits the following form

$$r_{t+1} = \alpha_{1,i} + \beta_{1,i}x_{i,t} + \varepsilon_{i,t+1} \qquad t=1,\ldots,k_1,$$
$$r_{t+1} = \alpha_{2,i} + \beta_{2,i}x_{i,t} + \varepsilon_{i,t+1} \qquad t=k_1,\ldots,T\text{-}1. \tag{4}$$

To examine whether structural breaks in the equity premium prediction regressions are present, we run three kinds of empirical break tests, following Rapach and Wohar (2006) and Paye and Timmermann (2006) in this respect. (1) Using in-sample predictive regressions, we employ Andrews (1993) *SupF* statistic, testing the null hypothesis of no structural break against the alternative of occasional change at unknown date. We impose a 15% trimming percentage to determine the minimum window length between breaks.[5] (2) Allowing for multiple breaks, we employ the Bai and Perron (1998) *UDmax* and *WDmax (5%)* statistics for testing the null hypothesis of no structural breaks against the alternative of multiple breaks of at most 5 occasional changes. Bai (1997) and Bai and Perron (2001) mention that the *UDmax* and *WDmax* statistics can be more powerful than Andrews *SupF* test in the case of multiple breaks. (3) Finally, we make use of Elliott and Müller (2006) $\widehat{qLL}$ which has good power and size properties even under heteroskedastic settings.

TABLE III about here

Results shown in Table III do not consistently provide evidence of structural instability. While empirical evidence is quite clear for technical indicators (nearly all tests do not reject the null hypothesis of no structural break), predictive regressions using economic indicators seem to be affected by breaks more intensively. Nevertheless, findings are mixed and strongly dependent on the selected break-test. Thus, neither previous evidence of structural instability can be confirmed nor is it obvious whether and when the predictive performance might offer major instability. Therefore, to highlight possible instability, we account for possible breaks in a more dynamic estimation setting in the following section.

3.4 Performance stability in a rolling-recursive setting

Motivated by concerns of Clark and McCracken (2005) and Pesaran and Timmermann (2007) on possible distortions of the earlier approaches we apply here a rolling-recursive setting. This is new in the literature on equity premium prediction and complements the other approaches.

Findings presented so far are based on recursive estimates over the full sample range which might strongly benefit from the specific sample period under analysis (see Clark and McCracken, 2005). Moreover, as there is no distinct evidence of structural breaks in the empirical relationship between the equity premium and predictor variables, it is less clear whether predictive ability is stable or may solely exists at specific point in time (i.e. at the beginning or at the end of the sample). Naturally, rolling window regressions might be well suited to account for such shifts, but this approach has several disadvantages. Concerning the bias-efficiency trade-off, rolling window regressions might reduce potential estimation bias but this approach suffers from increasing estimation uncertainty (see Pesaran and Timmermann, 2007). In addition, breaks seem to be frequent and in order to account for this fact, the initialization period should be comparably short which is opposite to the requirements of precise identification of common components.

Therefore, we account for these effects by using a rolling-recursive estimation setting where we allow the in-sample estimation period (15 years) to vary over time. In our case we shift the starting point of the out-of-sample period continuously forward by one month. Such a procedure is equal to different subsample analysis without choosing the sample start arbitrarily. In addition, we are able to examine whether the sample under analysis is responsible for obtained out-of-sample predictability results or whether the predictive ability remains even under more recent subsamples, i.e. forecasting stability over time.

In detail, Figure II shows the time-varying process of the R_{OS}^2 by starting with an estimation window over the evaluation period 1966:01-2013:12. Thus the first points of the three strategies, shown in Panels A to C, are exactly the R_{OS}^2 given in Table II, for example, -0.47% for PC_{ECON}. Next, we examine the out-of-sample predictability over the sample 1966:02-2013:12 using an initialization period from 1951:01 to 1966:01, and so on. To account for problems arising from small out-of-sample evaluation periods, our analysis ends concerning the evaluation period 1994:01-2013:12, i.e. covering at least 20 years.[6]

FIGURE II about here

Concerning our subsample stability analysis, Figure II shows large differences between the forecast performance of economic and technical indicators through time. The time-varying R_{OS}^2 of economic predictor variables (represented by principal component predictive regressions) do not consistently outperform the benchmark model. The contrary is the case, i.e. most of the time findings reveal higher prediction errors (negative R_{OS}^2) in comparison to forecasts made by the historical average. Remarkably, some economic predictor variables never exceed the zero line. In contrast, technical indicators seem to be much more robust predictors over time, even though at a low level of predictability. While most technical indicators exhibit a substantial decline in the R_{OS}^2 regarding the out-of-sample evaluation period during the 1990s (as has been mentioned by Park and Irwin, 2007), the predictive performance recovers to its previous level afterwards. This relative forecasting stability of technical indicators is conferred to forecasting strategies taking economic variables and technical indicators into account.

The figure also illustrates the time-varying process of predictability during recession and expansion periods. In line with earlier presented analyses, the predictive ability of indicators consistently exhibits higher prediction errors than the historical average in expansions, but profits from recession phases.

Overall, our analysis illustrates that in contrast to the literature's focus on economic variables (motivated by Cochrane, 1999, 2007), technical indicators exhibit clearly more stability over time.

---

[6] As has been mentioned by Inoue and Kilian (2004) and Hansen and Timmerman (2012), out-of-sample forecast evaluation results have reduced power under short sample periods. Thus, our last evaluation period covers at least 240 months which should avoid problems arising from small sample analysis.

4   Economic value of equity premium prediction

The quality of equity premium prediction is often assessed by the returns generated by forecasting strategies, as we do in Section 3 above. However, the high instability demonstrated provides a strong motivation to examine the economic value of such strategies. In Section 4.1 we introduce into established measures of economic value, in Section 4.2 we apply them to our data, and in Section 4.3 we examine their temporal stability.

4.1 Asset allocation

Statistical measures of forecast ability are informative but not necessarily decisive for investment and asset allocation decisions. Cenesizoglu and Timmermann (2012) show that statistical and economic measures of forecasting performance are only weakly positive correlated. Accordingly, low or even negative R^2_{OS}, such as the ones being documented in Section 3, may still provide economic value at the same time. In examining economic value of forecasting indicators, we follow Marquering and Verbeek (2004), Campbell and Thompson (2008) and Neely et al. (2014) in order to keep results comparable to these studies.

We consider an investor who optimally composes his portfolio by allocating across risky assets and a risk-free asset according to equation (5)

$$r_{p,s+1} = w_s r_{s+1} + rf_{s+1} \tag{5}$$

where $r_{p,s+1}$ represents the portfolio return at the end of period s+1, determined by allocating a share of w_s to the risky asset and $(1 - w_s)$ to the risk-free bill. For simplicity, we use simple (instead of log) returns to conduct asset allocation exercises. We postulate a mean-variance utility function of the following form

$$U(r_{p,s+1}) = E_s(r_{p,s+1}) - \frac{1}{2}\gamma Var_s(r_{p,s+1}) \tag{6}$$

where γ indicates investor's degree of relative risk-aversion. Maximizing the utility function with respect to w_s yields the optimal portfolio weight for the investor

$$w_s = \left(\frac{1}{\gamma}\right)\left(\frac{E_s(r_{s+1})}{Var_s(r_{s+1})}\right). \tag{7}$$

As can be seen from equation (7), and fully in line with conventional theory, optimal portfolio allocation depends positively on the equity risk premium forecast and negatively on the conditional variance. Because volatility is latent and has to be approximated, we follow the

recent literature (like Christiansen et al., 2012) relying on realized volatility forecasts.[7] In detail, realized volatility is defined by the sum of daily squared returns in month t

$$RV_t = \sum_{\tau=1}^{M_t} r_{t,\tau}^2 \qquad (8)$$

where Mt is the number of trading days and $r_{t,\tau}$ denotes the return at the day τ in month t. Due to the high persistency of RV_t, volatility forecasts are then obtained by using an AR(1)-process based on the log of the realized variance which shifts the distribution closer to normality (see Christiansen et al., 2012). Using the same volatility estimate for all models does rule out differences in portfolio allocations implied by model specifications. We also check whether equity premium prediction models additionally add economic value due to volatility forecasts, but results are nearly unchanged (reported in the robustness section). Figure III shows the realized variance and the predicted volatility over the out-of-sample evaluation period 1966:01-2013:12.

FIGURE III about here

4.2 The economic value of forecasting models

To determine the economic relevance, we use different measures to examine the performance of equity premium forecasts compared to predictions based on the historical average. In addition to the average realized portfolio return and the corresponding standard deviation, we show the difference in the realized utility using predictor variables instead of the historical average (i.e. the certainty equivalent return). This utility gain can be understood as a management fee that an investor is willing to pay to have access to the information of the prediction model compared to the information of the historical average. In the following, reported values are annualized such that they can be understood as an annually percentage management fee.

$$\Delta CER = ((\hat{\mu}_i - 0.5\gamma\hat{\sigma}_i^2) - (\hat{\mu}_0 - 0.5\gamma\hat{\sigma}_0^2)) * 1200 \qquad (9)$$

$\hat{\mu}_i$ ($\hat{\sigma}_i^2$) indicates the sample average (variance) of the portfolio return formed on prediction model i while $\hat{\mu}_0$ ($\hat{\sigma}_0^2$) denotes the sample average (variance) using the historical average

---

[7] The dynamic of the stock market volatility is an important factor for asset allocation decisions. In contrast to other studies which use constant or slightly time-varying volatility measures (based on rolling window estimates of monthly historical returns) we do not regard such approaches as an appropriate way to capture the true and latent volatility process (see Andersen et al., 2003).

forecast instead. Additionally, we report the annualized Sharpe ratio which is defined as the portfolio excess return divided by its volatility. We follow Campbell and Thompson (2008) and Cooper and Priestley (2009) and choose a relative risk aversion coefficient of three and constrain the optimal portfolio weight for the investor by preventing short sales of stocks and taking leverage of no more than 50% (variations are reported in the robustness section). Findings are documented in Table IV.

TABLE IV about here

In comparison to results based on the MSFE, Table IV shows that most predictive regressions add economic value beyond the historical average, even though the R_{OS}^2 had been very small. In the following, we focus on the annualized utility gains but findings are qualitatively the same for Sharpe ratios.

10 out of 14 economic variables outperform the historical average according to positive utility gains, while only two economic variables offer positive R_{OS}^2. However, we find large differences in realized utility gains. Four economic indicators perform comparatively well, with annualized gains above 1.50%. This means that the access to information in the predictive regression forecast compared to the historical average has a value of at least 150 basis points for investors. The highest utility gain is provided by the term spread with gains of 300 basis points.

Concerning technical indicators, results are more in line with previous evidence. While the maximum utility gain is limited to 205 basis points, all forecasts by using technical indicators are valuable. Similar to the limited R_{OS}^2, the added value is smaller compared to the best economic indicators. Nevertheless, 10 technical indicators generate utility gains of more than 50 basis points and 5 out of these indicators report average gains of over 100 basis points.

Portfolio performance measures by making use of principal component predictive regressions behave well (see Neely et al., 2014). Individual principal component predictive regressions add economic value (economic variables by 270 basis points; technical indicators by 205 basis points). Even better, PC$_{ALL}$ offers the highest Sharpe ratio (0.53) with an average utility gain of nearly 300 basis points.

4.3 Stability of economic values

Analogous to Section 3.3 we also investigate whether reported utility gains are stable over time, and whether they also exist in the more recent history. Given the time-varying nature of the R_{OS}^2 (Figure II), performance measures might face the same problems, i.e. the economic value could profit from an empirical relationship in the distant past. To account for possible instabilities, Figure IV shows the annualized Sharpe ratio for a mean-variance investor with a relative risk aversion coefficient of three and allocation constraints $0 \leq w_s \leq 1.5$ (no transaction costs imposed). To make our results easily comparable, we use the same expanding estimation window as in Section 3.3 allowing the initialization period to vary through time. For comparison purposes, this figure also shows the Sharpe ratios of using historical average forecasts and a simple buy-and-hold strategy in the S&P 500.

FIGURE IV about here

The resulting lines show very heterogeneous pattern across the three strategies indicated by the Sharpe ratios of investment strategies starting at different points in time. Principal component forecasts based on economic variables (Panel A) perform relatively well until the 1970s. Since then the Sharpe ratio predominantly declines and previously detected utility gains vanish, not only compared to the portfolio allocations based on the historical average forecast but also compared to a simple buy-and-hold strategy.

A completely different path is found for technical indicators reported in Panel B. While the R_{OS}^2 is even small in magnitude, the reported Sharpe ratio indicates a tentatively increasing slope. With the small exception of evaluation periods around the late 1970s, trading strategies based on PC_{Tech} forecasts are more valuable than using the historical average or a simple buy-and hold strategy instead. Comparing full sample Sharpe ratios (given in Table IV) with the average Sharpe ratio using our rolling-recursive estimation setting, we confirm previous findings of highly instable prediction performance concerning economic variables. While PC_{Econ} yields an annualized Sharpe of 0.52 over the full sample, the average Sharpe ratio shrinks to 0.49. In contrast, the average Sharpe ratio of PC_{Tech} is 0.60 which indicates a rise of 0.13 points compared to the full sample.

Again, closely related is the behavior of PC_{ALL}. Reported benefits strongly depend on the sample under consideration which is largely driven by economic information. While the economic outperformance is limited to the distant past, technical indicators stabilize the performance measure afterwards.

The performance of the three investment strategies deteriorates of course under moderate transaction costs of 50 basis points as shown in Figure V. Nevertheless, the PC_{Tech}-strategy is superior to the benchmark specification and mostly better than the buy-and-hold strategy which supports its relevance also from this new perspective.

FIGURE V about here

5   Robustness

This section presents robustness exercises in three directions: (1) We combine economic and technical indicators in various ways, (2) we demonstrate the small effect of various alternative specifications of volatility prediction models and (3) we examine the effect of alternative restrictions on portfolio formation, i.e. leveraged investments and shorting.

Alternative forecasting strategies.   As mentioned above, forecasting strategies that incorporate much information avoid problems of in-sample overfitting, model uncertainty and parameter instability. To check whether reported results are robust, we first use different in-sample based selection criteria to determine the optimal number of factors used for equity premium predictability. In the following we apply the Akaike information criterion (AIC) in addition to the adjusted R^2, proposed by Neely et al. (2014). In contrast to principal component prediction which combines individual predictors in a data-driven manner we also employ forecast combinations, emphasized by Rapach et al. (2010) to check whether results remain stable. Such a procedure uses different weighting schemes to combine individual forecasting models without a possible lack of economic interpretation and appropriate factor selection through time. Combined forecasts are obtained by using mean, median and trimmed mean combination approaches as well as combination weights based on discounted mean squared forecasting errors (DMSFE). According to the DMSFE combination forecasts we consider two discount factors θ=1 and θ=0.9. A detailed description about the weighing schemes is given by Rapach et al. (2010).

In a nutshell, previous mentioned findings still hold for various alternative forecasting strategies. Table V reports the R^2_{OS} over the evaluation period 1966:01-2013:12. Concerning the forecast performance of principal component predictive regressions based on economic variables or technical indicators (Panel A, B), we find only small differences compared to results given in Table II. In contrast, incorporating information from both predictor groups

(Panel C) using either the AIC or the adjusted R^2 indicate a large increase in the R^2_{OS} concentrated in recessions. The full-sample R^2_{OS} solely rise to 1.40% (compared to 0.71%) using the adjusted R^2 for factor selection. Forecast combinations on the other side, are extremely fruitful for economic information given by a statistically significant increase in the predictive performance located during expansions. Predictive performance of technical indicators on the other side is less affected due to the comparatively homogenous group variables.[8] The outperformance of economic variables during expansion periods is even visible concerning forecast combinations based on economic variables and technical indicators. Nevertheless, the increase in the prediction performance during expansions comes at the cost of decreasing R^2_{OS} during contraction periods.

TABLE V about here

Less affected is the time-varying stability of the equity premium prediction considered in Section 3.4. The time-varying process of the R^2_{OS} seems to be largely independent whether we use principal components or forecast combinations (see Figure VI). Regarding economic variables, the statistical performance measure indicates a continuously diminishing value. While principal component predictive regressions exhibit a permanent negative R^2_{OS}, the outperformance of forecast combinations even become negative through time. As mentioned above, the spread between prediction performance during recessions and expansions sharply converge under more recent evaluation periods. The outperformance of technical indicators, on the other side, is quite more stable and confirms previous mentioned findings. Results by incorporating information from both predictor groups are strongly affected by the pathway of economic variables. The high outperformance reported in Table V is largely determined by an empirical relationship concentrated in the past.

FIGURE VI about here

A closer look at volatility prediction models. Expected utility maximization approaches require conditional mean and volatility forecasts, according to formula 7. Recent empirical research shows that economic variables have predictive power above and beyond autoregressive models even for financial volatility (see, for example, Cenesizoglu and

---

[8] Rapach et al. (2010) mention that forecast combinations lead to substantial reduction in forecast volatility. A rational explanation for the usefulness of forecast combinations of economic variables is given by its negative correlated between individual forecasts. Empirical findings can be found in Zhu and Zhu (2013). Due to the comparatively high correlation of forecasts using individual technical indicators such a reduction is less obvious.

Timmermann, 2012, Christiansen et al., 2012, Marquering and Verbeek, 2004). To verify whether our findings depend on volatility prediction specifications, we account for this kind of research by including predictor variables additional to the autoregressive term. Referring to equity premium prediction, we evaluate the following forecasting models

$$log(RV_{t+1}) = \alpha_i + \beta_i x_{i,t} + \rho_i log(RV_t) + \varepsilon_{i,t+1}. \tag{10}$$

Thus, recursively estimated variance forecasts are obtained by $exp(\hat{\alpha}_i + \hat{\beta}_i x_{i,s-1} + \hat{\rho}_i log(RV_{s-1}))$ over the evaluation period 1966:01-2013:12. For convenience and comparison proposes, we use the same prediction models as for the equity premium and evaluate the forecast performance by the R^2_{OS} using log returns, while economic forecast performance measures are based on simple returns.[9] The responding benchmark model is based on a simple AR(1) specification, assuming $\beta_i = 0$. Results are given in Table VI.

TABLE VI about here

In contrast to results reported for the equity premium, we find evidence that economic variables as well as technical indicators have statistical significant predictive power above and beyond a first-order autoregressive term. In detail, seven economic variables and eleven technical indicators offer a positive R^2_{OS} over the full sample. The outperformance of five economic variables is statistically significant at the 1% level and four at the 5% level. Concerning technical indicators, only one variable generates significant smaller prediction errors than the benchmark model at the 1% level. Nevertheless, for seven indicators the R^2_{OS} is significant at the 5% level. More in line with previous results is the difference between the R^2_{OS} during recessions and expansions. Not solely equity premium predictability seems to be a recession phenomena, but also volatility prediction.

Principal component models highlight the beneficial use of such forecasting strategies. We find that combining information from economic variables and technical indicators yields the highest R^2_{OS} of 5.35% over the full evaluation period which is in line with previous findings. Even here the outperformance is mainly driven during recessions (18.24%) compared to expansions (1.44%).

However, looking at economic performance measures does not change our results considerably. Comparing Table VII with Table IV, we observe only mild deviations which do

---

[9] Full sample correlation between RV based on squared daily log returns and squared daily simple returns is above 99%. Thus, differences by using simple instead of log returns should not play any role.

not change our general findings. Concerning differences in the annualized Sharpe ratio, the overall discrepancy is between -0.02 up to 0.01.

TABLE VII about here

Alternative specifications in portfolio allocation. Next, we analyze, how results differ according to variations of imposed portfolio constraints, relative risk aversion and under transaction costs. To address this question, Table VIII reports differences in the realized utility. In addition, to shed light on any instability aspects, we also show differences in the Sharpe ratio over the evaluation period 1966:01-2013:12 and the average difference of the Sharpe ratio using our rolling-recursive estimation approach. All reported results are based on differences compared to asset allocations using the historical average forecast instead.

We allow for variation in the relative risk aversion, assuming coefficients of 3, 5, and 7. Portfolio allocation constraints are determined by: (a) short sales prevention and no leverage ($0 \leq w_s \leq 1$), (b) short sales prevention and taking leverage by no more than 50% ($0 \leq w_s \leq 1.5$), (c) allowing for short sales and taking leverage by 100% ($-1 \leq w_s \leq 2$). Last but not least, we also analyze portfolio performance measures net of transaction costs, assuming costs of 50bp for reallocation purposes.

TABLE VIII about here

Results reported in Table VIII indicate that alternative specifications do not change our general findings. In detail, the annualized Sharpe ratios of the prediction models compared with the benchmark specification indicate instabilities through time which is especially true for economic variables. The previously reported economic benefit of PC$_{Econ}$ disappears over time in all cases. In accordance with the time-varying R^2_{OS} it seems that the usefulness of economic variables is sample specific rather than systematic. We also find large variations for forecast combinations based on economic variables; however the average Sharpe ratio is mostly positive but small in magnitude.

On the other hand, imposing alternative allocation constraints and various relative risk aversion coefficients have only mild effects on the portfolio performance of technical indicators. While the economic benefit slightly decreases over time, the average Sharpe ratio is mostly positive which confirms the previously mentioned stability characteristics of technical indicators. Only if we allow for short selling, the economic importance vanishes through time using the adjusted R^2 as the model selection criterion.

Combining information from economic variables and technical indicators is also affected by instability pattern. The added economic value over the full sample largely declines over more recent subsamples which can be attributed to the poor performance of economic variables. Technical indicators on the other hand stabilize the added economic value, highlighting the importance of technical indicators as predictor variables.

6   Conclusions

The issue of equity premium prediction is long standing in the literature (see Spiegel, 2008). At least for standard economic variables, Goyal and Welch (2008) have demonstrated that high predictive ability of economic indicators until the 1970s has basically disappeared thereafter. From this perspective, it seems obvious to examine the predictive ability of technical indicators and, indeed, Neely et al. (2014) show their potential. We contribute to this literature by complementing the analysis of forecasting ability by focusing on possible instability. We confirm instability of economic indicators as indicated by Goyal and Welch (2008), here with more data and various methods. By contrast we find much less instability of technical indicators, supporting the favorable notion by Neely et al. (2014) and even indicating economic value of technical indicator-based forecasting.

As it is our ambition to complement earlier work, we closely follow the main approach in this literature (see Goyal and Welch, 2003, 2008, Neely et al., 2014). Thus, we consider the same set of economic and technical indicators as before and basically replicate earlier results, just extending the sample period by recent years. From that point on, we apply structural stability test with a somewhat disappointing result as various tests do not converge on the breakpoints. However, there seem to be breakpoints and this motivates to complement the standard recursive approach in demonstrating forecasting performance over time by a specific rolling-recursive approach.

We aim for simulating the fact that investors may enter the market at an ex ante unknown point of time and that they do not use infinite information from the past due to possible instability (which is real as we have shown). Thus, we propose a 15 years long in-sample estimation period and forecast the equity premium from that point on until the last month of our sample, i.e. December 2013. Accordingly, an investor may have invested at any month since January 1966 until January 1994 (in order to ensure a long out-of-sample period).

We find from this rolling-recursive approach that indeed economic indicators perform very instable over time, but technical indicators do much less.

Thus, we examine the possible economic value of such forecasting indicators. Largely confirming earlier findings, economic indicators do not provide economic value. By contrast, however, technical indicators deliver economic value. This value tends to increase over time – which is in stark contrast to economic indicators – and it is higher than the economic value of benchmark investment strategies over almost all subsamples. Providing a single figure on the predictive ability of technical indicators, their average annualized Sharpe ratio (over the rolling-recursive estimations) is about 0.6 and thus significantly higher than Sharpe ratio for buy and hold (0.47) or for applying a strategy relying on the historical mean (0.53).

Overall, predictive ability of economic and technical indicators seems to be of similar quality when assessed by their forecasting errors in the long-term. However, over time performance is completely different: economic indicators lose power but technical indicators remain powerful or even gain. Thus technical indicators perform more stable. When we complement statistical performance by economic values, which should be more important for the functioning of financial markets, the discrepancy between economic and technical indicators widens further. Only technical indicators provide economic values and they do this in a quite stable way. This may motivate further research in understanding the robustness of our finding and its origins.

References

Andersen, T.G., T. Bollerslev, F.X. Diebold, and P. Labys (2003), Modeling and Forecasting Realized Volatility, *Econometrica* 71 (2), 529-626.

Andrews, D.W.K. (1993), Tests for Parameter Instability and Structural Change with Unknown Change Point, *Econometrica* 61 (4), 821-856.

Ang, A., and G. Bekaert (2007), Stock Return Predictability: Is it There? *Review of Financial Studies* 20 (3), 651-707.

Bai, J. (1997), Estimation of a Change Point in Multiple Regressions, *Review of Economics and Statistics* 79 (4), 551-563.

Bai, J., and P. Perron (1998), Estimating and Testing Linear Models With Multiple Structural Changes, *Econometrica* 66 (1), 47-78.

Bai, J., and P. Perron (2006), Multiple Structural Change Models: A Simulation Analysis, in P.C.B. Phillips, D. Corbae, S. Durlauf, and B.E. Hansen, eds.: *Econometric Theory and Practice: Frontiers of Analysis and Applied Research*, Cambridge University Press, New York.

Bossaerts, P., and P. Hillion (1999), Implementing Statistical Criteria to Select Return Forecasting Models: What Do We Learn? *Review of Financial Studies* 12 (2), 405-428.

Brock, W., J. Lakonishok, and B. LeBaron (1992), Simple Technical Trading Rules and the Stochastic Properties of Stock Returns, *Journal of Finance* 47 (5), 1731-1764.

Brown, S.J., W.N. Goetzmann, and A. Kumar (1998), The Dow Theory: William Peter Hamilton's Track Record Reconsidered, *Journal of Finance* 53 (4), 1311-1333.

Campbell, J.Y. (2008), Viewpoint: Estimating the Equity Premium, *Canadian Journal of Economics* 41 (1), 1-21.

Campbell, J.Y., and R.J. Shiller (1988a), The Dividend-Price Ratio and Expectations of Future Dividends and Discount Factors, *Review of Financial Studies* 1 (3), 195-228.

Campbell, J.Y., and R.J. Shiller (1988b), Stock Prices, Earnings, and Expected Dividends, *Journal of Finance* 43 (3), 661-676.

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 (4), 1509-1531.

Cenesizoglu, T., and A. Timmermann (2012), Do Return Prediction Models Add Economic Value? *Journal of Banking and Finance* 36 (11), 2974-2987.

Christiansen, C., M. Schmeling, and A. Schrimpf (2012), A Comprehensive Look at Financial Volatility Prediction by Economic Variables, *Journal of Applied Econometrics* 27 (6), 956-977.

Clarke, T.E., and M.W. McCracken (2005), The Power of Tests of Predictive Ability in the Presence of Structural Breaks, *Journal of Econometrics* 124 (1), 1-31.

Clark, T.E., and K.D. West (2007), Approximately Normal Test for Equal Predictive Accuracy in Nested Models, *Journal of Econometrics* 138 (1), 291-311.

Cochrane, J.H. (1999), New Facts in Finance, Federal Reserve Bank of Chicago *Economic Perspectives* 23 (3), 36-58.

Cochrane, J.H. (2007), Financial Markets and the Real Economy, in R. Mehra, eds.: *Handbook of Equity Premium*, Elsevier, Amsterdam.

Cochrane, J.H. (2008), The Dog That Did Not Bark: A Defense of Return Predictability, *Review of Financial Studies* 21 (4), 1533-1575.

Cooper, I. and R. Priestley (2009), Time-Varying Risk Premiums and the Output Gap, *Review of Financial Studies* 22 (7), 2801-2833.

Elliott, G., and U.K. Müller (2006), Efficient Tests for General Persistent Time Variation in Regression Coefficients, *Review of Economic Studies* 73 (4), 907-940.

Fama, E.F., and K.R. French (1988), Dividend Yields and Expected Stock Returns, *Journal of Financial Economics* 22 (1), 3-25.

Fama, E.F., and K.R. French (1989), Business Conditions and Expected Returns on Stocks and Bonds, *Journal of Financial Economics* 25 (1), 23-49.

Goyal, A., and I. Welch (2003), Predicting the Equity Premium with Dividend Ratios, *Management Science* 49 (5), 639-654.

Goyal, A., and I. Welch (2008), A Comprehensive Look at the Empirical Performance of Equity Premium Prediction, *Review of Financial Studies* 21 (4), 1455-1508.

Granville, J. (1963), *Granville's New Key to Stock Market Profits*, Prentice-Hall, New York.

Hansen, B.E. (2000), Testing for Structural Change in Conditional Models, *Journal of Econometrics* 97 (1), 93-115.

Hansen, P.R., and A. Timmermann (2012), Choice of Sample Split in Out-of-Sample Forecast Evaluation, European University Institute Working Paper ECO 2012/10.

Henkel, S.J., J.S. Martin, and F. Nardari (2011), Time-Varying Short-Horizon Predictability, *Journal of Financial Economics* 99 (3), 560-580.

Inoue, A., and L. Kilian (2004), In-sample or Out-of-Sample Tests of Predictability: Which One Should We Use? *Econometric Reviews* 23 (4), 371-402.

Lettau, M., and S. Ludvigson (2001), Consumption, Aggregate Wealth, and Expected Stock Returns, *Journal of Finance* 56 (3), 815-849.

Lettau M., and S. Van Nieuwerburgh (2008), Reconciling the Return Predictability Evidence, *Review of Financial Studies* 21 (4), 1607-1652.

Lo, A.W., H. Mamaysky, and J. Wang (2000), Foundations of Technical Analysis: Computational Algorithms, Statistical Inference, and Empirical Implementation, *Journal of Finance* 55 (4), 1705-1765.

Marquering, W., and M. Verbeek (2004), The Economic Value of Predicting Stock Index Returns and Volatility, *Journal of Financial and Quantitative Analysis* 39 (2), 407-429.

Mele, A. (2007), Asymmetric Stock Market Volatility and the Cyclical Behavior of Expected Returns, *Journal of Financial Economics* 86 (2), 446-478.

Neely, C.J., D. Rapach, J. Tu, and G. Zhou (2014), Forecasting the Equity Risk Premium: The Role of Technical Indicators, *Management Science* 60 (7), 1772-1791.

Park, C.H., and S.H. Irwin (2007), What Do We Know about the Profitability of Technical Analysis? *Journal of Economic Surveys* 21 (4), 786-826.

Paye, B.S., and A. Timmermann (2006), Instability of Return Prediction Models, *Journal of Empirical Finance* 13 (3), 274-315.

Pesaran, M.H., and A. Timmermann (2002), Market Timing and Return Prediction under Model Instability, *Journal of Empirical Finance* 9 (5), 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 (3), 411-425.

Pesaran, M.H., and A. Timmermann (2007), Selection of Estimation Window in the Presence of Breaks, *Journal of Econometrics* 137 (1), 134-161.

Pettenuzzo, D., and A. Timmermann (2011), Predictability of Stock Returns and Asset Allocation under Structural Breaks, *Journal of Econometrics* 164 (1), 60-78.

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 (2), 238-274.

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 (2), 821-862.

Rapach, D.E., and G. Zhou (2013), Forecasting Stock Returns, in G. Elliott and A. Timmermann, eds.: *Handbook of Economic Forecasting*, Volume 2, Elsevier, Amsterdam.

Spiegel, M. (2008), Forecasting the Equity Premium: Where We Stand Today, *Review of Financial Studies* 21 (4), 1453-1454.

Stock, J.H., and M.W. Watson (2002a), Macroeconomic Forecasting Using Diffusion Indexes, *Journal of Business and Economic Statistics* 20 (2), 147-162.

Stock, J.H., and M.W. Watson (2002b), Forecasting Using Principal Components From a Large Number of Predictors, *Journal of the American Statistical Association* 97 (460), 1167-1179.

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

Zhu, X., and J. Zhu (2013), Predicting Stock Returns: A Regime-Switching Combination Approach and Economic Links, *Journal of Banking and Finance* 37 (11), 4120-4133.

Table I. Summary statistics

Variable	Mean	Std.	Skew.	Kurt.	AC(1)	AC(2)	AC(3)	Sharpe ratio
r_t	0.51	4.22	-0.66	5.38	0.06	-0.03	0.04	0.42
Economic variables								
DP	-3.50	0.42	-0.35	2.51	0.99	0.98	0.97	
DY	-3.50	0.42	-0.35	2.54	0.99	0.98	0.97	
EP	-2.78	0.43	-0.86	6.04	0.99	0.97	0.94	
DE	-0.72	0.30	2.54	18.10	0.99	0.95	0.90	
RVOL	0.14	0.05	0.79	3.83	0.96	0.92	0.88	
BM	0.53	0.25	0.49	2.58	0.99	0.99	0.98	
NTIS	0.01	0.02	-1.07	4.34	0.98	0.95	0.92	
TBL	4.53	3.02	0.90	4.26	0.99	0.97	0.95	
LTY	6.20	2.71	0.82	3.22	0.99	0.98	0.98	
LTR	0.53	2.76	0.52	6.36	0.05	-0.07	-0.02	
TMS	1.67	1.42	-0.08	2.82	0.96	0.90	0.86	
DFY	0.97	0.45	1.79	7.45	0.97	0.92	0.88	
DFR	0.02	1.39	-0.36	10.03	-0.08	-0.07	-0.03	
INFL	0.30	0.33	0.54	7.30	0.62	0.47	0.38	
Technical indicators								
MA(1,9)	0.69	0.46	-0.80	1.64	0.69	0.55	0.42	
MA(1,12)	0.71	0.45	-0.94	1.88	0.78	0.65	0.53	
MA(2,9)	0.69	0.46	-0.83	1.68	0.76	0.60	0.47	
MA(2,12)	0.71	0.45	-0.93	1.86	0.83	0.68	0.56	
MA(3,9)	0.70	0.46	-0.85	1.73	0.79	0.62	0.47	
MA(3,12)	0.71	0.45	-0.95	1.90	0.83	0.68	0.56	
MOM(9)	0.71	0.45	-0.92	1.85	0.76	0.69	0.58	
MOM(12)	0.73	0.45	-1.02	2.05	0.81	0.72	0.64	
VOL(1,9)	0.68	0.47	-0.75	1.56	0.60	0.54	0.41	
VOL(1,12)	0.70	0.46	-0.88	1.78	0.70	0.64	0.50	
VOL(2,9)	0.67	0.47	-0.73	1.53	0.75	0.56	0.46	
VOL(2,12)	0.70	0.46	-0.85	1.73	0.82	0.64	0.56	
VOL(3,9)	0.69	0.46	-0.81	1.66	0.76	0.58	0.45	
VOL(3,12)	0.70	0.46	-0.86	1.74	0.83	0.70	0.58	

Notes: The table reports summary statistics, including mean, standard deviation (Std.), Skewness (Skew.) and Kurtosis (Kurt.) of the monthly log equity premium (in percent) and predictor variables stemming from economic and technical indicators. We also report the first to third-order autocorrelation coefficient AC(.) and the annualized Sharpe ratio for the log equity premium. The sample period is December 1950 to December 2013. A full description of the data is given in the Data Appendix.

Table II. Out-of-sample forecasting evaluation

| Predictor	R^2_{OS}	R^2_{OS} expansion	R^2_{OS} recession	Predictor	R^2_{OS}	R^2_{OS} expansion	R^2_{OS} recession
Panel A: Bivariate predictive regression forecasts							
DP	-0.19%	-1.01%	1.55%*	MA(1,9)	0.29%	-0.71%	2.39%**
DY	-0.13%	-1.32%	2.39%**	MA(1,12)	0.68%*	-0.56%	3.30%**
EP	-0.58%	-0.30%	-1.17%	MA(2,9)	0.38%	-0.64%	2.54%**
DE	-0.88%	-1.71%	0.89%	MA(2,12)	0.83%**	-0.45%	3.53%**
RVOL	0.09%*	-0.12%*	0.53%	MA(3,9)	0.46%*	-0.71%	2.94%**
BM	-1.26%	-0.30%	-3.30%	MA(3,12)	0.08%	-0.45%	1.19%
NTIS	-0.92%	-0.18%	-2.47%	MOM(9)	0.11%	-0.47%	1.34%*
TBL	-0.88%**	-1.97%**	1.42%*	MOM(12)	0.15%	-0.44%	1.39%*
LTY	-0.80%*	-1.64%	0.98%	VOL(1,9)	0.46%*	-0.56%	2.61%**
LTR	0.20%**	-2.02%	4.90%***	VOL(1,12)	0.78%**	-0.23%	2.94%**
TMS	-0.87%**	-3.20%	4.06%**	VOL(2,9)	0.45%*	0.01%	1.37%*
DFY	-0.63%	-0.54%	-0.83%	VOL(2,12)	0.33%	0.16%	0.68%
DFR	-0.35%	0.46%*	-2.06%	VOL(3,9)	0.02%	-0.39%	0.89%
INFL	-0.27%	0.17%	-1.19%	VOL(3,12)	0.66%*	0.07%	1.90%*
Panel B: Principal component predictive regression forecasts							
PC_{Econ}	-0.47%**	-3.38%	5.69%**	PC_{Tech}	0.67%*	-0.36%	2.84%**
Panel C: Principal component predictive regressions forecast, all predictors							
PC_{ALL}	0.71%***	-2.06%	6.56%***				

Notes: Table reports the out-of-sample R^2_{OS} statistic proposed by Campbell and Thompson (2008) of log equity risk premium predictability over the evaluation period 1966:01 to 2013:12 and the predictive performance separately for NBER-dated expansions and recessions. Stars refer to significance level of 10% (*), 5% (**), and 1% (***) of the MSFE-adj. test statistic proposed by Clark and West (2007). Panel A reveals results for bivariate predictive models; Panel B shows out-of-sample results using principal component extracted from the full set of macroeconomic variables (PC_{Econ}) and technical indicators (PC_{Tech}); and Panel C indicates predictive performance by taking economic and technical indicators simultaneously into account (PC_{ALL}). The number of factor is selected according to the SIC information criterion.

Figure I: Dynamic predictive performance at any point of time
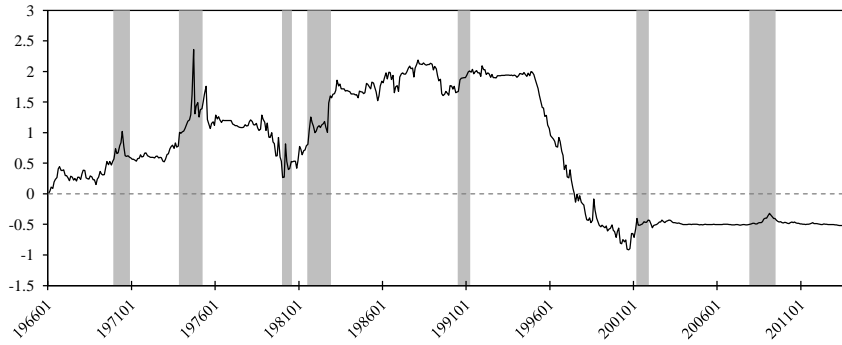
Panel A: PC$_{Econ}$
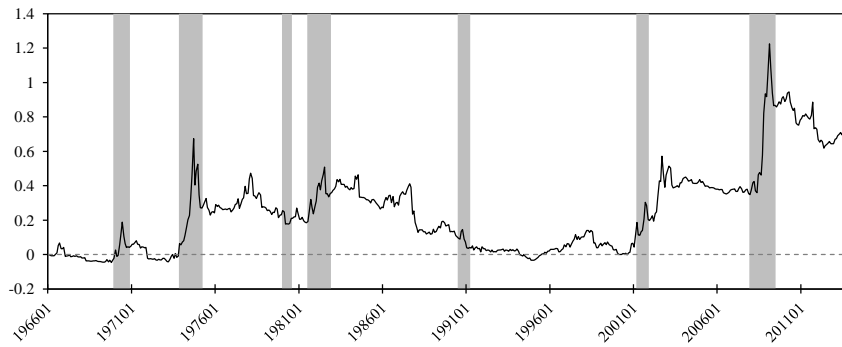
Panel B: PC$_{Tech}$
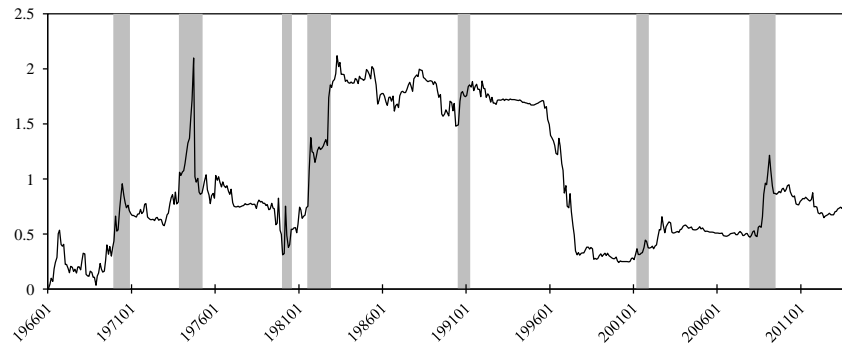
Panel C: PC$_{ALL}$

Notes: The figures plots the dynamic out-of-sample predictive performance of forecasts based principal component from the full set of macroeconomic variables (PC$_{Econ}$) and technical indicators (PC$_{Tech}$) as well as taking both predictor groups simultaneously into account (PC$_{ALL}$). Following Goyal and Welch (2003, 2008), the graphs show the cumulative sum of differences in the squared prediction errors under the benchmark specification and the squared prediction errors based on information regarding predictive variables: $CDSFE(s,i) = \sum_{s=1}^{T}((r_s - \bar{r}_s)^2 - (r_s - \hat{r}_{s,i})^2)$ where $(r_s - \bar{r}_s)^2$ are squared prediction errors of the historical average and $(r_s - \hat{r}_{s,i})^2$ are the forecasting errors of model i named in the headings. Shaded areas respond to NBER dated recessions. Overall, upward sloping curves characterize lower MSPE by make use of predictor variables.

Table III. Equity premium predictive regression and structural break tests

Predictor	$SupF$	Breakpoint	$UDmax$	$WDmax$ (5%)	$\widehat{qLL}$	Predictor	$SupF$	Breakpoint	$UDmax$	$WDmax$ (5%)	$\widehat{qLL}$
Panel A: Bivariate predictive regression forecasts											
DP	9.33	1994:10	9.27	13.57†	-10.59	MA(1,9)	7.59	2000:08	9.19	11.17	-10.40
DY	10.93	1994:10	10.47*	13.99†	-10.87	MA(1,12)	3.86	1998:09	6.79	8.65	-7.51
EP	3.53	1982:06	11.77**	14.87†	-10.20	MA(2,9)	4.20	2000:09	6.49	10.04	-7.04
DE	14.22**	1974:08	18.18***	18.18†	-7.40	MA(2,12)	3.38	1961:10	4.62	6.90	-6.78
RVOL	4.99	1961:10	6.89	8.04	-10.93	MA(3,9)	3.28	1961:10	5.81	8.08	-7.46
BM	6.56	1969:03	10.53*	12.60	-9.71	MA(3,12)	3.61	1961:10	5.31	7.83	-7.86
NTIS	9.86	2003:02	10.77*	13.82†	-16.38**	MOM(9)	3.68	1961:10	5.04	7.33	-7.65
TBL	18.92***	1974:08	23.71***	23.71†	-10.26	MOM(12)	3.85	1961:10	5.79	8.09	-7.35
LTY	16.19**	1974:08	20.53***	20.53†	-8.90	VOL(1,9)	4.24	1961:10	5.80	8.52	-6.35
LTR	4.88	1961:10	6.82	7.29	-8.82	VOL(1,12)	4.92	1964:06	6.71	9.21	-7.31
TMS	12.11**	1975:05	14.19**	14.19†	-12.00	VOL(2,9)	8.30	1969:04	10.41*	11.21	-8.64
DFY	4.78	1961:10	11.15*	13.12†	-13.43*	VOL(2,12)	7.74	1965:09	10.10	11.97	-9.30
DFR	9.99	1973:01	11.93**	11.93	-8.96	VOL(3,9)	4.52	2000:01	10.71*	13.74†	-8.73
INFL	7.76	1974:08	9.80	11.53	-9.33	VOL(3,12)	4.88	2000:01	7.67	10.36	-8.48
Panel B: Principal component predictive regression forecasts											
PC$_{Econ}$(1st)	6.95	1982:06	10.74*	12.63	9.44	PC$_{Tech}$(1st)	3.74	2000:02	6.47	8.81	-6.92
PC$_{Econ}$(1st-2nd)	10.21	1967:03	14.25**	17.48†	-19.88**	PC$_{Tech}$(1st-2nd)	6.21	1965:03	8.39	11.29	-12.18
PC$_{Econ}$(1st-3rd)	30.70***	1994:11	29.32***	29.32†	-21.19	PC$_{Tech}$(1st-3rd)	7.15	1965:03	9.35	12.09	-15.55
Panel C: Principal component predictive regression forecasts, all predictors											
PC$_{All}$(1st)	3.98	2000:02	6.39	10.20	-6.68						
PC$_{All}$(1st-2nd)	4.85	1982:06	9.69	13.50	-9.73						
PC$_{All}$(1st-3rd)	11.05	1994:11	14.25**	18.62†	-18.57						
PC$_{All}$(1st-4th)	20.30	1994:11	19.54***	20.01†	-18.28						

Notes: This table reports several test statistics to analyze whether an occasional change exists over the full sample period 1950:12-2013:12. We employ Andrews (1993) *SupF* statistic and estimated breakpoints with stars refer to significance levels of 10% (*), 5% (**), and 1% (***) based the heteroskedastic fixed-regressor bootstrap proposed by Hansen (2000). *SupF* statistic tests the null hypothesis of no structural break against the one-sided alternative that a structural change exists. We also account for multiple breaks (following Bai and Perron, 1998) by testing the null hypothesis of zero breaks against the alternative of at most 5 breaks. 10%, 5%, and 1% critical values equal 10.16, 11.70, 15.41 for *UDmax* (stars refer to significance at corresponding levels) and the 5% critical value for *WDmax* equals 12.81; significance is indicated by †. In addition, $\widehat{qLL}$ indicates the tests statistic proposed by Elliot and Müller (2006) with stars refer to significance levels of 10% (*), 5% (**), and 1% (***).

Figure II: Time-varying predictive performance
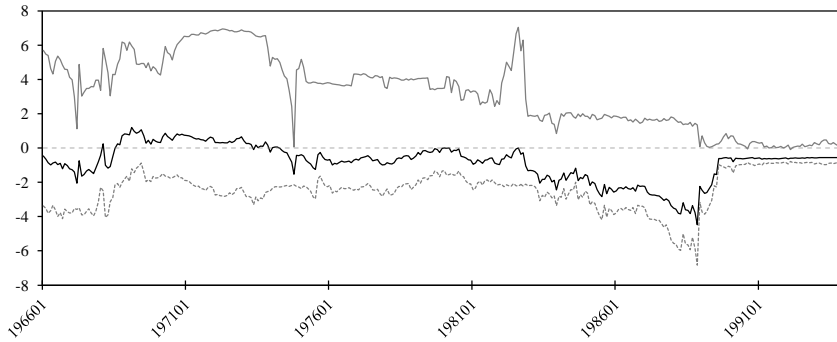
Panel A: PC$_{Econ}$

Panel B: PC$_{Tech}$
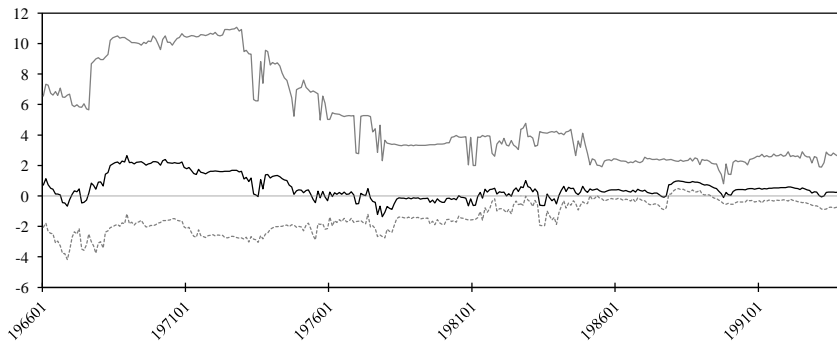
Panel C: PC$_{ALL}$

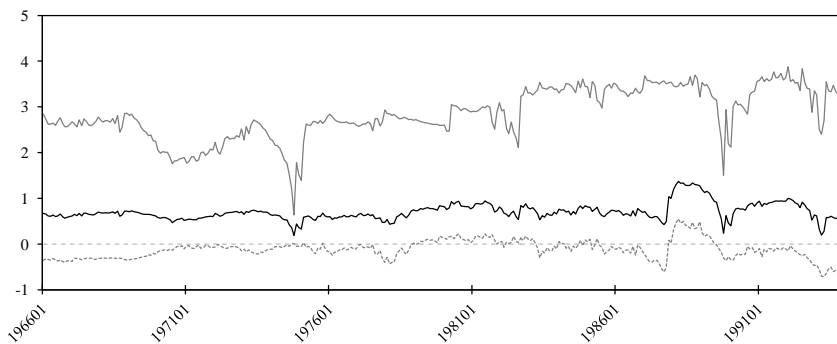*Notes:* These figures show the time-varying out-of-sample predictive performance measured by the R^2_{OS}, over different subsamples. Our analysis starts with recursive forecast estimation over an initial in-sample estimation period of 15 years (1950:12-1965:12) and conduct real-time forecasts up to 2013:12. Next we discard the most distant data (i.e. 1950:12), yielding an in-sample estimation sample 1951:01-1966:01 (15 years) and perform out-of-sample forecasts up to 2013:12, the most recent data of our sample period. The beginning of the out-of-sample evaluation period is given on the x-axis. Thus the last R^2_{OS} is obtained over the sample period 1994:01-2013:12. The black line shows the full sample R^2_{OS}, the grey solid line signals the R^2_{OS} regarding recessions and the grey dotted line corresponds to the R^2_{OS} over expansions. Corresponding predictive regressions are named in the headings.

Figure III: Equity market volatility
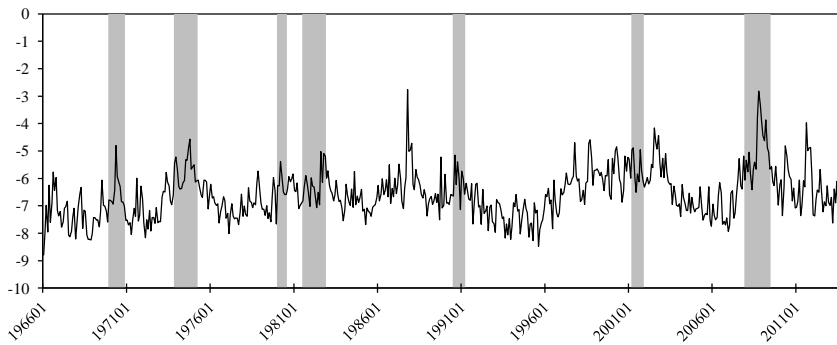
Panel A: Log realized volatility
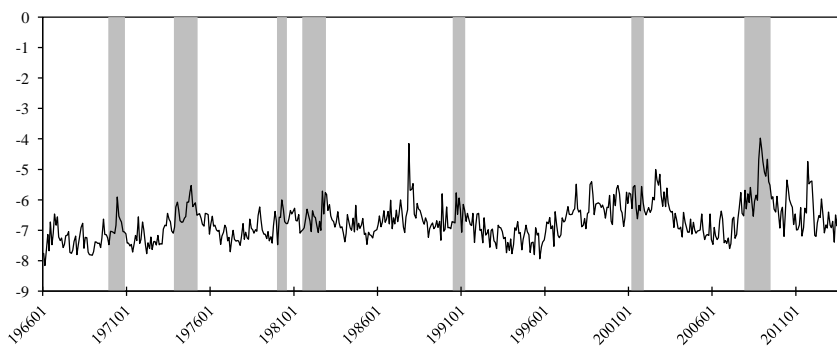
Panel B: Log realized volatility forecast, AR(1)

Notes: These figures show the realized volatility over the sample period 1966:01 through 2013:12. Panel A shows the log realized volatility while Panel B shows the corresponding out-of-sample forecast estimated recursively by an AR(1)-process. Shaded areas respond to NBER dated recessions.

Table IV. Economic measures of forecasting performance

Predictor	Mean	Std.	ΔCER (ann.)	SR	Predictor	Mean	Std.	ΔCER (ann.)	SR
HA	0.99%	5.35	6.73%	0.36					

Panel A: Bivariate predictive regression forecasts

Predictor	Mean	Std.	ΔCER (ann.)	SR	Predictor	Mean	Std.	ΔCER (ann.)	SR
DP	0.78%	4.33	-0.79%	0.28	MA(1,9)	0.97%	4.91	0.62%	0.38
DY	0.74%	4.25	-1.07%	0.25	MA(1,12)	1.09%	4.96	1.86%	0.46
EP	0.98%	5.03	0.49%	0.38	MA(2,9)	1.00%	4.97	0.77%	0.39
DE	0.84%	4.31	0.05%	0.33	MA(2,12)	1.10%	4.96	2.05%	0.47
RVOL	1.05%	5.57	0.25%	0.38	MA(3,9)	1.05%	4.88	1.53%	0.44
BM	0.93%	5.16	-0.40%	0.33	MA(3,12)	0.99%	5.06	0.49%	0.38
NTIS	0.99%	5.41	-0.15%	0.36	MOM(9)	0.99%	5.14	0.35%	0.37
TBL	0.95%	3.96	1.82%	0.45	MOM(12)	0.99%	5.18	0.37%	0.38
LTY	0.91%	3.67	1.73%	0.45	VOL(1,9)	1.01%	5.05	0.76%	0.39
LTR	1.08%	4.96	1.80%	0.45	VOL(1,12)	1.06%	4.90	1.65%	0.44
TMS	1.19%	5.05	3.00%	0.52	VOL(2,9)	1.03%	5.17	0.85%	0.40
DFY	1.02%	5.49	0.04%	0.37	VOL(2,12)	1.02%	5.13	0.72%	0.39
DFR	0.98%	5.25	0.05%	0.36	VOL(3,9)	0.96%	5.08	0.19%	0.36
INFL	0.98%	4.84	0.81%	0.39	VOL(3,12)	1.02%	4.98	1.10%	0.41

Panel B: Principal component predictive regression forecasts

Predictor	Mean	Std.	ΔCER (ann.)	SR	Predictor	Mean	Std.	ΔCER (ann.)	SR
PC_{Econ}	1.01%	3.89	2.70%	0.52	PC_{Tech}	1.11%	5.02	2.05%	0.47

Panel C: Principal component predictive regression forecasts, all predictors

Predictor	Mean	Std.	ΔCER (ann.)	SR
PC_{ALL}	1.10%	4.37	2.99%	0.53

Notes: The Table reports means and standard deviations (Std.) of portfolio returns for a mean-variance investor with relative risk aversion coefficient of three on the corresponding forecasting strategies over the sample period 1966:01-2013:12. Conditional variance forecasts are obtained by an AR(1) process of stock returns realized volatility (in logs). ΔCER denotes the annualized certainty equivalent return gain of predictive regression forecasts in comparison to the historical average forecast and SR is the annualized Sharpe ratio defined as the average portfolio excess return divided by the sample standard deviation. HA indicates the historical average forecast where portfolio performance measures are given in levels, Panel A reveals results for bivariate predictive models; Panel B shows out-of-sample results using principal component extracted from the full set of macroeconomic variables (PC_{Econ}) and technical indicators (PC_{Tech}); and Panel C indicates predictive performance by taking economic and technical indicators simultaneously into account (PC_{ALL}). The number of factor is selected according to the SIC information criterion.

Figure IV: Time-varying Sharpe ratios of forecasting strategies, (no transaction costs)
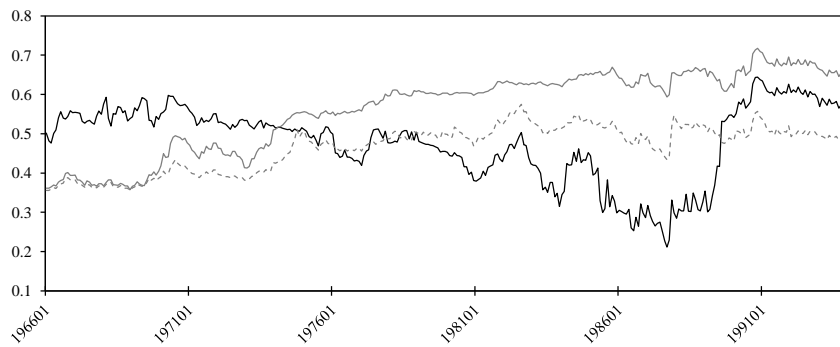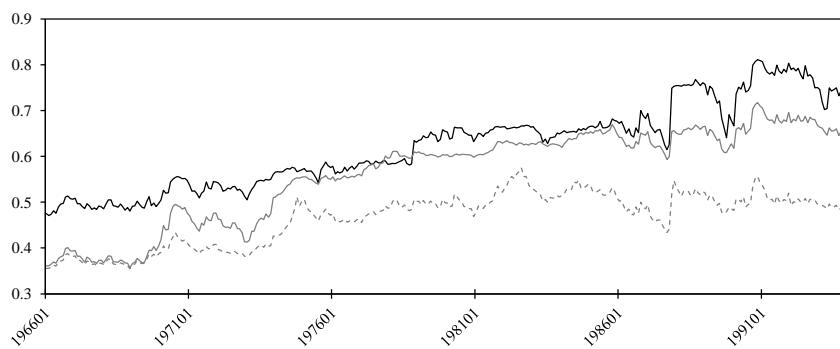

Panel A: PC$_{Econ}$

Panel B: PC$_{Tech}$

Panel C: PC$_{ALL}$

Notes: These figures show the time-varying annualized Sharpe ratio using a rolling-recursive estimation setting of forecasting strategies named in the headings. Our analysis starts with recursive forecast estimation over an initial in-sample estimation period of 15 years (1950:12-1965:12) and conduct real-time forecasts up to 2013:12. Next we discard the most distant data (i.e. 1950:12), yielding an in-sample estimation sample 1951:01-1966:01 (15 years) and perform out-of-sample forecasts up to 2013:12, the most recent data of our sample period. The beginning of the portfolio formation period is given on the x-axis. Thus the last Sharpe ratio corresponds to the sample period 1994:01-2013:12. We assume a relative risk-aversion coefficient of three and zero transactions. The black line shows the annualized Sharpe ratio of predictive regression forecasts, the grey solid line indicates the Sharpe ratio based on historical average forecasts and the grey dotted line corresponds to a simple buy-and-hold strategy.
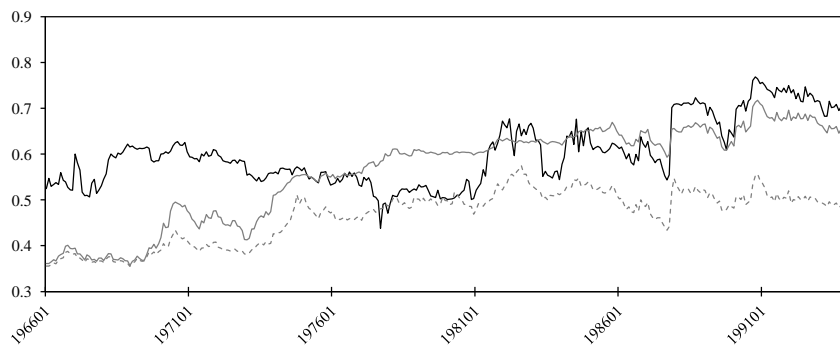
Figure V: Time-varying Sharpe ratios of forecasting strategies, (transaction costs = 50bp)
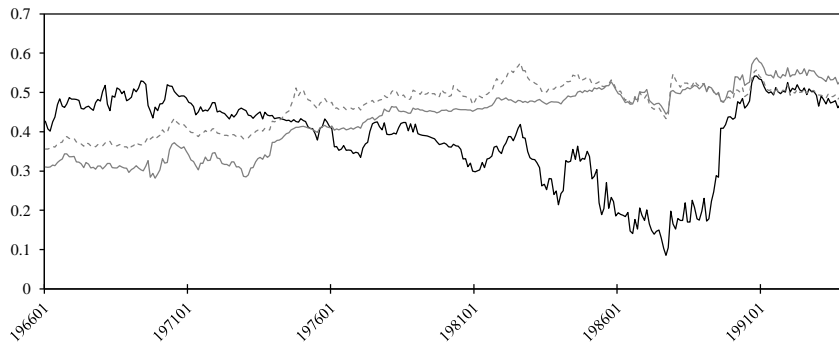

Panel A: PC$_{Econ}$
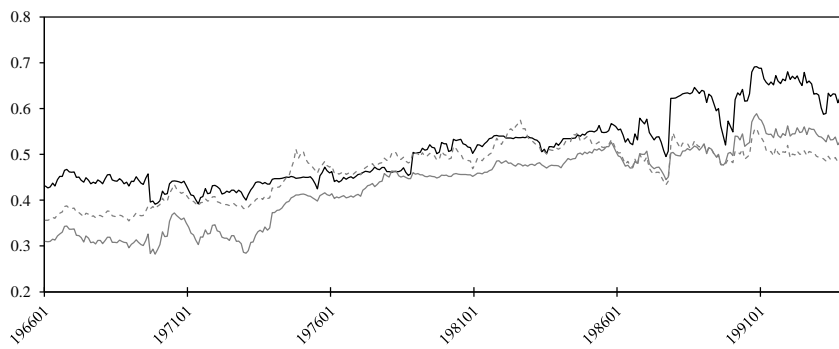

Panel B: PC$_{Tech}$

Panel C: PC$_{ALL}$

Notes: These figures show the time-varying annualized Sharpe ratio using a rolling-recursive estimation setting of forecasting strategies named in the headings. Our analysis starts with recursive forecast estimation over an initial in-sample estimation period of 15 years (1950:12-1965:12) and conduct real-time forecasts up to 2013:12. Next we discard the most distant data (i.e. 1950:12), yielding an in-sample estimation sample 1951:01-1966:01 (15 years) and perform out-of-sample forecasts up to 2013:12, the most recent data of our sample period. The beginning of the portfolio formation period is given on the x-axis. Thus the last Sharpe ratio corresponds to the sample period 1994:01-2013:12. We assume a relative risk-aversion coefficient of three and transaction costs of 50bp. The black line shows the annualized Sharpe ratio of predictive regression forecasts, the grey solid line indicates the Sharpe ratio based on historical average forecasts and the grey dotted line corresponds to a simple buy-and-hold strategy.
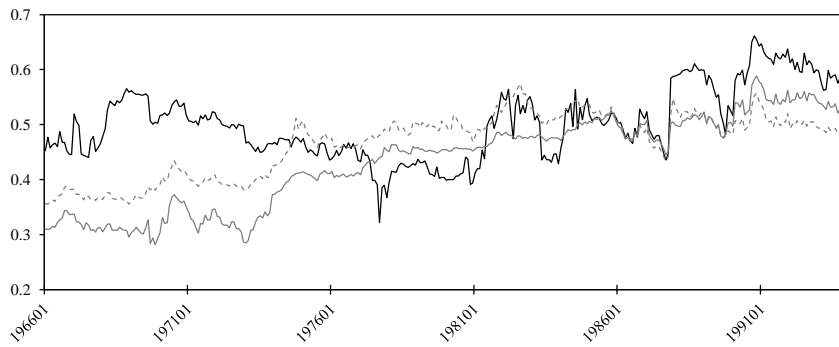
Table V. Out-of-sample forecasting evaluation using alternative pooling strategies

Pooling strategy	R^2_{OS}	R^2_{OS} expansion	R^2_{OS} recession
Panel A: Principal component predictive regression and combination forecasts; Econ			
Adj. R^2	-0.72%***	-3.73%*	5.65%**
AIC	-0.81%***	-4.02%*	5.99%**
Mean	1.16%***	0.75%***	2.00%**
Median	0.71%***	0.63%**	0.87%**
Trimmed Mean	1.07%***	0.74%***	1.77%**
DMSPE(1.0)	1.18%***	0.80%***	2.00%**
DMSPE(0.9)	1.20%***	0.73%**	2.19%**
Panel B: Principal component predictive regression and combination forecasts; Tech			
Adj. R^2	0.22%	-0.56%	1.87%*
AIC	0.59%*	-0.48%	2.84%**
Mean	0.55%*	-0.21%	2.15%**
Median	0.71%**	-0.04%	2.31%**
Trimmed Mean	0.61%*	-0.13%	2.17%**
DMSPE(1.0)	0.55%*	-0.21%	2.16%**
DMSPE(0.9)	0.55%*	-0.21%	2.17%**
Panel C: Principal component predictive regression and combination forecasts; ALL			
Adj. R^2	1.40%***	-3.22%	11.19%***
AIC	0.57%***	-3.78%	9.74%***
Mean	0.95%***	0.38%*	2.16%**
Median	0.91%**	0.29%	2.22%**
Trimmed Mean	0.90%***	0.36%*	2.05%**
DMSPE(1.0)	1.01%***	0.46%**	2.17%**
DMSPE(0.9)	1.01%***	0.41%*	2.27%**

Note: This table shows various specifications (Adj. R^2 and AIC) to determine the maximum number of common factors used for principal component predictive regressions. We also make use of forecast combinations (following Rapach et al., 2010) as an alternative to incorporate individual information. Presented results indicate the out-of-sample R^2_{OS} statistic proposed by Campbell and Thompson (2008) for the full sample period 1966:01 to 2013:12 and the predictive performance separately for NBER-dated expansions and recessions. Stars refer to significance level of 10% (*), 5% (**) and 1% (***) of the MSFE-adj. test statistic proposed by Clark and West (2007).

Figure VI: Time-varying predictive performance; alternative pooling strategies
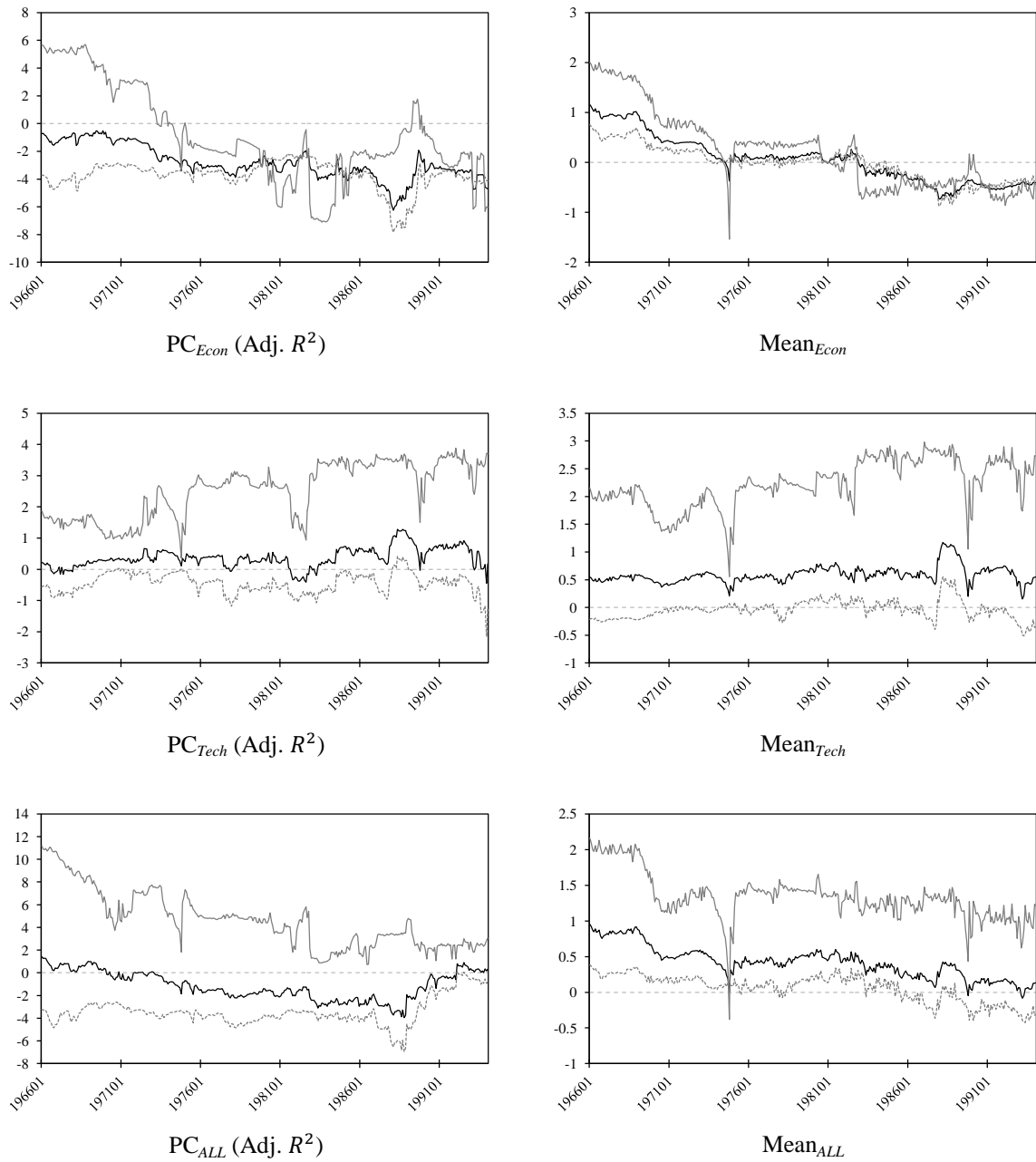
PC_{Econ} (Adj. R^2)

$Mean_{Econ}$

PC_{Tech} (Adj. R^2)

$Mean_{Tech}$

PC_{ALL} (Adj. R^2)

$Mean_{ALL}$

Notes: These figures show the time-varying out-of-sample predictive performance measured by the R^2_{OS}, using a rolling-recursive estimation setting. Our analysis starts with recursive forecast estimation over an initial in-sample estimation period of 15 years (1950:12-1965:12) and conduct real-time forecasts up to 2013:12. Next we discard the most distant data (i.e. 1950:12), yielding an in-sample estimation sample 1951:01-1966:01 (15 years) and perform out-of-sample forecasts up to 2013:12, the most recent data of our sample period. The beginning of the out-of-sample evaluation period is given on the x-axis. Thus the last R^2_{OS} is obtained over the sample period 1994:01-2013:12. The black line shows the full sample R^2_{OS}, the grey solid line signals the R^2_{OS} regarding recessions and the grey dotted line corresponds to the R^2_{OS} over expansions. Corresponding predictive regressions are named in the headings

37

Table VI. Out-of-sample forecasting evaluation (log realized volatility)

Predictor	R^2_{OS}	R^2_{OS} expansion	R^2_{OS} recession	Predictor	R^2_{OS}	R^2_{OS} expansion	R^2_{OS} recession
Panel A: Bivariate predictive regression forecasts							
DP	1.14%***	0.42%*	3.52%***	MA(1,9)	0.71%**	-0.12%	3.43%***
DY	1.72%***	1.31%***	3.07%**	MA(1,12)	0.61%**	-0.84%	5.40%***
EP	0.45%**	-2.05%	8.68%***	MA(2,9)	-0.00%	-0.90%	2.95%***
DE	-0.93%***	-2.79%**	5.21%**	MA(2,12)	0.37%**	-1.14%	5.35%***
RVOL	1.86%***	2.46%***	-0.12%	MA(3,9)	0.06%	-1.40%	4.86%***
BM	-0.09%**	-4.20%	13.44%***	MA(3,12)	0.42%**	-1.08%	5.39%***
NTIS	0.51%**	-0.42%*	3.59%**	MOM(9)	0.76%**	-0.91%	6.28%***
TBL	-0.96%**	-2.52%	4.17%**	MOM(12)	0.68%**	-0.74%	5.33%***
LTY	-0.29%***	-1.99%*	5.31%***	VOL(1,9)	0.61%**	-0.53%	4.39%***
LTR	-0.65%	-0.47%	-1.27%	VOL(1,12)	1.22%***	-0.18%	5.83%***
TMS	-0.28%	-0.60%	0.78%	VOL(2,9)	-0.39%	-0.79%	0.94%**
DFY	2.11%***	-0.09%*	9.37%***	VOL(2,12)	0.03%	-0.62%	2.17%***
DFR	0.21%	-0.12%	1.30%	VOL(3,9)	-0.04%	-0.68%	2.06%***
INFL	-0.24%	-1.27%	3.14%**	VOL(3,12)	0.28%*	-0.72%	3.60%***
Panel B: Principal component predictive regression forecasts							
PC$_{Econ}$	3.67%***	0.02%***	3.67%***	PC$_{Tech}$	0.75%**	-0.98%	6.45%***
Panel C: Principal component predictive regression forecasts, all predictors							
PC$_{ALL}$	5.35%***	1.44%***	18.24%***				

Notes: Table reports the out-of-sample R^2_{OS} statistic proposed by Campbell and Thompson (2008) of log realized volatility predictability over the evaluation period 1966:01 to 2013:12 and the predictive performance separately for NBER-dated expansions and recessions. Stars refer to significance level of 10% (*), 5% (**), and 1% (***) of the MSFE-adj. test statistic proposed by Clark and West (2007). Panel A reveals results for bivariate predictive models; Panel B shows out-of-sample results using principal component extracted from the full set of macroeconomic variables (PC$_{Econ}$) and technical indicators (PC$_{Tech}$); and Panel C indicates predictive performance by taking economic and technical indicators simultaneously into account (PC$_{ALL}$). The number of factor is selected according to the SIC information criterion.

Table VII. Economic value of equity premium and volatility forecasts

Predictor	Mean	Std.	ΔCER (ann.)	SR	Predictor	Mean	Std.	ΔCER (ann.)	SR
HA	0.99%	5.35	6.73%	0.36					

Panel A: Bivariate predictive regression forecasts

Predictor	Mean	Std.	ΔCER (ann.)	SR	Predictor	Mean	Std.	ΔCER (ann.)	SR
DP	0.77%	4.32	-0.82%	0.27	MA(1,9)	0.99%	4.91	0.75%	0.39
DY	0.74%	4.24	-1.11%	0.25	MA(1,12)	1.09%	4.97	1.97%	0.46
EP	0.99%	5.10	0.42%	0.38	MA(2,9)	1.00%	4.97	0.85%	0.40
DE	0.83%	4.28	-0.11%	0.32	MA(2,12)	1.11%	4.94	2.17%	0.47
RVOL	1.04%	5.61	0.09%	0.38	MA(3,9)	1.05%	4.87	1.63%	0.44
BM	0.93%	5.20	-0.41%	0.33	MA(3,12)	1.00%	5.07	0.62%	0.39
NTIS	0.99%	5.41	-0.08%	0.36	MOM(9)	1.00%	5.16	0.51%	0.38
TBL	0.96%	4.05	1.82%	0.45	MOM(12)	1.01%	5.21	0.49%	0.38
LTY	0.91%	3.72	1.71%	0.44	VOL(1,9)	1.01%	5.04	0.87%	0.40
LTR	1.08%	4.97	1.80%	0.45	VOL(1,12)	1.07%	4.91	1.82%	0.45
TMS	1.19%	5.04	2.97%	0.52	VOL(2,9)	1.04%	5.17	0.91%	0.41
DFY	1.01%	5.50	-0.05%	0.36	VOL(2,12)	1.02%	5.14	0.81%	0.40
DFR	0.98%	5.27	0.05%	0.36	VOL(3,9)	0.97%	5.10	0.24%	0.37
INFL	0.99%	4.86	0.84%	0.39	VOL(3,12)	1.03%	4.99	1.21%	0.42

Panel B: Principal component predictive regression forecasts

Predictor	Mean	Std.	ΔCER (ann.)	SR	Predictor	Mean	Std.	ΔCER (ann.)	SR
PC_{Econ}	1.00%	3.90	2.49%	0.50	PC_{Tech}	1.12%	5.01	2.19%	0.48

Panel C: Predictive regressions, all predictors taken together

Predictor	Mean	Std.	ΔCER (ann.)	SR
PC_{ALL}	1.09%	4.37	2.96%	0.52

Notes: The Table reports means and standard deviations (Std.) of portfolio returns for a mean-variance investor with relative risk aversion coefficient of three on the corresponding forecasting strategies over the sample period 1966:01-2013:12. Portfolio performance measures are based on both equity premium and volatility forecasts ΔCER denotes the annualized certainty equivalent return gain of predictive regression forecasts in comparison to the historical average forecast and SR is the annualized Sharpe ratio defined as the average portfolio excess return divided by the sample standard deviation. HA indicates the historical average forecast where portfolio performance measures are given in levels, Panel A reveals results for bivariate predictive models; Panel B shows out-of-sample results using principal component extracted from the full set of macroeconomic variables (PC_{Econ}) and technical indicators (PC_{Tech}); and Panel C indicates predictive performance by taking economic and technical indicators simultaneously into account (PC_{ALL}). The number of factor is selected according to the SIC information criterion.

Table VIII. Alternative settings for forecast performance evaluation

(a) Relative risk aversion: $\gamma=3$

| | $0 \leq w_s \leq 1$ | | | | | | $0 \leq w_s \leq 1.5$ | | | | | | $-1 \leq w_s \leq 2$ | | | | | |
| | c=0 | | | c=50 | | | c=0 | | | c=50 | | | c=0 | | | c=50 | | |
	ΔCER (ann.)	ΔSR	Avg. ΔSR	ΔCER (ann.)	ΔSR	Avg. ΔSR	ΔCER (ann.)	ΔSR	Avg. ΔSR	ΔCER (ann.)	ΔSR	Avg. ΔSR	ΔCER (ann.)	ΔSR	Avg. ΔSR	ΔCER (ann.)	ΔSR	Avg. ΔSR
PC_{Econ}(SIC)	0.61%	0.07	-0.02	0.29%	0.03	-0.02	2.70%	0.16	-0.04	2.61%	0.13	-0.03	2.66%	0.10	-0.12	2.37%	0.07	-0.11
PC_{Econ}(Adj.R^2)	0.74%	0.08	-0.11	0.36%	0.03	-0.12	2.43%	0.13	-0.14	2.22%	0.10	-0.14	2.37%	0.09	-0.25	1.75%	0.05	-0.25
$Mean_{Econ}$	0.64%	0.05	0.04	0.44%	0.03	0.03	1.62%	0.08	0.03	1.32%	0.06	0.01	2.60%	0.11	0.02	2.21%	0.08	0.00
PC_{Tech}(SIC)	1.31%	0.11	0.12	1.27%	0.10	0.14	2.05%	0.11	0.07	2.21%	0.11	0.09	1.25%	0.05	0.03	1.34%	0.06	0.04
PC_{Tech}(Adj.R^2)	1.00%	0.08	0.10	0.84%	0.07	0.11	1.73%	0.09	0.05	1.71%	0.08	0.07	0.79%	0.03	-0.01	0.56%	0.02	-0.00
$Mean_{Tech}$	0.66%	0.05	0.09	0.59%	0.04	0.10	1.29%	0.06	0.04	1.41%	0.07	0.06	1.06%	0.04	0.02	1.18%	0.05	0.03
PC_{ALL}(SIC)	1.33%	0.13	0.10	0.97%	0.09	0.11	2.99%	0.17	0.07	2.82%	0.14	0.10	3.42%	0.15	0.01	2.97%	0.12	0.03
PC_{ALL}(Adj.R^2)	1.63%	0.18	0.01	1.10%	0.12	-0.02	3.70%	0.23	-0.00	3.30%	0.18	-0.01	3.83%	0.16	-0.12	2.65%	0.09	-0.14
$Mean_{ALL}$	0.68%	0.05	0.07	0.59%	0.04	0.07	1.44%	0.07	0.05	1.38%	0.07	0.05	1.80%	0.07	0.04	1.70%	0.06	0.03

(b) Relative risk aversion: $\gamma=5$

| | $0 \leq w_s \leq 1$ | | | | | | $0 \leq w_s \leq 1.5$ | | | | | | $-1 \leq w_s \leq 2$ | | | | | |
| | c=0 | | | c=50 | | | c=0 | | | c=50 | | | c=0 | | | c=50 | | |
	ΔCER (ann.)	ΔSR	Avg. ΔSR	ΔCER (ann.)	ΔSR	Avg. ΔSR	ΔCER (ann.)	ΔSR	Avg. ΔSR	ΔCER (ann.)	ΔSR	Avg. ΔSR	ΔCER (ann.)	ΔSR	Avg. ΔSR	ΔCER (ann.)	ΔSR	Avg. ΔSR
PC_{Econ}(SIC)	1.98%	0.17	-0.05	1.97%	0.15	-0.03	3.22%	0.21	-0.08	3.47%	0.20	-0.05	1.91%	0.11	-0.15	1.83%	0.10	-0.12
PC_{Econ}(Adj.R^2)	1.77%	0.15	-0.15	1.69%	0.12	-0.14	2.77%	0.17	-0.17	2.94%	0.17	-0.15	1.54%	0.12	-0.27	1.04%	0.09	-0.26
$Mean_{Econ}$	1.18%	0.09	0.03	0.97%	0.07	0.01	2.12%	0.13	0.02	1.88%	0.10	0.01	2.82%	0.14	0.01	2.57%	0.11	-0.00
PC_{Tech}(SIC)	1.15%	0.09	0.05	1.29%	0.10	0.08	1.12%	0.08	0.06	1.17%	0.08	0.07	0.99%	0.07	0.03	0.79%	0.06	0.02
PC_{Tech}(Adj.R^2)	0.94%	0.07	0.04	0.95%	0.07	0.06	0.63%	0.04	0.04	0.50%	0.04	0.05	0.03%	0.02	-0.01	-0.50%	-0.01	-0.03
$Mean_{Tech}$	0.72%	0.05	0.04	0.82%	0.06	0.06	0.65%	0.04	0.05	0.67%	0.04	0.05	0.89%	0.05	0.03	0.76%	0.05	0.02
PC_{ALL}(SIC)	2.05%	0.17	0.05	1.99%	0.15	0.09	3.28%	0.22	0.04	3.32%	0.21	0.07	3.01%	0.20	0.02	2.47%	0.17	0.03
PC_{ALL}(Adj.R^2)	2.60%	0.24	-0.00	2.39%	0.20	-0.00	4.18%	0.29	-0.00	4.16%	0.27	0.00	3.48%	0.22	-0.12	2.55%	0.17	-0.15
$Mean_{ALL}$	0.95%	0.07	0.05	0.93%	0.07	0.05	1.33%	0.08	0.05	1.22%	0.06	0.04	1.78%	0.09	0.04	1.63%	0.07	0.03

Table VIII. Continued

	(c) Relative risk aversion: γ=7																	
	$0 \leq w_s \leq 1$						$0 \leq w_s \leq 1.5$						$-1 \leq w_s \leq 2$					
	c=0			c=50			c=0			c=50			c=0			c=50		
	ΔCER (ann.)	ΔSR	Avg. ΔSR	ΔCER (ann.)	ΔSR	Avg. ΔSR	ΔCER (ann.)	ΔSR	Avg. ΔSR	ΔCER (ann.)	ΔSR	Avg. ΔSR	ΔCER (ann.)	ΔSR	Avg. ΔSR	ΔCER (ann.)	ΔSR	Avg. ΔSR
PC_{Econ}(SIC)	2.19%	0.21	-0.08	2.33%	0.20	-0.05	2.84%	0.22	-0.09	3.08%	0.23	-0.05	0.76%	0.11	-0.15	0.63%	0.11	-0.13
PC_{Econ}(Adj.R^2)	1.91%	0.17	-0.17	2.00%	0.17	-0.15	2.35%	0.19	-0.17	2.46%	0.19	-0.14	0.61%	0.16	-0.26	0.05%	0.14	-0.26
$Mean_{Econ}$	1.40%	0.12	0.02	1.22%	0.10	0.01	2.12%	0.15	0.02	1.94%	0.11	0.00	2.57%	0.15	0.01	2.49%	0.12	-0.00
PC_{Tech}(SIC)	0.83%	0.08	0.06	0.90%	0.09	0.07	0.98%	0.09	0.06	0.87%	0.09	0.06	0.77%	0.08	0.03	0.51%	0.06	0.01
PC_{Tech}(Adj.R^2)	0.56%	0.05	0.04	0.50%	0.05	0.05	0.30%	0.04	0.03	0.00%	0.02	0.02	-0.27%	0.01	-0.02	-0.84%	-0.02	-0.04
$Mean_{Tech}$	0.49%	0.04	0.05	0.52%	0.05	0.06	0.72%	0.06	0.05	0.62%	0.05	0.04	0.70%	0.06	0.03	0.55%	0.05	0.01
PC_{ALL}(SIC)	2.18%	0.21	0.04	2.21%	0.20	0.07	3.07%	0.26	0.05	3.00%	0.25	0.06	1.93%	0.20	0.02	1.40%	0.18	0.01
PC_{ALL}(Adj.R^2)	2.79%	0.28	-0.00	2.75%	0.25	0.01	3.81%	0.32	-0.01	3.80%	0.30	-0.01	2.45%	0.24	-0.13	1.58%	0.19	-0.17
$Mean_{ALL}$	0.90%	0.08	0.04	0.84%	0.07	0.04	1.34%	0.09	0.05	1.22%	0.07	0.03	1.76%	0.10	0.04	1.67%	0.08	0.03

Notes: The Table reports portfolio performance measures for a mean-variance investor under alternative specifications. ΔCER denotes the annualized certainty equivalent return gain of predictive regression forecasts and ΔSR is the annualized Sharpe ratio defined as the average portfolio excess return divided by the sample standard deviation both in comparison to the historical average forecast over the evaluation period 1966:01-2013:12. Avg. ΔSR indicates differences in the generated Sharpe ratios using the rolling-recursive estimation setting described in section 3.4. Conditional volatility forecasts are based on AR(1) estimates of realized volatility measures.

Data Appendix

Technical indicators:
1. Moving Average Rules (MA(s,l)): Moving average trading rules are defined as

$$S_{i,t} = \begin{cases} 1 \ if \ MA_{s,t} \geq MA_{l,t} \\ 0 \ otherwise \end{cases} \quad s=(1,2,3); \ l=(9,12)$$

where

$$MA_{j,t} = \left(\frac{1}{j}\right)\sum_{i=0}^{j-1} P_{t-i} \quad for \ j = s, l$$

P_t is the level of the S&P 500 stock price index.

2. Momentum Rules (MOM(m)):Momentum based trading rules are defined as

$$S_{i,t} = \begin{cases} 1 \ if \ P_t \geq P_{t-m} \\ 0 \ otherwise \end{cases} \quad m=(9,12)$$

3. Volume Rules (VOL(s,l)) We follow Granville (1963) and make use of "on-balance" volume which is defined as

$$OBV_t = \sum_{k=1}^{t} VOL_k D_k$$

where VOL_k is the monthly trading volume on the S&P 500 index and D_k takes a value of 1 if $P_k - P_{k-1} \geq 0$ and 0 otherwise.[10] We then compare short- (s=1,2,3) and long-term (l=9,12) moving averages to the OBV indicators to design trading signals.

$$S_{i,t} = \begin{cases} 1 \ if \ MA_{s,t}^{OBV} \geq MA_{l,t}^{OBV} \\ 0 \ otherwise \end{cases}$$

Economic variables:
1. Dividend Price Ratio (d/p): Defined as the difference between the log of a twelve-month moving sum of dividends paid on the S&P 500 index and the log of stock prices.
2. Dividend Yield (d/y): Defined as the difference between the log of a twelve-month moving sum of dividends paid on the S&P 500 index and the log of lagged stock prices.
3. Earnings Price Ratio (e/p): Defined as the difference between the log of a twelve-month moving sum of earnings on the S&P 500 index and the log of stock prices.
4. Dividend Payout Ratio (d/e): Defined as the difference between the log of a twelve-month moving sum of dividends paid on the S&P 500 and the log of a twelve-month moving sum of earnings on the S&P 500.

---

[10] Volume data on the S&P 500 index is obtained and available at http://de.finance.yahoo.com

5. Equity Risk Premium Volatility (rvol): Following Neely et al. (2014) we make use of the volatility measure proposed by Mele (2007) to avoid problems arising from outliers in October 1987. Our equity risk premium volatility proxy is defined as

$$\hat{\sigma}_t = \frac{1}{12} \sum_{i=1}^{12} |r_{t+1-i}|$$

$$\widehat{rvol}_t \equiv \sqrt{\frac{\pi}{2}} \sqrt{12} \hat{\sigma}_t.$$

6. Book-to-Market Ratio (b/m): Defined as the ratio of book value to market value for the Dow Jones Industrial Average.
7. Net Equity Expansion (ntis): Defined as the ratio of a twelve-month moving sum of net equity issued by NYSE-listed stocks divided by the total end-of-year market capitalization of NYSE stocks.
8. Treasury Bill Rate (tbl): Defined as the 3-month Treasury bill rate (secondary market)
9. Long-term Yield (lty): Defined as the long-term government bond yield.
10. Long-term Return (ltr): Defined as the return on long-term government bonds.
11. Term Spread (tms): Defined as the difference between the long-term yield and the 3-month Treasury bill rate.
12. Default Yield Spread (dfy): Defined as the difference between Moody's BAA- and AAA- rated corporate bond yields.
13. Default Return Spread (dfr): Defined as the difference between the return on long-term corporate bonds and returns on long-term government bonds.
14. Inflation (infl): Calculated from the Consumer Price Index (CPI, All Urban Consumers). To account for a delay in CPI releases, we use the 1-month lagged inflation in the predictive regression.