Universität Trier

Predictability of Bull and Bear Markets: A New Look at Forecasting Stock Market Regimes (and Returns) in the US

Felix Haase
Matthias Neuenkirch

Research Papers in Economics
No. 1/20

Predictability of Bull and Bear Markets: A New Look at Forecasting Stock Market Regimes (and Returns) in the US*

Felix Haase
University of Trier

Matthias Neuenkirch†
University of Trier and CESifo

First Draft: March 23, 2020
This Version: November 19, 2021

Abstract

The empirical literature of stock market predictability mainly suffers from model uncertainty and parameter instability. To meet this challenge, we propose a novel approach that combines dimensionality reduction, regime-switching models, and forecast combination to predict the S&P 500. First, we aggregate the weekly information of 146 popular macroeconomic and financial variables using different principal component analysis techniques. Second, we estimate Markov-switching models with time-varying transition probabilities using the principal components as predictors. Third, we pool the models in forecast clusters to hedge against model risk and to evaluate the usefulness of different specifications. Our weekly forecasts respond to regime changes in a timely manner to participate in recoveries or to prevent losses. This is also reflected in an improvement of risk-adjusted performance measures as compared to several benchmarks. However, when considering stock market returns, our forecasts do not outperform common benchmarks. Nevertheless, they do add statistical and, in particular, economic value during recessions or in declining markets.

JEL Codes: C53; G11; G17.

Keywords: Forecast Combination; Markov-Switching Models; Shrinkage Methods; Stock Market Regimes; Time-Varying Transition Probabilities.

---

*We would like to thank seminar participants at Freie Universität Berlin for their helpful comments on an earlier version of the paper. The initial draft of the paper circulated under the title "Forecasting Stock Market Recessions in the US: Predictive Modeling using Different Identification Approaches."

†Corresponding author: University of Trier, Department of Economics, D-54286 Trier, Germany, Tel.: +49-651-201-2629, E-mail: neuenkirch@uni-trier.de.

1 Introduction

The existence of different stock market regimes is widely accepted among academics and practitioners. Stock market cycles typically precede business cycles and are caused by time-varying expectations of future cash flows and discount rates. In bullish periods, prices rise and fluctuate only mildly, whereas in bearish periods, prices decrease and volatility increases. Hence, anticipating regime changes and, in particular, contractions is of relevance for investors and corporate decision-makers. Furthermore, the state of the stock market as leading indicator is important for governments, (central) banks, and households. The global financial crisis (GFC) of 2007−2008 is the most recent example illustrating the danger of spill-over effects to the real economy.

Since stock market regimes are unobservable, their identification and prediction is challenging. Three methods have been established in the literature. First, observable measures that reflect the risk aversion of market participants are natural candidates to signal regime dynamics. Empirically, Coudert and Gex (2008) highlight the relevance of risk aversion proxies for stock crash predictions, whereas Chow et al. (1999) and Kritzman and Li (2010) underline the importance of market turbulence indices. Second, Markov-switching (MS) models are used to infer the probabilities of a latent state variable and to forecast returns or volatility (Ang and Bekaert 2002; Haas et al. 2004); the number of regimes in these models is still subject to debate (e.g., Guidolin and Timmermann 2007; Maheu et al. 2012; Hauptmann et al. 2014). Third, change point detection methods or dating rules are utilized in this context. The application of change point analysis to stock market data is similar to MS models (Pástor and Stambaugh 2001; Pettenuzzo and Timmermann 2011). However, the assumption that "history repeats" is neglected, so that each change point marks the beginning of a new regime. Dating rules, on the other hand, search for local extremes which are defined by period lengths (Pagan and Sossounov 2003) or by absolute price changes (Lunde and Timmermann 2004). The underlying algorithms need past and future prices for the dating of recessions and, consequently, delayed signals may occur. In addition, genuine backtesting cannot be performed in real-time when using dating rules.

Considering the empirical success of dimensionality reduction techniques (Neely et al. 2014; Çakmaklı and van Dijk 2016), regime-switching models (Guidolin and Hyde 2012; Maheu et al. 2012), and forecast combination (Rapach et al. 2010) in predicting stock market dynamics, we propose a novel procedure that combines these approaches. Confronted with a large real-time dataset of macroeconomic and financial market variables, we first reduce the dimensionality into a few latent factors by different principal component analysis (PCA) techniques. We employ a conventional PCA and a sparse PCA, where the loadings of some variables are set to zero. In addition, we apply a soft thresholding approach to both, conventional PCA and sparse

PCA, yielding two additional sets of targeted principal components. Second, using the principal components as predictors, we estimate MS models with time-varying transition probabilities (TVTP) to identify and predict regimes in a single step. For this purpose, we consider two specifications. On the one hand, we use a general specification, which models the conditional mean and the transitions (Specification A). And on the other hand, we rely on a restricted specification where only regimes are predictable while returns follow a (regime-dependent) random walk (Specification B). Since highly parameterized models tend to be inferior to parsimonious ones in terms of forecast accuracy, we limit the model size of each model to include only one principal component (or observable predictor). These different combinations of MS specifications and PCA techniques (or the usage of observable predictors) result in a large number of models that we combine into several forecast clusters (according to the shrinkage method and the model specification). In this third step, we also ensure robustness to different weighting decisions as we consider simple averaging and a continuous weighting approach. Throughout the procedure, we account for publication lags, data revisions, and consider transaction costs to ensure realistic forecasts in the backtest. Figure 1 (at the beginning of Section 2) provides an illustrative overview of our methodology and Table 1 (at the end of Section 2) summarizes the different specifications, clusters, models, and forecast combination techniques.

Our sample covers weekly data for the S&P 500 and spans the period from November 17, 1989 to May 7, 2021. We use weekly data since, at a higher frequency, regime forecasts would be too noisy and return forecasts virtually impossible. Moreover, the choice of weekly returns represents a good compromise between precision and data availability as fundamentals are usually updated monthly, while market data obviously changes on an intraday frequency. Our recursive out-of-sample real-time exercise focuses on the most recent 864 weeks. Accordingly, the first training set to estimate the MS models ends on October 15, 2004. For the evaluation of the different forecasts, we classify bull and bear markets using the dating rule of Lunde and Timmermann (2004), assuming knowledge of the full sample.

Our regime forecasts are suitable to respond to regime changes in a timely manner to participate in recoveries or to prevent losses. This is also reflected in an actual economic value added as many of our forecasts beat all benchmarks in risk-adjusted performance measures. However, when considering stock market returns, our forecasts do not statistically outperform common benchmarks. The fact that return forecasts perform worse than regime forecasts is not surprising since forecasting the broader trend of the stock market is obviously easier than providing point forecasts, in particular at a weekly frequency. Nevertheless, our return forecasts still provide some economic value added for risk-adverse investors as they generate a lower annualized standard deviation of the returns and better tail risk measures than the corresponding regime forecasts. Consistent with the literature (e.g., Henkel et al. 2011; Rapach and Zhou 2013), we find that much of the predictability comes from periods of market turmoil or recessions. Finally, we highlight that it is sufficient to model the time-varying conditional transitions in a Markov-switching model. We also propose to rely on dimensionality reduction techniques and to enhance the conventional principal component analysis with shrinkage methods such as sparsity and/or soft thresholding.

Our paper is, to the best of our knowledge, the first one to apply MS models with TVTP and several PCA techniques to predict bull and bear markets. We contribute to several strands of the stock market forecasting literature. First, we confirm the previous finding of predictable trends, in particular during recessions, in stock markets (Guidolin and Timmermann 2007; Chen 2009; Kritzman et al. 2012).

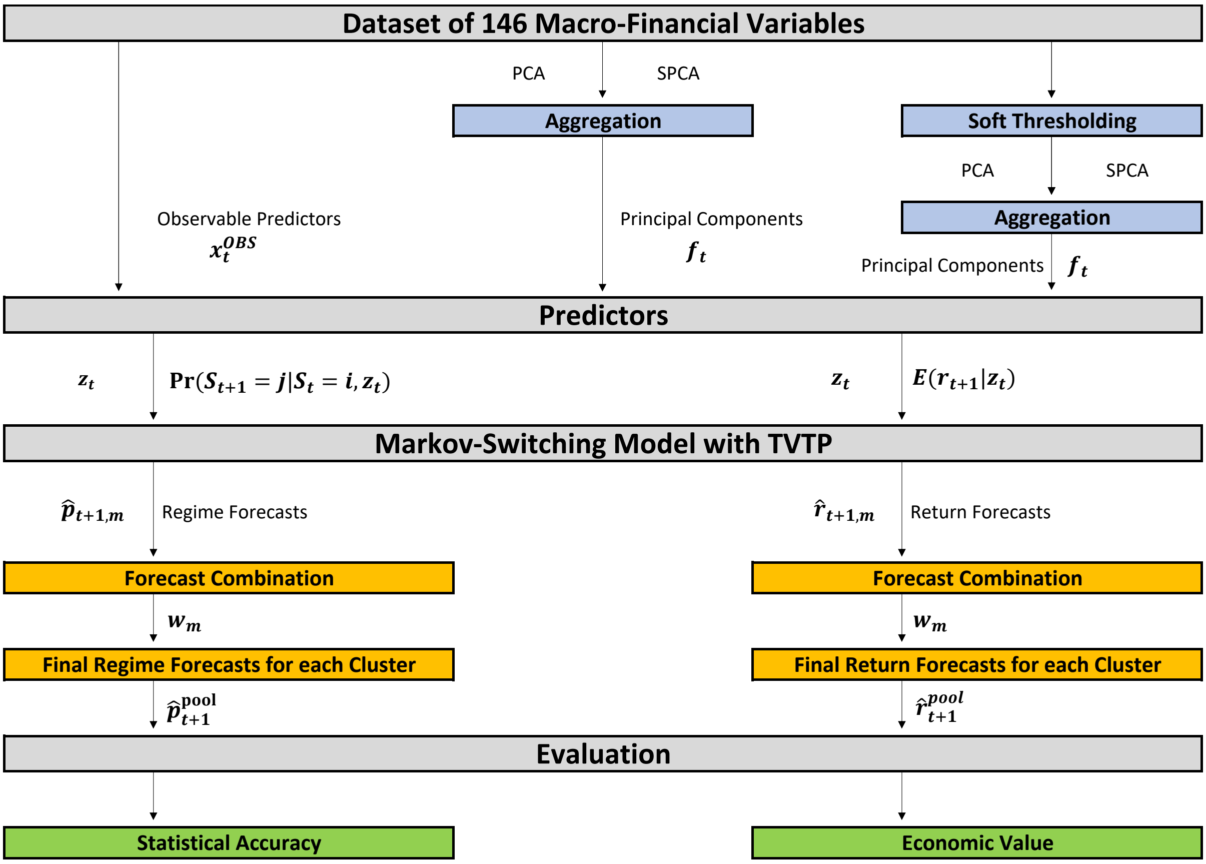
Second, we emphasize the benefits of MS models with principal components and TVTP. Although MS models with time-varying transitions have been developed more than 25 years ago (Diebold et al. 1994), there are only a handful of examples that apply these models in the context of bull and bear markets (e.g., Schaller and Norden 1997; Maheu and McCurdy 2000; Guidolin and Hyde 2012; Kole and van Dijk 2017; Focardi et al. 2019). The few existing papers that include macro-financial variables in the transition equation provide rather disappointing results. Guidolin and Hyde (2012) and Kole and van Dijk (2017) do not detect any advantage of modeling the transition with lagged returns, individual macro-financial variables, or a principal component based on seven popular predictors. Overly complex modeling of the switching process might cause their results since Guidolin and Hyde (2012) apply a multivariate three regime model and Kole and van Dijk (2017) consider multiple variables in the switching equation. We address these concerns in two ways. On the one hand, we follow the recommendation of Zens and Böck (2019) to include only a few latent factors (one, to be precise) into the transition equation and, on the other hand, we focus on a univariate setting with only two regimens to reduce possible identification problems and estimation uncertainty.

Third, for constructing stock predictors from "big data", we recommend using shrinkage methods. These provide a more straightforward interpretation of the extracted factors and, particularly appealing in forecasting, can reduce noise without losing much of the captured variance. This finding is also documented by Rapach and Zhou (2019) who emphasize the superiority of a sparse PCA. Finally, there are several other papers that use a high-dimensional dataset to predict financial variables. Mönch (2008) and Ludvigson and Ng (2009) predict bond yields. Ludvigson and Ng (2007), Neely et al. (2014), and Çakmaklı and van Dijk (2016) provide promising results and highlight the attractiveness of principal components as predictors for stock returns. Our empirical framework also resembles the ones used to forecast commodity returns and futures with large datasets and regime-switching models (Guidolin and Pedio 2020, 2021) or to exploit the business cycle to predict asset prices (e.g., Kaya et al. 2010; Hammerschmid and Lohre 2018; Sander 2018).

The remainder of this paper is organized as follows. Section 2 outlines our methodology and explains the necessary modeling choices. Section 3 introduces the dataset of macro-financial variables. Section 4 shows the classification of market regimes and discusses the aggregation of the predictors, both under the assumption of full-sample knowledge. Section 5 demonstrates how our approach works in a real-time situation with recursive out-of-sample forecasts. Section 6 concludes.

2 Methodology

Figure 1: Overview of the Methodology



We face the issues of model uncertainty and parameter instability when forecasting stock market regimes and returns (Pesaran and Timmermann 1995). Our approach combines dimensionality reduction, regime-switching models, and forecast combination to predict the S&P 500. In addition, we apply the MS specifications and the forecast combination schemes to a subset of directly observable popular predictors. Our aim is to evaluate whether a large dataset and the utilization of aggregation techniques (see, among others, Neely et al. 2014 and Çakmaklı and van Dijk 2016) provide an actual advantage over employing commonly used (simple) predictors.[1] Figure 1 provides an overview of the individual steps in our procedure that are explained in detail in the following subsections.

2.1 Step 1: Data Aggregation

Due to the increasing availability of data, an investor is confronted with the choice of the relevant predictors. Theoretical considerations might be helpful in this context, but even with certain restrictions there is a large pool of potential variables. Due to the substantial correlation of many covariates with unobserved state variables — such as the business cycle or investor sentiment — an efficient filtration of the variables is recommended to cover the co-movement and to eliminate potential noise. PCA is an appealing method to capture relevant information in a parsimonious way. A small number of components is usually sufficient to capture most of the variation in the data, allowing for a significant reduction in the dimensionality of the original dataset.

2.1.1 Conventional PCA and Sparse PCA

Conventional PCA: Principal components capture the co-movement of many (potentially) correlated predictors that are normalized to a mean of zero and a variance of one. Let X be a $T \times K$ matrix of potential predictors, where the number of rows T ($t = 1, 2, ..., T$) represents the time dimension and K ($k = 1, 2, ..., K$) the cross-sectional dimension. Using singular value decomposition of X, we can obtain the principal components as (Zou et al. 2006):

$$X = UDV^T \tag{1}$$

The principal components are $Z = UD$, with U representing a unitary matrix and D a diagonal matrix of singular values. V is a $K \times K$ matrix of eigenvectors, where the k-th column represents the loadings of the k-th component. Typically, a small positive number of q components is sufficient to aggregate the information in X, so that we achieve a substantial dimensionality reduction in exchange for a minimal loss of information ($q << min(K, T)$). In addition, the components are constructed in such a way as to be uncorrelated to each other. To determine q, we use the IC_{p2} information criterion by Bai and Ng (2002), where the upper bound is set according to an automatic

[1] We consider the lagged return R, the dividend-price-ratio DP (Campbell and Shiller 1988; Fama and French 1988; Schaller and Norden 1997), the volatility index VIX (Rubbaniy et al. 2014), the term spread TS and the credit spread CS (Fama and French 1989; Campbell and Yogo 2006), the Purchasing Managers Index PMI (Johnson and Watson 2011), and the variance risk premium VP (Bollerslev et al. 2009; Bekaert and Hoerova 2014). The term spread is defined as difference between the 10Y US treasury bond and the 3M treasury bill, the credit spread as excess yield of the Moody's seasoned Baa over the Aaa corporate bond yield, and the variance risk premium as difference between the squared VIX and the sum of the squared 5-minute returns of the last 22 trading days.

elbow procedure. Hence, we select the first q normalized principal components as relevant factors f to predict stock market regimes and returns.

Sparse PCA: One disadvantage of conventional PCA is that the components are based on all variables, which often leads to a lack of interpretability. A sparse PCA uses shrinkage methods to reduce the loadings of some variables to zero for a more straightforward interpretation without losing too much of the captured variance (Rapach and Zhou 2019). Following the illustration of Zou et al. (2006), we treat the optimization as regularized regression problem. Suppose we consider the first q principal components, and let x_t be the t-th row of X. We further denote A as $q \times K$ orthonormal matrix with elements $A = [\alpha_1, \alpha_2, ..., \alpha_K]$ and B as $q \times K$ sparse weight matrix with $B = [\beta_1, \beta_2, ..., \beta_K]$. Then we consider the following optimization problem for $\lambda > 0$:

$$\arg\min_{A,B} \left[\sum_{t=1}^{T} ||x_t - AB^T x_t||^2 + \lambda \sum_{p=1}^{q} ||\beta_p||_2^2 + \sum_{p=1}^{q} \lambda_{1,p} ||\beta_p||_1 \right] \quad s.t.\ A^T A = I \tag{2}$$

$||.||_1$ corresponds to the $L1$ and $||.||_2^2$ to the squared $L2$ norm. I represents the $q \times q$ identity matrix. The amount of ridge shrinkage λ is the same for all q components and the sparsity constraint $\lambda_{1,p}$ can vary over the components, where a higher value of $\lambda_{1,p}$ leads to more sparse loadings. If we restrict Eq. (2) by $B = A$ and set the LASSO (least absolute shrinkage and selection operator) penalty $\lambda_{1,p} = 0$, we obtain the conventional PCA (Zou et al. 2006). We solve Eq. (2) using the variable projection approach by Erichson et al. (2020).

For a better comparability, we do not apply the procedure of Bai and Ng (2002) on the adjusted sparse factors (see Zou et al. 2006). Instead, we assume the same number of components as the conventional PCA suggests. Additionally, we do not vary the degree of sparseness over the components and set $\lambda_{1,p} = \lambda_1$ due to missing economic arguments. For the $L1$ and $L2$ penalty, we follow Kristensen (2017) and tune the hyperparameters in every week such that a Bayesian information criterion (BIC) type problem is minimized.[2]

[2] For this purpose we use the grids $\lambda \in (1e^{-4}, 1e^{-3}, 1e^{-2})$ and $\lambda_1 \in (2e^{-3}, 4e^{-3}, 6e^{-3}, 8e^{-3}, 1e^{-2})$:

$$(\lambda_1^*, \lambda^*) = \arg\min_{\lambda_1, \lambda} log\left(\frac{1}{KT} \sum_{k=1}^{K} \sum_{t=1}^{T} [x_{k,t} - \hat{l}_{SPCA,k}(\lambda_1, \lambda) \hat{F}_{SPCA,t}(\lambda_1, \lambda)]^2 \right) + \phi(\lambda_1, \lambda) \frac{log(KT)}{KT} \tag{3}$$

$\phi(\lambda_1, \lambda)$ represents the number of non-zero PCA weights and $\hat{l}_{SPCA}$ and $\hat{F}_{SPCA}$ correspond to the loadings and adjusted scores (via QR decomposition) as suggested by Zou et al. (2006).

2.1.2 Soft Thresholding

Another drawback of the conventional PCA and the sparse PCA is that they do not consider the target variable during the construction of the factors. A soft thresholding approach (Bai and Ng 2008) conducts a pre-selection on the data to obtain targeted predictors and has already been applied in the return forecasting literature (e.g., Çakmaklı and van Dijk 2016). Our implementation of soft thresholding follows Bai and Ng (2008) and uses the elastic net (EN) methodology. The EN is a convex combination of LASSO and ridge regression that performs model selection and shrinkage simultaneously.[3] More formally, the EN optimization is a regularized regression to minimize the residual sum of squares (RSS) and can be written as follows:

$$\arg\min_{\beta}\left[RSS + \lambda_1 \sum_{k=1}^{K} |\beta_k| + \lambda_2 \sum_{k=1}^{K} \beta_k^2\right] \tag{4}$$

β corresponds to the EN estimate and λ_1 and λ_2 are non-negative hyperparameters, which balance the influence of LASSO and the ridge penalty. In our context, we use the non-zero β's to select relevant predictors.

The choice of the target variable depends on our objective. If we want to find targeted predictors for the return process, we rely on future excess returns r_{t+1}. However, if we want to select predictors to forecast regimes, our target cannot be observed. Here, we proceed with the VIX_{t+1}, which is a popular fear gauge in practice and, therefore, a good signal for shifts into a bearish regime. We follow Bai and Ng (2008) and use the least angle regression algorithm to solve the elastic net problem (LARS-EN). We obtain a ranking of selected predictors, such that we can substitute the LASSO penalty λ_1 with the size of the active set of predictors. We refrain from optimizing the size of the active set for simplicity and select the top 75 predictors, which is proportionally similar to the subset size in Çakmaklı and van Dijk (2016). With respect to the ridge penalty, we perform a grid search on the interval $[0, 0.25, 0.5, 0.75, 1]$ and choose the λ_2^* that optimizes Mallows's Cp. For the out-of-sample exercise, we repeat the hyperparameter search every week.

We apply the soft thresholding approach in combination with both conventional PCA and sparse PCA. Hence, as predictors, we utilize four different sets of principal components and, additionally, the subset of directly observable popular variables (see also Figure 1).

[3]The main benefit of EN over LASSO in soft thresholding is that in situations with a group of highly correlated predictors, LASSO selects only one variable of this group, whereas the EN approach stretches "the fishing net to retain all the *big fish*" (Bai and Ng 2008, p. 307).

2.2 Step 2: Markov-Switching Models

Since the pioneering work of Hamilton (1989), MS models have become increasingly popular in economics. MS models are able to reveal changes in the fundamental environment of financial markets in a timely manner, even if their interpretation is only possible ex post (Ang and Timmermann 2012). Thus, MS models help to account for time-varying risk premia and to uncover temporary trends in returns.

Starting with the basic switching model, r_t denotes the excess log-return of the S&P 500 over the 3 M Treasury Bill and S_t the unobservable state of the stock market. Then, the non-linear return dynamics can be described as:

$$\begin{aligned} r_t &= \mu_{S_t} + u_t \\ u_t &\sim i.i.d.\ N(0, \sigma^2_{S_t}) \\ Pr(S_t = j | S_{t-1} = i) &= p_{ij} \end{aligned} \tag{5}$$

Assuming that the mean μ_{S_t} and the variance $\sigma^2_{S_t}$ are dependent on the current market regime, the MS model is able to replicate stylized facts of financial time series such as fat tails, volatility clustering, and asymmetries (Ang and Timmermann 2012). In the basic time-homogeneous case, the regime variable S_t is assumed to follow a discrete first-order Markov chain, that is, the current market regime j depends only on the previous regime i. We will refer to this model, which later on serves as one of the benchmarks, as an MS model with time-constant transition probability (TCTP) and without external predictors.

The majority of papers treats the transition probabilities as constant over time, ignoring that these can be affected by changes in fundamental conditions. In this paper, we follow Diebold et al. (1994) and model the switching process as being dependent on macro-financial conditions z_{t-1}.

Specification A: In the general specification, we assume that the excess S&P 500 returns follow a MS model with predictable mean and regime processes:

$$\begin{aligned} r_t &= \mu_{S_t} + \beta_{S_t} z_{t-1} + u_t \\ u_t &\sim i.i.d.\ N(0, \sigma^2_{S_t}) \\ p_{i0,t} &= \frac{\exp(\upsilon_{i0} + \gamma_{i0} z_{t-1})}{1 + \exp(\upsilon_{i0} + \gamma_{i0} z_{t-1})} \end{aligned} \tag{6}$$

z_{t-1} are either observable predictors proposed by the literature or components approximated by the different PCA techniques described in the previous subsection. The intercept is denoted as μ_{S_t}, and u_t is the idiosyncratic error with a regime-dependent variance. To model the switching dynamics, we follow the standard in the literature by using a logit link function (Diebold et al. 1994), where the constant υ_{i0} and the slope γ_{i0} depend on the current regime. Finally, it has to be noted that all parameters are dependent on the regime variable S_t, allowing for parameter flexibility across regimes.

In our application, we consider two regimes, where regime 0 corresponds to bull markets and regime 1 to bear markets. The number of stock market regimes is certainly open to debate, and, since S_t is a latent variable, the "true" number is unknown. An approximation with econometric tests is also difficult (Hansen 1991; Ang and Timmermann 2012). Therefore, one usually relies on information criteria or theoretical arguments. Our decision to focus on two regimes is motivated by several reasons. First, a clear distinction can be made between (i) a volatile regime with a negative drift and (ii) a calm regime with positive average returns. Second, prominent dating rules (Pagan and Sossounov 2003; Lunde and Timmermann 2004) are available for two regimes. These ensure a transparent and straightforward regime classification and are helpful to evaluate our real-time regime predictions ex post. Finally, more than two regimes often lead to unstable estimations, particularly in our out-of-sample task with a variety of predictors and specifications.[4]

As highlighted by Zens and Böck (2019), only a small number of variables can be included in the transition probabilities to ensure a stable estimation process. Consequently, we rely on latent factors constructed from many variables to incorporate macro-financial information in a compact form and restrict the number of variables to avoid highly parameterized models. More precisely, we incorporate only one principal component (or observable variable) in the switching equation and in the conditional mean equation. This ensures a robust estimation process and reduces the variability of the forecasts. For simplicity, we assume that the external predictors are the same in both equations.

Specification B: Given the extensive regime dependency of *Specification A*, overfitting might be a problem. For this reason, we also consider a restricted model that focuses only on the switching process while returns follow a (regime-dependent) random walk. By setting the constraint $\beta_{S_t} = 0$ in *Specification A*, we obtain the restricted *Specification B*.

We estimate all models with maximum likelihood methods using the expectation maximization algorithm.[5]

[4]Note that some authors assume more than two regimes (Guidolin and Timmermann 2007; Guidolin and Hyde 2012; Maheu et al. 2012; Zhu and Zhu 2013). For example, Maheu et al. (2012) distinguish between two bullish regimes (normal and correction) and two bearish regimes (normal, rally).

[5]Hereby, we essentially follow Hamilton (1990). An alternative would be a Bayesian approach using the Gibbs sampler, in which the parameter uncertainty is explicitly incorporated (for an application, see Maheu et al. 2012). For further details about inference on regimes and the estimation procedure, we refer to Hamilton (1994).

Prediction: One appealing feature of MS models is that identification and prediction can be done in a single step. Using the filter proposed by Hamilton (1989), the one-step ahead regime prediction for j is:

$$\hat{p}_{t+1}^{j} = Pr(S_{t+1} = j|\Omega_t) = \sum_{i=0}^{1} p_{ij,t} Pr(S_t = i|\Omega_t) \tag{7}$$

Ω_t represents the information set in period t and $Pr(S_t = i|\Omega_t)$ the filtered probability, which is recursively updated using Bayes' rule. To simplify the notation, we define $\hat{p}_{t+1}^{1} = \hat{p}_{t+1}$ as predicted bear probability and $(1-\hat{p}_{t+1})$ as the corresponding bull probability.

Finally, the regime forecasts can be used to predict returns. Relying on the regime-dependent expectations $E[r_{t+1}|S_{t+1} = j]$, the return forecast $\hat{r}_{t+1}$ is given by the following probability-weighted average:

$$\hat{r}_{t+1} = (1-\hat{p}_{t+1})E[r_{t+1}|S_{t+1} = 0] + \hat{p}_{t+1}E[r_{t+1}|S_{t+1} = 1] \tag{8}$$

2.3 Step 3: Forecast Combination

Instead of using multiple predictors in one model, forecast combination uses multiple models with a restricted number of predictors in each model. Timmermann (2006) highlights that combined forecasts work particularly well in uncertain situations where the influence of relevant variables varies considerably over time. Hence, forecast combination is a promising strategy to hedge against model uncertainty and to increase the predictability of regimes (and returns). Compared to large multivariate regressions, forecast combination has the advantage that the estimation variability can be significantly reduced and that in-sample overfitting can be avoided (Rapach and Zhou 2013). In general, the forecast combination setting can be formulated as weighted average of individual forecasts for regimes and returns.

Regime Forecasts: Suppose we have M regime probability forecasts $\hat{p}_{t+1,m}$. This yields the following forecast combination problem:

$$\hat{p}_{t+1}^{pool} = \sum_{m=1}^{M} w_m \hat{p}_{t+1,m} \tag{9}$$

Return Forecast: The pooled return forecast, given M return forecasts $\hat{r}_{t+1,m}$, can be expressed as follows:

$$\hat{r}_{t+1}^{Pool} = \sum_{m=1}^{M} w_m \hat{r}_{t+1,m} \tag{10}$$

In this context, we have to make a decision about the number of included forecasts M and their weights w_m. In our application, the individual forecasts are combined within some pre-specified clusters. We form the clusters in such a way as to be able to evaluate the usefulness of the various aggregation techniques and the specification choices of the MS model. Consequently, we differentiate alongside two dimensions: (i) predictor choice (directly observable or estimated using the four different PCA techniques) and (ii) MS specification (*Specification A* or *Specification B*).

Next, we have to determine the individual weights of the forecasts w_m. For this purpose, we employ two different methods.

Simple Average (AVE) $$w_m = \frac{1}{M}$$

Bayesian Model Averaging (BMA) $$w_m = \frac{exp(-\Delta_m/2)}{\sum_{l=1}^{L} exp(-\Delta_l/2)}$$

The simple average forecast is straightforward and precludes any estimation risk. In addition, it often provides good results, which are difficult to outperform (Timmermann 2006). In addition, inspired by the results of Cremers (2002), we apply Bayesian model averaging. Since our estimation is not Bayesian, we approximate the posterior model probability with the observed data. We use Bayes' factors to avoid computational difficulties (overflow/underflow) and define $\Delta_m = BIC_m - BIC^*$, where BIC^* represents the model with the lowest BIC.

To summarize, we calculate a total of 20 forecast combinations. This number emerges from the *five* clusters of predictors (observable predictors and the four different PCA techniques), the *two* specifications of the MS model (mean and transitions versus transitions only), and the *two* different aggregation techniques (simple average versus BMA). Table 1 summarizes the different specifications, clusters, models, and forecast combination techniques.

3 Data

Our dataset consists of weekly data for the United States. The stock market is represented by the S&P 500 index, adjusted for dividends and stock splits. We consider a large set of 146 variables to predict regimes and returns. This includes several categories of variables: bond yields, term spreads and credit spreads, lagged returns, technical indicators, industry returns, market-based risk indicators, valuation ratios, survey-based expectations about macroeconomic variables/earnings and their dispersion, sentiment indicators, and macroeconomic fundamentals. All variables either are proven to be empirically relevant or can be recommended from a practical point of view.

Table 1: Specifications, Clusters, Models, and Forecast Combinations

Specification	Cluster	Models	Forecast Combinations	
Specification A: Conditional Mean and Transitions	OBS ($M = 7$)	A-R, A-DP, A-VIX, A-TS, A-CS, A-PMI, A-VP	A-OBS-AVE	A-OBS-BMA
	PC ($M^{max} = 8$)	A-PC1, A-PC2, ..., A-PCq	A-PC-AVE	A-PC-BMA
	SPC ($M^{max} = 8$)	A-SPC1, A-SPC2, ..., A-SPCq	A-SPC-AVE	A-SPC-BMA
	TPC ($M^{max} = 7$)	A-TPC1, A-TPC2, ..., A-TPCq	A-TPC-AVE	A-TPC-BMA
	TSPC ($M^{max} = 7$)	A-TSPC1, A-TSPC2, ..., A-TSPCq	A-TSPC-AVE	A-TSPC-BMA
Specification B: Transitions Only	OBS ($M = 7$)	B-R, B-DP, B-VIX, B-TS, B-CS, B-PMI, B-VP	B-OBS-AVE	B-OBS-BMA
	PC ($M^{max} = 8$)	B-PC1, B-PC2, ..., B-PCq	B-PC-AVE	B-PC-BMA
	SPC ($M^{max} = 8$)	B-SPC1, B-SPC2, ..., B-SPCq	B-SPC-AVE	B-SPC-BMA
	TPC ($M^{max} = 7$)	B-TPC1, B-TPC2, ..., B-TPCq	B-TPC-AVE	B-TPC-BMA
	TSPC ($M^{max} = 7$)	B-TSPC1, B-TSPC2, ..., B-TSPCq	B-TSPC-AVE	B-TSPC-BMA

Notes: All MS models are estimated with TVTP. Specification A contains predictors in the switching equation and the conditional mean equation according to Eq. (6). Specification B contains predictors in the switching equation only. OBS: observable predictors; PC: conventional PCA; SPC: sparse PCA; TPC: conventional PCA with soft thresholding; TSPC: sparse PCA with soft thresholding; R: lagged return; DP: dividend-price-ratio; TS: term spread; CS: credit spread; PMI: Purchasing Managers Index; VP: variance risk premium; AVE: simple average; BMA: Bayesian model averaging. For example, model A-SPC3 is estimated with the third sparse principal component as conditional mean predictor and transition predictor z_{t-1}. Model B-R uses lagged returns as transition predictor z_{t-1}. A-TSPC-AVE uses simple averaging to combine the forecasts from the cluster TSPC. The number of estimated principal components varies (i) for the full dataset of 146 macro-financial variables versus the targeted dataset of 75 variables and (ii) for the in-sample forecasts versus the recursive out-of-sample forecasts. See also Section 4.2 and Appendix B.

The bond market reflects expectations of market participants in terms of growth prospects, future interest rates, projected inflation, and current risk aversion. Among others, Estrella and Mishkin (1996, 1998) point out that information extracted from the yield curve and, in particular, term spreads are robust predictors for recessions in the real economy. Therefore, we consider government bond yields of all available maturities as well as various spreads over different maturities and the London Interbank Offered Rate (LIBOR). Since stock market contractions are often induced by an increase in risk aversion, credit spreads might also be useful in this context (Coudert and Gex 2008). Correspondingly, we take corporate bond spreads from Moody's and the TED (Treasury Bill Eurodollar) spread into account. As additional predictors, we consider the realized variance of the S&P 500 expressed as sum over the 5-minute squared returns of the previous 1, 5, and 22 trading days plus the close-to-open return (see Bollerslev et al. 2009). Furthermore, we use information from option markets by using the implied volatility index of the S&P500, the VIX. Following Bollerslev et al. (2009), the VIX can be decomposed in a component that reflects the expected future volatility and a risk premium. We extract the so-called variance risk premium by subtracting the squared VIX from the realized stock market variance of the last 22 trading days. Finally, we use additional indicators that capture changes in risk perception, like the gold price and the West Texas Intermediate (WTI) oil price.

We also utilize survey-based expectations as predictors. Consensus Economics asks analysts from banks and research institutes about their macroeconomic expectations at monthly intervals. As predictors, we employ the first and second moments of the individual one-year ahead expectations of macroeconomic variables and the three and twelve month ahead interest rate expectations. The macroeconomic expectations are complemented by sell-side analysts' earnings forecasts, their revisions, and their dispersion from the Institutional Brokers Estimate System. In addition, we employ sentiment measures, such as the surveys by the Conference Board. Following Chen (2012), we also consider several consumer confidence measures as predictors. To capture broader macroeconomic expectations, we utilize the leading composite index from the Conference Board and the PMI. Lastly, we roughly consider the same standard macroeconomic variables as Chen (2009) to incorporate previous findings into our analysis.[6]

The current valuation level is typically related to stock market turbulences (Campbell and Shiller 1988; Fama and French 1988; Lewellen 2004). Hence, we include the dividend price ratio, the earnings price ratio, the 10Y earnings price ratio, and the payout ratio in our dataset. Moreover, we use the same technical indicators as those proposed by Neely et al. (2014). In addition, we incorporate the short-run and long-run moving average of returns (1 M and 12 M) into our predictor set, which are either equally or exponentially weighted. It might be argued that price "excesses" are a major

[6]Industrial production, M1 and M2, the inflation rate, and the unemployment rate.

cause of future contractions, which suggests that valuation ratios or historical returns correlate positively with the risk of bear markets. Furthermore, signals from technical indicators are highly relevant in practice and reflect psychological aspects.

We also use the returns of 34 industry portfolios from the Center for Research in Security Prices Database. Hong et al. (2007) point out that the broad market often processes the information diffused in the industrial returns with a delay, which highlights the leading character of some industry returns. Additionally, we calculate the financial turbulence index (Chow et al. 1999; Kritzman et al. 2012) as well as the absorption ratio (Kritzman et al. 2011). Both measures are popular choices to detect anomalies. The financial turbulence index signals convergence and divergence regarding historical correlation structures and extreme price movements. The absorption ratio can be seen as proxy of systematic risk and encompasses the captured variance of a rolling PCA with a fixed number of components. Since this measure is relatively persistent, we rely on the standardized change in the absorption ratio. To calculate these two risk indicators, we follow the methodology of Kritzman et al. (2011, 2012).

Our sample spans the period between November 17, 1989 and May 7, 2021.[7] Our out-of-sample real-time exercise is conducted using the most recent 864 weeks. Correspondingly, the first training set to estimate the MS models ends on October 15, 2004. Starting from this date, we employ a recursive scheme with an expanding window to predict regimes and returns in the US. In all cases, we rely on end-of-week data (if the data is available at a higher frequency). Every variable is shifted to its publication date and we account for data revisions to ensure a real-time perspective. In addition, we apply common transformations to ensure stationary predictors as, for instance, we follow Rapach et al. (2005) and de-trend bond yields by their one-year moving average. Finally, all variables are centered and scaled before their inclusion in the prediction model. Appendix A lists all variables alongside their definitions (Table A1), provides the data sources (Table A2), and displays summary statistics (Table A3).

4 In-Sample Results

The focus of this paper is on the out-of-sample performance of our forecasts. Hence, we keep the discussion of the in-sample results as concise as possible. Accordingly, we focus on the identification of bull and bear markets that is necessary for an evaluation of the real-time forecasts and we illustrate the aggregation performance of the various PCA techniques assuming knowledge of the full sample. To preserve space, we do not present any in-sample forecasts of stock market regimes (and returns). We also

[7] The starting point is restricted by data availability for many of the predictors, such as the forecasts from Consensus Economics, but also the VIX, corporate bond yields, credit spreads, and sentiment indicators.

do not interpret the principal components since the number of components and their interpretation might be different when considering a training set that only covers a part of the full sample.

4.1 Classification of Bull and Bear Markets

Despite its practical importance and relevance, there is no uniform definition of what exactly characterizes a bull or bear market (Gonzalez et al. 2006). In general, a stock market contraction is a persistent price decline associated with higher fluctuations. However, there is no consensus on how long such a period should last or how strong the price decline should be. We follow the literature (e.g., Kole and van Dijk 2017) and use dating rules for an evaluation of our real-time forecasts.

The underlying idea is to identify local peaks and troughs in the stock price series P_t of the S&P 500 without any distributional assumptions. The identified extreme points mark the turning points of the stock market and the period between a high (low) point and a low (high) point reflects a bear (bull) market. We follow the dating rule of Lunde and Timmermann (2004), as it focuses on absolute price changes and, thus, allows for an intuitive and transparent distinction. Their identification procedure (LT, henceforth) can be summarized as follows:

1. Given that the last observed extreme was a local maximum, referred to as P^{max}, the subsequent price series is checked against the following criteria:

a) The peak is updated if the stock market has risen above the last peak.

b) A local minimum has been found if the stock market has fallen by 10% or more.

c) There are no updates if neither a) nor b) took place.

2. Given that the last observed extreme was a local minimum, referred to as P^{min}, the subsequent price series is checked against the following criteria:

a) The trough is updated if the stock market has dropped below the last minimum.

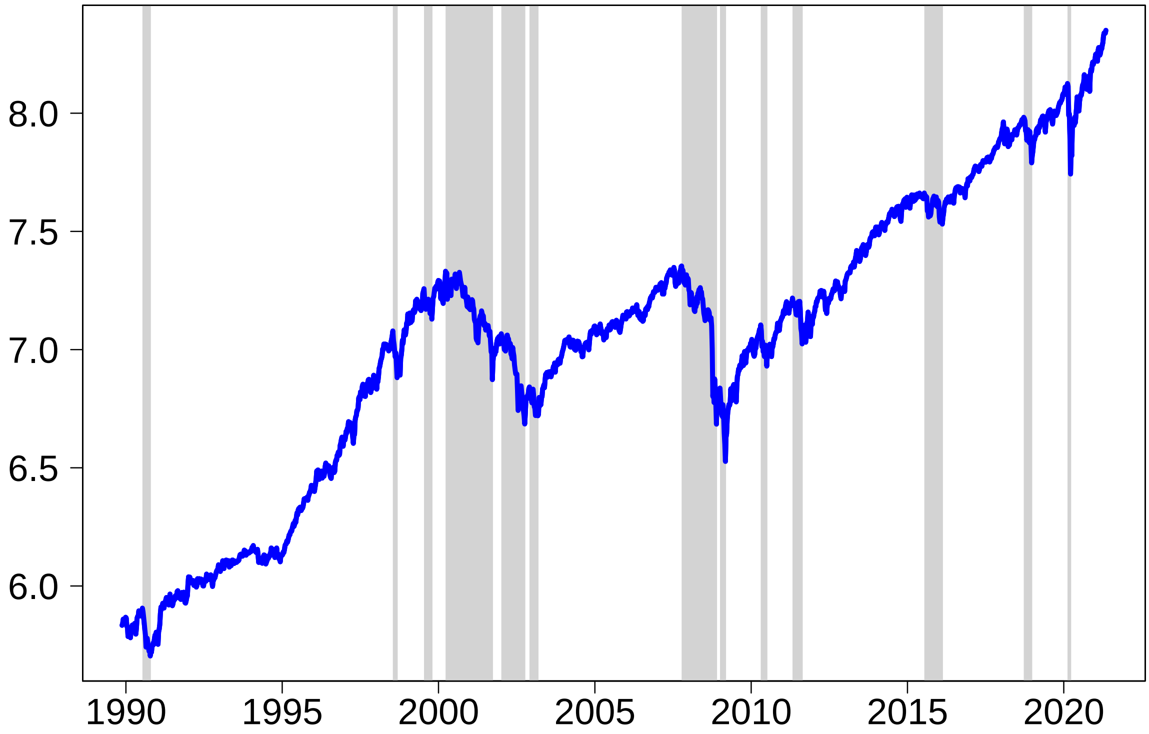
b) A peak has been found if the stock market has risen by 15% or more.

c) There are no updates if neither a) nor b) took place.

In simple terms, periods that result in at least a 10% drop in stock prices are classified as bearish. A switch to a bull market follows if the stock price increase from the low is at least 15%. The particular thresholds are indeed arbitrary, but common in practice.

Figure 2 and Table 2 show the performance of the S&P 500 (in logs) within bullish and bearish market regimes as identified by the LT filter. The biggest drop was caused by the GFC in 2007–2008 (–49%), whereas the bursting of the dotcom bubble (March 2000 to September 2001) marked the longest bear market with a duration of 78 weeks. The recent Covid-19 crash (February to March 2020) is historically the shortest contraction period, but the one with the third largest price slump.

Figure 2: Full-Sample Bull and Bear Market Classification



Notes: Figure shows the log S&P 500 price index and the identified bear markets as gray-shaded areas. The classification follows the dating rule of Lunde and Timmermann (2004).

During our evaluation period, the four economic recessions (according to the definition by the National Bureau of Economic Research, NBER) are always accompanied by a stock market contraction.[8] Despite the fact that the duration and the amplitude of bear markets vary considerably, we can confirm that the stock market acts as an important leading indicator for the business cycle (Hamilton and Lin 1996; Estrella and Mishkin 1998). However, the stock market would predict even more recessions (see Chauvet and Potter 2000), displaying the "excess" sensitivity of expectations and risk aversion to bad news. Overall, the LT dating rule is able to detect persistent downward and upward trends as well as temporary bear market rallies (or short-run bull markets). Hence, it serves as good proxy to evaluate the accuracy of the real-time predictions.

[8]The first recession lasted from August 1990 to March 1991, the second from April to November 2001, the third from January 2008 to June 2009, and the most recent one (as of November 2021) from March to April 2020 (https://fred.stlouisfed.org/series/USRECD).

Table 2: Bull and Bear Market Periods

Bull Markets			**Bear Markets**		
Dates	Durat.	Amplit.	Dates	Durat.	Amplit.
1989-11-17 to 1990-07-13	35	8	1990-07-20 to 1990-10-12	13	–18
1990-10-19 to 1998-07-17	405	296	1998-07-24 to 1998-09-04	7	–18
1998-09-11 to 1999-07-16	45	46	1999-07-23 to 1999-10-15	13	–12
1999-10-22 to 2000-03-24	23	22	2000-03-31 to 2001-09-21	78	–37
2001-09-28 to 2002-01-04	15	21	2002-01-11 to 2002-10-04	39	–32
2002-10-11 to 2002-11-29	8	17	2002-12-06 to 2003-03-07	14	–11
2003-03-14 to 2007-10-12	240	88	2007-10-19 to 2008-11-21	58	–49
2008-11-28 to 2009-01-02	6	16	2009-01-09 to 2009-03-06	9	–27
2009-03-13 to 2010-04-23	59	78	2010-04-30 to 2010-07-02	10	–16
2010-07-09 to 2011-04-29	43	33	2011-05-06 to 2011-08-19	16	–18
2011-08-26 to 2015-07-17	204	89	2015-07-24 to 2016-02-12	30	–12
2016-02-19 to 2018-09-21	136	57	2018-09-28 to 2018-12-21	13	–18
2018-12-28 to 2020-02-14	60	40	2020-02-21 to 2020-03-20	5	–32
2020-03-27 to 2021-05-07	59	84			

Notes: The classification follows the dating rule of Lunde and Timmermann (2004). The duration is measured in weeks and the amplitude as percentage price change between two subsequent extreme points.

4.2 Data Aggregation

To utilize the information from a high-dimensional dataset of potential predictors, we apply four different PCA techniques to aggregate the information into a few components and to filter out the noise: (i) conventional PCA, (ii) sparse PCA (SPCA), (iii) targeted PCA (TPCA), and (iv) targeted sparse PCA (TSPCA). Table 3 shows the number of selected components and the proportion of explained variance under full-sample knowledge.

Table 3: Cumulative Proportion of Explained Variance (In-Sample)

	Full Dataset ($N = 146$)		VIX: Targeted Dataset ($N = 75$)		ERP: Targeted Dataset ($N = 75$)	
	PCA	Sparse PCA	PCA	Sparse PCA	PCA	Sparse PCA
PC1	0.19	0.15 (58)	0.25	0.20 (44)	0.19	0.18 (28)
PC2	0.35	0.30 (42)	0.37	0.31 (30)	0.34	0.32 (40)
PC3	0.42	0.37 (51)	0.46	0.41 (47)	0.41	0.38 (41)
PC4	0.47	0.42 (36)				
PC5	0.52	0.47 (40)				
PC6	0.56	0.51 (54)				

Notes: The number of principal components is based on the selection procedure presented in Section 2.1 and for the full sample period. For the sparse PCA, the proportion of explained variance is calculated via the QR decomposition of the (correlated) principal component scores. "Targeted Dataset" refers to the subset of indicators obtained via soft thresholding. The number of non-zero coefficients of the sparse PCA are given in parentheses.

Six components are selected for the conventional PCA and the sparse PCA. These capture 56% (PCA) and 51% (SPCA) of the total variation. The benefits of soft thresholding becomes evident when targeting the (sparse) PCA on the VIX. In this case, three components are sufficient and explain more of the variation than the first three components of their non-targeted counterparts (TPCA: 46% vs. PCA: 42%; TSPCA: 41% vs. SPCA: 37%). When targeting on the equity risk premium (ERP), the explained variance of the first three components is similar to their non-targeted counterparts (TPCA: 41%; TSPCA: 38%).

Figures B1–B3 in Appendix B show the principal components over time. It is noticeable that the sparse PCA (right panel) achieves a more distinct smoothing over the indicators compared to the conventional PCA (left panel), irrespective of whether the set of predictors is unrestricted (Figure B1) or targeted (Figures B2 and B3). Hence, we can conclude that the sparse factors are more capable to filter out the noise, confirming the results of Rapach and Zhou (2019).

As mentioned before, the number of obtained principal components (and their interpretation) might vary when considering a training set that only covers a part of the full sample. Figure B4 in Appendix B provides an overview on the number of principal components used in the out-of-sample exercise in Section 5. The number of components used in the PCA and the SPCA varies between five and eight, whereas for the TPCA and TSPCA, three and seven mark the lower and upper bound.

5 Out-of-Sample Results

We use a recursive forecasting procedure with an expanding window to capture the stock market dynamics from October 22, 2004 to May 7, 2021, yielding a total of 864 forecasts. Our out-of-sample period starts with a prolonged bullish market (see Table 2). Starting from October 2007 onward, we have a total of 14 turning points that our models aim to predict in a real-time setting. Our entire methodology (estimation of PCAs, MS models, and forecast combination) is always applied on a weekly updated training sample. The first training set uses the available information from November 17, 1989 to October 15, 2004 to forecast regimes and returns for October 22, 2004. For the last forecast, information up to April 30, 2021 is used.

We evaluate the predictive power of our approach in terms of its statistical quality and its practical use for an investor. Our investment universe comprises a risky asset (SPDR S&P 500 ETF, Code: SPY) and an (almost) risk-free asset (3 M Treasury bill, secondary market rate). We resort to actually traded products to enable an assessment from an investor's perspective. For an evaluation of the economic value, we rely on two investment strategies. For regime forecasts, we employ a switching strategy, which allocates the total wealth either in the broad stock market or in Treasury bills, according to the different forecast clusters from Table 1. In the case of return forecasts, we utilize a mean-variance strategy where the stock market portfolio weight depends on the optimal conditional portfolio rule (Merton 1969). Short-selling and leverage are not allowed in both cases. As benchmarks, we utilize (simple) strategies based on the one-year moving average (MA_12M) for the regime forecasts. For the return forecasts, the historical average (HIST) of the equity risk premium serves as a benchmark. A MS model with TCTP according to Eq. (5) is applied for both types of forecasts. For an evaluation of the economic value, we additionally employ the straightforward buy-and-hold strategy (BH), the 50/50 strategy (50% equity and 50% risk-free), and the 60/40 strategy (60% equity and 40% risk-free) as further benchmarks.

Finally, we account for transaction costs to obtain a realistic perspective for an investor. Estimating transaction costs is not an easy task and depends on many factors (e.g., order size, market liquidity, and investor characteristics). We refrain from delving deeper into this topic and assume transaction costs of 20 basis points (bps) that are proportional to the size of the position change (e.g., Çakmaklı and van Dijk 2016) for the baseline case. The impact of alternative transaction cost assumptions (0 bps and 50 bps) as well as an ex ante consideration of transaction costs in spirit of Dal Pra et al. (2018) is investigated in Appendix E. All evaluation measures for the statistical quality and the economic performance are explained in Appendix C.

5.1 Regime Predictability

Statistical Performance: In the context of stock market regime identification, the timely detection of bear markets is particularly important for loss reduction. Put differently, the statistical evaluation follows the methodology of classification decisions. Hence, we have to handle the trade-off between the true positive rate (i.e., a bear market is correctly predicted) and the false positive rate (i.e., a bull market is misclassified as bear market; false alarm). Within a two regime case, one typically relies on a cut-off of 50% in the predicted probabilities to differentiate between regimes. This threshold appears to be the most intuitive choice at a first glance. The receiving operating characteristic (ROC) curve is a more nuanced approach of evaluating classifications as it considers a grid of thresholds and displays the benefits (true positive rate) and costs (false positive rate) of a classification model in a two-dimensional figure (Fawcett 2006). A popular way to aggregate the performance of the ROC curve into a single value is to calculate the area under the curve (hereafter AUC). Since the ROC curve is plotted on a unit square, the AUC takes values between 0 and 1, where 1 (0.5) corresponds to a perfect (random) classification.

As naïve benchmark for regime identification, we consider the one-year unweighted moving average (MA) of the excess stock returns. The MA is often applied as an indicator to signal trends and is therefore useful for market timing decisions (see, among others Brock et al. 1992). To separate the smoothed performance into two regimes, we define the binary variable D_t^{MA}:

$$D_t^{MA} = 0 \quad \text{if} \quad MA_t \geq 0 \quad \text{as bullish phase}$$
$$D_t^{MA} = 1 \quad \text{if} \quad MA_t < 0 \quad \text{as bearish phase}$$

The window length of one-year is indeed arbitrary, but common in practice. A shorter length might lead to too many turning points and very short-lived bullish and bearish periods, whereas a longer memory would not appropriately account for the most recent price dynamics.[9]

Table 4 shows the statistical performance of our forecasts against the moving average and the simple MS model. All proposed models can outperform the MA and the MS model with TCTP in terms of the quadratic probability score (QPS). In addition, the AUC statistics is better for all forecasts (exceptions: A-OBS-BMA, A-PC-BMA, B-OBS-BMA, and B-PC-BMA) than for the MS model with TCTP. The total accuracy reaches up to 81.0% (B-OBS-AVE and B-SPC-BMA) with bear market accuracy rates of up to 77.3% (B-PC-AVE). However, our forecasts cannot consistently outperform both benchmarks in the classification metrics. Nevertheless, they perform better than the MA (the MS model with TCTP) when it comes to predict bear (bull) markets.

For a formal identification of the best forecasts, we rely on an AUC test for the accuracy of regime predictions (DeLong et al. 1988). Table D1 in Appendix D displays the results. The best forecast according to this test is B-TSPC-AVE, which outperforms all but one forecast, followed by A-TSPC-AVE, which beats all but two forecasts. Hence, soft thresholding on a sparse PCA outperforms all other aggregation techniques (and the observable predictors). In general, the simple average performs better than the more complex BMA (exception: A-SPC-BMA). Finally, the richer Specification A (conditional transitions and mean) does not outperform the more parsimonious Specification B (conditional transitions only).

In addition to the average performance of the forecasts, it is of particular interest to see how timely recessions are detected. Figures D1–D5 in Appendix D show the bear market probabilities that the different forecasts predict. Overall, the predicted bear market probabilities respond promptly to regime turning points. In addition, the respective regime forecasts have a high degree of similarity across the different forecast combination clusters. Again, the simple average outperforms the BMA in terms of the R^2 for all forecast combinations except those using a sparse PCA. In terms of

[9] We also provide results with window lengths of 3, 6, 24, and 36 months as part of our robustness tests in Tables E1 and E2 of Appendix E. The 12-month MA performs best in the case of the statistical performance (exception: share of correctly predicted bear markets). In terms of the economic value, longer window lengths provide, on average, higher returns, while shorter lengths yield better tail risk measures and a lower standard deviation. Hence, we also rely on a 12 month window length in this case as it provides a good compromise between returns and tail risk.

this metric, forecast A-SPC-BMA ($R^2 = 0.287$) and B-TSPC-AVE ($R^2 = 0.277$) perform slightly better than the remaining forecasts.

Table 4: Regime Forecasts: Statistical Performance

Forecast	QPS	AUC	Accuracy	Bear	Bull
MA_12M (D_t^{MA})	0.310		0.836	0.475	0.906
TCTP	0.324	0.830	0.788	0.773	0.791
A-OBS-AVE	**0.250**	**0.853**	0.800	0.738	0.812
A-PC-AVE	**0.281**	**0.834**	0.795	0.766	0.801
A-SPC-AVE	**0.268**	**0.845**	0.793	0.745	0.802
A-TPC-AVE	**0.259**	**0.850**	0.808	0.766	0.816
A-TSPC-AVE	**0.240**	**0.861**	0.808	0.738	0.822
A-OBS-BMA	**0.290**	0.820	0.785	0.624	0.816
A-PC-BMA	**0.275**	0.813	0.803	0.681	0.827
A-SPC-BMA	**0.233**	**0.859**	0.818	0.688	0.844
A-TPC-BMA	**0.276**	**0.842**	0.796	0.695	0.816
A-TSPC-BMA	**0.277**	**0.836**	0.796	0.660	0.823
B-OBS-AVE	**0.245**	**0.853**	0.810	0.731	0.826
B-PC-AVE	**0.274**	**0.838**	0.802	0.773	0.808
B-SPC-AVE	**0.262**	**0.846**	0.799	0.752	0.808
B-TPC-AVE	**0.253**	**0.851**	0.808	0.752	0.819
B-TSPC-AVE	**0.238**	**0.863**	0.808	0.731	0.823
B-OBS-BMA	**0.302**	0.803	0.774	0.582	0.812
B-PC-BMA	**0.285**	0.810	0.800	0.688	0.822
B-SPC-BMA	**0.256**	**0.846**	0.810	0.738	0.824
B-TPC-BMA	**0.279**	0.824	0.788	0.638	0.817
B-TSPC-BMA	**0.232**	0.824	0.819	0.638	0.855

Notes: All forecasts except D_t^{MA} (naïve 12-month moving average) and TCTP (MS model with TCTP and without external predictors) are estimated as MS-TVTP models. Specification A contains predictors in the switching equation and the conditional mean equation, Specification B contains predictors in the switching equation only. OBS: observable predictors; PC: conventional PCA; SPC: sparse PCA; TPC: conventional PCA with soft thresholding; TSPC: sparse PCA with soft thresholding; AVE: simple average; BMA: Bayesian model averaging; QPS: quadratic probability score; AUC: area under the curve; Accuracy: share of correctly predicted regimes overall (50% threshold); Bear/Bull: share of correctly predicted bearish/bullish regimes (50% threshold). See Appendix C for a detailed description of the evaluation measures. Forecasts that outperform both benchmarks are highlighted in bold.

Table 5 aggregates the information from Figures D1–D5 and shows how quickly turning points are detected over the course of the different recessions. The MS model with TCTP and without external predictors serves as benchmark. As an illustration, the best model can identify the start and end of the GFC without a delay. The Covid-19 crash is also classified as a bear market from end of February 2020 onwards, with the re-entry taking place in mid-April. Confirming the impression from Table 4, the TCTP specification is well-suited to detect bear markets. Even the best of our forecasts is only able to match its performance (see column “Bull → Bear”). However, a key advantage of our approach is to identify the turning point from bear to bull markets in a timely manner as our best model never exceeds a delay of four weeks when classifying the switch into a bull market (see column "Bear → Bull").

Table 5: Identification of Turning Points in Bull and Bear Markets

	Bull → Bear			Bear → Bull		
	Best	Worst	TCTP	Best	Worst	TCTP
Global Financial Crisis I (2007-10-19 to 2008-11-21)	0	0	0	0	+6	+6
Global Financial Crisis II (2009-01-09 to 2009-03-06)	0	+1	0	0	+40	+41
Flash Crash Aftermath (2010-04-30 to 2010-07-02)	+1	+2	+1	+4	+13	+14
Debt Crisis (2011-05-06 to 2011-08-19)	+8	+12	+8	+3	+22	+23
Chinese Market Crash (2015-07-24 to 2016-02-12)	+4	+4	+4	0	+5	+5
Economic Slowdown Fear (2018-09-28 to 2018-12-21)	+2	+4	+2	+1	+7	+7
COVID-19 Crash (2020-02-21 to 2020-03-20)	+1	+1	+1	+3	+36	+18

Notes: Table shows the out-of-sample delay (in weeks) when identifying regime switches from bull to bear markets and from bear to bull markets. The dating rule of Lunde and Timmermann (2004) (assuming full-sample knowledge) is used for the classification of bull and bear markets. Across all forecast combinations, the performance of the best model and the worst model is reported with the delay of the MS-TCTP model as benchmark. The threshold for the bear market probability is 50%.

Economic Value: All the metrics so far have tested for (sometimes nuanced) differences in the statistical performance of the different forecasts. For an investor, however, it is important to see if these statistical differences turn into an economic value added, in particular when considering transaction costs. Hence, we evaluate the profitability of regime forecasts by translating the regime probabilities into a binary investment strategy that either allocates the total wealth to the stock market (risk-on) or to short-term government bonds (risk-off). If a bear (bull) market is predicted, we avoid (go long in) the stock market. Accordingly, the optimal stock market weight w_t^* goes hand in hand with a threshold dependent indicator $\hat{I}_t(\tau)$:

$$w_t^* = \hat{I}_t(\tau) \tag{11}$$

In our baseline scenario, we consider transaction costs ex post when calculating the performance metrics and switch the indicator to zero if the bear market probability $\hat{p}_t$

exceeds a certain threshold τ (in our baseline scenario, we assume $\tau = 0.5$):[10]

$$\hat{I}_t = 0 \quad \text{if} \quad \hat{p}_t \geq \tau$$
$$\hat{I}_t = 1 \quad \text{if} \quad \hat{p}_t < \tau$$

Table 6 shows the economic value of the forecasts in the baseline scenario against five benchmarks (BH, 50/50, 60/40, MA, and TCTP). The buy-and-hold strategy performs best in terms of final wealth (R^{cum}) and annualized average returns ($\bar{R}$). The 50/50 strategy yields the lowest annualized standard deviation ($\bar{\sigma}$) and the best (conditional) value-at-risk (VaR and CVaR). Our forecasts, in turn, perform particularly well when considering the annualized Sharpe ratio (SR) and the annualized certainty equivalent return (CER). In particular, eight forecasts that rely on the restricted Specification B perform better than all benchmarks in these two risk-adjusted measures. Hence, together with the similar statistical performance of Specification A and Specification B, these results indicate that modeling the conditional transitions might be sufficient when forecasting regimes.

Turning to the forecast combination scheme, we find that more combinations based on the simple average — as opposed to the BMA — outperform the five benchmarks. On the other hand, the best risk-adjusted metrics are found for the forecast combination B-SPC-BMA. The latter finding is reassuring since B-SPC-BMA is also the forecast with the best statistical accuracy. In addition, the models with the best AUC from Table 4 (A-TSPC-AVE and B-TSPC-AVE) outperform all benchmarks in terms of their SR and CER. Hence, their statistical accuracy is also reflected in an actual value added.

[10] As part of our robustness tests, we also consider a more (less) recession-averse agent and set $\tau = 0.25$ ($\tau = 0.75$). In our second scenario, we consider transaction costs ex ante in spirit of Dal Pra et al. (2018) in combination with varying threshold levels:

$$\hat{I}_t = 0 \quad \text{if} \quad \hat{p}_t \geq \tau \quad \textit{and} \quad -\hat{r}_{e,t+1} \geq c_t$$
$$\hat{I}_t = 1 \quad \text{if} \quad \hat{p}_t < \tau \quad \textit{and} \quad \hat{r}_{e,t+1} \geq c_t$$
$$\hat{I}_t = I_{t-1} \quad \textit{else}$$

Hence, an investor switches from risk-off (risk-on) to risk-on (risk-off) only if the predicted stock return (loss) that results from the regime forecast exceeds the transaction costs c_t. In all other cases, no trade is executed. The economic performance for an ex ante consideration of different transaction costs is documented in Table E3 in Appendix E. In Table E4, we vary the amount of ex post transaction costs and the recession thresholds. To conserve space, we show only the results for the annualized certainty equivalent return as important risk-adjusted measure.

We can infer a couple of interesting results from this robustness test. First, obviously more trades are executed in the absence of transaction costs, which also leads to higher annualized certainty equivalent returns. Conversely, higher transaction costs lead to fewer trades and a lower CER. Second, when considering transactions costs ex ante, the CER is, on average, lower as compared to the case of ex post trading costs as fewer trades are conducted. 50 bps appear to be prohibitive for the case of ex ante transaction costs. Third, lowering the threshold makes the positioning “too cautious” and leads to a disproportionate decline in returns relative to the risk improvement. A cut-off point of 75%, however, leads to higher annualized CER returns.

Table 6: Regime Forecast: Economic Value

Forecast	R^{cum}	$\bar{R}$	$\bar{\sigma}$	SR	CER	MaxDD	VaR	CVaR
BH	5.28	10.53	17.90	0.51	4.38	−54.6	−3.87	−6.04
50/50	2.76	6.30	8.57	0.58	3.90	−29.2	−1.95	−2.90
60/40	3.18	7.22	10.35	0.57	4.31	−34.7	−2.34	−3.51
MA_12M (D_t^{MA})	4.03	8.75	12.92	0.57	4.92	−22.2	−2.79	−4.58
TCTP	3.96	8.63	10.25	0.71	5.74	−12.5	−2.11	−3.55
A-OBS-AVE	3.87	8.49	10.43	0.69	5.53	−12.5	−2.16	−3.66
A-PC-AVE	3.77	8.32	10.30	0.68	5.41	−17.7	−2.12	−3.59
A-SPC-AVE	4.16	8.95	10.40	**0.73**	**6.01**	−12.5	−2.14	−3.61
A-TPC-AVE	4.21	9.03	10.58	**0.73**	**6.03**	−17.7	−2.14	−3.65
A-TSPC-AVE	4.18	8.99	10.57	**0.72**	**5.99**	−14.4	−2.16	−3.64
A-OBS-BMA	2.48	5.61	10.72	0.40	2.61	−23.7	−2.39	−3.93
A-PC-BMA	4.70	9.76	10.92	**0.77**	**6.64**	−17.5	−2.16	−3.73
A-SPC-BMA	4.29	9.16	11.39	0.69	**5.89**	−26.5	−2.39	−3.97
A-TPC-BMA	3.89	8.52	11.14	0.65	5.34	−17.5	−2.31	−3.88
A-TSPC-BMA	4.08	8.84	10.78	0.70	**5.77**	−18.9	−2.22	−3.71
B-OBS-AVE	4.24	9.08	10.55	**0.74**	**6.08**	−12.5	−2.14	−3.65
B-PC-AVE	3.99	8.69	10.22	**0.72**	**5.80**	−12.5	−2.12	−3.50
B-SPC-AVE	4.37	9.29	10.49	**0.76**	**6.31**	−12.5	−2.14	−3.61
B-TPC-AVE	4.30	9.18	10.60	**0.74**	**6.17**	−15.3	−2.14	−3.64
B-TSPC-AVE	4.45	9.40	10.61	**0.76**	**6.38**	−13.2	−2.16	−3.64
B-OBS-BMA	2.64	6.01	10.64	0.44	3.02	−18.0	−2.30	−3.86
B-PC-BMA	4.75	9.84	10.70	**0.80**	**6.79**	−13.0	−2.14	−3.63
B-SPC-BMA	5.06	10.25	10.86	**0.82**	**7.14**	−12.6	−2.16	−3.62
B-TPC-BMA	3.32	7.49	11.01	0.56	4.36	−17.5	−2.41	−3.94
B-TSPC-BMA	4.82	9.92	11.39	**0.75**	**6.64**	−24.3	−2.33	−3.83

Notes: Table shows the economic value of different investment strategies. BH: Buy & Hold Strategy of the S&P 500; 50/50 and 60/40: mixed strategy S&P 500 and 3 M Treasury Bill; D_t^{MA}: naïve 12-month moving average; TCTP: MS model with TCTP and without external predictors. Specification A contains predictors in the switching equation and the conditional mean equation, Specification B contains predictors in the switching equation only. OBS: observable predictors; PC: conventional PCA; SPC: sparse PCA; TPC: conventional PCA with soft thresholding; TSPC: sparse PCA with soft thresholding; AVE: simple average; BMA: Bayesian model averaging; R^{cum}: final wealth of strategy assuming a 1\$ investment; $\bar{R}$: annualized average returns; $\bar{\sigma}$: annualized standard deviation; SR: annualized Sharpe ratio; CER: annualized certainty equivalent return with $\gamma = 3$; MaxDD: maximum drawdown; VaR: value-at-risk; CVaR: conditional value-at-risk. See Appendix C for a detailed description of the evaluation measures. Forecasts that outperform all benchmarks are highlighted in bold.

Another takeaway is the worse performance of observable predictors in comparison to the principal components, confirming the findings of Neely et al. (2014) and Çakmaklı and van Dijk (2016). Within the PCA models, introducing sparsity in the PCA helps to improve the predictability with, on average, a better risk-adjusted performance. In addition to a more straightforward interpretation, the sparse factors provide a sharper distinction between signal and noise, which is in line with the results of Rapach and Zhou (2019). Finally, it is worth noting that the best of our forecasts (B-SPC-BMA) in terms of its risk-adjusted performance comes close to the return metrics of the buy-and-hold strategy, while at the same time having a much lower annualized standard deviation ($\bar{\sigma}$) and a maximum drawdown that is close to the best benchmark value (MS model with TCTP).

Another way of illustrating the economic performance of the different forecasts is to plot the cumulative returns of the different strategies over time. This provides another view on the ability to detect turning points in a timely manner. Figure D6 in Appendix D shows the cumulative returns over time. As indicated by the results in Table 6, the BH strategy performs best when considering the final wealth. However, in particular during the GFC and the Covid-19 crash, losses can be reduced and re-entry points can be found in a timely manner when relying on our forecast methodology. Finally, it also becomes evident that our best forecast (B-SPC-BMA) performs even better than the BH strategy for almost the entire out-of-sample period and is only outperformed during the booming post-Covid crash stock market, which shifts the overall results in favor of the BH strategy. All these findings suggest that it pays off to model the switching process with TVTP when evaluating returns on a risk-adjusted basis.

5.2 Return Predictability

Statistical Performance: When it comes to forecasting stock market returns, accurate point forecasts in terms of a low mean squared prediction error (MSPE) are difficult to find, in particular at a weekly frequency. It is therefore common to compare the forecast quality relative to the historical average. Hence, we rely on the R^2_{OS} proposed by Campbell and Thompson (2008). The historical average is calculated with an expanding window, so that the period from November 17, 1989 to October 15, 2004 is used for the first forecast.[11]

Table 7 shows the statistical performance of the return forecasts. There is no added value with regard to the R^2_{OS} and root mean squared error ($RMSE$) over the entire out-of-sample period. The values for R^2_{OS} are negative (except for A-TPC-AVE and A-TPC-BMA) and the null hypothesis of $R^2_{OS} \leq 0$ cannot be rejected. The sign predictability varies between 55% and 58%; positive returns are correctly predicted in 76% to 83% of the cases and the negative return accuracy rate ranges from 21% to 30% (when excluding the outlier forecast B-OBS-BMA that predicts 94% of the positive returns and 7% of the negative returns).

For a formal identification of the best forecasts, we rely on the Diebold and Mariano (1995) test for the accuracy of return predictions. Table D2 in Appendix D displays the results. A-TPC-AVE is the best forecast as it outperforms nine other forecasts, followed

[11]The assumption of an expanding window is required for an evaluation of our forecasts within a nested framework and for the same training sample. Obviously, such a benchmark model implies that returns (and regimes) are unpredictable. However, when comparing the RMSE of the historical average (2.49) to moving averages with varying lengths, it is never outperformed (3 M: 2.59; 6 M: 2.53; 12 M: 2.51; 24 M: 2.50; 36 M: 2.49).

by B-TSPC-BMA (8/20). The best forecast for regime predictions (B-SPC-BMA) ranks third (6/20). According to this test, Specification B, which models only the conditional transitions, performs slightly better than specification A, which also models the conditional mean process. Applying sparsity and/or soft thresholding on the PCA appears to outperform the observable predictors and a conventional PCA. Finally, considering the forecast combination scheme, Specification A (B) performs worse when relying on the BMA (simple average).[12]

Table 7: Return Forecasts: Statistical Performance

Forecast	RMSE	CW R^2_{OS}	p-val.	Direction	R^+	R^-
HIST	2.4866			0.582	1.000	0.000
TCTP	2.4929	−0.511	0.60	0.549	0.740	0.283
A-OBS-AVE	2.4918	−0.418	0.47	0.567	0.789	0.258
A-PC-AVE	2.4931	−0.527	0.53	0.553	0.779	0.238
A-SPC-AVE	2.4940	−0.596	0.58	0.552	0.783	0.230
A-TPC-AVE	**2.4854**	**0.096**	0.18	0.576	0.825	0.230
A-TSPC-AVE	2.4952	−0.695	0.71	0.571	0.831	0.208
A-OBS-BMA	2.5120	-2.058	0.90	0.558	0.807	0.211
A-PC-BMA	2.5035	-1.371	0.60	0.565	0.807	0.227
A-SPC-BMA	2.5132	-2.152	0.87	0.551	0.763	0.255
A-TPC-BMA	**2.4826**	**0.318**	0.20	0.566	0.759	**0.296**
A-TSPC-BMA	2.5038	-1.388	0.72	0.571	0.807	0.241
B-OBS-AVE	2.4917	−0.412	0.61	0.566	0.797	0.244
B-PC-AVE	2.4941	−0.611	0.65	0.552	0.759	0.263
B-SPC-AVE	2.4926	−0.483	0.58	0.556	0.758	0.274
B-TPC-AVE	2.4924	−0.474	0.57	0.560	0.777	0.258
B-TSPC-AVE	2.4926	−0.486	0.61	0.560	0.781	0.252
B-OBS-BMA	2.4906	−0.322	0.88	0.579	0.944	0.069
B-PC-BMA	2.4949	−0.676	0.66	0.559	0.775	0.258
B-SPC-BMA	2.4900	−0.275	0.47	0.559	0.767	0.269
B-TPC-BMA	2.4949	−0.675	0.70	0.565	0.779	0.266
B-TSPC-BMA	2.4896	−0.243	0.48	0.565	0.817	0.213

Notes: HIST: historical average of excess stock returns; TCTP: MS model with TCTP and without external predictors. Specification A contains predictors in the switching equation and the conditional mean equation, Specification B contains predictors in the switching equation only. OBS: observable predictors; PC: conventional PCA; SPC: sparse PCA; TPC: conventional PCA with soft thresholding; TSPC: sparse PCA with soft thresholding; AVE: simple average; BMA: Bayesian moving average; R^2_{OS}: out-of sample R^2; CW: CW test statistic; Direction: correctly predicted forecast direction; R^+: true positive forecasts; R^-: true negative forecasts. See Appendix C for a detailed description of the evaluation measures. Forecasts that outperform both benchmarks are highlighted in bold.

Economic Value: As a final step, we test if the dispiriting statistical performance of the return forecasts is also reflected in their economic value. We assume a risk-averse agent

[12]To conserve space, we do not show a graphical representation of the excess return forecasts over time. These are very noisy and not informative due to their low R^2 that never exceeds 0.005. All omitted results are available on request.

with mean-variance preferences. Solving the standard expected utility maximization, we obtain the following optimal stock market weight that is restricted between 0% and 100%:

$$w_t^* = \frac{1}{\gamma} \frac{\hat{r}_{t+1}}{\hat{\sigma}_{t+1}^2} \tag{12}$$

$\hat{r}_{t+1}$ represents the one-step ahead return forecast and $\hat{\sigma}_{t+1}^2$ the expected variance. The coefficient of relative risk aversion γ is set to 3, and we use the historical 5-year variance as risk proxy.[13] We consider proportional transaction costs of 20 basis points (bps) ex post when calculating the performance metrics.[14]

Table 8 shows the economic value of the forecasts against five benchmarks (BH, 50/50, 60/40, HIST, and TCTP). Again, the buy-and-hold strategy performs best in terms of final wealth (R^{cum}) and annualized average returns ($\bar{R}$). The 50/50 strategy yields the lowest annualized standard deviation ($\bar{\sigma}$) and the best conditional value-at-risk (CVaR). A few of our forecasts outperform all benchmarks in terms of the risk-adjusted metrics (SR and CER) and tail risk measures (MaxDD and VaR). These all rely on the restricted Specification B. Hence, it appears that modeling conditional returns does not improve the economic value of regime or return forecasts.

Turning to the combination schemes, we find that forecast combinations using the BMA (again including B-SPC-BMA) produce slightly better values for the risk-adjusted performance measures than the simple average. Again, the aggregation of information in principal components is more helpful than just relying on observable predictors. Finally, and perhaps the most important takeaway, is that the return forecasts provide a lower final wealth, lower annualized returns, a lower SR, and a lower CER when compared to the regime forecasts (see Table 6). On the other hand, return forecasts based on the restricted Specification B provide a lower standard deviation and bet-

[13]It has to be noted that the forecast results are also influenced by the expected variance proxy. A rolling window of 5 years implies a high degree of persistence. Consequently, the stock exposure might be biased downwards, in particular after large shocks. An extension for future research would be to explicitly include a variance forecast.

[14]Although the assumption of $\gamma = 3$ is common in the empirical literature, we provide results for $\gamma = 2$ and $\gamma = 5$ in Table E5 of Appendix E. In addition, we vary the amount of ex post transaction costs. Including ex ante transaction costs in a mean-variance optimization is a more complex task that we leave open for future studies. Again, we show only the results for the annualized certainty equivalent return as important risk-adjusted measure. We can infer a couple of interesting results from this robustness test. First, obviously more trades are executed in the absence of transaction costs, which also leads to higher annualized certainty equivalent returns. Yet, we do not find a substantial improvement of our forecasts in terms of outperforming the benchmarks. Second, increasing the degree of risk-aversion (to 5) makes the positioning "too cautious" and leads to a disproportionate decline in returns relative to the risk improvement.

Finally, we also use our excess return prediction as trigger for the switching strategy. Following Dal Pra et al. (2018), we fully invest in the stock market whenever the excess returns are positive or at least zero. For negative predictions (i.e., expected stock returns are smaller than the risk-free rate), we only invest in Treasury Bills. Table E6 in Appendix E shows the results. Here, we find an outperformance of the benchmarks in the absence of transaction costs or when transaction costs are only considered ex post. In particular, Specification B that models only the conditional transitions performs well in that regard.

ter tail risk measures than the corresponding regime forecasts. Hence, there is some economic value added in forecasting returns, in particular for a risk-adverse investor.

Table 8: Return Forecast: Economic Value

Forecast	R^{cum}	$\bar{R}$	$\bar{\sigma}$	SR	CER	MaxDD	VaR	CVaR
BH	5.28	10.53	17.90	0.51	4.38	−54.6	−3.87	−6.04
50/50	2.76	6.30	8.57	0.58	3.90	−29.2	−1.95	−2.90
60/40	3.18	7.22	10.35	0.57	4.31	−34.7	−2.34	−3.51
HIST	2.29	5.11	12.86	0.30	1.35	−43.5	−2.51	−4.77
TCTP	3.09	7.03	9.46	0.61	4.39	−12.9	−1.97	−3.30
A-OBS-AVE	2.21	4.88	11.39	0.32	1.66	−34.0	−2.07	−4.06
A-PC-AVE	2.81	6.42	9.91	0.52	3.65	−19.7	−2.11	−3.55
A-SPC-AVE	2.51	5.70	10.14	0.44	2.88	−28.0	−2.10	−3.70
A-TPC-AVE	2.27	5.06	12.32	0.31	1.51	−37.6	−2.16	−4.45
A-TSPC-AVE	1.86	3.81	11.82	0.22	0.45	−42.0	−2.12	−4.40
A-OBS-BMA	1.69	3.21	13.57	0.14	-0.81	−43.6	−2.12	−4.88
A-PC-BMA	2.30	5.13	12.08	0.32	1.66	−39.7	−2.32	−4.37
A-SPC-BMA	1.52	2.57	12.61	0.10	−1.06	−43.8	−2.23	−4.72
A-TPC-BMA	1.96	4.13	13.03	0.22	0.31	−36.5	−2.02	−4.53
A-TSPC-BMA	2.06	4.44	11.76	0.27	1.10	−40.5	−2.34	−4.39
B-OBS-AVE	3.04	6.93	9.26	0.61	4.34	**−12.4**	**−1.91**	−3.25
B-PC-AVE	3.03	6.90	9.39	0.60	4.28	−13.6	−1.95	−3.29
B-SPC-AVE	3.02	6.87	9.35	0.60	4.26	−13.4	**−1.91**	−3.30
B-TPC-AVE	3.09	7.03	9.43	0.61	**4.40**	−13.9	**−1.91**	−3.31
B-TSPC-AVE	2.93	6.68	9.40	0.57	4.06	−15.1	−1.95	−3.34
B-OBS-BMA	2.02	4.31	10.22	0.30	1.47	−30.4	−2.22	−3.89
B-PC-BMA	3.38	7.61	9.66	**0.65**	**4.90**	**−12.4**	**−1.91**	−3.43
B-SPC-BMA	3.08	7.01	9.28	**0.62**	**4.42**	−16.2	**−1.90**	−3.29
B-TPC-BMA	2.86	6.54	9.28	0.56	3.95	−13.5	**−1.87**	−3.32
B-TSPC-BMA	2.81	6.42	9.63	0.53	3.73	−25.7	−2.00	−3.44

Notes: Table shows the economic value of different investment strategies. BH: Buy & Hold Strategy of the S&P 500; 50/50 and 60/40: mixed strategy S&P 500 / 3 M Treasury Bill; HIST: historical average of excess stock returns; TCTP: MS model with TCTP and without external predictors. Specification A contains predictors in the switching equation and the conditional mean equation, Specification B contains predictors in the switching equation only. OBS: observable predictors; PC: conventional PCA; SPC: sparse PCA; TPC: conventional PCA with soft thresholding; TSPC: sparse PCA with soft thresholding; AVE: simple average; BMA: Bayesian model averaging; R^{cum}: final wealth of strategy assuming a 1$ investment; $\bar{R}$: annualized average returns; $\bar{\sigma}$: annualized standard deviation; SR: annualized Sharpe ratio; CER: annualized certainty equivalent return with $\gamma = 3$; MaxDD: maximum drawdown; VaR: value-at-risk; CVaR: conditional value-at-risk. See Appendix C for a detailed description of the evaluation measures. Forecasts that outperform all benchmarks are highlighted in bold.

As next exercise, we graphically inspect the ability of the forecasts to detect turning points in a timely manner. Figure D7 in Appendix D shows the cumulative returns over time. As mentioned before, the BH strategy performs best when considering the final wealth. However, in particular during the GFC and the Covid-19 crash, losses can be reduced and re-entry points can be found in a timely manner when relying on our forecast methodology and the restricted Specification B. But, it does become evident once more that the cumulative performance of our forecasts is worse in predicting returns in comparison to predicting regimes (see Figure D6).

In a final step, we explore the performance of our forecast methodology conditional on the state of the stock market or the state of the business cycle. For this purpose, we separate the forecasts into two subsamples depending on whether the observation in $t+1$ is assigned to a bear market (recession) or a bull market (expansion). To classify the market state, we use the dating rule of Lunde and Timmermann (2004) (see Section 4.1) and for the business cycle, we rely on the binary recession indicator of the NBER. Table 9 shows the results where the historical mean is used as benchmark for calculating the R^2_{OS} and the Δ_{CER}.

Table 9: Market and Economic State Dependency of Return Forecasts

	LT Dating Rule				**NBER**			
	Bull		**Bear**		**Expansion**		**Recession**	
	$\mathbf{R^2_{OS}}$	$\mathbf{\Delta_{CER}}$	$\mathbf{R^2_{OS}}$	$\mathbf{\Delta_{CER}}$	$\mathbf{R^2_{OS}}$	$\mathbf{\Delta_{CER}}$	$\mathbf{R^2_{OS}}$	$\mathbf{\Delta_{CER}}$
TCTP	−2.77	−4.52	**2.06**	27.86	−1.22	−1.44	0.42	36.17
A-OBS-AVE	−3.29	−5.47	**2.86**	17.94	−0.44	−1.93	−0.38	15.60
A-PC-AVE	−3.15	−4.45	**2.47**	24.16	−0.91	−1.32	−0.03	28.68
A-SPC-AVE	−3.05	−4.66	**2.20**	21.27	−0.85	−1.48	−0.26	23.22
A-TPC-AVE	−2.04	−4.16	**2.53**	12.59	−0.42	−0.90	0.78	7.21
A-TSPC-AVE	−2.24	−4.33	**1.07**	9.22	−0.53	−1.08	−0.92	1.34
A-OBS-BMA	**0.60**	−2.60	−5.09	−1.19	0.00	−2.30	−4.78	−3.13
A-PC-BMA	−4.21	−3.20	**1.86**	10.63	−1.51	−0.76	−1.19	7.81
A-SPC-BMA	−3.15	−5.38	−1.01	6.32	−1.23	−3.13	−3.37	2.06
A-TPC-BMA	−0.91	−4.62	1.71	9.26	−0.76	−3.47	1.74	14.48
A-TSPC-BMA	−2.80	−4.05	0.22	11.45	−1.25	−1.80	−1.57	10.25
B-OBS-AVE	−2.63	−4.56	**2.12**	27.77	−0.96	−1.31	0.31	34.67
B-PC-AVE	−3.05	−4.57	**2.17**	27.54	−1.28	−1.53	0.27	35.91
B-SPC-AVE	−3.00	−4.68	**2.39**	27.85	−1.21	−1.61	0.48	36.35
B-TPC-AVE	−2.94	−4.32	**2.33**	27.19	−1.14	−1.32	0.40	35.24
B-TSPC-AVE	−2.87	−4.45	**2.23**	26.07	−1.04	−1.33	0.25	32.28
B-OBS-BMA	−1.34	−4.95	**0.84**	15.84	−0.52	−2.52	−0.06	19.03
B-PC-BMA	−3.24	−3.88	**2.25**	27.76	−1.42	−1.05	0.31	37.65
B-SPC-BMA	−2.77	−4.69	**2.57**	28.66	−1.07	−1.72	0.78	38.76
B-TPC-BMA	−3.11	−5.51	**2.11**	29.40	−1.26	−2.29	0.10	39.10
B-TSPC-BMA	−2.21	−4.20	**2.00**	23.69	−0.84	−1.32	0.54	29.33

Notes: Table shows the statistical performance and the economic value of the return forecasts in different states of the stock market and the economy (all in %). Market states are classified with the LT dating rule described in Section 4.1. Economic states are defined by the NBER classification. R^2_{OS}: out-of-sample R^2; Δ_{CER}: difference in the annualized certainty equivalent return (with $\gamma = 3$) between the forecast-based strategy and the historical average. TCTP: MS model with TCTP and without external predictors. Specification A contains predictors in the switching equation and the conditional mean equation, Specification B contains predictors in the switching equation only. OBS: observable predictors; PC: conventional PCA; SPC: sparse PCA; TPC: conventional PCA with soft thresholding; TSPC: sparse PCA with soft thresholding; AVE: simple average; BMA: Bayesian model averaging. See Appendix C for a detailed description of the evaluation measures. Forecasts that outperform the historical average according to the Clark and West (2007) statistic with a 10% significance level are highlighted in bold.

Confirming the broad consensus in the literature, return predictability is especially prevalent in "bad times." Here, the forecast-based strategy generates statistical and, in particular, economic value. A risk-averse mean-variance investor is willing to pay an annualized management fee (as indicated by Δ_{CER}) of up to 29.40% (in bear markets) and 39.10% (in recessions) to participate in the forecast-based strategy (B-TPC-BMA). The economic performance of B-SPC-BMA ranks second in that regard. In addition, the statistical performance is better during bear markets with significant values for the R^2_{OS} of more than 2% (mostly for Specification B). During economic recessions, we find positive but insignificant values for the R^2_{OS} (again, mostly for Specification B). However, we have to conclude that all forecast combinations are clearly inferior to the historical average during bull markets or expansions. All these results are in line with previous findings for regime switching models (Henkel et al. 2011) and for return predictions in general (Rapach and Zhou 2013).

5.3 Discussion

Before concluding, we need to revisit our testing framework and, in particular, shed light on the difference in performance between regime forecasts and return forecasts. It is not possible to directly compare the statistical goodness of the two forecasts due to the different scaling of the target variables (binary classification versus continuous scale). However, it is noticeable that the regime forecasts outperform their benchmarks more often than the return forecasts do. A better comparability can be achieved through an utility or profit analysis.

From an investor's point of view, it is striking that the strategy based on regime probabilities has a more attractive risk-return structure. Obviously, predicting trends is much easier than generating point forecasts. However, since regime probabilities are also a significant factor in the return forecast setting due to the weighting of the regime-dependent averages in Eq. (8), the difference in the performance cannot be entirely caused by this. Another reason might be excessive trading and large transaction costs. Table D3 in Appendix D shows the average annual turnover and the cumulative transaction costs of the two forecast strategies. These differ significantly only for Specification A. For Specification B, the cumulative transaction costs are similar. But even if we set transaction costs to zero, significant differences in the CERs remain (see column TC = 0 bps of Tables E4 and E5 in Appendix E). Hence, transaction costs cannot fully explain the differences in the economic performance of regime forecasts and return forecasts.

The profile of the respective investment strategy might be another key factor. To evaluate regime forecasts, a switching strategy is applied, which either invests only in Treasury Bills or in the stock market. Such a strategy (e.g., Pesaran and Timmermann

1995; Dal Pra et al. 2018) tends to reflect the investment behavior of a risk-neutral investor. In contrast, the mean-variance strategy explicitly takes risk aversion into account. Since there is no direct link between the regime probability and the classical utility function of an investor, we can only implicitly account for this fact by reducing the threshold to classify a bear market (or by reducing the relative risk aversion of the mean-variance strategy). Our robustness checks (see Tables E4 and E5 in Appendix E) show that lowering the threshold or the relative risk aversion narrows the gap between the performance of regime forecasts and return forecasts.

Another source affecting the degree of the stock allocation is the variance proxy. We rely on a common benchmark (sample variance of a rolling 5-year window) since we are not aiming at volatility forecasting. The resulting high variance persistence might bias the stock exposure downwards for the mean-variance strategy and the return forecasts, in particular after significant shocks. Conversely, when calculating the regime probabilities to obtain inference about the current state, an estimate of the regime-dependent variance is used to determine the density functions. Since the second moment is crucial to identify regime shifts (Kole and van Dijk 2017), a certain proportion of the difference in performance could be attributed to this procedure. However, if the explicit consideration of risk aversion as well as the proxy of the variance were to explain the different performance of regime and return forecast strategies, a switching strategy based on return forecasts should show a very similar performance to its counterpart based on regime forecasts. Table E6 in Appendix E shows the result of a switching strategy (Dal Pra et al. 2018), which invests in stocks (Treasury Bills) whenever the prediction is positive or zero (negative). We find that the CERs are decreasing in the vast majority of cases when compared to Tables E3 and E4.

Even if the difference in performance can be partly explained by the considerations above, regime forecasts remain superior compared to return forecasts in terms of their economic value. They are able to capture the trend-changing behavior of markets so that tail risks are reduced without sacrificing large returns. However, it is worth noting that the return predictions add statistical and, in particular, economic value during recessions or in declining markets.

6 Conclusions

Using a high-dimensional dataset of macro-financial variables, this paper offers a promising approach to predict stock market regimes on a weekly basis. Since stock market predictions suffer particularly from parameter instability and model uncertainty, our approach combines the merits of dimensionality reduction techniques, regime-switching models, and forecast combination. We provide a comprehensive overview of the empirical usefulness of Markov-switching models with principal components and time-varying transition probabilities.

Our best weekly regime forecasts use a (targeted) sparse principal component Markov-switching model and time-varying transition probabilities. They are suitable to respond to trend changes in a timely manner either to participate in recoveries or to prevent losses. This is also reflected in an actual economic value added as many of our forecasts excel all benchmarks in risk-adjusted performance measures. However, when considering stock market returns, our forecasts do not statistically outperform common benchmarks. The fact that return forecasts perform worse than regime forecasts is not surprising. Predicting the broader trend of the stock market is obviously easier than providing point forecasts, in particular on a weekly basis. This outperformance can also — to some extent — be explained by differences in the testing procedure and the investment strategy. Nevertheless, our return forecasts still provide some economic value added for risk-adverse investors as they generate a lower annualized standard deviation of the returns and better tail risk measures than the corresponding regime forecasts. We also confirm the previous findings that return predictability is limited to recessions or to periods of market turmoils.

In addition, we find that it is sufficient to model the time-varying conditional transitions in a Markov-switching model. Additionally modeling the conditional mean introduces further noise into the forecasts and particularly harms the economic performance of our forecasts. Based on our results, we propose to rely on dimensionality reduction techniques (instead of just relying on observable predictors) and to enhance the conventional principal component analysis with shrinkage methods such as sparsity and/or soft thresholding. Concerning the forecast combination technique, we do not find a clear advantage of Bayesian model averaging over the simple average.

Our results offer a variety of starting points for future work. First, modeling intra-regime dynamics in greater detail in the context of our forecasting model could be a promising extension — if practically feasible. Thereby, incorporating more than two regimes directly (Maheu et al. 2012) or a sequential partition (Hauptmann et al. 2014) are potential avenues. Second, despite the success of the regime forecasts, we do not consider the underlying forecast uncertainty in the economic application. A confidence measure for the probabilities could be useful for various applications, such as portfolio optimization or asset pricing. In this context, Alvarez et al. (2019) provide a foundation for future work. Finally, our approach can be extended to volatility forecasts and density forecasts. In addition, one could study international stock market indices or portfolios formed on industries or styles with the help of our forecasting model.

References

Alvarez, R., Camacho, M., and Ruiz, M. (2019). Inference on filtered and smoothed probabilities in Markov-switching autoregressive models. *Journal of Business and Economic Statistics*, 37(3):484–495.

Ang, A. and Bekaert, G. (2002). International asset allocation with regime shifts. *Review of Financial Studies*, 15(4):1137–1187.

Ang, A. and Timmermann, A. (2012). Regime changes and financial markets. *Annual Review of Financial Economics*, 4(1):313–337.

Bai, J. and Ng, S. (2002). Determining the number of factors in approximate factor models. *Econometrica*, 70(1):191–221.

Bai, J. and Ng, S. (2008). Forecasting economic time series using targeted predictors. *Journal of Econometrics*, 146(2):304–317.

Bekaert, G. and Hoerova, M. (2014). The VIX, the variance premium and stock market volatility. *Journal of Econometrics*, 183(2):181–192.

Bollerslev, T., Tauchen, G., and Zhou, H. (2009). Expected stock returns and variance risk premia. *Review of Financial Studies*, 22(11):4463–4492.

Brock, W., Lakonishok, J., and LeBaron, B. (1992). Simple technical trading rules and the stochastic properties of stock returns. *Journal of Finance*, 47(5):1731–1764.

Çakmaklı, C. and van Dijk, D. (2016). Getting the most out of macroeconomic information for predicting excess stock returns. *International Journal of Forecasting*, 32(3):650–668.

Campbell, J. Y. and Shiller, R. J. (1988). The dividend-price ratio and expectations of future dividends and discount factors. *Review of Financial Studies*, 1(3):195–228.

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

Campbell, J. Y. and Yogo, M. (2006). Efficient tests of stock return predictability. *Journal of Financial Economics*, 81(1):27–60.

Chauvet, M. and Potter, S. (2000). Coincident and leading indicators of the stock market. *Journal of Empirical Finance*, 7(1):87–111.

Chen, S.-S. (2009). Predicting the bear stock market: Macroeconomic variables as leading indicators. *Journal of Banking and Finance*, 33(2):211–223.

Chen, S.-S. (2012). Consumer confidence and stock returns over market fluctuations. *Quantitative Finance*, 12(10):1585–1597.

Chow, G., Jacquier, E., Kritzman, M., and Lowry, K. (1999). Optimal portfolios in good times and bad. *Financial Analysts Journal*, 55(3):65–73.

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

Coudert, V. and Gex, M. (2008). Does risk aversion drive financial crises? Testing the predictive power of empirical indicators. *Journal of Empirical Finance*, 15(2):167–184.

Cremers, K. M. (2002). Stock return predictability: A Bayesian model selection perspective. *Review of Financial Studies*, 15(4):1223–1249.

Dal Pra, G., Guidolin, M., Pedio, M., and Vasile, F. (2018). Regime shifts in excess stock return predictability: An out-of-sample portfolio analysis. *Journal of Portfolio Management*, 44(3):10–24.

DeLong, E. R., DeLong, D. M., and Clarke-Pearson, D. L. (1988). Comparing the areas under two or more correlated receiver operating characteristic curves: A nonparametric approach. *Biometrics*, 44(3):837–845.

Diebold, F. X., Lee, J.-H., and Weinbach, G. C. (1994). Regime switching with time-varying transition probabilities. *Business Cycles: Durations, Dynamics, and Forecasting*, 1:144–165.

Diebold, F. X. and Mariano, R. S. (1995). Comparing predictive accuracy. *Journal of Business and Economic Statistics*, 13(3):253–263.

Diebold, F. X. and Rudebusch, G. D. (1989). Scoring the leading indicators. *Journal of Business*, 62(3):369–391.

Erichson, N. B., Zheng, P., Manohar, K., Brunton, S. L., Kutz, J. N., and Aravkin, A. Y. (2020). Sparse principal component analysis via variable projection. *SIAM Journal on Applied Mathematics*, 80(2):977–1002.

Estrella, A. and Mishkin, F. S. (1996). The yield curve as a predictor of US recessions. *Current Issues in Economics and Finance*, 2(7):1–6.

Estrella, A. and Mishkin, F. S. (1998). Predicting US recessions: Financial variables as leading indicators. *Review of Economics and Statistics*, 80(1):45–61.

Fama, E. F. and French, K. R. (1988). Dividend yields and expected stock returns. *Journal of Financial Economics*, 22(1):3–25.

Fama, E. F. and French, K. R. (1989). Business conditions and expected returns on stocks and bonds. *Journal of Financial Economics*, 25(1):23–49.

Fawcett, T. (2006). An introduction to ROC analysis. *Pattern Recognition Letters*, 27(8):861–874.

Focardi, S. M., Fabozzi, F. J., and Mazza, D. (2019). Modeling local trends with regime shifting models with time-varying probabilities. *International Review of Financial Analysis*, 66:101368.

Gonzalez, L., Hoang, P., Powell, J. G., and Jing, S. (2006). Defining and dating bull and bear markets: Two centuries of evidence. *Multinational Finance Journal*, 10(1/2):81–116.

Guidolin, M. and Hyde, S. (2012). Can VAR models capture regime shifts in asset returns? A long-horizon strategic asset allocation perspective. *Journal of Banking and Finance*, 36(3):695–716.

Guidolin, M. and Pedio, M. (2020). Distilling large information sets to forecast commodity returns: Automatic variable selection or hidden Markov models? *BAFFI CAREFIN Working Papers*, 20140.

Guidolin, M. and Pedio, M. (2021). Forecasting commodity futures returns with stepwise regressions: Do commodity-specific factors help? *Annals of Operations Research*, 299(1):1317–1356.

Guidolin, M. and Timmermann, A. (2007). Asset allocation under multivariate regime switching. *Journal of Economic Dynamics and Control*, 31(11):3503–3544.

Haas, M., Mittnik, S., and Paolella, M. S. (2004). A new approach to Markov-switching GARCH models. *Journal of Financial Econometrics*, 2(4):493–530.

Hamilton, J. D. (1989). A new approach to the economic analysis of nonstationary time series and the business cycle. *Econometrica*, 57(2):357–384.

Hamilton, J. D. (1990). Analysis of time series subject to changes in regime. *Journal of Econometrics*, 45(1-2):39–70.

Hamilton, J. D. (1994). *Time Series Analysis*, volume 2. Princeton University Press Princeton, NJ.

Hamilton, J. D. and Lin, G. (1996). Stock market volatility and the business cycle. *Journal of Applied Econometrics*, 11(5):573–593.

Hammerschmid, R. and Lohre, H. (2018). Regime shifts and stock return predictability. *International Review of Economics and Finance*, 56(C):138–160.

Hansen, B. E. (1991). The likelihood test under non-standard conditions: Testing the Markov trend model of GNP. *Journal of Applied Econometrics*, 7:61–82.

Harvey, D., Leybourne, S., and Newbold, P. (1997). Testing the equality of prediction mean squared errors. *International Journal of Forecasting*, 13(2):281–291.

Hauptmann, J., Hoppenkamps, A., Min, A., Ramsauer, F., and Zagst, R. (2014). Forecasting market turbulence using regime-switching models. *Financial Markets and Portfolio Management*, 28(2):139–164.

Henkel, S. J., Martin, J. S., and Nardari, F. (2011). Time-varying short-horizon predictability. *Journal of Financial Economics*, 99(3):560–580.

Hong, H., Torous, W., and Valkanov, R. (2007). Do industries lead stock markets? *Journal of Financial Economics*, 83(2):367–396.

Johnson, M. A. and Watson, K. J. (2011). Can changes in the purchasing managers' index foretell stock returns? An additional forward-looking sentiment indicator. *Journal of Investing*, 20(4):89–98.

Kaya, H., Lee, W., and Pornrojnangkool, B. (2010). Regimes: Nonparametric identification and forecasting. *Journal of Portfolio Management*, 36(2):94–105.

Kole, E. and van Dijk, D. (2017). How to identify and forecast bull and bear markets? *Journal of Applied Econometrics*, 32(1):120–139.

Kristensen, J. T. (2017). Diffusion indexes with sparse loadings. *Journal of Business and Economic Statistics*, 35(3):434–451.

Kritzman, M. and Li, Y. (2010). Skulls, financial turbulence, and risk management. *Financial Analysts Journal*, 66(5):30–41.

Kritzman, M., Li, Y., Page, S., and Rigobon, R. (2011). Principal components as a measure of systemic risk. *Journal of Portfolio Management*, 37(4):112–126.

Kritzman, M., Page, S., and Turkington, D. (2012). Regime shifts: Implications for dynamic strategies (corrected). *Financial Analysts Journal*, 68(3):22–39.

Lewellen, J. (2004). Predicting returns with financial ratios. *Journal of Financial Economics*, 74(2):209–235.

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

Ludvigson, S. C. and Ng, S. (2009). Macro factors in bond risk premia. *Review of Financial Studies*, 22(12):5027–5067.

Lunde, A. and Timmermann, A. (2004). Duration dependence in stock prices: An analysis of bull and bear markets. *Journal of Business and Economic Statistics*, 22(3):253–273.

Maheu, J. M. and McCurdy, T. H. (2000). Identifying bull and bear markets in stock returns. *Journal of Business and Economic Statistics*, 18(1):100–112.

Maheu, J. M., McCurdy, T. H., and Song, Y. (2012). Components of bull and bear markets: Bull corrections and bear rallies. *Journal of Business and Economic Statistics*, 30(3):391–403.

Merton, R. C. (1969). Lifetime portfolio selection under uncertainty: The continuous-time case. *Review of Economics and Statistics*, 51(3):247–257.

Mönch, E. (2008). Forecasting the yield curve in a data-rich environment: A no-arbitrage factor-augmented VAR approach. *Journal of Econometrics*, 146(1):26–43.

Neely, C. J., Rapach, D. E., Tu, J., and Zhou, G. (2014). Forecasting the equity risk premium: The role of technical indicators. *Management Science*, 60(7):1772–1791.

Pagan, A. R. and Sossounov, K. A. (2003). A simple framework for analysing bull and bear markets. *Journal of Applied Econometrics*, 18(1):23–46.

Pástor, L. and Stambaugh, R. F. (2001). The equity premium and structural breaks. *Journal of Finance*, 56(4):1207–1239.

Pesaran, M. H. and Timmermann, A. (1995). Predictability of stock returns: Robustness and economic significance. *Journal of Finance*, 50(4):1201–1228.

Pettenuzzo, D. and Timmermann, A. (2011). Predictability of stock returns and asset allocation under structural breaks. *Journal of Econometrics*, 164(1):60–78.

Rapach, D. and Zhou, G. (2013). Forecasting stock returns. In Elliot, G. and Timmermann, A., editors, *Handbook of economic forecasting*, volume 2A, pages 328–383. Amsterdam: Elsevier.

Rapach, D. and Zhou, G. (2019). Sparse macro factors. *http://dx.doi.org/10.2139/ssrn.3259447. Working Paper*.

Rapach, D. E., Strauss, J. K., and Zhou, G. (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., Wohar, M. E., and Rangvid, J. (2005). Macro variables and international stock return predictability. *International Journal of Forecasting*, 21(1):137–166.

Rubbaniy, G., Asmerom, R., Rizvi, S. K. A., and Naqvi, B. (2014). Do fear indices help predict stock returns? *Quantitative Finance*, 14(5):831–847.

Sander, M. (2018). Market timing over the business cycle. *Journal of Empirical Finance*, 46(C):130–145.

Schaller, H. and Norden, S. V. (1997). Regime switching in stock market returns. *Applied Financial Economics*, 7(2):177–191.

Timmermann, A. (2006). Forecast Combinations. In Elliott, G., Granger, C., and Timmermann, A., editors, *Handbook of economic forecasting*, volume 1, pages 135–196. Amsterdam: Elsevier.

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

Zens, G. and Böck, M. (2019). A factor-augmented Markov switching (FAMS) model. *https://arxiv.org/abs/1904.13194. Working Paper.*

Zhu, X. and Zhu, J. (2013). Predicting stock returns: A regime-switching combination approach and economic links. *Journal of Banking and Finance*, 37(11):4120–4133.

Zou, H., Hastie, T., and Tibshirani, R. (2006). Sparse principal component analysis. *Journal of Computational and Graphical Statistics*, 15(2):265–286.

Appendix A: Data Description

Table A1: Variable Description (continued on next two pages)

Variable Description	Abbreviation	Transformation
Log excess returns on the S&P 500 index, incl. dividends over 3 M Treasury Bill	ERP	
4-Week Simple Moving Average of S&P 500 Returns	SMA_4	
52-Week Simple Moving Average of S&P 500 Returns	SMA_52	
4-Week Exponential Moving Average of S&P 500 Returns	EMA_4	
52-Week Exponential Moving Average of S&P 500 Returns	EMA_52	
Realized Variance (Squared S&P 500 Intraday Returns) 22 Days	RV_22	
Realized Variance (Squared S&P 500 Intraday Returns) 5 Days	RV_5	
Realized Variance (Squared S&P 500 Intraday Returns) 1 Day	RV_1	
Variance Risk Premium, $VIX^2 - RV_{22}$	VRP	
Implied Volatility Index for options based on the S&P 500	VIX	Log
Moving Average 5 days - 100 d, S&P 500 Price	MA_5_100	
Moving Average 20 days - 100 days, S&P 500 Price	MA_20_100	
Moving Average 50 days - 100 days, S&P 500 Price	MA_50_100	
Moving Average 5 days - 200 days, S&P 500 Price	MA_5_200	
Moving Average 20 days - 200 days, S&P 500 Price	MA_20_200	
Moving Average 50 days - 200 days, S&P 500 Price	MA_50_200	
Price Momentum 100 days, S&P 500 Price	MOM_100	
Price Momentum 200 days, S&P 500 Price	MOM_200	
On-balance Volume 5 days - 100 days, S&P 500 Price and Volume	OBV_5_100	
On-balance Volume 20 days - 100 days, S&P 500 Price and Volume	OBV_20_100	
On-balance Volume 50 days - 100 days, S&P 500 Price and Volume	OBV_50_100	
On-balance Volume 5 days - 200 days, S&P 500 Price and Volume	OBV_5_200	
On-balance Volume 20 days - 200 days, S&P 500 Price and Volume	OBV_20_200	
On-balance Volume 50 days - 200 days, S&P 500 Price and Volume	OBV_50_200	
Log Dividend Price Ratio	DP	De-trend
Earnings over Price	EP	Log
Log 10 Y Earnings Price Ratio	E10P	De-trend
Payout Ratio: Dividends over Earnings	Payout	Log
3 M Moving Average of Earnings Revision Ratio (No. Companies)	EPS_REV	
3 M Percentage Change in forward 12 EPS Expectations	EPS_MOM	
Long-term Earning per Share Growth Expectation	EPS_LTG	Log
Expected forward 12 M Earning per Share divided by current price	EY_fwd	De-trend
Long Term EPS Growth Uncertainty scaled by trailing 12 M EPS	SD_LTG	QQ
12 M EPS Uncertainty scaled by trailing 12 M EPS	SD_NTM	QQ
Excess Industry Return: Agriculture, forestry, and fishing	Agric	
Excess Industry Return: Mining	Mines	
Excess Industry Return: Oil and Gas Extraction	Oil	
Excess Industry Return: Nonmetallic Minerals Except Fuels	Stone	
Excess Industry Return: Construction	Cnstr	
Excess Industry Return: Food and Kindred Products	Food	
Excess Industry Return: Tobacco Products	Smoke	
Excess Industry Return: Textile Mill Products	Txtls	
Excess Industry Return: Apparel and other Textile Products	Apprl	
Excess Industry Return: Lumber and Wood Products	Wood	
Excess Industry Return: Furniture and Fixtures	Chair	
Excess Industry Return: Paper and Allied Products	Paper	
Excess Industry Return: Printing and Publishing	Print	
Excess Industry Return: Chemicals and Allied Products	Chems	
Excess Industry Return: Petroleum and Coal Products	Ptrlm	
Excess Industry Return: Rubber and Miscellaneous Plastics Products	Rubbr	
Excess Industry Return: Leather and Leather Products	Lethr	
Excess Industry Return: Stone, Clay and Glass Products	Glass	
Excess Industry Return: Primary Metal Industries	Metal	
Excess Industry Return: Fabricated Metal Products	MtlPr	
Excess Industry Return: Machinery, Except Electrical	Machn	
Excess Industry Return: Electrical and Electronic Equipment	Elctr	
Excess Industry Return: Transportation Equipment	Cars	
Excess Industry Return: Instruments and Related Products	Instr	
Excess Industry Return: Miscellaneous Manufacturing Industries	Manuf	
Excess Industry Return: Transportation	Trans	
Excess Industry Return: Telephone and Telegraph Communication	Phone	
Excess Industry Return: Radio and Television Broadcasting	TV	
Excess Industry Return: Electric, Gas, and Water Supply	Utils	
Excess Industry Return: Sanitary Services	Garbg	

Table A1: Variable Description (continued from previous page and on next page)

Variable Description	Abbreviation	Transformation
Excess Industry Return: Wholesale	Whlsl	
Excess Industry Return: Retail Stores	Rtail	
Excess Industry Return: Finance, Insurance, and Real Estate	Money	
Excess Industry Return: Services	Srvc	
30 Y Treasury Bonds	T30Y	De-trend
10 Y Treasury Bonds	T10Y	De-trend
7 Y Treasury Bonds	T7Y	De-trend
5 Y Treasury Bonds	T5Y	De-trend
3 Y Treasury Bonds	T3Y	De-trend
2 Y Treasury Bonds	T2Y	De-trend
1 Y Treasury Bonds	T1Y	De-trend
6 M Treasury Bonds	T6M	De-trend
3 M Treasury Bonds	T3M	De-trend
Effective Federal Funds Rate	FEDR	De-trend
Corporate Bonds Yield AAA rated	AAA	De-trend
Corporate Bonds Yield BAA rated	BAA	De-trend
LIBOR USD	LIBOR	De-trend
Term Spread 30 Y and 10 Y	TS_30Y10Y	
Term Spread 10 Y and 1 Y	TS_10Y1Y	
Term Spread 10 Y and 3 M	TS_10Y3M	
Term Spread 5 Y and 3 M	TS_5Y3M	
Spread between 3 M USD LIBOR and 3 M Treasury Bill	TED	
Credit Spread AAA rated Corp. Bonds and Gov. Bonds	CS_AAA10Y	
Credit Spread BAA rated Corp. Bonds and Gov. Bonds	CS_BAA10Y	
Credit Spread BAA rated Corp. Bonds and AAA Corp. Bonds	CS_BAAAAA	
Turbulence Index of Industry Returns	Turb_Index	
Standardized change in the Absorption Ratio of Industry Returns	AR	
Return Gold Price	Gold	
Return WTI Oil Price	WTI	
GDP Growth Rate, Mean Forecast	gdp	
Consumption Growth Rate, Mean Forecast	cons	
Investments Growth Rate, Mean Forecast	inv	
Profit Growth Rate, Mean Forecast	profit	
Production Growth Rate, Mean Forecast	prod	
CPI Inflation Rate, Mean Forecast	cpi	
PPI Inflation Rate, Mean Forecast	ppi	
Employment Cost Growth Rate, Mean Forecast	emp.cost	MM
Car Sales in Millions, Mean Forecast	csales	mm%
Housing Starts in Million, Mean Forecast	housep	MM
Unemployment Rate, Mean Forecast	unemp	MM
Current Account in Billion USD, Mean Forecast	ca	MM
Fiscal Balance in Billion USD, Mean Forecast	fb	MM
Mean Forecast of 3 M Interest Rate (in 3 Months) Minus 3 M Yield	i3m.3m	
Mean Forecast of 3 M Interest Rate (in 12 Months) Minus 3 M Yield	i3m.12m	
Mean Forecast of 10 Y Interest Rate (in 3 Months) Minus 10 Y Yield	i10y.3m	
Mean Forecast of 10 Y Interest Rate (in 12 Months) Minus 10 Y Yield	i10y.12m	
Mean Forecast term spread (in 3 Months) Minus Current 10Y3M spread	term_spread.3m	
Mean Forecast term spread (in 12 Months) Minus Current 10Y3M spread	term_spread.12m	
GDP Growth Rate, Forecast Standard Deviation	gdp.sd	
Consumption Growth Rate, Forecast Standard Deviation	cons.sd	
Investments Growth Rate, Forecast Standard Deviation	inv.sd	
S&P 500 Profits Growth Rate, Forecast Standard Deviation	profit.sd	
Production Growth Rate, Forecast Standard Deviation	prod.sd	
Inflation (CPI), Forecast Standard Deviation	cpi.sd	
Inflation (PPI), Forecast Standard Deviation	ppi.sd	
Employment Cost, Forecast Standard Deviation	emp.cost.sd	
Car Sales, Forecast Standard Deviation	csales.sd	
Housing Starts, Forecast Standard Deviation	housep.sd	
Unemployment Rate, Forecast Standard Deviation	unemp.sd	
Current Account, Forecast Standard Deviation	ca.sd	
Fiscal Balance, Forecast Standard Deviation	fb.sd	MM
3 M Interest rate (in 3 Months), Forecast Standard Deviation	i3m.3m.sd	
3 M Interest rate (in 12 Months), Forecast Standard Deviation	i3m.12m.sd	
10 Y Interest rate (in 3 Months), Forecast Standard Deviation	i10y.3m.sd	
10 Y Interest rate (in 12 Months), Forecast Standard Deviation	i10y.12m.sd	
Term Spread (in 3 Months), Forecast Standard deviation	term_spread.3m.sd	
Term Spread (in 12 Months), Forecast Standard deviation	term_spread.12m.sd	

Table A1: Variable Description (continued from previous two pages)

Variable Description	Abbreviation	Transformation
Consumer Climate Survey, TCB	Cons_Cli_Conf	YY
Consumer Situation Survey, TCB	Cons_Sit_Conf	YY
Consumer Expectation Survey, TCB	Cons_Exp_Conf	YY
Consumer Inflation Expectation Survey, TCB	Cons_Inf_Conf	MM
12 M Expectation FED Rate Increase, TCB	Int_exp_higher	
12 M Expectation FED Rate Decrease, TCB	Int_exp_lower	
12 M Expectation Stock Price Increase, TCB	stock_exp_higher	
12 M Expectation Stock Price Decrease, TCB	stock_exp_lower	
Leading Economic Index for the US, TCB	Lead_Conf	yy%
Purchasing Manager Index, ISM	PMI	
Industrial Production	Indpro	yy%
Consumer Price Index	Inflation	yy%
Unemployment Rate	Unemp	De-trend
Monthly Growth Rate Money Stock M2	M2	mm%
Monthly Growth Rate Money Stock M1	M1	mm%

Notes: YY: annual change; QQ: quarterly change; MM: monthly change; yy%: annual log change; mm%: monthly log change; De-trend: de-trended by 1-Y moving average; Log: log transformation. For three of the predictors (fb, fb.sd, and VIX), we face the problem of missing values and solve this by employing appropriate proxies. First, the missing values of the fiscal balance forecast series are substituted by the realized fiscal balance data of the previous year. Accordingly, we presume a standard deviation (fb.sd) of zero during that time. Second, since the implied volatility of the option market (VIX) is an important risk aversion measure and, therefore, a promising candidate to predict stock market crashes (Coudert and Gex 2008), we re-fill the missing values before 1990 with values of the CBOE S&P 100 Volatility Index.

Table A2: Data Sources

Data	**Source**
Market Data	
S&P 500 Prices	Datastream
S&P 500 Trading Volume	Yahoo Finance
S&P 500 Intraday Return Data	Oxford MAN Institute
34 US Industry Portfolios	Kenneth R. French Data Library
Implied Volatility Data	Chicago Board Options Exchange
Int. Rates and Treasury Bonds Yields	Federal Reserve System
Corporate Bonds Yields	Moody´s Investor Services
WTI Oil Price	Chicago Mercantile Exchange
Gold Price	London Bullion Market Assoc.
Fundamental and Macroeconomic Data	
S&P 500 Dividends and Earnings	Online Data Robert Shiller
Industrial Production Index	Federal Reserve System
Consumer Price Index	Bureau of Labor Statistics
Unemployment Rate	Bureau of Labor Statistics
Money Stock M1	Federal Reserve System
Money Stock M2	Federal Reserve System
Leading Economic Indicator	The Conference Board
NBER Recession Indicator	Federal Reserve of St. Louis
Survey Data	
Macroeconomic Expectations	Consensus Economics Inc.
Sell-Side Analyst Earnings Expectations	Institutional Brokers Estimate System
PMI (Manufacturing)	Institute for Supply Management
Consumer Survey Data	The Conference Board

Table A3: Summary Statistics (continued on next two pages)

	Mean	Std. Dev.	Skewness	Kurtosis	Minimum	Maximum
ERP	0.14	2.33	−0.85	7.61	−20.03	11.46
SMA_52	0.15	0.30	−1.34	2.91	−1.23	1.02
SMA_4	0.15	1.12	−1.51	9.81	−9.26	5.52
EMA_52	0.15	0.30	−1.84	5.65	−1.63	0.74
EMA_4	0.15	1.12	−1.99	12.22	−10.93	3.39
RV_22	23.74	42.33	7.26	66.64	1.60	536.67
RV_5	23.59	49.58	9.08	112.86	0.11	885.61
RV_1	23.93	66.39	14.77	323.03	0.00	1765.34
VIX	2.89	0.35	0.69	0.46	2.21	4.37
VRP	1.52	2.55	−6.12	97.69	−40.06	14.75
DP	−0.01	0.09	0.89	3.89	−0.31	0.49
EP	−3.14	0.36	−2.36	8.56	−4.99	−2.61
E10P	−0.01	0.09	1.47	4.29	−0.22	0.48
Payout	−0.80	0.39	2.79	11.11	−1.24	1.38
EPS_REV	−0.09	0.22	−0.36	0.03	−0.79	0.48
EPS_MOM	0.02	0.04	−2.43	10.73	−0.23	0.12
EY_fwd	−0.00	0.01	0.76	2.99	−0.02	0.03
EPS_LTG	2.51	0.15	0.99	0.87	2.20	3.17
SD_LTG	−0.00	0.01	0.26	8.88	−0.04	0.05
SD_NTM	−0.00	0.01	3.67	24.46	−0.03	0.08
MA_5_100	0.71	0.45	−0.92	−1.15	0.00	1.00
MA_20_100	0.71	0.45	−0.94	−1.11	0.00	1.00
MA_50_100	0.72	0.45	−0.96	−1.09	0.00	1.00
MA_5_200	0.75	0.43	−1.15	−0.69	0.00	1.00
MA_20_200	0.75	0.44	−1.13	−0.71	0.00	1.00
MA_50_200	0.76	0.43	−1.19	−0.59	0.00	1.00
MOM_100	0.71	0.45	−0.93	−1.13	0.00	1.00
MOM_200	0.78	0.42	−1.33	−0.23	0.00	1.00
OBV_5_100	0.74	0.44	−1.09	−0.80	0.00	1.00
OBV_20_100	0.74	0.44	−1.11	−0.76	0.00	1.00
OBV_50_100	0.74	0.44	−1.09	−0.81	0.00	1.00
OBV_5_200	0.79	0.40	−1.45	0.10	0.00	1.00
OBV_20_200	0.79	0.41	−1.45	0.09	0.00	1.00
OBV_50_200	0.79	0.41	−1.39	−0.06	0.00	1.00
Agric	0.14	3.34	0.04	5.46	−23.21	20.66
Mines	0.07	4.65	−0.36	3.43	−30.01	22.60
Oil	0.03	4.44	−1.07	8.81	−40.08	21.00
Stone	0.11	3.69	−0.10	7.58	−25.46	31.92
Cnstr	0.15	4.02	−0.19	8.15	−27.67	34.11
Food	0.15	2.10	−0.91	7.14	−18.68	9.18
Smoke	0.17	3.33	−0.72	5.74	−21.00	22.44
Txtls	0.06	4.28	−0.84	14.62	−43.34	30.06
Apprl	0.10	3.39	−0.34	4.28	−18.86	19.10
Wood	0.14	4.19	−0.36	6.33	−27.53	30.22
Chair	0.12	3.42	−0.47	6.15	−24.49	20.10
Paper	0.10	2.72	−0.47	3.61	−17.96	12.57
Print	0.06	3.04	−0.39	7.30	−21.61	21.00
Chems	0.16	2.26	−0.66	4.77	−18.41	9.62
Ptrlm	0.12	2.99	−1.00	8.48	−26.26	15.08
Rubbr	0.18	2.89	−0.27	3.69	−15.22	17.85
Lethr	0.14	4.20	−0.34	4.70	−27.60	26.45
Glass	0.11	3.84	−0.55	8.24	−30.13	24.03
Metal	0.05	4.33	−0.48	4.93	−30.97	25.88
MtlPr	0.17	2.94	−0.63	6.37	−21.71	16.84
Machn	0.16	3.51	−0.65	3.90	−23.79	14.23
Elctr	0.19	3.71	−0.50	3.56	−21.95	18.80
Cars	0.17	3.18	−0.85	9.50	−27.02	20.53
Instr	0.17	2.50	−1.00	6.21	−20.03	12.23
Manuf	0.08	3.19	−0.27	5.57	−19.75	23.04
Trans	0.15	2.96	−0.51	4.63	−22.38	14.50
Phone	0.06	2.63	−0.43	5.49	−21.65	15.18
TV	0.14	3.12	−0.62	6.11	−25.85	15.29
Utils	0.12	2.30	−1.25	14.99	−23.41	16.16
Garbg	0.06	3.24	−1.10	11.19	−33.61	16.96

Table A3: Summary Statistics (continued from previous page and on next page)

	Mean	Std. Dev.	Skewness	Kurtosis	Minimum	Maximum
Whlsl	0.13	2.54	−0.81	6.75	−18.87	14.60
Rtail	0.18	2.59	−0.33	3.12	−15.96	12.83
Money	0.14	3.20	−0.33	10.21	−24.19	23.40
Srvc	0.17	2.92	−0.79	5.36	−24.25	12.31
T10Y	−0.11	0.49	0.17	−0.35	−1.50	1.42
T30Y	−0.10	0.41	−0.01	0.01	−1.67	1.03
T7Y	−0.12	0.52	0.14	−0.37	−1.50	1.54
T5Y	−0.12	0.56	0.04	−0.22	−1.66	1.65
T3Y	−0.13	0.61	−0.07	0.22	−2.12	1.81
T2Y	−0.13	0.63	−0.14	0.39	−2.26	1.80
T1Y	−0.13	0.66	−0.30	0.71	−2.53	2.01
T6M	−0.13	0.68	−0.54	0.86	−2.69	1.86
T3M	−0.12	0.65	−0.70	0.94	−2.90	1.59
FEDR	−0.14	0.71	−0.75	0.79	−2.61	1.85
AAA	−0.10	0.38	0.35	−0.01	−1.22	0.99
BAA	−0.10	0.45	0.41	2.31	−1.77	2.41
LIBOR	−0.14	0.70	−0.58	0.58	−2.36	1.84
TS_30Y10Y	0.55	0.36	0.19	−0.86	−0.32	1.54
TS_10Y1Y	1.43	1.04	0.12	−1.18	−0.47	3.47
TS_10Y3M	1.75	1.11	−0.01	−1.03	−0.60	3.90
TS_5Y3M	1.21	0.83	0.12	−0.68	−0.68	3.21
TED	0.48	0.37	3.39	21.47	0.10	4.58
CS_AAA10Y	1.40	0.45	0.31	−0.57	0.57	2.87
CS_BAA10Y	2.36	0.73	1.69	5.12	1.26	6.12
CS_BAAAAA	0.96	0.38	3.21	14.65	0.52	3.47
Turb_Index	40.65	29.52	2.82	14.52	8.94	308.70
AR	−0.22	1.38	0.35	−1.04	−2.50	3.72
WTI	0.07	5.67	−0.19	13.92	−48.76	60.54
Gold	0.09	2.20	−0.13	4.45	−13.79	14.69
gdp	2.48	1.00	−1.40	4.28	−1.69	5.24
cons	2.50	0.94	−1.26	4.27	−1.65	5.76
inv	4.68	4.04	−1.40	2.86	−12.36	10.90
profit	5.06	4.65	−0.24	0.71	−10.63	15.85
prod	2.55	1.88	−1.96	5.65	−6.59	6.01
cpi	2.42	0.83	0.28	1.68	−0.50	5.15
ppi	1.90	1.12	−0.75	2.45	−3.12	4.87
empcost	−0.01	0.09	−0.70	2.79	−0.50	0.33
csales	0.02	0.32	8.38	162.94	−2.35	5.10
house	0.00	0.02	−2.08	9.71	−0.17	0.07
unemp	0.00	0.28	11.11	163.00	−0.64	4.23
ca	−1.80	16.73	0.92	6.80	−65.30	92.29
fb	−12.04	81.99	−9.49	125.04	−1178.01	182.33
i3m.3m	0.15	0.16	0.64	0.65	−0.28	0.66
i3m.12m	0.49	0.42	0.37	−0.33	−0.35	1.75
i10y.3m	0.13	0.17	−0.18	−0.49	−0.34	0.52
i10y.12m	0.42	0.33	−0.23	−0.27	−0.60	1.13
term_spread.3m	−0.02	0.19	−0.09	−0.11	−0.64	0.47
term_spread.12m	−0.06	0.40	0.03	−0.52	−1.15	0.88
gdp.sd	0.32	0.15	1.91	5.24	0.09	1.06
cons.sd	0.33	0.16	2.89	11.25	0.10	1.31
inv.sd	1.46	0.59	0.63	0.42	0.46	3.73
profit.sd	3.08	1.33	1.78	4.57	1.02	10.91
prod.sd	0.72	0.33	2.92	12.71	0.30	2.92
cpi.sd	0.29	0.10	1.92	8.76	0.09	0.97
ppi.sd	0.65	0.28	2.41	12.12	0.22	2.72
empcost.sd	0.31	0.12	0.88	0.80	0.09	0.75
csales.sd	0.36	0.13	3.01	22.40	0.09	1.54
house.sd	0.06	0.02	0.90	0.66	0.01	0.16
unemp.sd	0.21	0.16	7.70	71.11	0.05	1.99
ca.sd	43.85	25.18	0.94	1.64	6.38	172.57
fb.sd	−0.87	42.50	−0.59	36.25	−329.73	376.57
i3m.3m.sd	0.18	0.09	0.88	2.28	0.03	0.64
i3m.12m.sd	0.37	0.16	0.13	−0.34	0.07	0.84
i10y.3m.sd	0.23	0.07	0.98	1.57	0.11	0.57
i10y.12m.sd	0.40	0.10	0.29	−0.30	0.16	0.68
term_spread.3m.sd	0.24	0.07	1.13	2.59	0.09	0.59
term_spread.12m.sd	0.38	0.11	0.35	−0.48	0.14	0.70

Table A3: Summary Statistics (continued from previous two pages)

	Mean	**Std. Dev.**	**Skewness**	**Kurtosis**	**Minimum**	**Maximum**
Cons_Cli_Conf	−0.72	19.36	−0.98	0.67	−60.00	33.90
Cons_Sit_Conf	−1.40	32.23	−1.14	0.43	−102.30	49.60
Cons_Exp_Conf	−0.26	16.87	−0.48	0.41	−51.70	40.20
Cons_Inf_Conf	0.00	0.28	0.83	5.63	−1.05	1.75
Int_exp_higher	54.32	11.80	−0.04	−0.70	23.40	79.20
Int_exp_lower	15.18	8.82	1.30	1.20	5.20	45.80
stock_exp_higher	35.59	6.06	−0.38	−0.11	18.10	51.00
stock_exp_lower	27.91	7.53	1.04	1.05	15.30	54.90
Indpro	1.48	4.29	−2.01	5.43	−17.96	8.48
Inflation	2.38	1.25	0.02	1.21	−1.98	6.66
Unemp	0.03	0.92	6.31	59.72	−2.64	10.91
M2	0.50	0.53	5.41	47.19	−0.46	6.23
M1	0.93	7.44	16.01	258.96	−3.37	122.66
Lead_Conf	1.30	5.83	−1.68	3.44	−22.50	10.40
PMI	52.32	4.88	−0.73	0.83	34.50	64.70

Appendix B: Aggregation Results

Figure B1: Factors of Ordinary PCA and Sparse PCA (continued on next page)

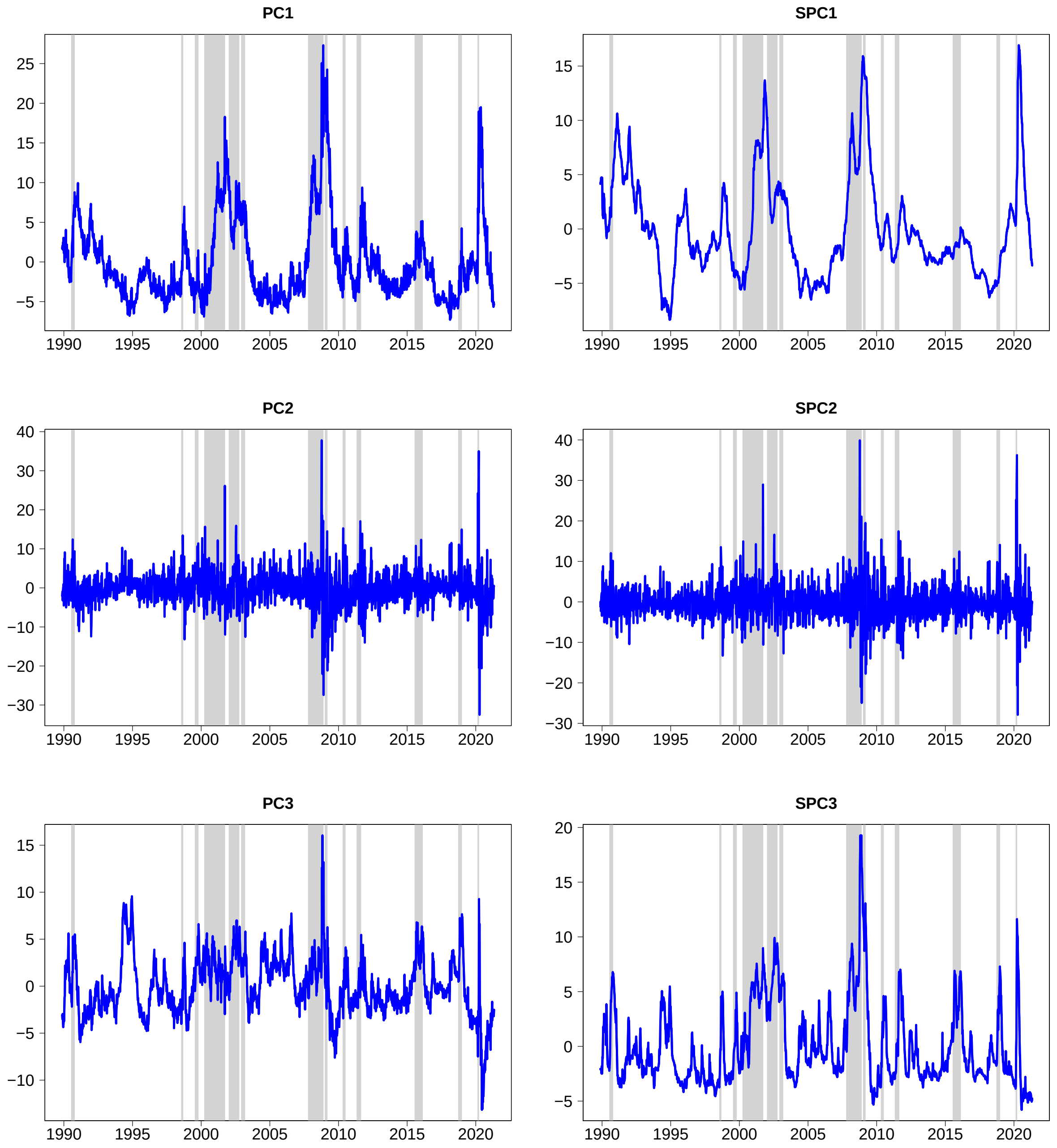
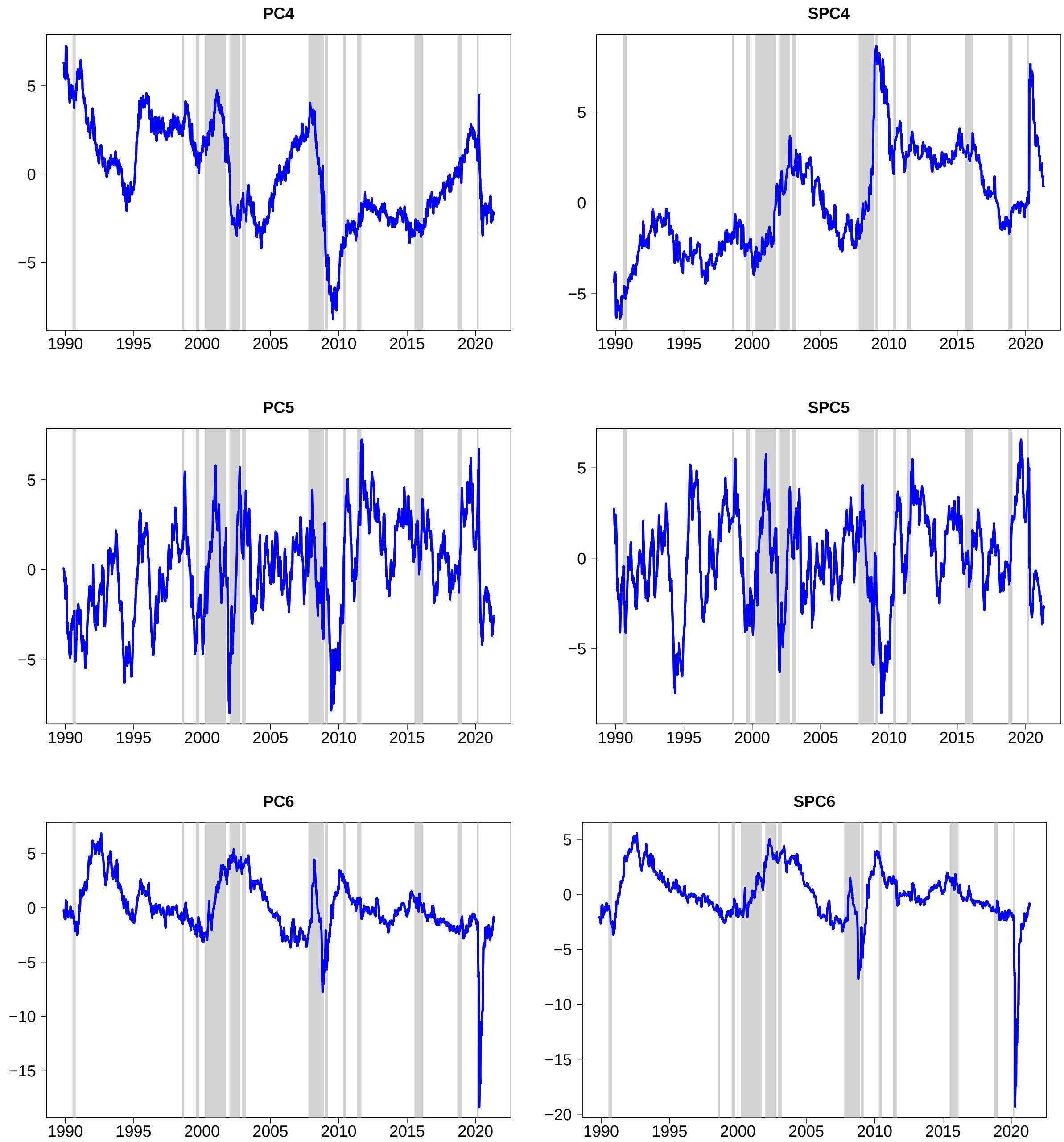
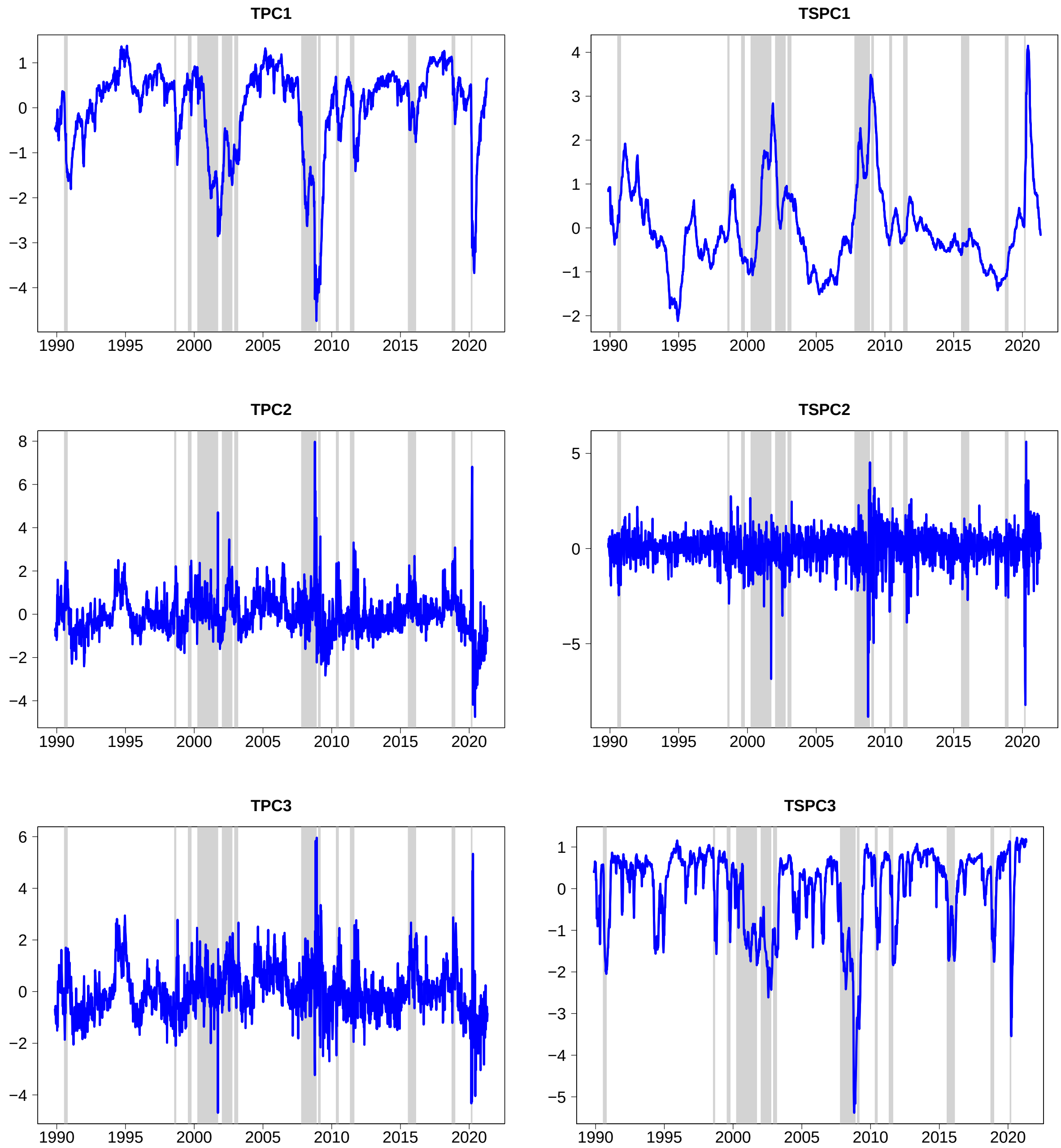


Figure B1: Factors of Ordinary PCA and Sparse PCA (continued from previous page)



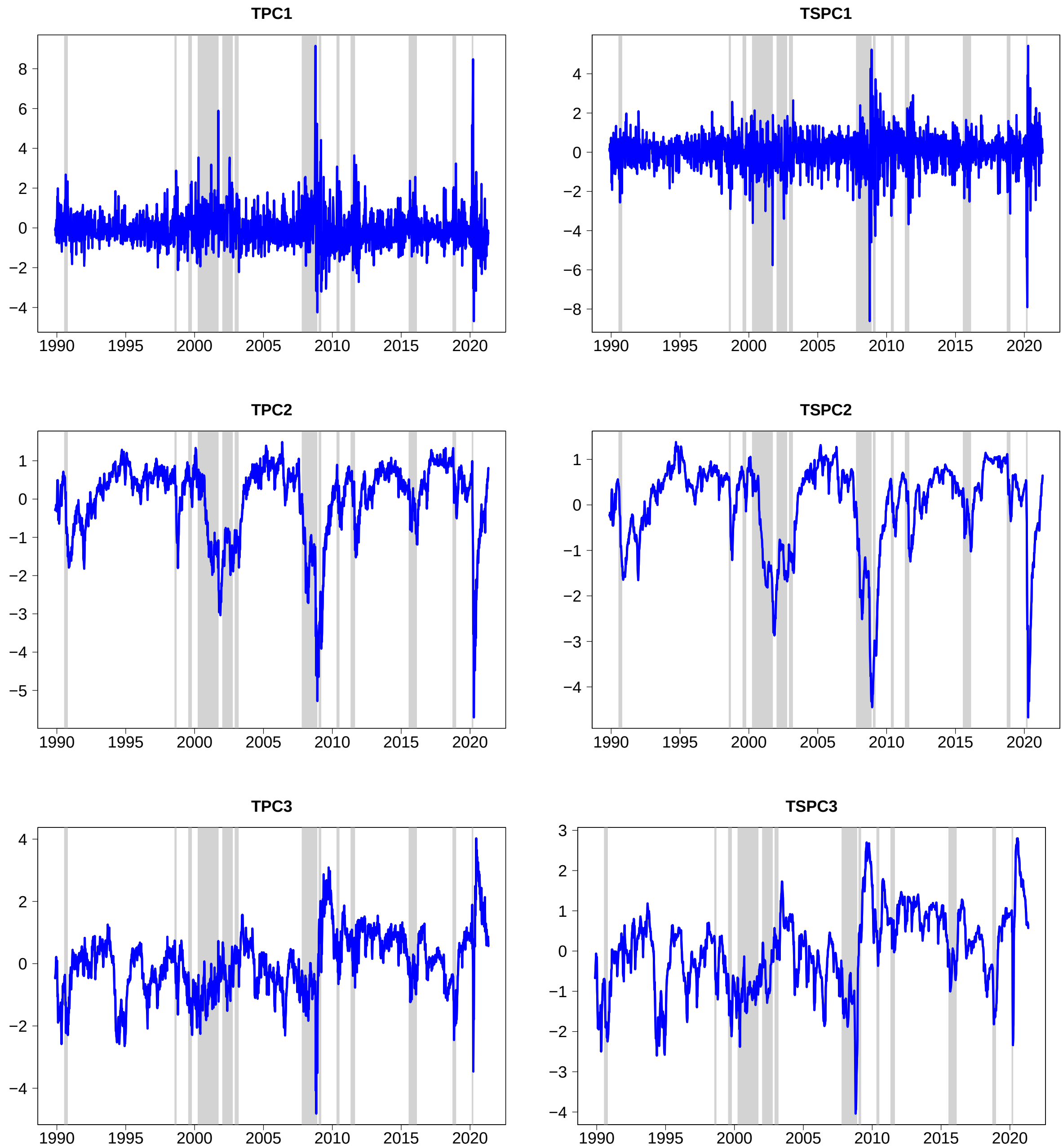
Notes: Figure shows the principal components obtained by using conventional PCA (left panel) and sparse PCA (right panel), assuming knowledge of the full sample. For further details, see also Section 2.1 and Table 3 in Section 4. Gray-shaded areas indicate actual bear market periods according to an ex post detection using the dating rule of Lunde and Timmermann (2004).

Figure B2: Factors of Targeted (Sparse) PCA Based on the VIX as Target



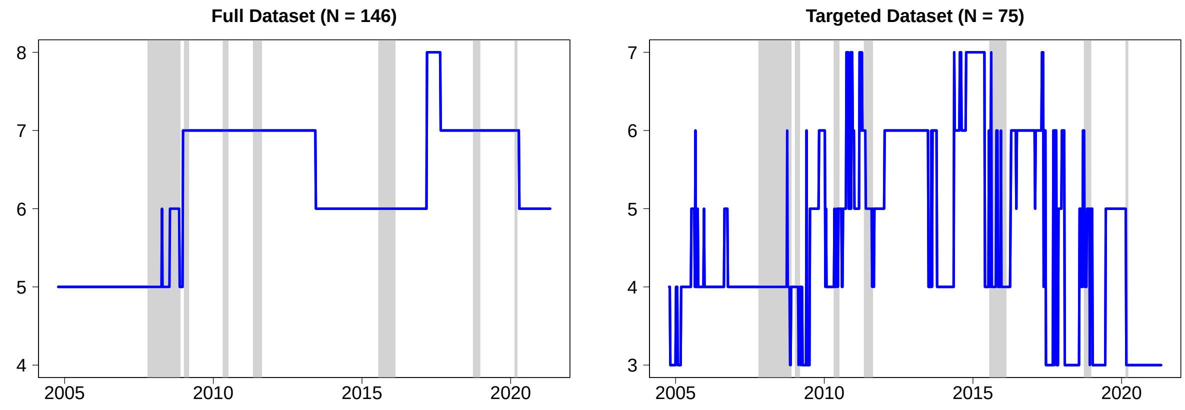
Notes: Figure shows the principal components obtained by using conventional PCA (left panel) and sparse PCA (right panel), employing the VIX as target for the soft thresholding and assuming knowledge of the full sample. For further details, see also Section 2.1 and Table 3 in Section 4. Gray-shaded areas indicate actual bear market periods according to an ex post detection using the dating rule of Lunde and Timmermann (2004).

Figure B3: Factors of Targeted (Sparse) PCA Based on the ERP as Target



Notes: Figure shows the principal components obtained by using conventional PCA (left panel) and sparse PCA (right panel), employing the ERP as target for the soft thresholding and assuming knowledge of the full sample. For further details, see also Section 2.1 and Table 3 in Section 4. Gray-shaded areas indicate actual bear market periods according to an ex post detection using the dating rule of Lunde and Timmermann (2004).

Figure B4: Number of Principal Components (Out-of-Sample)



Notes: Figure shows the number of principal components in the out-of-sample exercise, according to the IC_2 criterion of Bai and Ng (2002) and an upper bound according to an automatic elbow procedure. "Targeted Dataset" refers to the subset of indicators obtained via soft thresholding. In this case, we choose the minimum of proposed components for both targeting variables (VIX and ERP). Gray-shaded areas indicate actual bear market periods according to an ex post detection using the dating rule of Lunde and Timmermann (2004).

Appendix C: Evaluation Measures

To assess the out-of-sample performance, we consider two different dimensions: (i) statistical accuracy and economic value; (ii) regime predictions and return predictions. The economic value of a forecast-based investment strategy is evaluated with the same measures, irrespective of whether we forecast regimes or returns. For the statistical performance, however, the metrics are different, since the identification of bullish and bearish states is based on a binary classification decision, whereas returns are given on a continuous scale. In the following, the entire sample $t = 1, 2, ..., T$ is divided into an in-sample period with length T_0 and an out-of-sample period T_1 (and $T_0 + T_1 = T$).

C1: Statistical Accuracy

Regime Predictability: We measure the deviation from the actual regime with the quadratic probability score (QPS) proposed by Diebold and Rudebusch (1989):

$$QPS = \frac{1}{T_1} \sum_{t=T_0}^{T_1} 2[\hat{p}_{t+1} - S_{t+1}]^2 \tag{C1}$$

The QPS is defined for the interval between 0 (perfect accuracy) and 2 (worst possible accuracy). S_{t+1} corresponds to the regime indicator obtained with the dating rule of Lunde and Timmermann (2004) and $\hat{p}_{t+1}$ indicates the predicted bear market probability.

To measure the regime classification ability, we follow Fawcett (2006) and consider the receiving operating characteristic (ROC) curve, which displays the relationship between the true positive rate (TPR; on the y-axis) and the false positive rate (FPR; on the x-axis) of a classifier depending on a grid of thresholds. The two (threshold-dependent) quantities are defined as follows:

$$\text{TPR} = \frac{\text{True positive (TP)}}{\text{True positive (TP)} + \text{False negative (FN)}} \tag{C2}$$

$$\text{FPR} = \frac{\text{False positive (FP)}}{\text{False positive (FP)} + \text{True negative (TN)}} \tag{C3}$$

Hence, the ROC curve visualizes the trade-off between the threshold choice and the benefits/costs of bear market identification. For instance, a low threshold improves the accuracy of bear market predictions, but also causes an increase in false alarms. Although the ROC is a very flexible and robust visualization tool, it can be impractical when comparing different classifiers.[15] In this context, a straightforward solution is

[15]This is also the reason why we do not show results for the ROC curve in an effort to conserve space. The figures are available on request.

to compute the area under the ROC curve (AUC) that aggregates the classifier's performance into a single measure (Fawcett 2006). The AUC takes values between 0 and 1, where higher values indicate a better classification and a value of 0.5 corresponds to a random classifier. DeLong et al. (1988) provide a nonparametric approach to calculate confidence intervals for the AUC. Assuming two i.i.d. samples $X_1, ..., X_m$ and $Y_1, ..., Y_n$, they exploit the link to the Mann-Whitney U-statistic and obtain an unbiased estimator for the AUC:

$$\hat{\Theta}_{AUC} = \frac{1}{mn} \sum_{i=1}^{m} \sum_{j=1}^{n} H(X_i - Y_j) \tag{C4}$$

H is the Heaviside function, which is equal to 1 (0) for strictly positive (negative) values and 0.5 for zero values. With the use of a consistent estimator for the variance-covariance matrix $V[.]$, we are able two compare two classifiers, $\hat{\Theta}^1_{AUC}$ and $\hat{\Theta}^2_{AUC}$, using the following standard normally distributed test statistic:

$$\frac{\hat{\Theta}^1_{AUC} - \hat{\Theta}^2_{AUC}}{\sqrt{V[\hat{\Theta}^1_{AUC} - \hat{\Theta}^2_{AUC}]}} \tag{C5}$$

Finally, we are interested in the hit ratios, given a threshold of 50% in the MS model. This shows how often we can classify the state of the stock market correctly in general (Accuracy) as well as in bearish (Bear) and bullish weeks (Bull).

Return Predictability: It is common practice to assess the accuracy of stock market return predictions with the mean squared prediction error (MSPE). Denoting the point forecast of a model i as $\hat{R}^i_{t+1}$, the $MSPE_i$ is as follows:

$$MSPE_i = \frac{1}{T_1} \sum_{t=T_0}^{T_1} (R_{t+1} - \hat{R}^i_{t+1})^2 \tag{C6}$$

To provide a measure with the same unit as the predicted series, we build the root mean squared error (RMSE):

$$RMSE_i = \sqrt{MSPE_i} \tag{C7}$$

The $RMSE$ depends on the scale of the variable and is no clear yardstick for the assessment of whether a forecast is good or not. Furthermore, forecasters usually do not ascribe much importance to the absolute (squared) deviation since stock returns have a high noise-to-signal ratio. Instead, an often applied evaluation measure is the relative added value in comparison to the historical average as this is based on the assumption of non-predictability. Most empirical studies show that it is very difficult to outperform the historical average (e.g., Welch and Goyal 2008). Hence, we prefer to utilize the out-of-sample R^2 from Campbell and Thompson (2008), which is defined as follows:

$$R^2_{OS} = 1 - \frac{MSPE_i}{MSPE_0} \tag{C8}$$

$MSPE_0$ denotes the MSPE of the historical average. A positive R^2_{OS} signals a lower MSPE of model i and, thus, an improvement of the predictability relative to the historical benchmark.

To determine if the improvement is significant, we test the hypothesis $H_0 : R^2_{OS} \leq 0$ against $H_1 : R^2_{OS} > 0$. For this purpose, we rely on the adjusted MSPE by Clark and West (2007) as we always compare nested forecasts. Since a larger model produces additional noise in the prediction, the ordinary MSPE is adjusted in the test statistic. As $\hat{R}_{1,t+1}$ and $\hat{R}_{2,t+1}$ denote the one-step ahead forecasts from the restricted and the unrestricted model and the corresponding forecasting errors are $\hat{e}_{1,t+1}$ and $\hat{e}_{2,t+1}$, the adjusted MSPE is given by $\hat{f}_{t+1} = \hat{e}_{1,t+1} - [\hat{e}_{2,t+1} - (\hat{R}_{1,t+1} - \hat{R}_{2,t+1})^2]$. Using the sample average $\bar{f} = 1/T_1 \sum_{t=T_0}^{T_1} \hat{f}_{t+1}$ and the sample variance $\hat{V} = 1/(T_1 - 1)(\hat{f}_{t+1} - \bar{f})$, the CW test statistic is as follows:

$$CW = \frac{\sqrt{T_1}\bar{f}}{\sqrt{\hat{V}}} \tag{C9}$$

The CW statistic is approximately standard normal distributed. Hence, we can directly apply the standard critical values for a one-sided hypothesis test.

To evaluate forecasts from non-nested models, we compare them pairwise with the test by Diebold and Mariano (1995). $\hat{e}_{1,t+1}$ and $\hat{e}_{2,t+1}$ denote the forecasts errors of models 1 and 2. The corresponding quadratic loss functions are $g(\hat{e}_{1,t+1})$ and $g(\hat{e}_{2,t+1})$. The loss differential $d = g(\hat{e}_{1,t+1}) - g(\hat{e}_{2,t+1})$ is assumed to have an expected value of zero, to be covariance stationary, and to be asymptotically normal distributed. Then, the resulting DM test statistic is as follows:

$$DM = \frac{\bar{d}}{\sqrt{2\pi \hat{f}_d(0)/T}} \tag{C10}$$

$\bar{d}$ denotes the average loss differential and the denominator represents the standard deviation of d. $\hat{f}_d(0)$ indicates an estimate of the spectral density at frequency zero.[16] The DM test statistic relies on the null hypothesis of equal prediction accuracy with the alternative that the forecast of model 2 is more accurate.

Finally, since point forecasts for returns are usually very difficult, we also examine to what extent we can at least forecast the their direction correctly. For this purpose, the evaluation can be considered as a binary classification problem. Here, we are interest in the overall accuracy (Direction) as well as the true-positive rate (R^+) and the true-negative rate (R^-).

[16] To account for small sample properties, we use the corrected version of Harvey et al. (1997).

C2: Economic Value

As Rapach and Zhou (2013) or Dal Pra et al. (2018) note, a very small or negative R^2_{OS} does not necessarily mean that the forecasts are useless for investors. Consequently, we present evaluation measures that consider the utility from an investor's point of view. The settings on which the backtests are based and the corresponding investment strategies are described in Section 5. We evaluate the resulting time series of returns of the different strategies according to various return and risk metrics.

The final wealth of the strategy (assuming an investment of 1$ at the beginning) is denoted as R^{CUM}, the annualized average return as $\bar{R}$, and the annualized standard deviation as $\bar{\sigma}$. We also calculate two risk-adjusted performance measures: one profit-based metric and one utility-based metric. The Sharpe ratio (SR) expresses the ratio between the excess return of a strategy over the risk-free interest rate (measured by the returns of the 3 M Treasury Bill) and the strategy's standard deviation. The certainty equivalent return (CER) measures the average utility gain for a mean-variance investor with relative risk aversion γ:

$$CER = \hat{\mu}_i - \frac{1}{2}\gamma\hat{\sigma}_i^2 \tag{C11}$$

$\hat{\mu}_i$ ($\hat{\sigma}_i^2$) represents the strategy's average return (variance) for the out-of-sample period. The CER can also be interpreted as how large the risk-free rate should be so that the investor is indifferent to the risky investment opportunity. Furthermore, it is also common to calculate the CER gain as the difference between the strategy's CER and the CER of the historical average (henceforth Δ_{CER}). The Δ_{CER} has the practical interpretation as the management fee an investor is willing to pay to participate in the forecast-based strategy (Rapach and Zhou 2013). Regarding the degree of risk aversion, we follow the standard in the literature and assume $\gamma = 3$ (e.g., Zhu and Zhu 2013) and set $\gamma = 2$ and $\gamma = 5$ in the robustness tests in Appendix E.

For most equity investors, an exclusive focus on the mean and the variance is not sufficient. Since the stylized distribution of stock returns has a negative skewness and fat tails, tail risk measures are particularly important. Hence, we consider three popular downside-risk measures to assess whether the strategy protects the investor from significant losses: (i) the maximum drawdown (MaxDD), (ii) the value-at-risk (VaR), and (iii) the conditional value-at-risk (CVaR). The MaxDD quantifies the largest loss suffered by an investor during a particular period. The VaR indicates the maximum loss that will not be exceeded with a certain probability α and over a certain time horizon. More formally, the VaR can be expressed as:

$$VaR_{1-\alpha}(R) = F_R^{-1}(1-\alpha) = \inf\{r \in \mathbb{R} : F_R^r \geq 1-\alpha\} \tag{C12}$$

Our calculation relies on a confidence level of 95% and a one-week horizon. We utilize the historical return distribution F_R to calculate the VaR. In addition to the VaR, the CVaR answers the question about the expected average loss of an investor if the loss indeed exceeds the VaR:

$$CVaR_{1-\alpha}(R) = E[r|r \leq F_R^{-1}(1-\alpha)] \tag{C13}$$

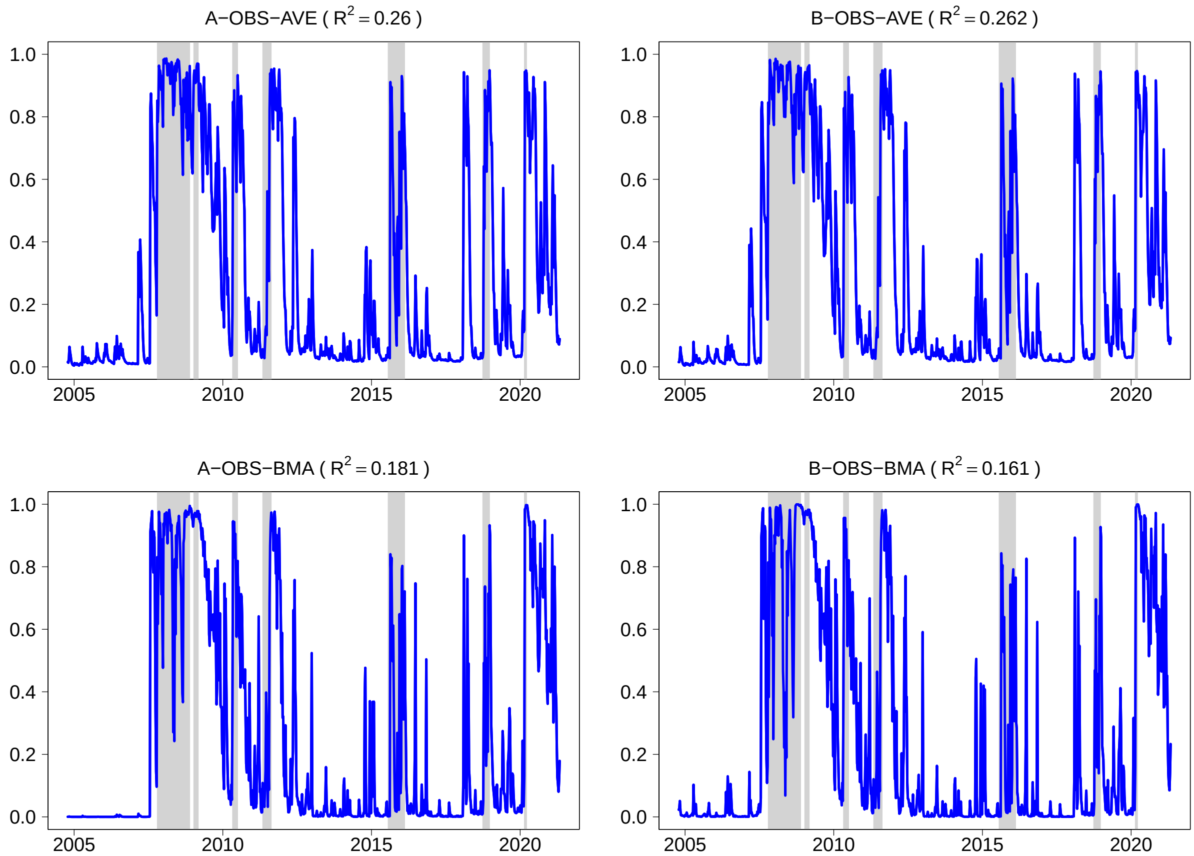
Appendix D: Additional Out-of-Sample Results

Table D1: AUC Test for the Accuracy of Regime Predictions

	TCTP	A-OBS-AVE	A-PC-AVE	A-SPC-AVE	A-TPC-AVE	A-TSPC-AVE	A-OBS-BMA	A-PC-BMA	A-SPC-BMA	A-TPC-BMA	A-TSPC-BMA	B-OBS-AVE	B-PC-AVE	B-SPC-AVE	B-TPC-AVE	B-TSPC-AVE	B-OBS-BMA	B-PC-BMA	B-SPC-BMA	B-TPC-BMA	B-TSPC-BMA
TCTP		1.00	0.79	1.00	1.00	1.00	0.21	**0.05**	1.00	0.92	0.74	1.00	0.93	1.00	1.00	1.00	**0.02**	**0.02**	0.99	0.21	0.33
A-OBS-AVE	**0.00**		**0.00**	**0.06**	0.26	0.99	**0.00**	**0.00**	0.77	**0.06**	**0.03**	0.52	**0.01**	0.11	0.37	1.00	**0.00**	**0.00**	0.23	**0.00**	**0.02**
A-PC-AVE	0.21	1.00		1.00	1.00	1.00	0.15	**0.01**	0.99	0.79	0.56	1.00	1.00	1.00	1.00	1.00	**0.01**	**0.00**	0.93	0.10	0.23
A-SPC-AVE	**0.00**	0.94	**0.00**		0.90	1.00	**0.02**	**0.00**	0.95	0.39	0.19	0.95	**0.01**	0.84	0.95	1.00	**0.00**	**0.00**	0.58	**0.00**	**0.05**
A-TPC-AVE	**0.00**	0.74	**0.00**	0.10		0.99	**0.00**	**0.00**	0.82	0.15	**0.08**	0.75	**0.00**	0.16	0.76	1.00	**0.00**	**0.00**	0.31	**0.00**	**0.03**
A-TSPC-AVE	**0.00**	**0.01**	**0.00**	**0.00**	**0.01**		**0.00**	**0.00**	0.39	**0.01**	**0.00**	**0.01**	**0.00**	**0.00**	**0.02**	0.92	**0.00**	**0.00**	**0.03**	**0.00**	**0.00**
A-OBS-BMA	0.79	1.00	0.85	0.98	1.00	1.00		0.34	1.00	0.98	0.92	1.00	0.90	0.98	1.00	1.00	**0.00**	0.28	0.97	0.59	0.59
A-PC-BMA	0.95	1.00	0.99	1.00	1.00	1.00	0.66		1.00	0.98	0.95	1.00	1.00	1.00	1.00	1.00	0.29	0.14	1.00	0.87	0.76
A-SPC-BMA	**0.00**	0.23	**0.01**	**0.05**	0.18	0.61	**0.00**	**0.00**		**0.08**	**0.02**	0.24	**0.02**	**0.08**	0.22	0.72	**0.00**	**0.00**	**0.09**	**0.00**	**0.01**
A-TPC-BMA	**0.08**	0.94	0.21	0.61	0.85	0.99	**0.02**	**0.02**	0.92		0.29	0.93	0.32	0.65	0.86	1.00	**0.00**	**0.01**	0.64	**0.02**	0.13
A-TSPC-BMA	0.26	0.97	0.44	0.81	0.92	1.00	**0.08**	**0.05**	0.98	0.71		0.97	0.56	0.84	0.93	1.00	**0.00**	**0.03**	0.83	0.14	0.14
B-OBS-AVE	**0.00**	0.48	**0.00**	**0.05**	0.25	0.99	**0.00**	**0.00**	0.76	**0.07**	**0.03**		**0.01**	0.10	0.35	1.00	**0.00**	**0.00**	0.22	**0.00**	**0.02**
B-PC-AVE	**0.07**	0.99	**0.00**	0.99	1.00	1.00	0.10	**0.00**	0.98	0.68	0.44	0.99		1.00	1.00	1.00	**0.01**	**0.00**	0.85	**0.05**	0.16
B-SPC-AVE	**0.00**	0.89	**0.00**	0.16	0.84	1.00	**0.02**	**0.00**	0.92	0.35	0.16	0.90	**0.00**		0.92	1.00	**0.00**	**0.00**	0.51	**0.00**	**0.05**
B-TPC-AVE	**0.00**	0.63	**0.00**	**0.05**	0.24	0.98	**0.00**	**0.00**	0.78	0.14	**0.07**	0.65	**0.00**	**0.08**		1.00	**0.00**	**0.00**	0.25	**0.00**	**0.02**
B-TSPC-AVE	**0.00**	**0.00**	**0.00**	**0.00**	**0.00**	**0.08**	**0.00**	**0.00**	0.28	**0.00**	**0.00**	**0.00**	**0.00**	**0.00**	**0.00**		**0.00**	**0.00**	**0.01**	**0.00**	**0.00**
B-OBS-BMA	0.98	1.00	0.99	1.00	1.00	1.00	1.00	0.71	1.00	1.00	1.00	1.00	0.99	1.00	1.00	1.00		0.66	1.00	0.92	0.88
B-PC-BMA	0.98	1.00	1.00	1.00	1.00	1.00	0.72	0.86	1.00	0.99	0.97	1.00	1.00	1.00	1.00	1.00	0.34		1.00	0.93	0.82
B-SPC-BMA	**0.01**	0.77	**0.07**	0.42	0.69	0.97	**0.03**	**0.00**	0.91	0.36	0.17	0.78	0.15	0.49	0.75	0.99	**0.00**	**0.00**		**0.00**	**0.05**
B-TPC-BMA	0.79	1.00	0.90	1.00	1.00	1.00	0.41	0.13	1.00	0.98	0.86	1.00	0.95	1.00	1.00	1.00	**0.08**	**0.07**	1.00		0.52
B-TSPC-BMA	0.67	0.98	0.77	0.95	0.97	1.00	0.41	0.24	0.99	0.87	0.86	0.98	0.84	0.95	0.98	1.00	0.12	0.18	0.95	0.48	

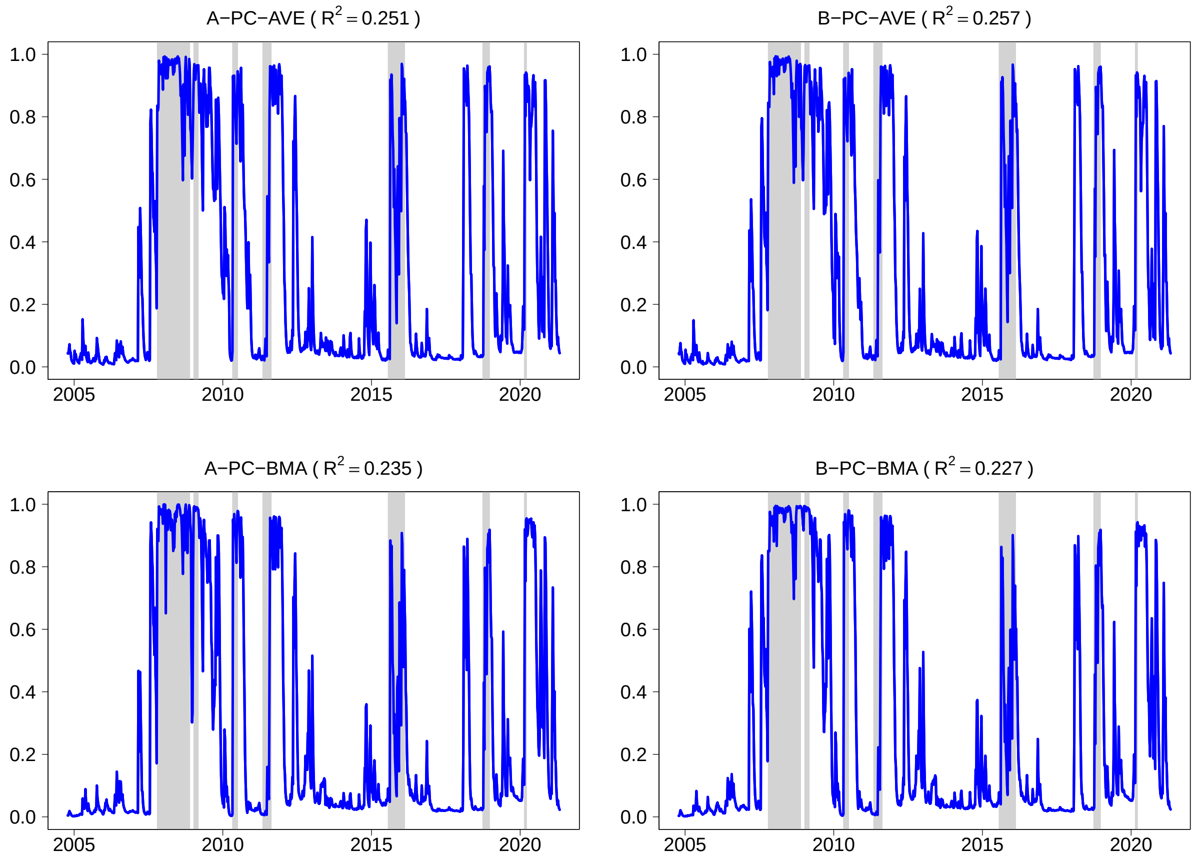
Notes: Table shows p-values for the AUC test (DeLong et al. 1988). The alternative hypothesis is that a column forecast is less accurate than a row forecast. Hence, "good" ("bad") forecasts are indicated by low p-values in the rows (columns). p-values lower than 10% are highlighted by bold entries.

Figure D1: Regime Forecasts Using Observable Predictors: Bear Market Probability



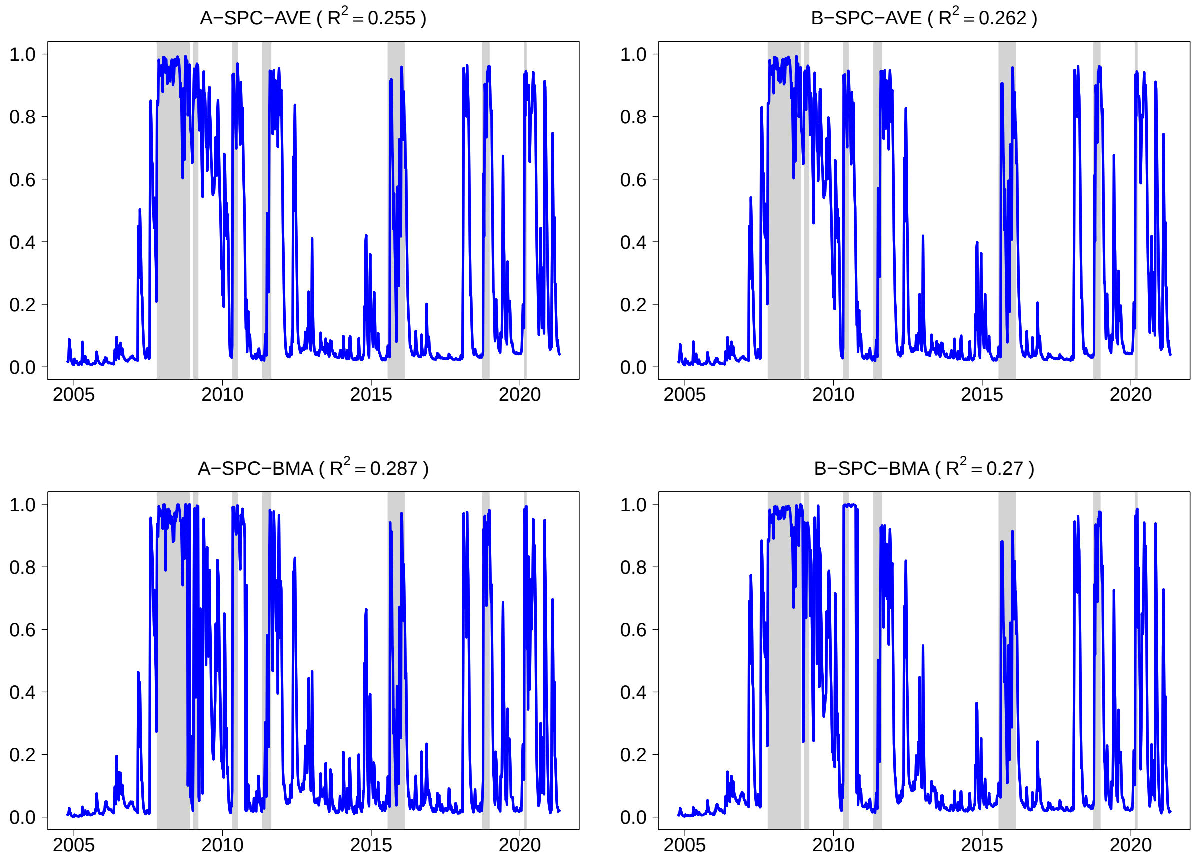
Notes: Figure shows the predicted out-of-sample probabilities of forecasting a bear market. Gray-shaded areas indicate actual bear market periods according to an ex post detection using the dating rule of Lunde and Timmermann (2004). The R^2 is obtained from a linear probability model where the dummy for a bearish period in period $t+1$ is explained with the predicted probability in period t.

Figure D2: Regime Forecasts Using an Ordinary PCA: Bear Market Probability



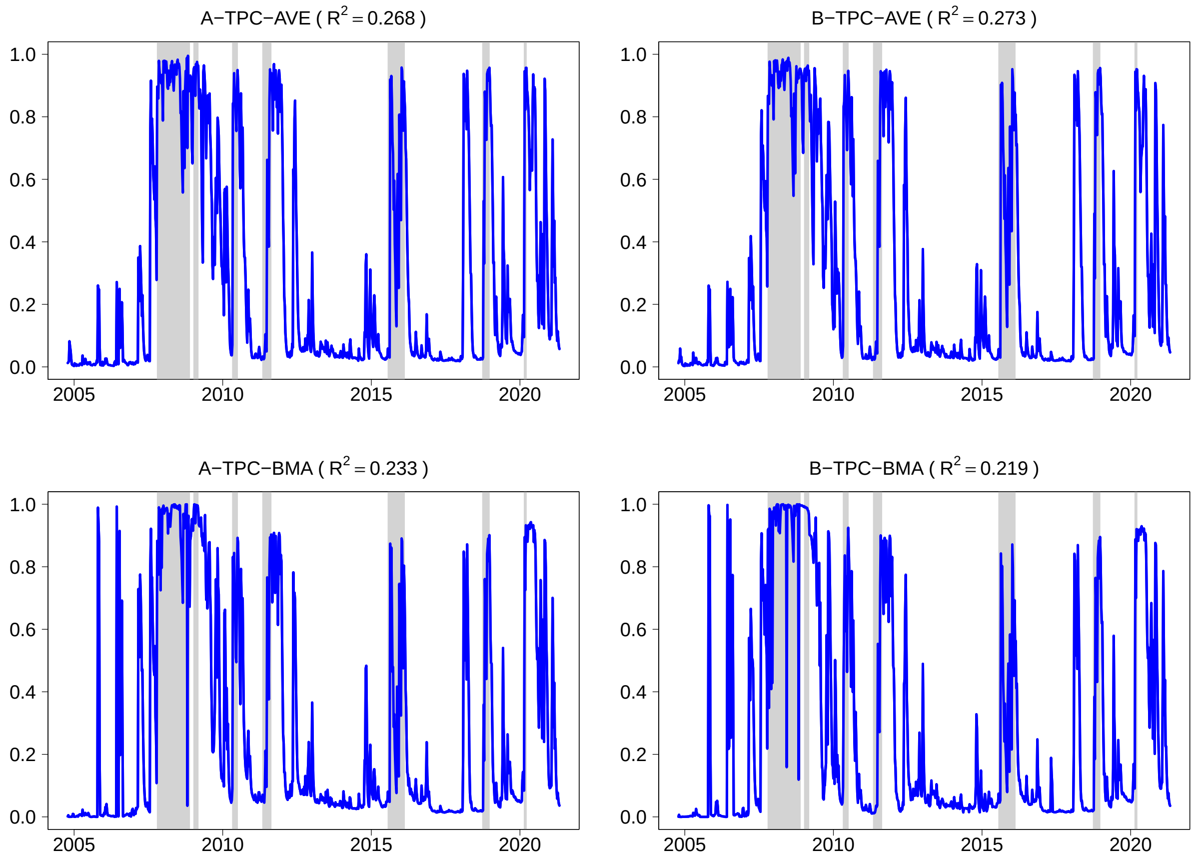
Notes: Figure shows the predicted out-of-sample probabilities of forecasting a bear market. Gray-shaded areas indicate actual bear market periods according to an ex post detection using the dating rule of Lunde and Timmermann (2004). The R^2 is obtained from a linear probability model where the dummy for a bearish period in period $t+1$ is explained with the predicted probability in period t.

Figure D3: Regime Forecasts Using a Sparse PCA: Bear Market Probability



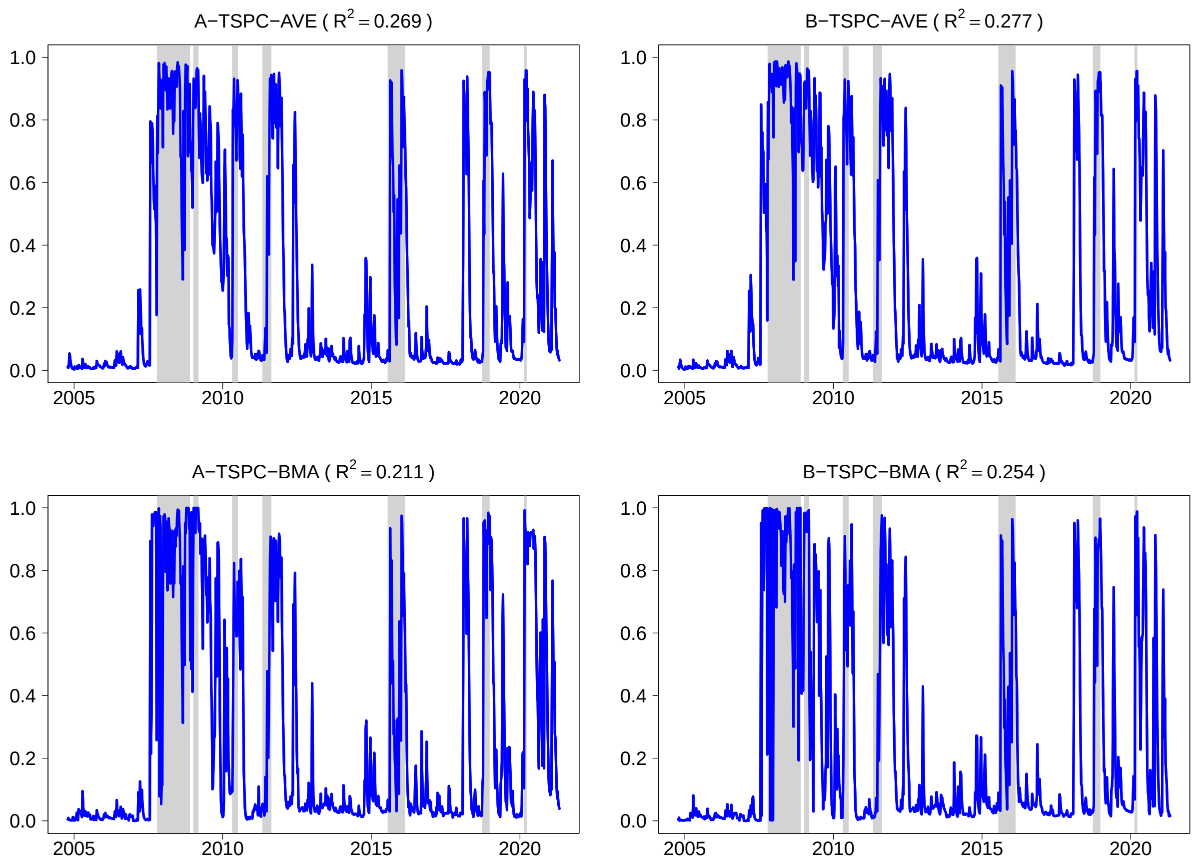
Notes: Figure shows the predicted out-of-sample probabilities of forecasting a bear market. Gray-shaded areas indicate actual bear market periods according to an ex post detection using the dating rule of Lunde and Timmermann (2004). The R^2 is obtained from a linear probability model where the dummy for a bearish period in period $t+1$ is explained with the predicted probability in period t.

Figure D4: Regime Forecasts Using a Targeted PCA: Bear Market Probability



Notes: Figure shows the predicted out-of-sample probabilities of forecasting a bear market. Gray-shaded areas indicate actual bear market periods according to an ex post detection using the dating rule of Lunde and Timmermann (2004). The R^2 is obtained from a linear probability model where the dummy for a bearish period in period $t+1$ is explained with the predicted probability in period t.

Figure D5: Regime Forecasts Using a Targeted Sparse PCA: Bear Market Probability



Notes: Figure shows the predicted out-of-sample probabilities of forecasting a bear market. Gray-shaded areas indicate actual bear market periods according to an ex post detection using the dating rule of Lunde and Timmermann (2004). The R^2 is obtained from a linear probability model where the dummy for a bearish period in period $t+1$ is explained with the predicted probability in period t.

Figure D6: Regime Forecasts: Economic Performance over Time (continued on next page)

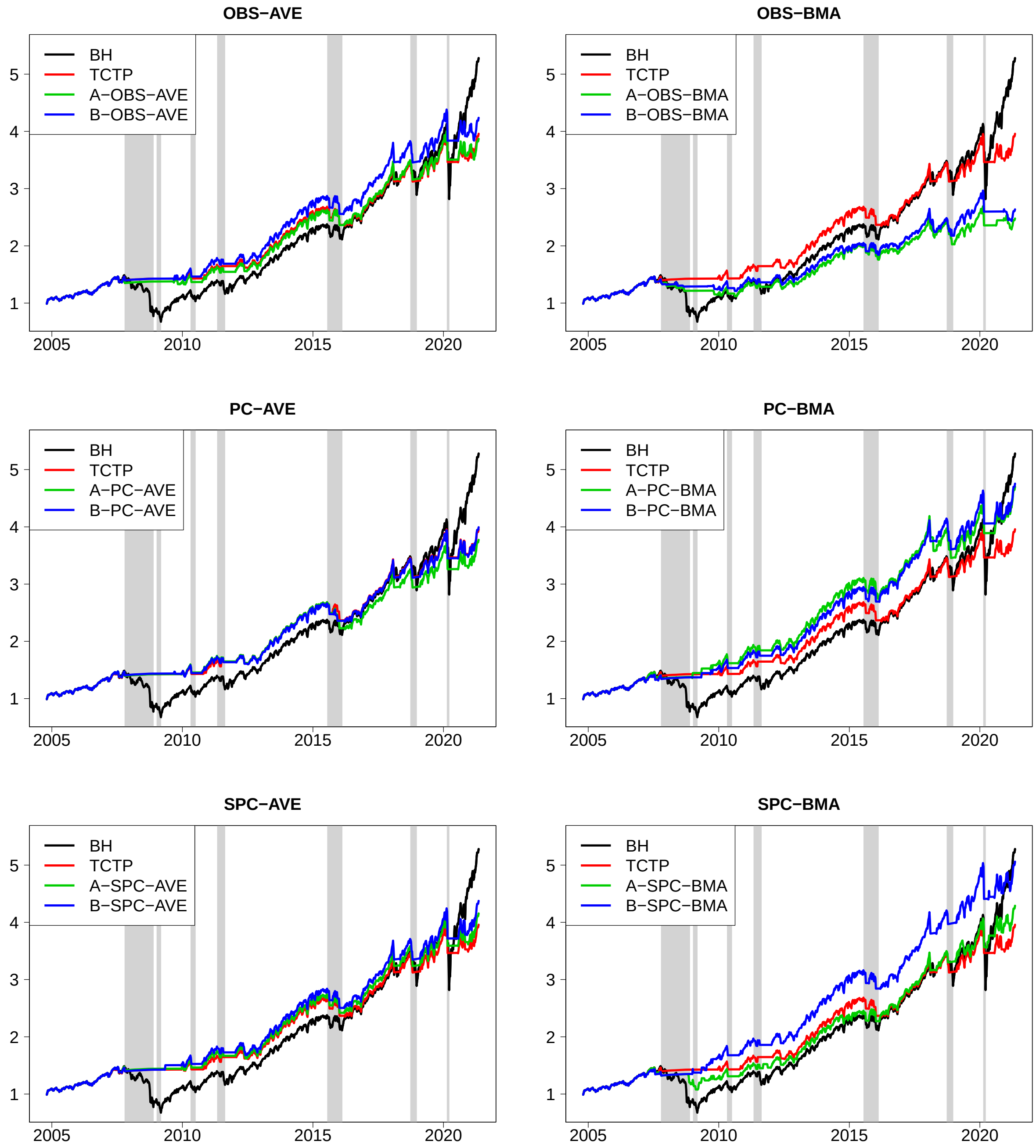
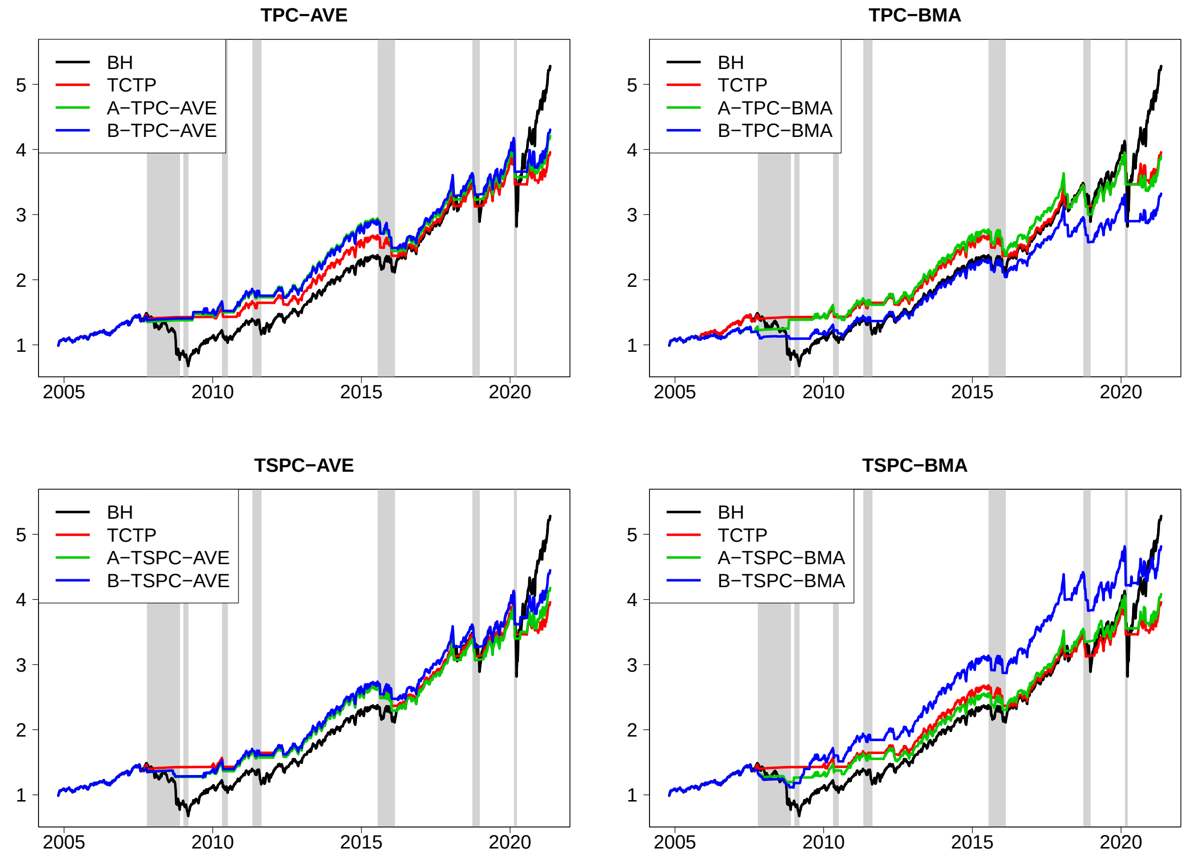


Figure D6: Regime Forecasts: Economic Performance over Time (continued from previous page)



Notes: Figure shows the cumulative performance of the forecasts compared to two benchmarks (BH, TCTP). All strategies are binary, where the total wealth is invested either in the stock market or in the risk-free proxy. See Appendix C for a detailed description of the evaluation measures. Gray-shaded areas indicate actual bear market periods according to an ex post detection using the dating rule of Lunde and Timmermann (2004).

Table D2: Diebold Mariano Test for the Accuracy of Return Predictions

	TCTP	A-OBS-AVE	A-PC-AVE	A-SPC-AVE	A-TPC-AVE	A-TSPC-AVE	A-OBS-BMA	A-PC-BMA	A-SPC-BMA	A-TPC-BMA	A-TSPC-BMA	B-OBS-AVE	B-PC-AVE	B-SPC-AVE	B-TPC-AVE	B-TSPC-AVE	B-OBS-BMA	B-PC-BMA	B-SPC-BMA	B-TPC-BMA	B-TSPC-BMA
TCTP		0.55	0.48	0.40	0.86	0.32	0.18	0.20	**0.08**	0.68	0.19	0.78	0.13	0.61	0.63	0.58	0.68	0.14	0.89	0.15	0.93
A-OBS-AVE	0.45		0.43	0.40	0.84	0.35	0.15	0.17	**0.06**	0.69	0.19	0.50	0.39	0.46	0.47	0.46	0.55	0.36	0.58	0.36	0.60
A-PC-AVE	0.52	0.57		0.36	0.94	0.28	0.18	0.15	**0.05**	0.69	0.17	0.68	0.37	0.57	0.59	0.57	0.67	0.29	0.81	0.32	0.81
A-SPC-AVE	0.60	0.60	0.64		0.91	0.34	0.19	0.17	**0.05**	0.69	0.19	0.74	0.48	0.65	0.67	0.65	0.71	0.41	0.85	0.42	0.84
A-TPC-AVE	0.14	0.16	**0.06**	**0.09**		**0.04**	**0.08**	**0.05**	**0.01**	0.56	**0.04**	0.16	0.10	0.16	0.14	0.14	0.22	**0.09**	0.26	**0.09**	0.28
A-TSPC-AVE	0.68	0.65	0.72	0.66	0.96		0.19	0.21	**0.04**	0.72	0.19	0.79	0.59	0.71	0.72	0.71	0.81	0.52	0.85	0.52	0.87
A-OBS-BMA	0.82	0.85	0.82	0.81	0.92	0.81		0.67	0.47	0.95	0.65	0.84	0.80	0.82	0.82	0.82	0.88	0.79	0.85	0.79	0.86
A-PC-BMA	0.80	0.83	0.85	0.83	0.95	0.79	0.33		0.21	0.83	0.49	0.84	0.78	0.82	0.82	0.82	0.85	0.77	0.87	0.75	0.87
A-SPC-BMA	0.92	0.94	0.95	0.95	0.99	0.96	0.53	0.79		0.92	0.82	0.94	0.91	0.93	0.93	0.93	0.96	0.90	0.95	0.90	0.95
A-TPC-BMA	0.32	0.31	0.31	0.31	0.44	0.28	**0.05**	0.17	**0.08**		0.17	0.34	0.31	0.33	0.33	0.33	0.35	0.29	0.37	0.29	0.38
A-TSPC-BMA	0.81	0.81	0.83	0.81	0.96	0.81	0.35	0.51	0.18	0.83		0.84	0.78	0.82	0.82	0.82	0.87	0.76	0.86	0.76	0.87
B-OBS-AVE	0.22	0.50	0.32	0.26	0.84	0.21	0.16	0.16	**0.06**	0.66	0.16		**0.04**	0.26	0.28	0.20	0.60	**0.05**	0.83	**0.05**	0.86
B-PC-AVE	0.87	0.61	0.63	0.52	0.90	0.41	0.20	0.22	**0.09**	0.69	0.22	0.96		0.98	0.96	0.92	0.75	0.30	0.98	0.34	0.98
B-SPC-AVE	0.39	0.54	0.43	0.35	0.84	0.29	0.18	0.18	**0.07**	0.67	0.18	0.74	**0.02**		0.54	0.48	0.65	**0.05**	0.93	0.11	0.94
B-TPC-AVE	0.37	0.53	0.41	0.33	0.86	0.28	0.18	0.18	**0.07**	0.67	0.18	0.72	**0.04**	0.46		0.43	0.64	**0.07**	0.88	0.10	0.91
B-TSPC-AVE	0.42	0.54	0.43	0.35	0.86	0.29	0.18	0.18	**0.07**	0.67	0.18	0.80	**0.08**	0.52	0.57		0.66	0.11	0.93	0.13	0.96
B-OBS-BMA	0.32	0.45	0.33	0.29	0.78	0.19	0.12	0.15	**0.04**	0.65	0.13	0.40	0.25	0.35	0.36	0.34		0.22	0.54	0.20	0.58
B-PC-BMA	0.86	0.64	0.71	0.59	0.91	0.48	0.21	0.23	0.10	0.71	0.24	0.95	0.70	0.95	0.93	0.89	0.78		0.98	0.50	0.98
B-SPC-BMA	0.11	0.42	0.19	0.15	0.74	0.15	0.15	0.13	**0.05**	0.63	0.14	0.17	**0.02**	**0.07**	0.12	**0.07**	0.46	**0.02**		**0.04**	0.59
B-TPC-BMA	0.85	0.64	0.68	0.58	0.91	0.48	0.21	0.25	0.10	0.71	0.24	0.95	0.66	0.89	0.90	0.87	0.80	0.50	0.96		0.97
B-TSPC-BMA	**0.07**	0.40	0.19	0.16	0.72	0.13	0.14	0.13	**0.05**	0.62	0.13	0.14	**0.02**	**0.06**	**0.09**	**0.04**	0.42	**0.02**	0.41	**0.03**	

Notes: Table shows p-values for the Diebold and Mariano (1995) test. The alternative hypothesis is that a column forecast is less accurate than a row forecast. Hence, "good" ("bad") forecasts are indicated by low p-values in the rows (columns). p-values lower than 10% are highlighted by bold entries.

Figure D7: Return Forecasts: Performance over Time (continued on next page)

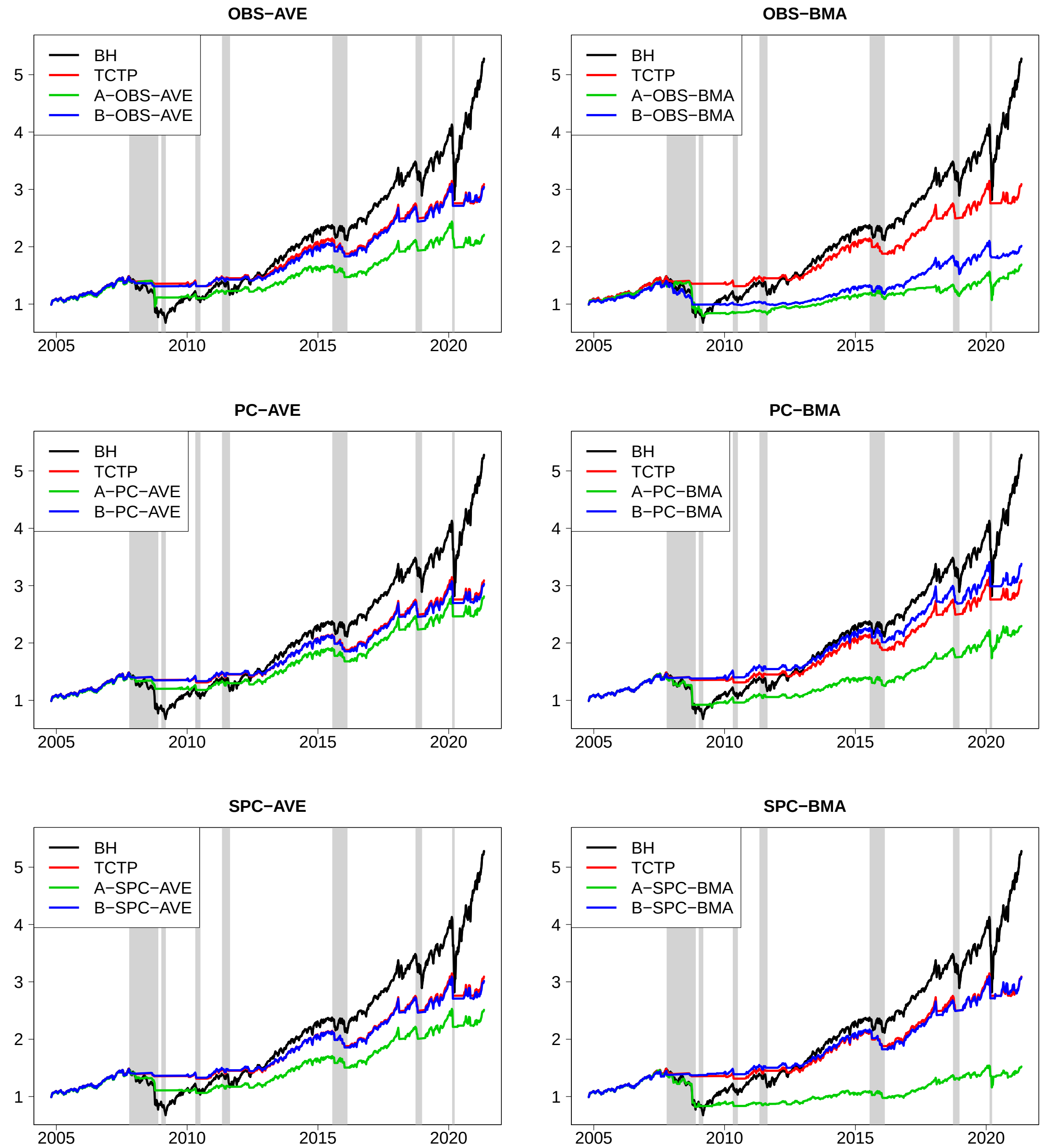
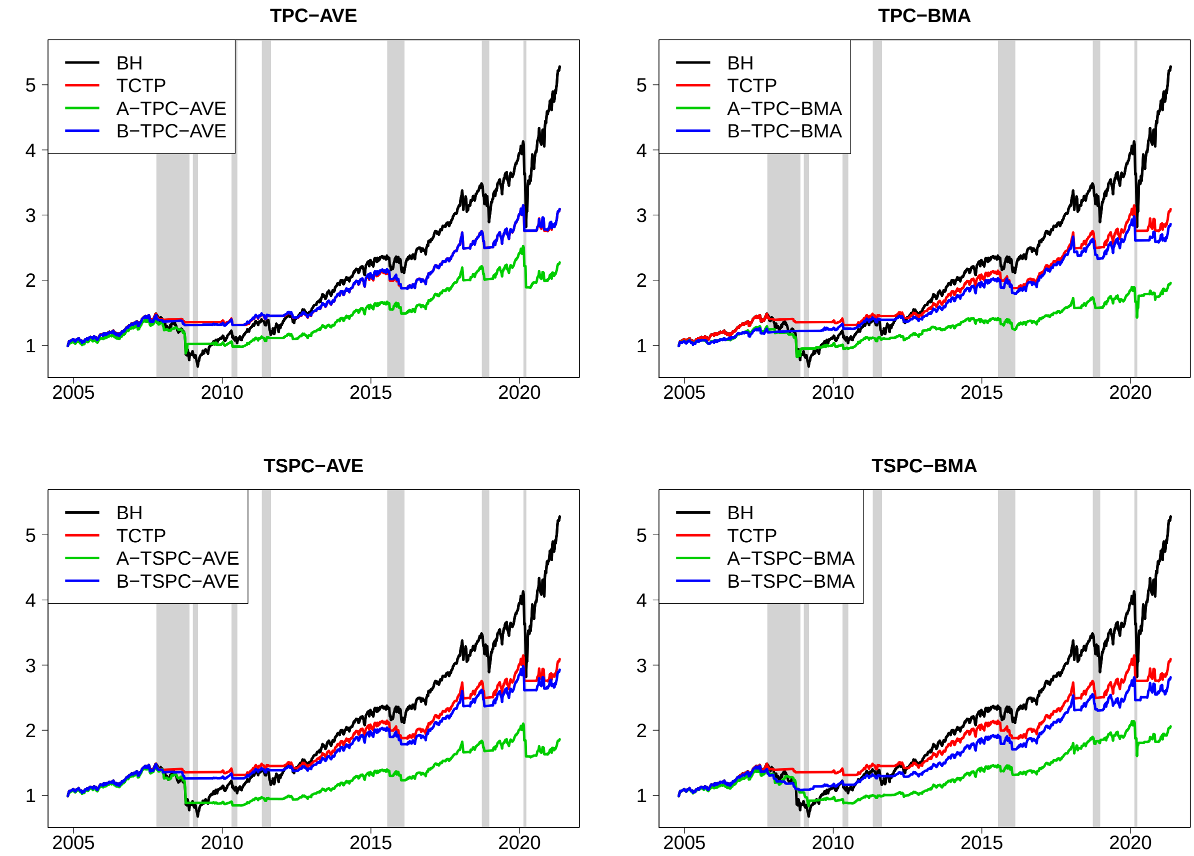


Figure D7: Return Forecasts: Performance over Time (continued from previous page)



Notes: Figure shows the cumulative performance of the models according to each model specification compared to several benchmarks. All strategies are binary, where the total wealth is invested either in the stock market or in the risk-free proxy. See Appendix C for a detailed description of the evaluation measures. Gray-shaded areas indicate actual bear market periods according to an ex post detection using the dating rule of Lunde and Timmermann (2004).

Table D3: Average Annual Turnover and Transaction Costs in the Baseline Scenario

	Switching Strategy		**Mean-Variance Strategy**	
	Turnover	**Transaction Costs**	**Turnover**	**Transaction Costs**
TCTP	253.07	8.40	249.94	8.30
A-OBS-AVE	253.07	8.40	554.37	18.40
A-PC-AVE	241.02	8.00	410.27	13.62
A-SPC-AVE	265.12	8.80	448.72	14.89
A-TPC-AVE	265.12	8.80	610.42	20.26
A-TSPC-AVE	265.12	8.80	546.03	18.12
A-OBS-BMA	409.73	13.60	941.40	31.25
A-PC-BMA	265.12	8.80	710.21	23.57
A-SPC-BMA	361.53	12.00	1272.16	42.23
A-TPC-BMA	385.63	12.80	1420.85	47.16
A-TSPC-BMA	373.58	12.40	1070.82	35.54
B-OBS-AVE	253.07	8.40	269.08	8.93
B-PC-AVE	228.97	7.60	258.35	8.58
B-SPC-AVE	289.22	9.60	265.21	8.80
B-TPC-AVE	265.12	8.80	263.02	8.73
B-TSPC-AVE	289.22	9.60	283.18	9.40
B-OBS-BMA	445.89	14.80	193.62	6.43
B-PC-BMA	265.12	8.80	281.06	9.33
B-SPC-BMA	337.43	11.20	284.03	9.43
B-TPC-BMA	469.99	15.60	438.90	14.57
B-TSPC-BMA	421.78	14.00	400.66	13.30

Notes: Table shows the average annual portfolio turnover and the corresponding cumulative proportional transaction costs (both in %). We assume transaction costs of 20 bps in the baseline scenario. TCTP: MS model with TCTP and without external predictors. Specification A contains predictors in the switching equation and the conditional mean equation, Specification B contains predictors in the switching equation only. OBS: observable predictors; PC: ordinary PCA; SPC: sparse PCA; TPC: ordinary PCA with soft thresholding; TSPC: sparse PCA with soft thresholding; AVE: simple average; BMA: Bayesian model averaging.

Appendix E: Robustness Tests

Table E1: Statistical Performance of Regime Forecasts: Sensitivity to Moving Average Window Lengths

Forecast	QPS	Accuracy	Bear	Bull
MA_3M	0.410	0.781	**0.695**	0.798
MA_6M	0.343	0.814	**0.695**	0.837
MA_12M (D_t^{MA})	**0.310**	**0.836**	0.475	**0.906**
MA_24M	0.384	0.803	0.369	0.888
MA_36M	0.447	0.770	0.291	0.863

Notes: Table shows the statistical performance of regime forecasts when using different moving average window lengths. QPS: quadratic probability score; Accuracy: share of correctly predicted regimes overall; Bear/Bull: share of correctly predicted bearish/bullish regimes. See Appendix C for a detailed description of the evaluation measures. Forecasts with the best performance are highlighted in bold.

Table E2: Economic Value of Regime Forecasts: Sensitivity to Moving Average Window Lengths

Forecast	$\mathbf{R^{cum}}$	$\bar{\mathbf{R}}$	$\bar{\sigma}$	SR	CER	MaxDD	VaR	CVaR
MA_3M	2.65	6.03	**11.25**	0.42	2.84	**−19.7**	−2.48	**−3.99**
MA_6M	2.58	5.87	11.34	0.40	2.65	−21.0	**−2.47**	−4.05
MA_12M (D_t^{MA})	4.03	8.75	12.92	0.57	4.92	−22.2	−2.79	−4.58
MA_24M	**4.69**	**9.75**	13.04	**0.65**	**5.86**	−27.8	−2.82	−4.54
MA_36M	4.01	8.71	13.78	0.54	4.54	−33.5	−2.86	−4.77

Notes: Table shows the economic value of regime forecasts when using different moving average window lengths. R^{cum}: final wealth of strategy assuming a 1$ investment; $\bar{R}$: annualized average returns; $\bar{\sigma}$: annualized standard deviation; SR: annualized Sharpe ratio; CER: annualized certainty equivalent return with $\gamma = 3$; MaxDD: maximum drawdown; VaR: value-at-risk; CVaR: conditional value-at-risk. See Appendix C for a detailed description of the evaluation measures. Forecasts with the best performance are highlighted in bold.

Table E3: Economic Value of Regime Forecasts: Certainty Equivalent Return for Different Ex Ante Transaction Costs and Different Thresholds

	TC = 0 bps			**TC = 20 bps**			**TC = 50 bps**		
	τ			τ			τ		
Forecast	**25%**	**50%**	**75%**	**25%**	**50%**	**75%**	**25%**	**50%**	**75%**
BH	4.38	4.38	4.38	4.38	4.38	4.38	4.38	4.38	4.38
50/50	3.91	3.91	3.91	3.90	3.90	3.90	3.89	3.89	3.89
60/40	4.31	4.31	4.31	4.31	4.31	4.31	4.29	4.29	4.29
MA_12M (D_t^{MA})	5.36	5.36	5.36	4.92	4.92	4.92	4.26	4.26	4.26
TCTP	4.51	6.31	7.46	0.58	0.58	0.58	4.38	4.38	4.38
A-OBS-AVE	5.31	6.12	**8.08**	1.75	3.31	4.24	−2.82	−2.82	−0.82
A-PC-AVE	4.74	5.96	**7.85**	−0.21	0.29	0.98	−3.14	−3.14	−3.14
A-SPC-AVE	4.98	**6.61**	**7.81**	−1.24	−0.45	0.03	−3.74	−3.74	−3.74
A-TPC-AVE	5.22	**6.63**	**8.67**	0.67	1.97	3.36	−2.51	−2.51	−2.04
A-TSPC-AVE	5.13	**6.59**	7.20	2.85	0.41	1.03	−3.14	−3.14	−4.42
A-OBS-BMA	1.30	3.53	5.19	2.68	2.68	1.05	−0.34	−0.34	−0.34
A-PC-BMA	5.31	**7.24**	7.00	3.05	3.76	4.14	0.02	0.02	0.02
A-SPC-BMA	**5.96**	**6.71**	**7.92**	3.57	4.10	**7.50**	−2.46	−2.98	**4.95**
A-TPC-BMA	**5.38**	6.21	6.91	4.51	**6.64**	**7.23**	2.49	2.49	2.89
A-TSPC-BMA	**5.92**	**6.61**	6.22	3.92	**5.16**	**7.58**	−0.54	−0.54	0.55
B-OBS-AVE	**5.48**	**6.67**	**7.64**	−0.78	−0.78	−0.78	4.38	4.38	4.38
B-PC-AVE	4.71	**6.32**	7.42	−0.12	−0.12	−0.12	4.38	4.38	4.38
B-SPC-AVE	4.96	**6.96**	**8.11**	0.05	0.05	0.38	4.38	4.38	4.38
B-TPC-AVE	**5.65**	**6.77**	**8.18**	0.47	0.47	0.47	4.38	4.38	4.38
B-TSPC-AVE	5.20	**7.04**	**7.96**	0.31	0.31	0.20	4.38	4.38	4.38
B-OBS-BMA	1.25	4.02	4.64	4.38	4.38	4.38	4.38	4.38	4.38
B-PC-BMA	5.17	**7.39**	6.39	0.44	0.44	0.44	4.38	4.38	4.38
B-SPC-BMA	**5.72**	**7.91**	**8.59**	0.32	0.32	0.32	−4.92	−4.92	−4.92
B-TPC-BMA	4.37	5.41	5.55	0.02	0.02	0.02	4.38	4.38	4.38
B-TSPC-BMA	**7.10**	**7.59**	6.55	1.76	1.76	1.76	4.38	4.38	4.38

Notes: Table shows the annualized certainty equivalent return (in % and with $\gamma = 3$) of the switching strategy for different transaction costs and regime probability thresholds τ. BH: Buy & Hold Strategy of the S&P 500; 50/50 and 60/40: mixed strategy S&P 500 and 3 M Treasury Bill; D_t^{MA}: naïve 12-month moving average; TCTP: MS model with TCTP and without external predictors. Specification A contains predictors in the switching equation and the conditional mean equation, Specification B contains predictors in the switching equation only. OBS: observable predictors; PC: conventional PCA; SPC: sparse PCA; TPC: conventional PCA with soft thresholding; TSPC: sparse PCA with soft thresholding; AVE: simple average; BMA: Bayesian model averaging. See Appendix C for a detailed description of the evaluation measures. Forecasts that outperform all benchmarks are highlighted in bold.

Table E4: Economic Value of Regime Forecasts: Certainty Equivalent Return for Different Ex Post Transaction Costs and Different Thresholds

	TC = 0 bps			**TC = 20 bps**			**TC = 50 bps**		
	τ			τ			τ		
Forecast	**25%**	**50%**	**75%**	**25%**	**50%**	**75%**	**25%**	**50%**	**75%**
BH	4.38	4.38	4.38	4.38	4.38	4.38	4.38	4.38	4.38
50/50	3.91	3.91	3.91	3.90	3.90	3.90	3.89	3.89	3.89
60/40	4.31	4.31	4.31	4.31	4.31	4.31	4.29	4.29	4.29
MA_12M (D_t^{MA})	5.36	5.36	5.36	4.92	4.92	4.92	4.26	4.26	4.26
TCTP	4.51	6.31	7.46	3.82	5.74	6.92	2.79	4.88	6.11
A-OBS-AVE	5.31	6.12	**8.08**	4.62	5.53	**7.26**	3.58	4.65	6.04
A-PC-AVE	4.74	5.96	**7.85**	4.08	5.41	**7.21**	3.09	4.59	**6.25**
A-SPC-AVE	4.98	**6.61**	**7.81**	4.36	**6.01**	**7.09**	3.43	**5.10**	6.01
A-TPC-AVE	5.22	**6.63**	**8.67**	4.61	**6.03**	**7.97**	3.69	**5.12**	**6.92**
A-TSPC-AVE	5.13	**6.59**	7.20	4.54	**5.99**	6.29	3.65	**5.08**	4.93
A-OBS-BMA	1.30	3.53	5.19	0.41	2.61	4.41	−0.93	1.23	3.24
A-PC-BMA	5.31	**7.24**	7.00	4.63	**6.64**	6.38	3.62	**5.73**	5.45
A-SPC-BMA	**5.96**	**6.71**	**7.92**	**5.04**	**5.89**	**7.07**	3.66	4.65	5.79
A-TPC-BMA	**5.38**	6.21	6.91	4.59	5.34	5.92	3.39	4.04	4.44
A-TSPC-BMA	**5.92**	**6.61**	6.22	**5.20**	**5.77**	5.17	4.12	4.50	3.60
B-OBS-AVE	**5.48**	**6.67**	**7.64**	4.79	**6.08**	6.89	3.75	**5.20**	5.75
B-PC-AVE	4.71	**6.32**	7.42	4.08	**5.80**	6.76	3.13	**5.02**	5.76
B-SPC-AVE	4.96	**6.96**	**8.11**	4.35	**6.31**	**7.37**	3.42	**5.33**	**6.25**
B-TPC-AVE	**5.65**	**6.77**	**8.18**	**4.98**	**6.17**	**7.46**	3.98	**5.27**	**6.37**
B-TSPC-AVE	5.20	**7.04**	**7.96**	4.55	**6.38**	**7.11**	3.58	**5.40**	5.83
B-OBS-BMA	1.25	4.02	4.64	0.23	3.02	3.76	−1.30	1.53	2.44
B-PC-BMA	5.17	**7.39**	6.39	4.47	**6.79**	5.85	3.42	**5.88**	5.03
B-SPC-BMA	**5.72**	**7.91**	**8.59**	**4.97**	**7.14**	**7.65**	3.85	**5.99**	**6.24**
B-TPC-BMA	4.37	5.41	5.55	3.55	4.36	4.57	2.32	2.80	3.10
B-TSPC-BMA	**7.10**	**7.59**	6.55	**6.24**	**6.64**	5.47	**4.96**	**5.21**	3.85

Notes: Table shows the annualized certainty equivalent return (in % and with $\gamma = 3$) of the switching strategy for different transaction costs and regime probability thresholds τ. BH: Buy & Hold Strategy of the S&P 500; 50/50 and 60/40: mixed strategy S&P 500 and 3 M Treasury Bill; D_t^{MA}: naïve 12-month moving average; TCTP: MS model with TCTP and without external predictors. Specification A contains predictors in the switching equation and the conditional mean equation, Specification B contains predictors in the switching equation only. OBS: observable predictors; PC: conventional PCA; SPC: sparse PCA; TPC: conventional PCA with soft thresholding; TSPC: sparse PCA with soft thresholding; AVE: simple average; BMA: Bayesian model averaging. See Appendix C for a detailed description of the evaluation measures. Forecasts that outperform all benchmarks are highlighted in bold. The middle column (with $\tau = 50\%$) replicates the results of Table 6. Note that the results without transaction costs are inherently the same in Tables E3 and E4.

Table E5: Economic Value of Return Forecasts: Certainty Equivalent Return for Different Ex Post Transaction Costs and Different Degrees of Risk Aversion

	TC = 0 bps			**TC = 20 bps**			**TC = 50 bps**		
		γ			γ			γ	
Forecast	**2**	**3**	**5**	**2**	**3**	**5**	**2**	**3**	**5**
BH	5.98	4.38	1.17	5.98	4.38	1.17	5.98	4.38	1.17
50/50	4.28	3.91	3.17	4.27	3.90	3.17	4.26	3.89	3.15
60/40	4.85	4.31	3.24	4.84	4.31	3.23	4.83	4.29	3.22
HIST	3.82	1.41	0.45	3.78	1.35	0.38	3.73	1.26	0.28
TCTP	5.87	4.94	3.04	5.30	4.39	2.51	4.45	3.56	1.71
A-OBS-AVE	4.02	2.81	1.22	2.93	1.66	0.03	1.32	−0.04	−1.72
A-PC-AVE	5.26	4.53	3.16	4.42	3.65	2.31	3.17	2.34	1.03
A-SPC-AVE	4.32	3.83	2.83	3.39	2.88	1.88	2.00	1.45	0.45
A-TPC-AVE	4.25	2.78	1.71	3.12	1.51	0.38	1.44	−0.37	−1.60
A-TSPC-AVE	2.87	1.59	0.83	1.80	0.45	−0.44	0.21	−1.24	−2.31
A-OBS-BMA	2.89	1.12	−2.01	0.87	−0.81	−3.76	−2.08	−3.66	−6.34
A-PC-BMA	4.50	3.15	1.11	3.15	1.66	−0.52	1.15	−0.54	−2.95
A-SPC-BMA	3.16	1.50	−0.63	0.73	−1.06	−3.15	−2.84	−4.81	−6.85
A-TPC-BMA	4.98	3.23	0.67	2.12	0.31	−2.09	−2.06	−3.93	−6.12
A-TSPC-BMA	4.83	3.28	1.11	2.67	1.10	−1.09	−0.49	−2.11	−4.32
B-OBS-AVE	5.85	4.94	2.98	5.20	4.34	2.42	4.21	3.46	1.60
B-PC-AVE	5.85	4.85	2.97	5.25	4.28	2.42	4.37	3.42	1.61
B-SPC-AVE	5.81	4.85	3.01	5.18	4.26	2.46	4.25	3.38	1.64
B-TPC-AVE	5.87	**4.98**	2.95	5.25	**4.40**	2.40	4.31	3.53	1.57
B-TSPC-AVE	5.72	4.69	2.82	5.04	4.06	2.25	4.03	3.12	1.40
B-OBS-BMA	3.37	1.89	0.40	2.99	1.47	−0.02	2.43	0.85	−0.66
B-PC-BMA	**6.81**	**5.52**	**3.29**	**6.19**	**4.90**	2.71	5.27	3.96	1.85
B-SPC-BMA	**6.21**	**5.05**	2.94	5.52	**4.42**	2.36	4.48	3.48	1.49
B-TPC-BMA	**6.25**	4.90	2.74	5.26	3.95	1.94	3.78	2.52	0.75
B-TSPC-BMA	**6.00**	4.61	2.76	5.04	3.73	2.00	3.61	2.42	0.86

Notes: Table shows the annualized certainty equivalent return (in %) of the mean-variance strategy for different transaction costs and varying levels of risk aversion γ. BH: Buy & Hold Strategy of the S&P 500; 50/50 and 60/40: mixed strategy S&P 500 and 3 M Treasury Bill; HIST: historical average of excess stock returns; TCTP: MS model with TCTP and without external predictors. Specification A contains predictors in the switching equation and the conditional mean equation, Specification B contains predictors in the switching equation only. OBS: observable predictors; PC: conventional PCA; SPC: sparse PCA; TPC: conventional PCA with soft thresholding; TSPC: sparse PCA with soft thresholding; AVE: simple average; BMA: Bayesian model averaging. See Appendix C for a detailed description of the evaluation measures. Forecasts that outperform all benchmarks are highlighted in bold.

Table E6: Economic Value of Return Forecasts with Switching Strategy: Certainty Equivalent Return for Different Ex Post and Ex Ante Transaction Costs

	Ex Post Transaction Costs			**Ex Ante Transaction Costs**		
	0 bps	**20 bps**	**50 bps**	**0 bps**	**20 bps**	**50 bps**
BH	4.38	4.38	4.38	4.38	4.38	4.38
50/50	3.91	3.90	3.89	3.91	3.90	3.89
60/40	4.31	4.31	4.29	4.31	4.31	4.29
MA_12M (D_t^{MA})	5.36	4.92	4.26	5.36	4.92	4.26
TCTP	4.96	4.34	3.40	4.96	0.58	4.38
A-OBS-AVE	4.23	3.14	1.51	4.23	0.10	−2.43
A-PC-AVE	3.46	2.75	1.68	3.46	0.29	−3.14
A-SPC-AVE	3.21	2.32	1.00	3.21	-0.45	−3.74
A-TPC-AVE	5.31	4.25	2.68	5.31	1.50	−2.04
A-TSPC-AVE	3.91	2.98	1.58	3.91	0.86	−3.14
A-OBS-BMA	2.16	0.59	−1.75	2.16	**5.86**	4.38
A-PC-BMA	4.41	3.31	1.68	4.41	2.34	−1.29
A-SPC-BMA	3.47	1.29	−1.93	3.47	-0.54	−4.28
A-TPC-BMA	**5.54**	3.11	−0.45	**5.54**	2.73	−0.05
A-TSPC-BMA	**5.93**	3.89	0.88	**5.93**	0.13	−1.80
B-OBS-AVE	**6.10**	**5.50**	**4.58**	**6.10**	-0.78	4.38
B-PC-AVE	5.11	4.49	3.56	5.11	-0.12	4.38
B-SPC-AVE	**5.59**	**4.96**	4.02	**5.59**	0.05	4.38
B-TPC-AVE	**6.23**	**5.53**	**4.48**	**6.23**	0.47	4.38
B-TSPC-AVE	**6.31**	**5.65**	4.67	**6.31**	0.31	4.38
B-OBS-BMA	4.19	4.04	3.80	4.19	4.38	4.38
B-PC-BMA	5.08	4.42	3.43	5.08	0.44	4.38
B-SPC-BMA	**7.02**	**6.26**	**5.12**	**7.02**	0.32	−4.92
B-TPC-BMA	**6.45**	**5.61**	4.34	**6.45**	0.02	4.38
B-TSPC-BMA	**6.08**	**5.12**	3.67	**6.08**	1.76	4.38

Notes: Table shows the annualized certainty equivalent return (in % and with $\gamma = 3$) of the switching strategy for different ex post and ex ante transaction costs. We use the same switching strategy as in Dal Pra et al. (2018), where the sign of the predicted equity risk premium is used to allocate between Treasury Bills (strictly negative) and the stock market (positive or zero). BH: Buy & Hold Strategy of the S&P 500; 50/50 and 60/40: mixed strategy S&P 500 and 3 M Treasury Bill; D_t^{MA}: naïve 12-month moving average; TCTP: MS model with TCTP and without external predictors. Specification A contains predictors in the switching equation and the conditional mean equation, Specification B contains predictors in the switching equation only. OBS: observable predictors; PC: conventional PCA; SPC: sparse PCA; TPC: conventional PCA with soft thresholding; TSPC: sparse PCA with soft thresholding; AVE: simple average; BMA: Bayesian model averaging. See Appendix C for a detailed description of the evaluation measures. Forecasts that outperform all benchmarks are highlighted in bold. Note that the results without transaction costs are naturally the same.