CESifo WORKING PAPERS

8828 2021

January 2021

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

Felix Haase, Matthias Neuenkirch

CESifo

Impressum:

CESifo Working Papers
ISSN 2364-1428 (electronic version)
Publisher and distributor: Munich Society for the Promotion of Economic Research - CESifo GmbH
The international platform of Ludwigs-Maximilians University's Center for Economic Studies and the ifo Institute
Poschingerstr. 5, 81679 Munich, Germany
Telephone +49 (0)89 2180-2740, Telefax +49 (0)89 2180-17845, email office@cesifo.de
Editor: Clemens Fuest
https://www.cesifo.org/en/wp
An electronic version of the paper may be downloaded
· from the SSRN website: www.SSRN.com
· from the RePEc website: www.RePEc.org
· from the CESifo website: https://www.cesifo.org/en/wp

CESifo *Working Paper No. 8828*

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

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 the documented merits of diffusion indices, regime-switching models, and forecast combination to predict the dynamics in the S&P 500. First, we aggregate the weekly information of 115 popular macroeconomic and financial variables through an interaction of principal component analysis and shrinkage methods. Second, we estimate one-step Markov-switching models with time-varying transition probabilities using the diffusion indices as predictors. Third, we pool the forecasts in clusters to hedge against model risk and to evaluate the usefulness of different specifications. Our results show that we can adequately predict regime dynamics. Our forecasts provide a statistical improvement over several benchmarks and generate economic value by boosting returns, improving the certainty equivalent return, and reducing tail risk. Using the same approach for return forecasts, however, does not lead to a consistent outperformance of the historical average.

JEL-Codes: C530, G110, G170.

Keywords: forecast combination, Markov-Switching Models, shrinkage methods, stock market regimes, time-varying transition probabilities.

Felix Haase
University of Trier
Department of Economics
Germany – 54286 Trier
felixhaase.mesenich@web.de

*Matthias Neuenkirch**
University of Trier
Department of Economics
Germany – 54286 Trier
neuenkirch@uni-trier.de

*corresponding author

First Draft: March 23, 2020
This Version: January 8, 2021
An earlier version of the paper circulated under the title "Forecasting Stock Market Recessions in the US: Predictive Modeling using Different Identification Approaches."

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 jumps 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 2007−2009 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). Hereby, the number of regimes is still subject to debate (for instance, 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.

Considering the empirical success of diffusion indices (Neely et al. 2014; Çakmaklı and van Dijk 2016), regime-switching models (Guidolin and Timmermann 2007; Maheu et al. 2012), and forecast combination (Rapach et al. 2010) in predicting stock market dynamics, we propose a novel procedure that combines these three approaches. Confronted with a large real-time dataset of macroeconomic and financial market variables, we first reduce the dimension into a few latent factors by principal component analysis (PCA). Thereby, we utilize shrinkage methods either to select targeted predictors or to introduce sparsity on the factor loadings. Second, using the factors 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 either a general specification, where the conditional mean and the transitions are modeled, or we rely on a restricted model that focuses on the switching process only. Since highly parameterized models tend to be inferior to parsimonious ones in terms of forecast accuracy, we limit the model size to just one exogenous factor. This results in a large number of models and forecasts that we combine into several 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, an ordinal ranking, 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.

Our sample covers weekly data for the S&P 500 and the period from November 24, 1989 to July 3, 2020. Our out-of-sample real-time exercise focuses on the most recent 820 weeks, that is, the first training set to estimate the MS models ends on October 15, 2004. Our results suggest that bull and bear markets can be identified in a timely manner to participate in recoveries or to prevent losses. We show that modeling regime dynamics improves the risk-adjusted performance, reduces the tail risks, and achieves substantial utility gains relative to a simple MS model or other common benchmarks. More precisely, the models achieve accuracy rates of more than 80%, whereby bearish weeks are correctly classified in 70–75% of the cases. We also demonstrate that pooled forecasts of factor-augmented MS models are more useful than a combination of standard stock predictors proposed by the literature. Concerning the factor aggregation techniques, the sparse PCA creates more valuable predictors compared to a conventional PCA, highlighting the benefits of shrinkage methods. However, the advantages of our procedure cannot be replicated when forecasting returns. Despite certainty equivalent gains of more than 1% relative to the historical average, we do not find a consistent statistical superiority of our approach.

Our paper is, to the best of our knowledge, the first one to apply factor-augmented MS models with TVTP to predict bull and bear market. We contribute to three strands of the stock market forecasting literature. First, we confirm the previous finding of predictable trends in stock markets (Guidolin and Timmermann 2007; Chen 2009; Kritzman et al. 2012). These trends are recurrent and persistent, whereby a timely detection of turning points determines success. Second, we emphasize the benefits of factor-augmented MS models with TVTP. Although MS models with time-varying transitions have been developed over 25 years ago (Diebold et al. 1994), there are only a few examples that apply these models in the context of bull and bear markets (for instance, Maheu and McCurdy 2000; Kole and van Dijk 2017; Focardi et al. 2019). Hereby, Kole and van Dijk (2017) do not detect any advantage of modeling the transition with macro-financial variables. Overly complex modeling of the switching process might cause their results. Hence, we follow the suggestion of Zens and Böck

(2019) to include only a few latent factors (one, to be precise) into the transition equation. Finally, 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.

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 results of an ex post identification of market regimes and discusses the aggregation of the predictors assuming knowledge of the full sample. Section 5 demonstrates how our approach works in a real-time situation with recursive out-of-sample forecasts. Section 6 concludes.

2 Methodology

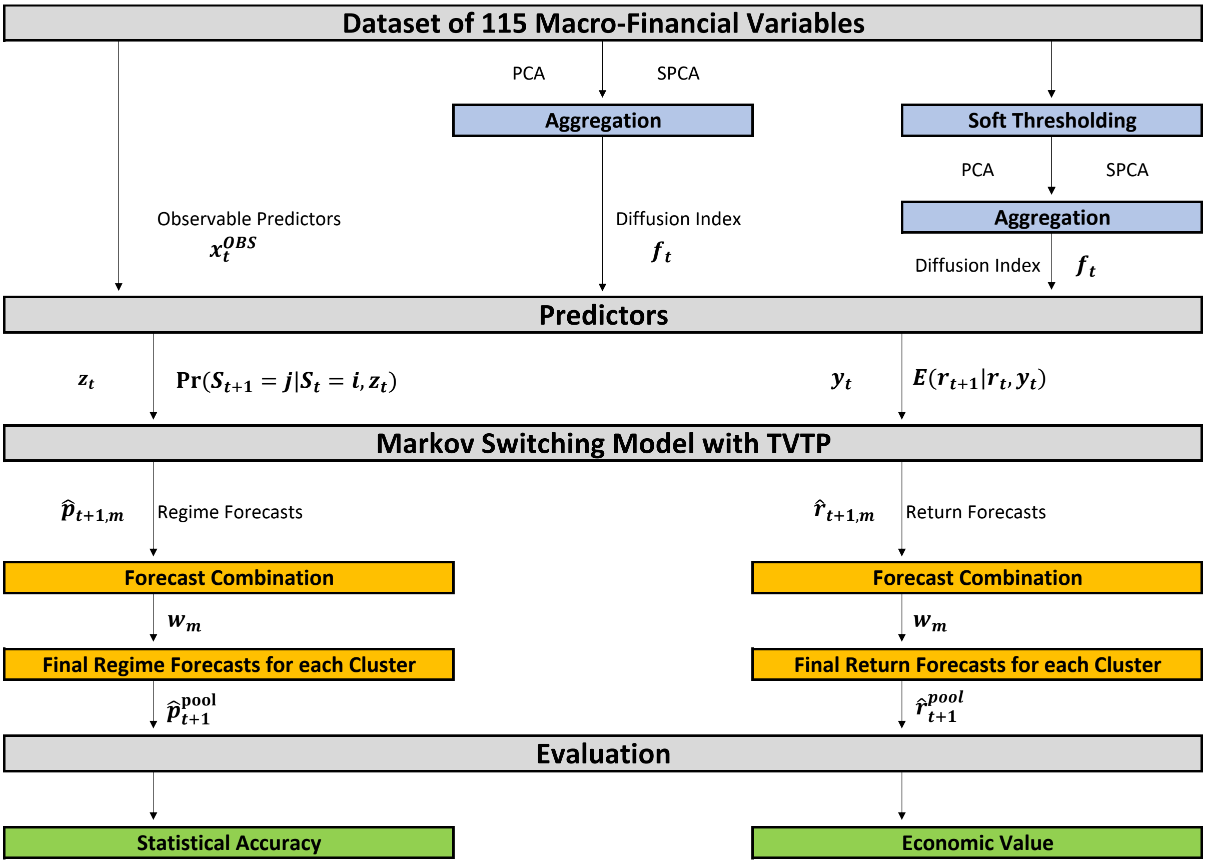
We face the issues of parameter instability and model uncertainty when forecasting stock market regimes and returns (Pesaran and Timmermann 1995). Rapach and Zhou (2013) summarize four different approaches that have proven to be empirically helpful in this context: (i) theoretical model restrictions, (ii) diffusion indices, (iii) regime-switching models, and (iv) forecast combination. Our procedure builds on the latter three approaches and tests for the extent to which a combination of these adds statistical value and provides economic gains.

To evaluate the usefulness of aggregation techniques, we also apply the MS models and the forecast combination scheme to a subset of directly observable popular predictors as proposed by the literature. Hereby, we consider the lagged returns R, the dividend-price-ratio DP (Campbell and Shiller 1988; Fama and French 1988; Schaller and Norden 1997), the 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).[1]

Figure 1 provides an overview of the individual steps in our procedure that are explained in detail in the following subsections.

[1] Term spread: difference between the 10Y US treasury bond and the 3M treasury bill. Credit spread: excess yield of the Moody's seasoned Baa corporate bond yield over the 10Y treasury bond. Variance risk premium: difference between the squared VIX and the sum of the squared daily (realized) returns of the last 22 trading days.

Figure 1: Overview of the Methodology



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 huge pool of potential variables. Due to the substantial correlation of many variables with unobserved state variables — such as the business cycle, investor sentiment, or market movements — an efficient filtration of the variables is recommended to cover the co-movement and to eliminate potential noise. In this spirit, PCA is an appealing method to capture relevant information in a parsimonious way. Common factors or diffusion indices are constructed so that these are linear combinations of observable predictors and are ranked in descending order according to their explained variance. A small number of diffusion indices is usually sufficient to capture most of the variation in the data, allowing for a significant reduction in the dimension of the original dataset.

There are two main problems when using PCA in economic applications. First, the diffusion indices are constructed unsupervised (without considering the variable that should be predicted) and second, the indices are a combination of all variables, which often leads to a lack of interpretability. To tackle these problems, we use shrinkage methods either to select targeted predictors (Bai and Ng 2008) or/and to restrict the weight of some predictors to zero (Rapach and Zhou 2019). In brief, we apply a conventional PCA and a sparse PCA on the total dataset and we use both methods on a targeted subset of sixty predictors, whereby we conduct a soft thresholding using the Least Angle Regression Elastic Net (LARS-EN) algorithm.

2.1.1 Conventional Principal Component Analysis

Among others, Stock and Watson (2002) propose a PCA to approximate unobserved common factors to forecast macroeconomic variables. These factors capture the comovement of many (potentially) correlated predictors and can be interpreted as diffusion indices. In the context of the stock market forecasting literature, the diffusion index approach delivers promising results. Neely et al. (2014) and Çakmaklı and van Dijk (2016) show that diffusion indices significantly outperform individual macro-financial variables and also beat the historical average.

Let X be a $T \times K$ matrix of potential predictors, where the number of rows T ($t = 1, 2, ..., T$) covers the time dimension and K ($k = 1, 2, ..., K$) the cross-sectional dimension of predictors. Since the scale of our data is very heterogeneous, we normalize all variables to a mean of zero and a variance of one. Using singular value decomposition of X, we can obtain the principal components as follows (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 dimension reduction in exchange for a minimal loss of information ($q << min(K, T)$). In addition, the components are constructed in such a way so that these are uncorrelated to each other. To determine q, we use the IC_{p2} information criterion by Bai and Ng (2002), where the maximum is obtained from a scree plot. Hence, we select the first q normalized principal components as relevant factors F to predict stock market regimes and returns.

2.1.2 Soft Thresholding

One major drawback of the PCA is that it does 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 (for instance, 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 (least absolute shrinkage and selection operator) and ridge regression that performs model selection and shrinkage simultaneously.[2] More formally, the EN optimization is a regularized regression to minimize the residual sum of squares (RSS) and can be written as follows:

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

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 returns. 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 is therefore a good signal for shifts into a bearish regime. We follow Bai and Ng (2008) and use the LARS-EN algorithm to obtain a ranking of selected predictors. We set $\lambda_2 = 0.25$ and select the top 60 predictors, which is identical to the subset size in Çakmaklı and van Dijk (2016).

2.1.3 Sparse Principal Component Analysis

Another disadvantage of conventional PCA is that the components are based on all variables. A sparse PCA uses shrinkage methods to reduce the loadings of some variables to zero. Rapach and Zhou (2019) summarize two advantages of sparse PCA as compared to conventional PCA. In addition to a more straightforward interpretation of the factors, the noise can be filtered out more adequately without losing much of the captured variance.

Following the illustration of Zou et al. (2006), we treat the optimization as regularized regression problem. Suppose that 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 $\lambda > 0$:

$$\arg\min_{A,B}\left[\sum_{t=1}^{T} \|x_t - AB^T x_t\|^2 + \lambda \sum_{p=1}^{q} \|\beta_p\|_2^2 + \sum_{p=1}^{q} \lambda_{1,p} \|\beta_p\|_1\right] \tag{3}$$
$$s.t.\ A^T A = I$$

[2]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).

$\|.\|_1$ 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, whereby a higher value of $\lambda_{1,p}$ leads to more sparse loadings. If we restrict (3) by $B = A$ and set the LASSO penalty $\lambda_{1,p} = 0$, the ordinary PCA results (Zou et al. 2006). We solve (3) using the variable projection approach by Erichson et al. (2020) and set the penalization factor $\lambda_{1,p} = 0.01$ and $\lambda = 0.0001$ based on different trails before the first forecast is done.[3]

2.2 Step 2: Markov-Switching Models

Since the pioneering work of Hamilton (1989), MS models became increasingly popular in economics. MS models are able to reveal structural 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 log-return of the S&P 500 and S_t the unobservable state of the stock market. Then, the non-linear return dynamics can be described as follows:

$$\begin{gathered} r_t = \mu_{S_t} + u_t \\ u_t \sim i.i.d.\ N(0, \sigma^2_{S_t}) \\ Pr(S_t = j | S_{t-1} = i) = p_{ij} \end{gathered} \tag{4}$$

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.

The majority of papers treats the transition probabilities as constant over time, ignoring that these can be affected by changes in fundamental conditions. Diebold et al. (1994) suggest to relate the switching process to economic variables via a logit link function. Despite the appealing character of TVTP, there are only a few applications in the empirical literature of stock market predictability (Schaller and Norden 1997; Maheu and McCurdy 2000; Kole and van Dijk 2017; Focardi et al. 2019). In this paper, we follow Diebold et al. (1994) and model the switching process as being dependent

[3] In the case where we apply the sparse PCA on the targeted subset, the hyperparameter setting is slightly different ($\lambda_{1,p} = 0.01$ and $\lambda = 0.01$) due to the dimensional reduction already achieved by using soft thresholding.

on macro-financial conditions z_{t-1}, so that the transition probabilities can expressed as $Pr(S_t = j|S_{t-1} = i, z_{t-1}) = p_{ij,t}$. In our application, we consider two regimes, where regime 0 corresponds to bull markets and regime 1 to bear markets. Consequently, the transition matrix $\mathbf{P}$ is as follows:

$$\mathbf{P} = \begin{bmatrix} p_{00,t} & p_{10,t} \\ p_{01,t} & p_{11,t} \end{bmatrix} = \begin{bmatrix} p_{00,t} & p_{10,t} \\ 1 - p_{00,t} & 1 - p_{10,t} \end{bmatrix}$$

The number of stock market regimes is certainly open to debate. Since S_t is a latent variable, the true number of regimes is unknown. An approximation with econometric tests is also difficult (Hansen 1991; Ang and Timmermann 2012), so that 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 (exogenous) variable in the switching equation and in the conditional mean equation. This ensures a robust estimation process and reduces the variability of the forecasts.

Model A: In the general framework, we assume that the S&P 500 returns follow a MS autoregressive (MS-AR) model of order one with TVTP:

$$\begin{gathered} r_t = \mu_{S_t} + \phi_{S_t} r_{t-1} + \beta_{S_t} y_{t-1} + u_t \\ u_t \sim i.i.d.\ N(0, \sigma^2_{S_t}) \\ p_{ij,t} = \frac{\exp(\upsilon_{ij} + \gamma_{ij} z_{t-1})}{1 + \exp(\upsilon_{ij} + \gamma_{ij} z_{t-1})} \end{gathered} \tag{5}$$

y_{t-1} and z_{t-1} are either observable predictors proposed by the literature or diffusion indices approximated by the different PCA techniques described in the previous subsec-

[4]Note that some authors assume more than two regimes (Guidolin and Timmermann 2007; 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).

tion. In addition, z_{t-1} can also represent the lagged returns r_{t-1}, implying an endogenous switching process. ϕ_{S_t} represents the regime-dependent AR(1) coefficients and captures the degree of persistence in the returns. 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_{ij} and the slope γ_{ij} 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.

Model B: Given the rich regime dependency of *Model A*, overfitting might be a problem. For this reason, we also consider a restricted model that focuses only on the switching process. By setting the constraints $\phi_{S_t} = 0$ and $\beta_{S_t} = 0$ in *Model A*, we obtain *Model B*:

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

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

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:

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

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 $\hat{p}^1_{t+1} = \hat{p}_{t+1}$ 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 $E[r_{t+1}|S_t = j]$, the return forecast is given by the following probability-weighted average:

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

[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).

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. The conceptional idea is based on the seminal paper of Bates and Granger (1969) and works in a similar manner as the portfolio theory in finance. 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). Among others, Rapach et al. (2010) demonstrate its added value and superiority over the historical average in forecasting the equity risk premium. 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. 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 so that we are 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 various PCA techniques) and (ii) model specification (*Model A* or *Model B*). Table 1 provides an overview of the clusters and their components.

Next, we have to determine the weights of the forecasts. For this purpose, we employ three different methods and account for the different characteristics of regime forecasts and return forecasts.

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

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

The individual weights w_m are calculated according to the following three approaches:

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

Inverse Rank (IVR) $\quad w_m = \frac{Rank_m^{-1}}{\sum_{l=1}^{L} Rank_l^{-1}}$

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

The simple average forecast is straightforward and precludes any estimation risk. In addition, it often provides good results, which are difficult to beat (Timmermann 2006). When setting the weights proportional to its inverse rank, we have to rely on a criterion. We choose the BIC, where the model with the lowest value gets rank one, the one with the second lowest gets rank two, and so on. As third option, 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.

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

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

Similar to the regime forecasts, we apply three approaches to obtain the combination weights:

Simple Average (AVE)	$w_m = \frac{1}{M}$
Inverse Rank (IVR)	$w_m = \frac{Rank_m^{-1}}{\sum_{l=1}^{L} Rank_l^{-1}}$
Discounted Mean Squared Prediction Error (DMSPE)	$w_m = \frac{\Phi_m^{-1}}{\sum_{l=1}^{L} \Phi_l^{-1}}$

Again, we use the simple average and the inverse rank approach. In the third weighting scheme, we employ the discounted mean squared prediction error (DMSPE) to determine the weights with $\Phi = \sum_{s=l+1}^{T} D^{t-s}(r_s - \hat{r}_{s,m})^2$ and a discount factor of $D = 0.9$ (Rapach et al. 2010).

To summarize, we aggregate each cluster in Table 1 using three different techniques, yielding a total of 30 forecast combinations. In the discussion of the results, we replace the placeholder $COMB$ in Table 1 with $\{AVE, IVR, BMA\}$ for regime forecasts and with $\{AVE, IVR, DMSPE\}$ for return forecasts.

Table 1: Forecast Combination Clusters

Specification	Cluster	Models	
	OBS-COMB	A-R	
	M = 13	A-R-DP	A-DP-DP
		A-R-VIX	A-VIX-VIX
		A-R-TS	A-TS-TS
		A-R-CS	A-CS-CS
		A-R-PMI	A-PMI-PMI
		A-R-VP	A-VP-VP
	PC-COMB	A-R-PC1	A-PC1-PC1
	M = 10	A-R-PC2	A-PC2-PC2
		A-R-PC3	A-PC3-PC3
		A-R-PC4	A-PC4-PC4
		A-R-PC5	A-PC5-PC5
Model A	**SPC-COMB**	A-R-SPC1	A-SPC1-SPC1
	M = 10	A-R-SPC2	A-SPC2-SPC2
		A-R-SPC3	A-SPC3-SPC3
		A-R-SPC4	A-SPC4-SPC4
		A-R-SPC5	A-SPC5-SPC5
	TPC-COMB	A-R-TPC1	A-TPC1-TPC1
	M = 8	A-R-TPC2	A-TPC2-TPC2
		A-R-TPC3	A-TPC3-TPC3
		A-R-TPC4	A-TPC4-TPC4
	TSPC-COMB	A-R-TSPC1	A-TSPC1-TSPC1
	M = 8	A-R-TSPC2	A-TSPC2-TSPC2
		A-R-TSPC3	A-TSPC3-TSPC3
		A-R-TSPC4	A-TSPC4-TSPC4
	OBS-COMB	B-R	B-DP
	M = 7	B-VIX	B-TS
		B-CS	B-PMI
		B-VP	
	PC-COMB	B-PC1	B-PC2
	M = 5	B-PC3	B-PC4
		B-PC5	
Model B	**SPC-COMB**	B-SPC1	B-SPC2
	M = 5	B-SPC3	B-SPC4
		B-SPC5	
	TPC-COMB	B-TPC1	B-TPC2
	M = 4	B-TPC3	B-TPC4
	TSPC-COMB	B-TSPC1	B-TSPC2
	M = 4	B-TSPC3	B-TSPC4

Notes: All MS models are estimated with TVTP. Model A contains predictors in the switching equation and the conditional mean equation according to Eq. (5). Model B contains predictors in the switching equation only according to Eq. (6). OBS: observable predictors; PC: ordinary PCA; SPC: sparse PCA; TPC: ordinary PCA with soft thresholding; TSPC: sparse PCA with soft thresholding; R: lagged returns; DP: dividend-price-ratio; VIX: VIX; TS: term spread; CS: credit spread; PMI: Purchasing Managers Index; VP: variance risk premium. For example, A-R-SPC3 is estimated with lagged returns as transition predictor z_{t-1} and the third sparse principal component as conditional mean predictor y_{t-1}. B-R only uses lagged returns as transition predictor z_{t-1}.

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 115 variables to predict regimes and returns. This includes bonds yields, term spreads and credit spreads, lagged returns, technical indicators, industry returns, market-based risk indicators, valuation ratios, survey-based expectations about macroeconomic variables and their dispersion, sentiment indicators, and macroeconomic fundamentals. All variables either have proved to be empirically relevant or can be recommended from a practical point of view. There are several papers that use a high-dimensional dataset to predict financial variables. Mönch (2008) and Ludvigson and Ng (2009) predict bonds yields. For stock market applications, Ludvigson and Ng (2007), Neely et al. (2014), and Çakmaklı and van Dijk (2016) provide promising results and highlight the attractiveness of diffusion indices as predictors for stock returns. Compared to these papers, our work is differentiated in at least two ways. We consider weekly data and, perhaps more important, the latent regimes are explicitly the subject of our forecast.

First, 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 or inflation-linked bonds, and the LIBOR inter-banking rate. 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 spread into account. As additional predictors, we consider the realized variance of the S&P 500 expressed as sum over the squared returns of the previous 1, 5, and 22 trading days. Furthermore, we use information from option markets by using the implied volatility index of the S&500, the VIX. According to Bollerslev et al. (2009) and Bekaert and Hoerova (2014), the VIX can be decomposed in a component that reflects the expected future volatility and in 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 also use the de-trended trading volume of the S&P500 and additional indicators that capture changes in risk perception, like the gold price and the WTI oil price.

Second, we utilize survey-based expectations as predictors. Consensus Economics asks analysts from banks and research institutes about their macroeconomic expectations at a monthly frequency. We utilize 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 as predictors.[6] 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, the PMI, and the manufactures business condition measured by the Federal Reserve Bank of Philadelphia. Lastly, we roughly consider the same standard macroeconomic variables as Chen (2009) to nest previous findings into our analysis.[7]

Third, 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 yield ratio, the earnings yield ratio, the 10Y earnings yield ratio, and the payout ratio in our dataset. Moreover, we incorporate the same technical indicators as proposed by Neely et al. (2014).[8] In addition, we incorporate the short-run and long-run moving average of returns into our predictor set, which are either equally or exponentially weighted. It might be argued that price "excesses" are a major 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.

Fourth, we use the returns of ten 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).

[6] Batchelor (2001) provides evidence that forecasts by Consensus Economics are superior to forecasts by the IMF or the OECD. In addition, these forecasts are available on a monthly basis (instead of quarterly or annually), which is particularly helpful for our analysis of changes in stock market regimes (and returns).

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

[8] We adjust the period length of calculation to account for the weekly frequency and use 1, 4, and 8 weeks as short-run horizons and 26 and 52 weeks as long-run horizons.

Our sample covers the period between November 24, 1989 and July 3, 2020.[9] Our out-of-sample real-time exercise is executed with the help of the most recent 820 weeks. Correspondingly, the first training set to estimate the MS models end on October 15, 2004. Starting from this date, we employ a recursive scheme 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. Table A1 in Appendix A lists all variables, alongside their definitions and sources.

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 ex post evaluation of the real-time forecasts and we illustrate the aggregation performance of the various PCA techniques assuming knowledge of the full sample. To conserve space, we do not present any in-sample estimations of stock market regimes (and returns).

4.1 Identification 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. Chauvet and Potter (2000) emphasize that stock market "recessions" occur more frequently than economic recessions and that an economic recession is always accompanied by a stock market contraction. In the end, the approaches to identify regimes in stock markets are adopted from the business cycle literature where parametric MS models (Hamilton 1989, 2003) and dating rules are utilized (Harding and Pagan 2003). The focus of our paper lies on the evaluation of real-time out-of-sample forecasts. Hence, we follow the literature and use dating rules (Kole and van Dijk 2017) for an ex post 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

[9] 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 bonds yields, credit spreads, and sentiment indicators.

thus allows for an intuitive 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:

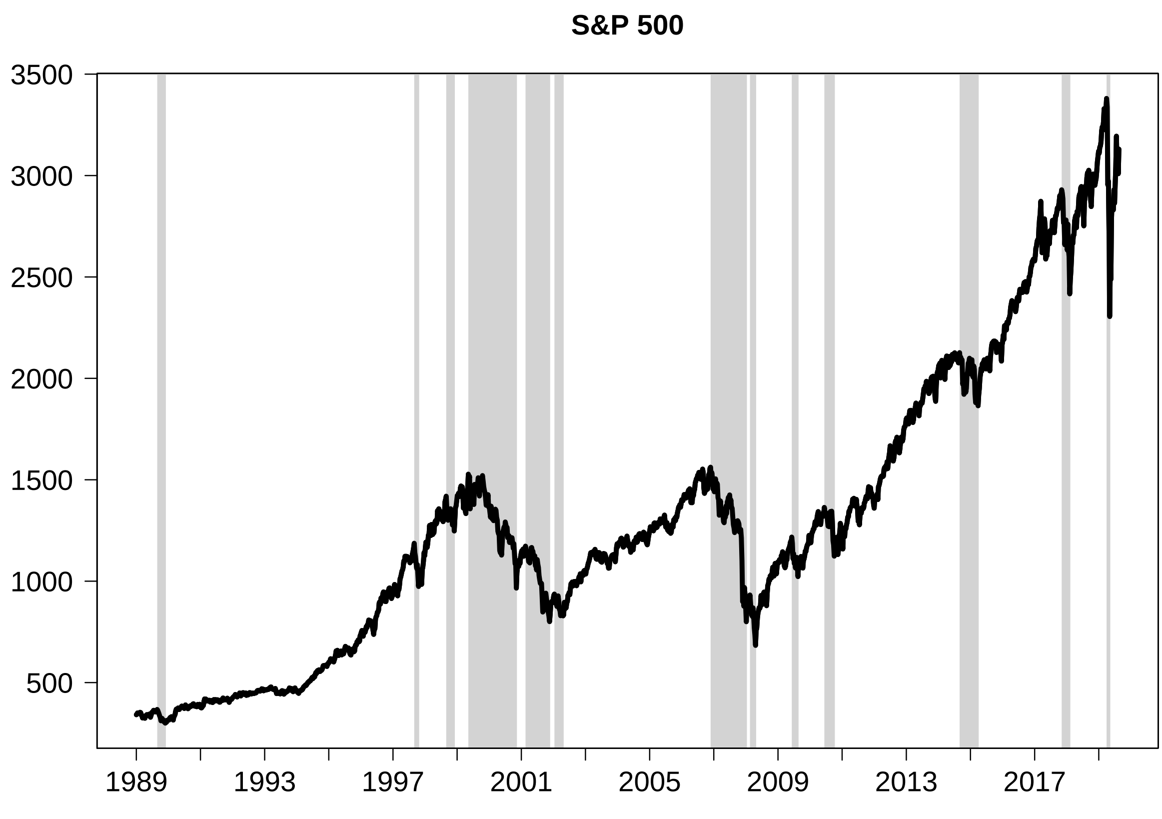
a) The trough is updated if the stock market has dropped below the last minimum.

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: Full-Sample Bull and Bear Market Identification



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

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 2020-07-03	15	36			

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.

Figure 2 and Table 2 shows the performance of the S&P 500 within corresponding 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.

During our evaluation period, the four economic recessions (according to the NBER definition) are always accompanied by a stock market contraction.[10] Despite the fact that the duration and the amplitude of bear markets varies 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 ex post proxy to evaluate the accuracy of the real-time predictions.

4.2 Data Aggregation

To utilize the information from a high-dimensional dataset of potential predictors, we apply different PCA techniques to aggregate the information into a few diffusion indices and to filter out the noise. An interpretation of the factors is only possible if

[10]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 current one (at the time of this writing) exists since March 2020 (`https://fred.stlouisfed.org/series/USRECD`).

the loadings are concentrated on variables with similar content. Therefore, we focus on the interpretation of the sparse PCA without and with soft thresholding (based on a preselection of 60 targeted predictors). The results are shown in Appendix B.

Based on the total dataset, five components seem to be sufficient for the conventional PCA and the sparse PCA (see Figure B1). These capture 60% (PCA) and 53% (sparse PCA) of the total variation (see Table B1). Considering the factor loadings, the benefit of sparse PCA in terms of a straightforward interpretation becomes evident in Table B2. The first factor can be interpreted as business cycle measure with high exposure to the leading indicator of the Conference Board, consumer confidence, the PMI, as well as expected and actual industrial production. The second factor captures information from the yield curve and the third represents industry returns and valuation ratios. The fourth factor summarizes technical indicators and the final factor corresponds to the negative slope of the yield curve (TS_5Y3M and TS_10Y3M).[11]

Turning to the targeted subset of 60 predictors, four components are selected (see also Figure B1). If the VIX is used as target, the captured variance is 54% (targeted PCA) and 49% (targeted sparse PCA), respectively. The first component of the targeted sparse PCA mainly captures historical averages and technical indicators. The second factor emphasizes corporate bonds yields and (expected) interest rates. Earnings yield and the industrial production form the third factor and some industry returns the final one. If the stock return is used as the target variable, the PCA captures 45% (targeted PCA) and 40% (targeted sparse PCA) of the data's variation. The first factor of the targeted sparse PCA summarizes technical indicators and the second the payout ratio and earnings yield. The dividend yield and the turbulence index form the third factor and in the final factor the credit spread is pooled together with the long-term term spread (TS_30Y10Y).

Figures B2–B4 show the diffusion indices over time. It is noticeable that the sparse PCA (right panel) achieves a more distinct smoothing over the indices compared to the ordinary PCA (left panel), irrespective of whether the set of predictors is unrestricted (Figure B2) or targeted (Figures B3 and B4). Hence, we can conclude that the sparse factors are more capable to filter out the noise, confirming the results of Rapach and Zhou (2019).

One caveat has to be emphasized. The preceding discussion of the different PCA approaches and the interpretation of the different factors is based on knowledge of the full sample. Hence, the performance and the interpretation might be different when considering a training set that only covers a part of the sample. Nevertheless, we

[11] To conserve space, we only show the loadings of the sparse PCAs in Tables B2–B4. All omitted results are available on request. Is has to be noted that the general structure of the first four components is similar to the conventional PCA, although the interpretation of the individual loading is clearer in the case of the sparse PCA. In contrast to the sparse PCA, the fifth factor of the conventional PCA corresponds to the dispersion of macroeconomic expectations.

will use the same amount of factors for the creation of diffusion indices for all out-of-sample exercises, that is, five when we consider the full dataset of 115 macro-financial variables and four when we target the PCA on the top-60 predictors.

5 Out-of-Sample Results

We use a recursive forecasting procedure to capture the stock market dynamics from October 22, 2004 to July 3, 2020, yielding a total of 820 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. To handle the trade-off between estimation speed and accuracy, we expand the estimation window of the training set every 13 weeks, which corresponds to a quarterly update.[12] In weeks during which the training set is not renewed, the investor updates the regime probabilities using the filter proposed by Hamilton (1989) to make one-step ahead forecasts.

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 (iShares 1Y-3Y Treasury Bond ETF, Code: SHY). Hereby, we resort to actually traded products to enable an assessment from an investor's perspective. We implement a threshold-based rebalancing strategy. If an asset weight deviates from the target weight by 10%, an adjustment of the portfolio follows according to the forecast-based asset allocation. This ensures an appropriate evaluation of the economic value of the model-based strategy and precludes an "excessive" number of transactions. We do not allow short-selling and leverage, so that the portfolio weights are always between $[0,1]$. Finally, we consider transaction costs of 10 basis points per trade for a realistic evaluation.[13] All evaluation measures for the statistical quality and the economic performance are explained in detail 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

[12] Note that, for instance, Kole and van Dijk (2017) renew their training set every 52 weeks.

[13] Two trades are necessary (sell one asset and buy the other) at each transaction date. Therefore, we have transaction costs of 20 basis points per position change. Typically, transaction costs are difficult to estimate, but our assumption corresponds to the middle case of Çakmaklı and van Dijk (2016).

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). 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}:

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

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.[14]

The out-of-sample results in Table 3 show that our modeling approach is able to distinguish between stock market regimes. Regarding the quadratic probability score (QPS) and the AUC, all proposed models can beat the MA rule and the MS model with TCTP. The total accuracy is usually higher than 80% with a bear market classification rate between 70% and 75%.

We find a clear benefit of factor augmentation as compared to the specifications using only observable predictors (*A-OBS-...*) when considering the AUC for *Model A*, where we have predictors in the mean equation and the switching equation. This, however, is not replicated for *Model B*, where we just have predictors in the switching equation, except for those models where we use BMA as forecast combination technique. Considering the different forecast combination approaches, we can conclude that Bayesian model averaging works quite good for *Model B* (with the exception of *B-OBS-BMA*). Using *Model A*, we do not find a “best” combination scheme. The high-

[14]We also provide results with window lengths of 3, 6, 18, and 36 months as part of our robustness tests in Tables E1 and E2 of Appendix E.

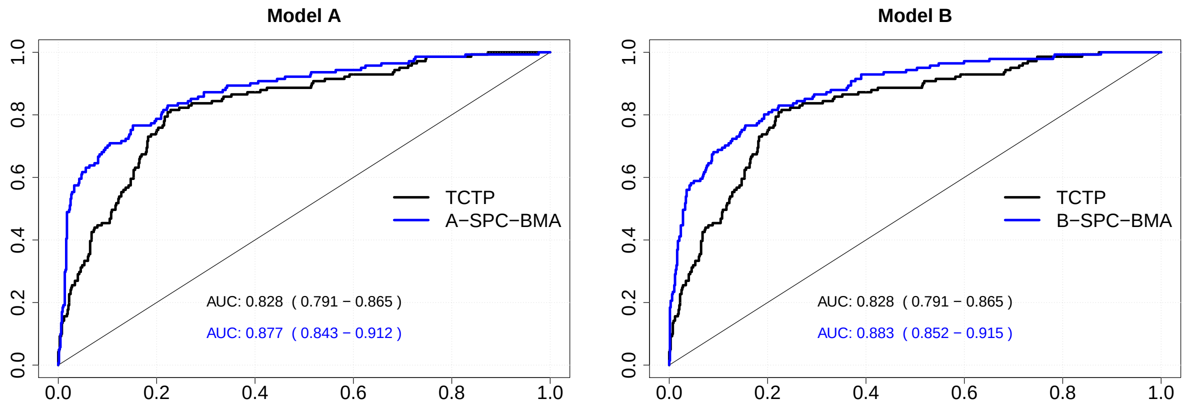
est accuracy over all models delivers *B-SPC-BMA* with an AUC of 0.8834. This is also evident in the ROC curve (see Figure 3).

Table 3: Regime Forecasts: Statistical Performance

Specification	Cluster	QPS	AUC	Accuracy	Bear	Bull
MA_12M (D_t^{MA})		0.3439		0.8183	0.5177	0.8807
TCTP		0.3493	0.8278	0.7854	0.7589	0.7909
Model A	A-OBS-AVE	0.2483	0.8514	0.8268	0.7376	0.8454
	A-PC-AVE	0.2494	0.8646	0.8220	0.7305	0.8409
	A-SPC-AVE	0.2399	0.8631	0.8305	0.7305	0.8513
	A-TPC-AVE	0.2613	0.8568	0.8073	0.7447	0.8203
	A-TSPC-AVE	0.2419	0.8690	0.8207	0.7234	0.8409
	A-OBS-IVR	0.2542	0.8599	0.8012	0.7376	0.8144
	A-PC-IVR	0.2392	0.8624	0.8256	0.7092	0.8498
	A-SPC-IVR	0.2249	0.8716	0.8354	0.7305	0.8571
	A-TPC-IVR	0.2510	0.8656	0.8098	0.7447	0.8233
	A-TSPC-IVR	0.2469	0.8635	0.8146	0.7092	0.8365
	A-OBS-BMA	0.2931	0.8413	0.7780	0.7092	0.7923
	A-PC-BMA	0.2488	0.8537	0.8293	0.7021	0.8557
	A-SPC-BMA	0.2205	0.8774	0.8390	0.7234	0.8630
	A-TPC-BMA	0.2593	0.8661	0.8159	0.7447	0.8306
	A-TSPC-BMA	0.2783	0.8527	0.8085	0.7163	0.8277
Model B	B-OBS-AVE	0.2648	0.8631	0.8037	0.7305	0.8189
	B-PC-AVE	0.2825	0.8436	0.8061	0.7234	0.8233
	B-SPC-AVE	0.2670	0.8574	0.8073	0.7447	0.8203
	B-TPC-AVE	0.2960	0.8266	0.7976	0.7376	0.8100
	B-TSPC-AVE	0.2861	0.8404	0.7976	0.7305	0.8115
	B-OBS-IVR	0.2658	0.8556	0.8073	0.7021	0.8292
	B-PC-IVR	0.2577	0.8579	0.8183	0.7163	0.8395
	B-SPC-IVR	0.2396	0.8735	0.8293	0.7305	0.8498
	B-TPC-IVR	0.2814	0.8398	0.8012	0.7092	0.8203
	B-TSPC-IVR	0.2711	0.8498	0.8012	0.7163	0.8189
	B-OBS-BMA	0.2900	0.8328	0.8098	0.5887	0.8557
	B-PC-BMA	0.2303	0.8791	0.8402	0.7305	0.8630
	B-SPC-BMA	0.2179	0.8834	0.8451	0.7163	0.8719
	B-TPC-BMA	0.2672	0.8454	0.8085	0.6312	0.8454
	B-TSPC-BMA	0.2611	0.8587	0.8207	0.7092	0.8439

Notes: All models except D_t^{MA} (naïve 12-month moving average) and $TCTP$ (MS model with TCTP) are estimated as MS-TVTP models. Model A contains predictors in the switching equation and the conditional mean equation according to (5), Model B contains predictors in the switching equation only according to (6). OBS: observable predictors; PC: ordinary PCA; SPC: sparse PCA; TPC: ordinary PCA with soft thresholding; TSPC: sparse PCA with soft thresholding; AVE: simple average; IVR: inverse rank; 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.

Figure 3: Regime Forecasts: ROC Curve



Notes: Figure shows the ROC curve of regime forecasts. The y-axis represents the true positive rate (correctly predicted bear markets) and the x-axis the false positive rate (false alarms of bear markets). For each specification choice, the model with the best AUC is displayed. For comparison, we additionally report the classification performance of the MS-TCTP model. Confidence intervals according to DeLong et al. (1988) are given in parentheses. See Appendix C for a detailed description of the evaluation measures.

Overall, it should be noted that the predicted bear market probabilities respond promptly to regime turning points. Figures D1–D5 in Appendix D also show that the respective regime forecasts have a high degree of similarity across the different forecast combination clusters. As Table D1 demonstrates, a key advantage of our approach over the TCTP model is to identify the turning point from bear to bull markets in a timely manner. The delay of the change into a bear market, however, is comparable to the benchmark. As an illustration, the best models can identify the start and end of the GFC with a delay of one week. The Covid-19 crash is also classified as a bear market from end of February 2020 onward, with the re-entry taking place in mid-April.

Economic Value: 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.

Starting with the performance of our benchmarks, it is worth noting that the simple MS model with TCTP substantially outperforms all others on a risk-adjusted basis (see Table 4). The advantage of using regime probabilities as a timing tool compared to a buy-and-hold strategy (S&P 500), an equally-weighted mixed portfolio (50/50), or a naïve MA rule is obvious when considering the Sharpe ratio and the certainty equivalent return. In most cases, our approach yields higher returns, lower risk, and large utility gains. The maximum drawdown is 12.25% for most of the models as compared to a loss of more than 50% in the broad stock market.

Table 4: Regime Forecast: Economic Value

	R^{cum}	$\bar{R}$	$\bar{\sigma}$	SR	CER	MaxDD	VaR	CVaR
S&P 500	3.86	8.94	18.00	0.37	5.35	54.61	–3.90	–6.16
Rf	1.39	2.10	1.20	0.00	2.06	1.90	–0.19	–0.31
50/50	2.51	6.02	9.12	0.42	5.02	28.76	–2.03	–3.09
MA_12M (D_t^{MA})	2.68	6.45	12.27	0.34	4.75	24.46	–2.66	–4.50
TCTP	3.67	8.60	10.05	0.63	7.25	14.53	–2.11	–3.50
A-OBS-AVE	5.58	11.51	10.81	0.85	9.74	12.25	–2.13	–3.56
A-PC-AVE	5.31	11.17	10.61	0.83	9.47	12.25	–2.11	–3.51
A-SPC-AVE	4.82	10.49	10.73	0.76	8.84	12.25	–2.14	–3.56
A-TPC-AVE	5.03	10.78	10.46	0.81	9.16	12.25	–2.11	–3.50
A-TSPC-AVE	4.84	10.51	10.74	0.76	8.85	12.25	–2.14	–3.61
A-OBS-IVR	3.91	9.04	10.34	0.65	7.60	12.25	–2.13	–3.60
A-PC-IVR	4.24	9.60	10.99	0.66	7.97	12.25	–2.22	–3.76
A-SPC-IVR	5.73	11.71	10.86	0.86	9.91	12.25	–2.13	–3.54
A-TPC-IVR	4.38	9.82	10.50	0.72	8.28	12.25	–2.13	–3.56
A-TSPC-IVR	5.23	11.06	10.61	0.82	9.38	12.25	–2.11	–3.58
A-OBS-BMA	2.99	7.19	10.23	0.48	5.90	14.71	–2.14	–3.67
A-PC-BMA	5.37	11.25	11.66	0.76	9.31	14.46	–2.22	–3.81
A-SPC-BMA	5.22	11.04	11.06	0.79	9.27	12.25	–2.17	–3.65
A-TPC-BMA	4.67	10.27	11.08	0.72	8.56	17.70	–2.14	–3.69
A-TSPC-BMA	5.82	11.82	10.59	0.90	10.06	12.25	–2.11	–3.53
B-OBS-AVE	4.24	9.60	10.25	0.71	8.13	12.25	–2.11	–3.56
B-PC-AVE	4.23	9.58	10.48	0.69	8.06	12.25	–2.13	–3.55
B-SPC-AVE	4.60	10.16	10.24	0.77	8.64	12.25	–2.11	–3.50
B-TPC-AVE	3.89	9.00	10.19	0.66	7.59	12.25	–2.11	–3.53
B-TSPC-AVE	4.29	9.68	10.25	0.72	8.20	12.25	–2.11	–3.54
B-OBS-IVR	3.52	8.31	10.48	0.58	6.89	20.17	–2.22	–3.76
B-PC-IVR	4.83	10.51	10.71	0.76	8.86	12.25	–2.13	–3.55
B-SPC-IVR	5.18	10.99	10.74	0.81	9.29	12.25	–2.13	–3.56
B-TPC-IVR	3.81	8.85	10.29	0.64	7.43	12.25	–2.14	–3.58
B-TSPC-IVR	4.68	10.28	10.37	0.77	8.73	12.25	–2.11	–3.53
B-OBS-BMA	3.05	7.33	10.93	0.47	5.89	18.58	–2.42	–3.89
B-PC-BMA	5.14	10.94	10.95	0.79	9.20	12.25	–2.16	–3.59
B-SPC-BMA	5.44	11.34	11.13	0.81	9.51	12.25	–2.21	–3.65
B-TPC-BMA	3.36	7.99	10.84	0.53	6.52	15.26	–2.40	–3.82
B-TSPC-BMA	4.44	9.92	10.70	0.71	8.32	17.98	–2.22	–3.66

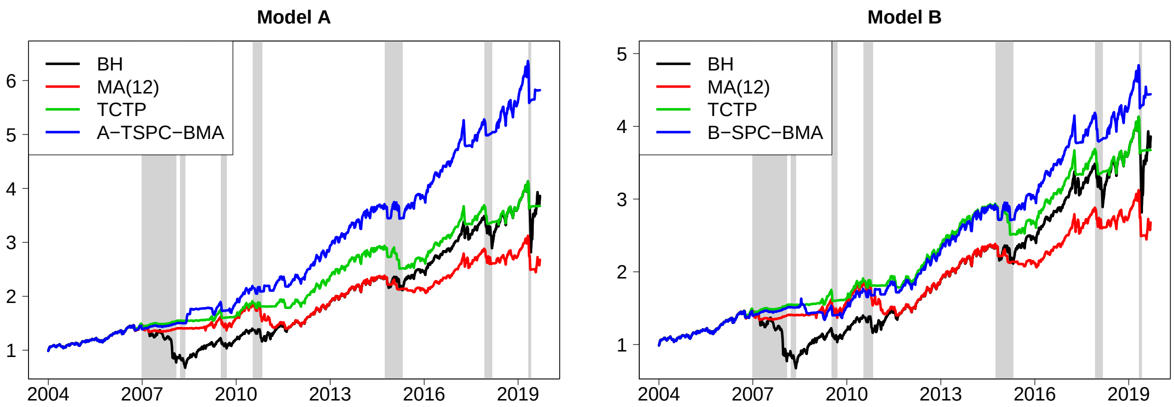
Notes: Table shows the economic value of different investment strategies. S&P 500: stocks only; Rf: short-term government bonds only; 50/50: mixed strategy; D_t^{MA}: naïve 12-month moving average; $TCTP$: MS model with TCTP. Model A contains predictors in the switching equation and the conditional mean equation, Model 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; IVR: inverse rank; 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; MaxDD: maximum drawdown; VaR: value-at-risk; CVaR: conditional value-at-risk. See Appendix C for a detailed description of the evaluation measures.

On average, a simultaneous modeling of the conditional mean (*Model A*) improves the economic performance as compared to *Model B*. Considering the different forecasting combination schemes, we observe that the BMA slightly increases the volatility, but allows for a higher upside potential. Hence, it is not surprising that the model with the highest utility gain (certainty equivalent return, CER, of 10.06% p.a.) uses the BMA as weighting scheme (*A-TSPC-BMA*). However, we cannot find a globally dominating weighting scheme. Consequently, a simple equal-weighting of all model seems to be a good choice supported by the highest robustness (smallest variation) and its simplicity.

Looking at the quality of the predictors, diffusion indices seem to outperform the standard set of observable predictors (with the exception of *A-OBS-AVE*), confirming the findings of Neely et al. (2014) and Çakmaklı and van Dijk (2016). Within the factor-augmented 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, a pre-selection of variables based on soft thresholding does not generally lead to a significant improvement, even if the best model is based on the targeted set of predictors.[15]

The ability to detect turning points in a timely manner is also documented in the cumulative return plots (see Figure 4). For both extreme events (GFC and Covid-19 shock) as well as for the remaining four contraction periods, losses can be reduced and re-entry points can be found in a timely manner. All these findings suggest that it pays off to model the switching process with TVTP.

Figure 4: Regime Forecasts: Performance over Time



Notes: Figure shows the cumulative performance of the best models according to each model specification compared to several benchmarks. All strategies follow a binary strategy, which invest the total wealth either in the stock market or in the risk-free proxy. The CER is used as ranking criteria to select the MS-TVTP model with the best performance. See Appendix C for a detailed description of the evaluation measures.

[15]Tables E3 and E4 in Appendix E show results for a less risk-averse agent (threshold of 25%) and a more risk-averse agent (threshold of 75%), respectively. 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% provides mixed results. The volatility increases and this can be overcompensated by higher profits only in some cases.

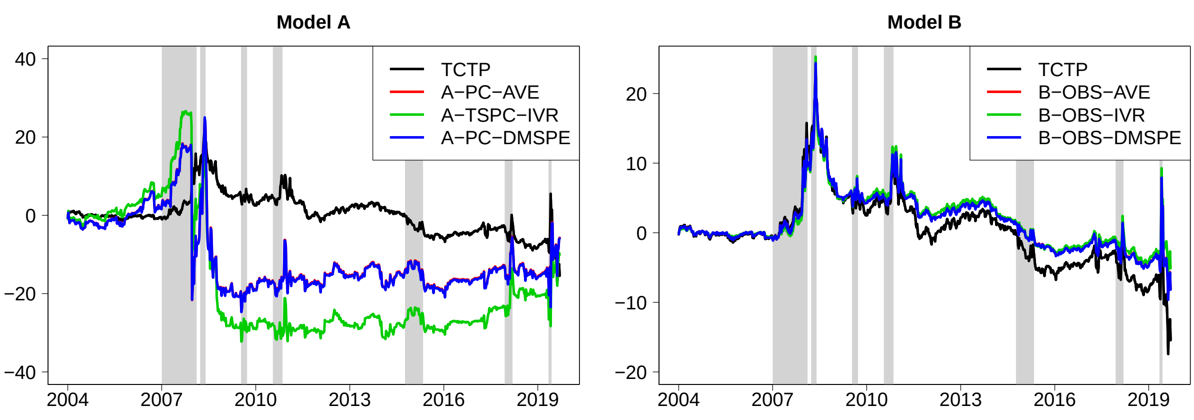
A final supporting factor of the economic value of our results is the limited number of transactions. As noted by Table D2 in Appendix D, around 50 position changes are required over the 820 weeks, which corresponds to three to four position changes per year.

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. 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 24, 1989 to October 15, 2004 is used for the first forecast.

Table 5 shows that there is no added value with regard to the R^2_{OS} and $RMSE$ over the entire out-of-sample period. The values for R^2_{OS} are negative and the null hypothesis of $R^2_{OS} \leq 0$ cannot be rejected. However, when comparing the difference in the MSPE between the forecasts and the historical mean over time, it is noticeable the forecasting models are valuable during contractions (see Figure 5 for a plot of the cumulative difference squared forecast error, CDSFE). The empirical observation that returns are more predictable in economic recessions or during turbulences has been highlighted in the literature (Henkel et al. 2011). Nevertheless, the TVTP models cannot systematically beat the simple MS model with TCTP when forecasting returns. Finally, the sign predictability varies between 55% and 57%, whereby positive returns are correctly predicted in 76% to 84% of the cases and the negative return accuracy rate ranges from 20% to 31%.

Figure 5: Return Forecasts: Cumulative Difference Squared Forecast Error



Notes: Figure shows the CDSFE (in %) of the best TVTP models for the different aggregation techniques and the TCTP model. The reference model for calculating the CDSFE is the specification using the historical mean. See Appendix C for a detailed description of the evaluation measures.

Table 5: Return Forecasts: Statistical Performance

Specification	Cluster	RMSE	CW R^2_{OS}	p-val.	Direction	R^+	R^-
Hist_Mean		2.4995			0.5768	1	0
TCTP		2.5032	–0.3013	0.4989	0.5476	0.7822	0.2277
Model A	A-OBS-AVE	2.5221	–1.8207	0.5061	0.5610	0.7611	0.2882
	A-PC-AVE	2.5009	–0.1117	0.4865	0.5683	0.7844	0.2738
	A-SPC-AVE	2.5012	–0.1400	0.4873	0.5671	0.7780	0.2795
	A-TPC-AVE	2.5054	–0.4738	0.4934	0.5634	0.7738	0.2767
	A-TSPC-AVE	2.5015	–0.1645	0.4887	0.5549	0.7696	0.2622
	A-OBS-IVR	2.5205	–1.6924	0.5043	0.5695	0.7611	0.3084
	A-PC-IVR	2.5028	–0.2681	0.4874	0.5695	0.7822	0.2795
	A-SPC-IVR	2.5036	–0.3286	0.4896	0.5622	0.7717	0.2767
	A-TPC-IVR	2.5074	–0.6399	0.4946	0.5707	0.7801	0.2853
	A-TSPC-IVR	2.5019	–0.1963	0.4891	0.5707	0.7801	0.2853
	A-OBS-DMSPE	2.5198	–1.6341	0.5054	0.5585	0.7590	0.2853
	A-PC-DMSPE	2.5009	–0.1155	0.4865	0.5646	0.7801	0.2709
	A-SPC-DMSPE	2.5013	–0.1462	0.4874	0.5671	0.7780	0.2795
	A-TPC-DMSPE	2.5054	–0.4723	0.4934	0.5634	0.7738	0.2767
	A-TSPC-DMSPE	2.5016	–0.1690	0.4888	0.5573	0.7696	0.2680
Model B	B-OBS-AVE	2.5014	–0.1580	0.4980	0.5610	0.8266	0.1988
	B-PC-AVE	2.5027	–0.2574	0.4979	0.5549	0.8076	0.2104
	B-SPC-AVE	2.5027	–0.2601	0.4984	0.5537	0.8034	0.2133
	B-TPC-AVE	2.5039	–0.3585	0.5013	0.5524	0.8013	0.2133
	B-TSPC-AVE	2.5033	–0.3067	0.4998	0.5537	0.8055	0.2104
	B-OBS-IVR	2.5007	–0.0998	0.4965	0.5646	0.8351	0.1960
	B-PC-IVR	2.5026	–0.2520	0.4986	0.5549	0.8161	0.1988
	B-SPC-IVR	2.5022	–0.2214	0.4975	0.5598	0.8161	0.2104
	B-TPC-IVR	2.5039	–0.3595	0.5011	0.5512	0.7992	0.2133
	B-TSPC-IVR	2.5033	–0.3054	0.5003	0.5585	0.8118	0.2133
	B-OBS-DMSPE	2.5015	–0.1599	0.4980	0.5610	0.8288	0.1960
	B-PC-DMSPE	2.5027	–0.2587	0.4979	0.5561	0.8076	0.2133
	B-SPC-DMSPE	2.5027	–0.2610	0.4984	0.5537	0.8034	0.2133
	B-TPC-DMSPE	2.5039	–0.3589	0.5013	0.5524	0.8013	0.2133
	B-TSPC-DMSPE	2.5033	–0.3071	0.4998	0.5537	0.8055	0.2104

Notes: Hist_Mean: historical mean; $TCTP$: MS model with TCTP. Model A contains predictors in the switching equation and the conditional mean equation, Model 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; IVR: inverse rank; DMSPE: discounted mean squared prediction error; RMSE: root mean squared error; 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.

Economic Value: To evaluate the economic value of the stock return forecasts, we assume a risk-averse agent with mean-variance preferences. Solving the standard expected utility maximization, we obtain the following optimal stock market weight:

$$w_t^* = \frac{1}{\gamma} \frac{\hat{r}_{t+1}}{\hat{\sigma}_{t+1}^2} \tag{11}$$

$\hat{r}_{t+1}$ 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.[16] Allowing for a tolerance level of 10%, the investor dynamically updates the return forecasts and the stock allocation.

The summary statistics in Table 6 and the cumulative performance in Figure 6 show that the buy-and-hold strategy and the 50/50 benchmark are superior regarding the SR and the CER over all other approaches, although the tail risk as represented by the maximum drawdown can be substantially reduced with the forecasts. Nevertheless, we can conclude that the predictions of the more parsimonious specification (*Model B*) are able to produce utility gains. These models are very similar to the TCTP case, since the conditional mean is not directly modeled. Alternatively, these models can roughly be described as probability-weighted version of the historical mean. The annualized CER gain of these strategies, defined as $CER^{Model} - CER^{Hist\_Mean}$, is between 0.35% (*B-TPC-IVR*) and 1.37% (*B-PC-AVE* and *B-PC-DMPSE*). This corresponds to a maximum management fee that an investor would accept for participating in the respective investment strategy (see also Appendix C).[17]

The significant underperformance of *Model A* comes from the fact that additional variables in the conditional mean produce noisy forecasts. This leads to many transactions and drastically reduces the profitability after accounting for transaction costs (see Table D2 in Appendix D). If we compare the different combination approaches, no pooling method seems to be superior in general. This again favors the simple equal-weighting or, if at all, the slightly modified DMSPE rule. Although the use of the aggregated factors instead of directly observable variables is not as evidently superior as in the regime forecasting case, it still provides some added value. Finally, we can conclude that following the return forecasts is less profitable than following the regime forecasts.

[16] Although the assumption of $\gamma = 3$ is common in the empirical literature, we provide results for $\gamma = 1, 2, 5$ in Tables E5–E7 of Appendix E. 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.

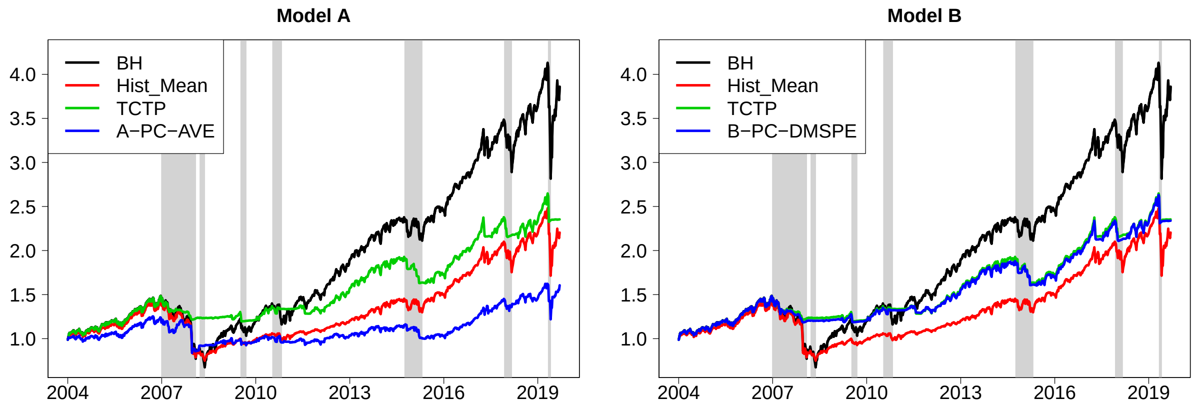
[17] For a less risk-averse investor ($\gamma = 1$ or $\gamma = 2$), the utility gain vanishes, whereas for more risk-averse agents ($\gamma = 5$), the maximum accepted management fee increases to 2%. See Tables E5–E7 in Appendix E.

Table 6: Return Forecast: Economic Value

	R^{cum}	$\bar{R}$	$\bar{\sigma}$	SR	CER	MaxDD	VaR	CVaR
S&P 500	3.86	8.94	18.00	0.37	5.35	54.61	–3.90	–6.16
Rf	1.39	2.10	1.20	0.00	2.06	1.90	–0.19	–0.31
50/50	2.51	6.02	9.12	0.42	5.02	28.76	–2.03	–3.09
Hist_Mean	2.20	5.14	14.05	0.21	3.07	47.77	–2.84	–5.07
TCTP	2.35	5.58	9.74	0.35	4.49	19.36	–2.11	–3.46
A-OBS-AVE	1.13	0.76	13.28	–0.10	–0.98	42.43	–2.32	–5.03
A-PC-AVE	1.61	3.05	13.24	0.07	1.28	33.18	–2.35	–4.83
A-SPC-AVE	1.45	2.36	12.94	0.02	0.69	31.27	–2.33	–4.69
A-TPC-AVE	1.40	2.16	13.66	0.00	0.29	39.19	–2.34	–4.96
A-TSPC-AVE	1.38	2.08	13.26	–0.01	0.33	32.93	–2.31	–4.81
A-OBS-IVR	1.07	0.42	13.20	–0.13	–1.29	43.24	–2.32	–4.93
A-PC-IVR	1.40	2.17	13.14	0.00	0.45	34.53	–2.33	–4.93
A-SPC-IVR	1.28	1.58	13.24	–0.04	–0.15	38.05	–2.26	–4.90
A-TPC-IVR	1.36	1.98	13.15	–0.01	0.26	36.22	–2.34	–4.78
A-TSPC-IVR	1.33	1.85	13.14	–0.02	0.13	42.00	–2.21	–4.84
A-OBS-DMSPE	1.02	0.10	13.39	–0.15	–1.66	46.96	–2.33	–5.13
A-PC-DMSPE	1.59	2.98	13.24	0.06	1.22	33.24	–2.35	–4.83
A-SPC-DMSPE	1.43	2.31	12.92	0.01	0.64	31.27	–2.33	–4.69
A-TPC-DMSPE	1.37	2.04	13.57	–0.01	0.20	39.93	–2.34	–4.96
A-TSPC-DMSPE	1.42	2.25	13.25	0.01	0.50	32.10	–2.31	–4.79
B-OBS-AVE	2.21	5.15	9.48	0.31	4.13	23.67	–2.01	–3.42
B-PC-AVE	2.34	5.53	9.77	0.34	4.44	20.19	–2.01	–3.51
B-SPC-AVE	2.32	5.49	9.68	0.34	4.41	21.03	–2.01	–3.46
B-TPC-AVE	2.05	4.65	9.94	0.25	3.56	27.90	–2.15	–3.58
B-TSPC-AVE	2.29	5.40	9.71	0.33	4.33	23.05	–2.05	–3.50
B-OBS-IVR	2.22	5.18	9.73	0.30	4.11	25.60	–2.02	–3.54
B-PC-IVR	2.27	5.35	9.51	0.33	4.31	19.34	–1.94	–3.43
B-SPC-IVR	2.30	5.43	9.77	0.33	4.34	23.26	–2.02	–3.50
B-TPC-IVR	2.00	4.48	9.85	0.23	3.42	28.20	–2.13	–3.56
B-TSPC-IVR	2.24	5.26	9.72	0.31	4.18	25.26	–2.07	–3.50
B-OBS-DMSPE	2.23	5.21	9.49	0.32	4.19	23.62	–2.01	–3.42
B-PC-DMSPE	2.34	5.53	9.76	0.34	4.44	20.15	–2.01	–3.51
B-SPC-DMSPE	2.32	5.50	9.68	0.34	4.42	20.98	–2.01	–3.46
B-TPC-DMSPE	2.05	4.67	9.93	0.25	3.58	27.77	–2.15	–3.58
B-TSPC-DMSPE	2.31	5.46	9.69	0.33	4.38	22.96	–2.05	–3.49

Notes: Table shows the economic value of different investment strategies. S&P 500: stocks only; Rf: short-term government bonds only; 50/50: mixed strategy; Hist_Mean: historical mean; $TCTP$: MS model with TCTP. Model A contains predictors in the switching equation and the conditional mean equation, Model 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; IVR: inverse rank; DMSPE: discounted mean squared prediction error; 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; MaxDD: maximum drawdown; VaR: value-at-risk; CVaR: conditional value-at-risk. See Appendix C for a detailed description of the evaluation measures.

Figure 6: Return Forecasts: Performance over Time



Notes: Figure shows the cumulative performance of the best models according to each model specification compared to several benchmarks. All strategies follow an expected utility optimization to dynamically determine the weight of the risky asset. The CER is used as ranking criteria to select the MS-TVTP model with the best performance. See Appendix C for a detailed description of the evaluation measures.

6 Conclusions

Using a high-dimensional dataset of macro-financial variables, this paper offers a promising outlook on the predictability of 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 diffusion indices, regime-switching models, and forecast combination. We provide a comprehensive overview of the empirical usefulness of factor-augmented MS models with time-varying transition probabilities.

Regime forecasts using a (targeted) sparse PCA MS-TVTP model are suitable to respond to trend changes in a timely manner, while increasing the utility of investors. In addition, these are able to classify bull and bear markets more accurately than common benchmarks. Our results are robust with respect to different aggregation methods or alternative forecast weights. Nevertheless, we propose to enhance the conventional PCA with shrinkage methods and to use Bayesian model averaging in this context. However, when considering stock market return forecasts, we have to conclude that the benefits of the factor augmented MS-TVTP regime prediction models cannot be translated into a significant outperformance relative to the historical average of stock returns.

Our results offer a variety of starting points for future work. First, modeling intra-regime dynamics in greater detail in the context of factor augmented MS-TVTP models could be a promising extension — if practically feasible. Thereby, incorporating more than two regimes directly (Maheu et al. 2012) or a sequential partition are potential avenues (Hauptmann et al. 2014). 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 or density forecasts. In addition, one could study international stock market indices or portfolios formed on industries or styles with the help of a factor augmented MS-TVTP 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.

Batchelor, R. (2001). How useful are the forecasts of intergovernmental agencies? The IMF and OECD versus the Consensus. *Applied Economics*, 33(2):225–235.

Bates, J. M. and Granger, C. W. (1969). The combination of forecasts. *Journal of the Operational Research Society*, 20(4):451–468.

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.

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 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 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. (2003). Comment on "A comparison of two business cycle dating methods". *Journal of Economic Dynamics and Control*, 27(9):1691–1693.

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

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.

Harding, D. and Pagan, A. (2003). A comparison of two business cycle dating methods. *Journal of Economic Dynamics and Control*, 27(9):1681–1690.

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' indexforetell stock returns? An additional forward-looking sentiment indicator. *Journal of Investing*, 20(4):89–98.

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

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.

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. and Rapach, D. E. (2008). Real interest rate persistence: Evidence and implications. *Federal Reserve Bank of St. Louis Review*, 90(6):609–641.

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.

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.

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

Stock, J. H. and Watson, M. W. (2002). Macroeconomic forecasting using diffusion indexes. *Journal of Business and Economic Statistics*, 20(2):147–162.

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: Data Description and Sources

Variable Description	Abbreviation	Source
Log Returns of S&P 500 (Adj. Close Price)	R	Yahoo Finance
5-Week Simple Moving Average of S&P 500 Returns	SMA_5	Yahoo Finance
52-Week Simple Moving Average of S&P 500 Returns	SMA_52	Yahoo Finance
5-Week Exponential Moving Average of S&P 500 Returns	EMA_5	Yahoo Finance
52-Week Exponential Moving Average of S&P 500 Returns	EMA_52	Yahoo Finance
Log of Detrended Weekly Trading Volume of S&P 500	Volume	Yahoo Finance
22-Day Sum of Daily Squared Returns	RV_22	Yahoo Finance
5-Day Sum of Daily Squared Returns	RV_5	Yahoo Finance
1-Day Sum of Daily Squared Returns	RV_1	Yahoo Finance
Consumer Nondurables Industry Returns	NoDur	Kenneth French Data
Consumer Durables Industry Returns	Durbl	Kenneth French Data
Manufacturing Industry Returns	Manuf	Kenneth French Data
Energy Industry Returns	Enrgy	Kenneth French Data
Business Equipment Industry Returns	HiTec	Kenneth French Data
Telecommunication Industry Returns	Telcm	Kenneth French Data
Consumer Nondurables Industry Returns	Shops	Kenneth French Data
Health Care Industry Returns	Hlth	Kenneth French Data
Utilities Industry Returns	Utils	Kenneth French Data
Other Industry Returns	Other	Kenneth French Data
Moving Average Crossing Indicator 1–52 Weeks	MA_1_52	Yahoo Finance
Moving Average Crossing Indicator 4–52 Weeks	MA_4_52	Yahoo Finance
Moving Average Crossing Indicator 8–52 Weeks	MA_8_52	Yahoo Finance
Moving Average Crossing Indicator 1–26 Weeks	MA_1_26	Yahoo Finance
Moving Average Crossing Indicator 4–26 Weeks	MA_4_26	Yahoo Finance
Moving Average Crossing Indicator 8–26 Weeks	MA_8_26	Yahoo Finance
Momentum Indicator 26 Weeks	MOM_26	Yahoo Finance
Momentum Indicator 52 Weeks	MOM_52	Yahoo Finance
On Balance Volume Indicator 1–52 Weeks	OBV_1_52	Yahoo Finance
On Balance Volume Indicator 4–52 Weeks	OBV_4_52	Yahoo Finance
On Balance Volume Indicator 8–52 Weeks	OBV_8_52	Yahoo Finance
On Balance Volume Indicator 1–26 Weeks	OBV_1_26	Yahoo Finance
On Balance Volume Indicator 4–26 Weeks	OBV_4_26	Yahoo Finance
On Balance Volume Indicator 8–26 Weeks	OBV_8_26	Yahoo Finance
Log (Dividends / Price) (S&P 500)	DP	Robert Shiller Data
Log (Earnings / Price) (S&P 500)	EP	Robert Shiller Data
Log (Price / 10Y Average Earnings) (S&P 500)	E10P	Robert Shiller Data
Log (Dividends / Earnings) (S&P 500)	Payout	Robert Shiller Data
Implied Volatility Index for S&P 500 Options	VIX	CBOE
Variance Risk Premium	VP	CBOE and Yahoo Finance
Financial Turbulence Index	Turb_Index	Kenneth French Data
Delta of Absorption Ratio	Delta_AR	Kenneth French Data
Gold Price Returns	Gold	LBMA
Effective Federal Funds Rate	FEDR	Federal Reserve System
3M Treasury Bill	T3M	Federal Reserve System
6M Treasury Bill	T6M	Federal Reserve System
1Y Treasury Bonds	T1Y	Federal Reserve System
2Y Treasury Bonds	T2Y	Federal Reserve System
3Y Treasury Bonds	T3Y	Federal Reserve System
5Y Treasury Bonds	T5Y	Federal Reserve System
7Y Treasury Bonds	T7Y	Federal Reserve System
10Y Treasury Bonds	T10Y	Federal Reserve System
30Y Treasury Bonds	T30Y	Federal Reserve System
Term Spread 30Y and 10Y	TS_30Y10Y	Federal Reserve System
Term Spread 10Y and 2Y	TS_10Y2Y	Federal Reserve System
Term Spread 10Y and 1Y	TS_10Y1Y	Federal Reserve System
Term Spread 10Y and 3M	TS_10Y3M	Federal Reserve System
Term Spread 5Y and 3M	TS_5Y3M	Federal Reserve System
5Y 5Y Forward Inflation Expectation	T5Y_IE	Federal Reserve System
10Y Break-Even Inflation Rate	T10Y_IE	Federal Reserve System
Corporate Bonds Yield AAA	AAA	Moody's
Corporate Bonds Yield BAA	BAA	Moody's
Credit Spread AAA Corp. Bond and Gov. Bond	CS_AAA10Y	Moody's
Credit Spread BAA Corp. Bond and Gov. Bond	CS_BAA10Y	Moody's
Credit Spread BAA Corp. Bond and AAA Corp. Bond	CS_BAAAAA	Moody's
3M USD LIBOR Inter-Banking Rate	LIBOR	Federal Reserve System
3M USD LIBOR and 3M T-Bill Spread	TED	Federal Reserve System

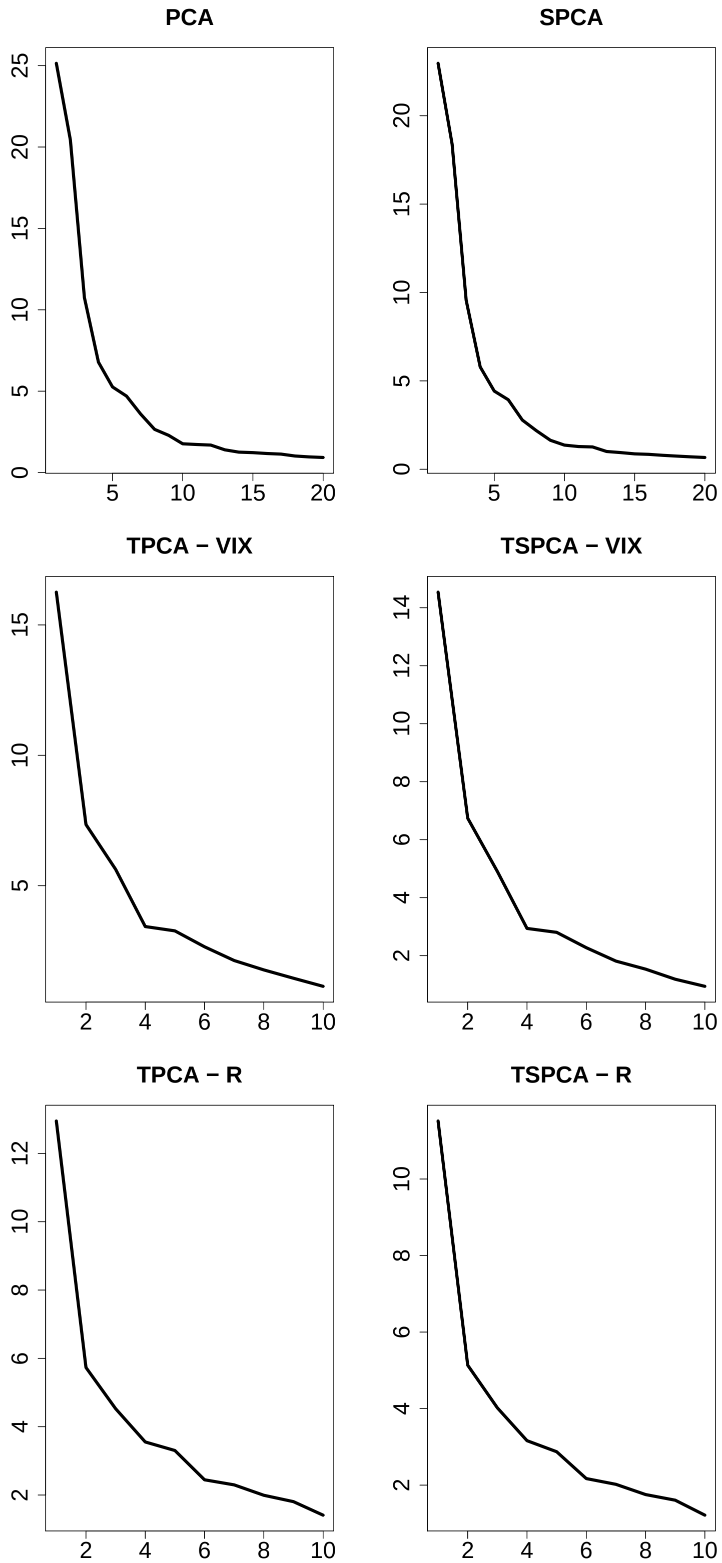
Table A1: Data Description and Sources (Continued)

Variable Description	Abbreviation	Source
GDP Mean Forecast	gdp	Consensus Economics
Investments Mean Forecast	inv	Consensus Economics
S&P 500 Profits Mean Forecast	profit	Consensus Economics
Production Mean Forecast	prod	Consensus Economics
CPI Inflation Mean Forecast	cpi	Consensus Economics
PPI Inflation Mean Forecast	ppi	Consensus Economics
Consumption Mean Forecast	cons	Consensus Economics
Employment Cost Mean Forecast	emp.cost	Consensus Economics
Car Sales Mean Forecast	csales	Consensus Economics
Housing Starts Mean Forecast	housep	Consensus Economics
Unemployment Rate Mean Forecast	unemp	Consensus Economics
Current Account Mean Forecast	ca	Consensus Economics
Fiscal Balance Mean Forecast	fb	Consensus Economics
Term Spread (in 3 Months) Mean Forecast	term.spread.exp.3m.	Consensus Economics
Term Spread (in 12 Months) Mean Forecast	term.spread.exp.12m.	Consensus Economics
10Y Int. Rate (in 3 Months) Mean Forecast	i10y.3m	Consensus Economics
10Y Int. Rate (in 12 Months) Mean Forecast	i10y.12m	Consensus Economics
3M Int. Rate (in 3 Months) Mean Forecast	i3m.3m	Consensus Economics
3M Int. Rate (in 12 Months) Mean Forecast	i3m.12m	Consensus Economics
GDP Forecast Std. Dev.	gdp.sd	Consensus Economics
Consumption Forecast Std. Dev.	cons.sd	Consensus Economics
Investment Forecast Std. Dev.	inv.sd	Consensus Economics
S&P 500 Profits Forecast Std. Dev.	profit.sd	Consensus Economics
Production Forecast Std. Dev.	prod.sd	Consensus Economics
CPI Inflation Forecast Std. Dev.	cpi.sd	Consensus Economics
PPI Inflation Forecast Std. Dev.	ppi.sd	Consensus Economics
Employment Cost Forecast Std. Dev.	emp.cost.sd	Consensus Economics
Car Sales Forecast Std. Dev.	csales.sd	Consensus Economics
Housing Starts Forecast Std. Dev.	housep.sd	Consensus Economics
Unemployment Rate Forecast Std. Dev.	unemp.sd	Consensus Economics
Current Account Forecast Std. Dev.	ca.sd	Consensus Economics
Fiscal Balance Forecast Std. Dev.	fb.sd	Consensus Economics
3M Int. Rate (in 3 Months) Forecast Std. Dev.	i3m.3m.sd	Consensus Economics
3M Int. Rate (in 12 Months) Forecast Std. Dev.	i3m.12m.sd	Consensus Economics
10Y Int. Rate (in 3 Months) Forecast Std. Dev.	i10y.3m.sd	Consensus Economics
10Y Int. Rate (in 12 Months) Forecast Std. Dev.	i10y.12m.sd	Consensus Economics
12M Expectation Fed Rate Increase	Int_exp_higher	The Conference Board
12M Expectation Fed Rate Decrease	Int_exp_lower	The Conference Board
12M Expectation Stock Price Increase	stock_exp_higher	The Conference Board
12M Expectation Stock Price Decrease	stock_exp_lower	The Conference Board
Change of Consumer Climate Survey TCB (YoY)	Cons_Cli_Conf	The Conference Board
Change of Consumer Situation Survey TCB (YoY)	Cons_Sit_Conf	The Conference Board
Consumer Expectation Survey TCB	Cons_Exp_Conf	The Conference Board
Change of Inflation Expectation Survey TCB (YoY)	Cons_Inf_Conf	The Conference Board
Change of the Leading Economic Index for the US (YoY)	Lead_Conf	The Conference Board
Purchasing Managers Index	PMI	ISM
Business Condition Philadelphia Fed	BusCondPhil	Philadelphia Fed
WTI Oil Price Returns	Oil	NY Mercantile Exchange
M1 Growth Rate	M1	Federal Reserve System
M2 Growth Rate	M2	Federal Reserve System
CPI Inflation Rate (YoY)	Inflation	Bureau of Labor Statistics
Industrial Production Growth (YoY)	IP	Federal Reserve System
Unemployment Rate (Centered by 1-Year Rolling Mean)	Unemp	Bureau of Labor Statistics

Notes: ISM: Institute for Supply Management; LBMA: London Bullion Market Association; CBOE: Chicago Board Options Exchange; NAHB: National Association of Home Builders. For five of the predictors (fb, fb.sd, T5Y_IE, T10Y_IE, 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, inflation expectations from the bond market are only available since 2003 when the US government started to issue inflation-linked bonds (TIPS). These capture real interest rate expectations and the corresponding risk premium. Assuming constant real interest rate expectations of 2% (Neely and Rapach 2008) and a risk premium of 0%, we can replace the TIPS yields with a simple proxy for market-based expectations before 2003. Finally, 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.

Appendix B: Aggregation Results

Figure B1: Scree Plots



Notes: (T)PCA: ordinary PCA (with soft thresholding); (T)SPCA: sparse PCA (with soft thresholding). Soft thresholding is either based on the future VIX or future returns.

Table B1: Proportion of Explained Variance

	Total Sample		Targeted Sample – VIX		Targeted Sample – R	
	PCA	Sparse PCA	PCA	Sparse PCA	PCA	Sparse PCA
PC1	0.2186	0.2000	0.2709	0.2420	0.2158	0.1920
PC2	0.3963	0.3600	0.3934	0.3550	0.3113	0.2770
PC3	0.4897	0.4430	0.4872	0.4360	0.3868	0.3440
PC4	0.5487	0.4930	**0.5444**	**0.4850**	**0.4460**	**0.3970**
PC5	**0.5945**	**0.5320**	0.5988	0.5320	0.5010	0.4450
PC6	0.6353	0.5660	0.6431	0.5700	0.5418	0.4810
PC7	0.6666	0.5900	0.6786	0.6000	0.5801	0.5150
PC8	0.6897	0.6090	0.7081	0.6260	0.6133	0.5440
PC9	0.7096	0.6230	0.7323	0.6460	0.6434	0.5710
PC10	0.7250	0.6350	0.7513	0.6610	0.6669	0.5910

Table B2: Loadings of Sparse PCA

Predictor	SPC1	SPC2	SPC3	SPC4	SPC5
SMA_5	0.2189				
EMA_5			0.1592		
EMA_52				−0.4261	
NoDur			0.1834		
Durbl			0.2923		
Manuf			0.4187		
Telcm			0.0333		
Shops			0.2593		
Other			0.2794		
MA_1_52				−0.5056	
MA_4_52				−0.5189	
MA_8_52				−0.4473	
MA_1_26				−0.0833	
MOM_26				−0.3503	
MOM_52				−0.2308	
OBV_1_52				−0.1256	
OBV_4_52				−0.0421	
DP			−0.5634		
E10P			−0.6523		
T30Y		0.1302			
T10Y		0.3642			
T7Y		0.4340			
T5Y		0.4767			
T3Y		0.4100			
T2Y		0.3408			
T1Y		0.1928			
T6M		0.1472			
T3M		0.1672			
FEDR		0.1835			
AAA		0.1088			
LIBOR		0.0937			
TS_10Y3M					−1.2381
TS_5Y3M					−0.5618
T5Y_IE		0.1467			
T10Y_IE		0.2127			
CS_BAAAAA	−0.3499				

Table B2: Loadings of Sparse PCA (Continued)

Predictor	SPC1	SPC2	SPC3	SPC4	SPC5
inv	0.2692				
prod	0.4793				
cpi		0.0937			
i3m.3m		0.1572			
i3m.12m		0.3198			
i10y.3m		0.2823			
i10y.12m		0.1729			
ca.sd		–0.1011			
Lead_Conf	0.5103				
PMI	0.2082				
Cons_Cli_Conf	0.3909				
Cons_Sit_Conf	0.2916				
IP	0.3025				

Table B3: Loadings of Targeted Sparse PCA with VIX as Target

	TSPC1	TSPC2	TSPC3	TSPC4
VIX	0.0544			
MA_1_52	–0.4744			
MA_4_52	–0.5460			
MOM_26	–0.0998			
OBV_1_52	–0.0415			
MA_8_52	–0.4987			
OBV_4_52	–0.0087			
MOM_52	–0.2116			
EP			–0.6019	
BAA		0.3154		
AAA		0.5048		
FEDR		0.3165		
IP			–0.4756	
SMA_5	–0.2624			
Manuf				–0.6409
Other				–0.6544
LIBOR		0.3309		
EMA_52	–0.4347			
i10y.12m		0.4669		
T30Y		0.4347		
Payout			0.5618	

Table B4: Loadings of Targeted Sparse PCA with R as Target

	TSPC1	TSPC2	TSPC3	TSPC4
Telcm			–0.2982	
E10P			0.6667	
TS_30Y10Y				0.5949
EP		–0.7663		
MA_8_52	–0.2077			
MOM_52	–0.8399			
EMA_52	–0.2430			
Payout		0.4469		
MA_1_52	–0.2271			
DP			0.6282	
CS_AAA10Y				0.6783

Figure B2: Factors of Ordinary PCA and Sparse PCA

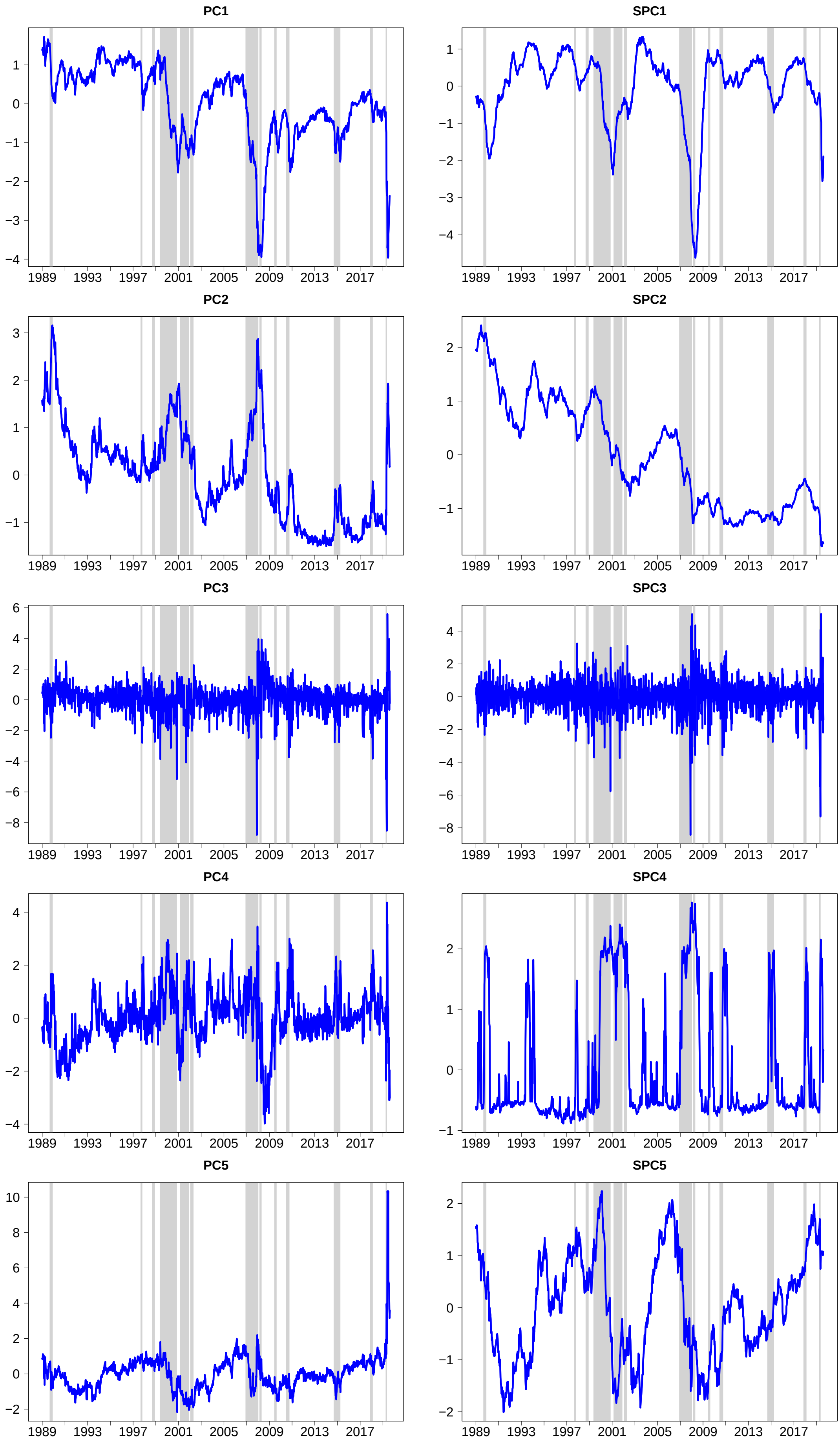


Figure B3: Factors of Targeted (Sparse) PCA Based on the VIX as Target

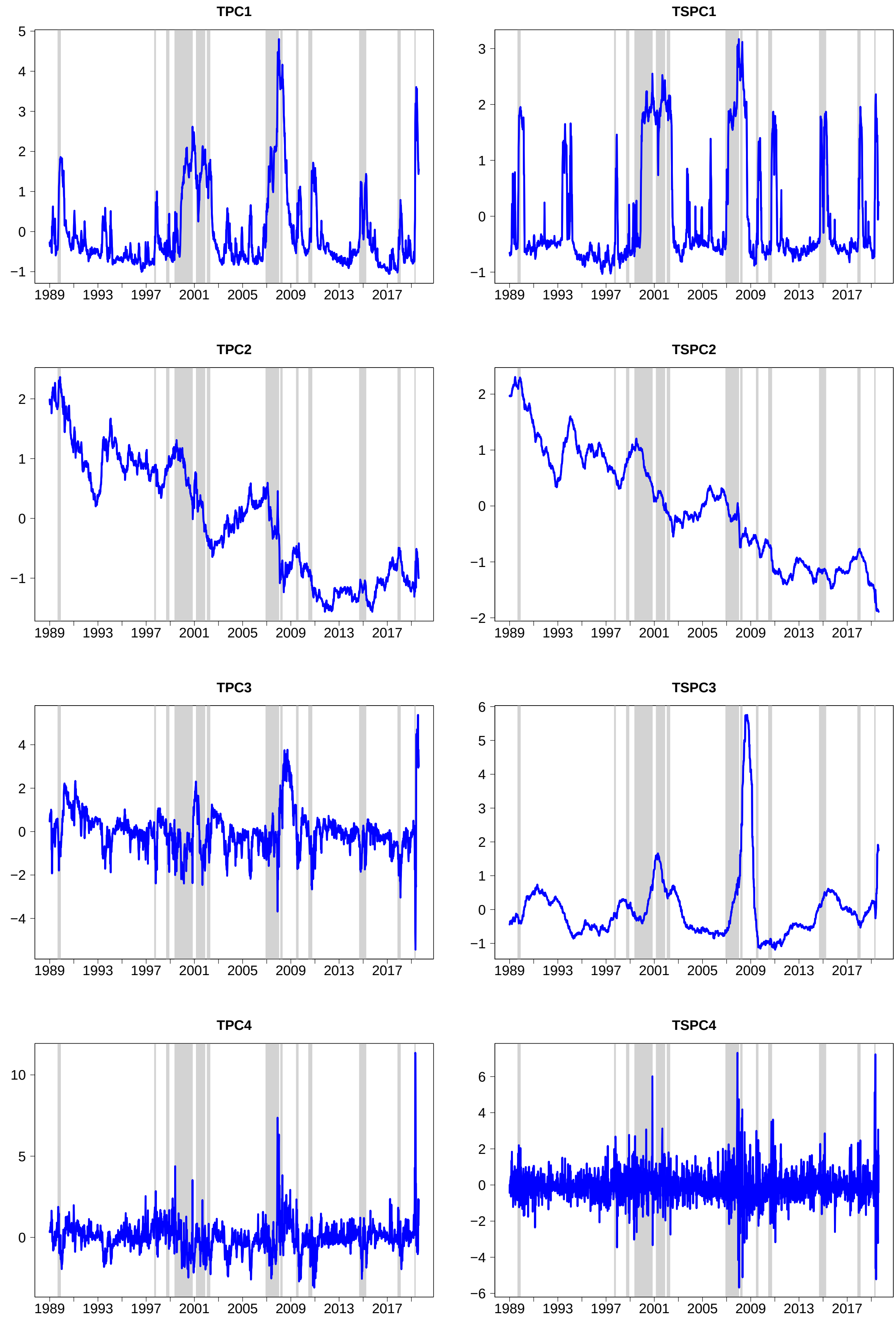
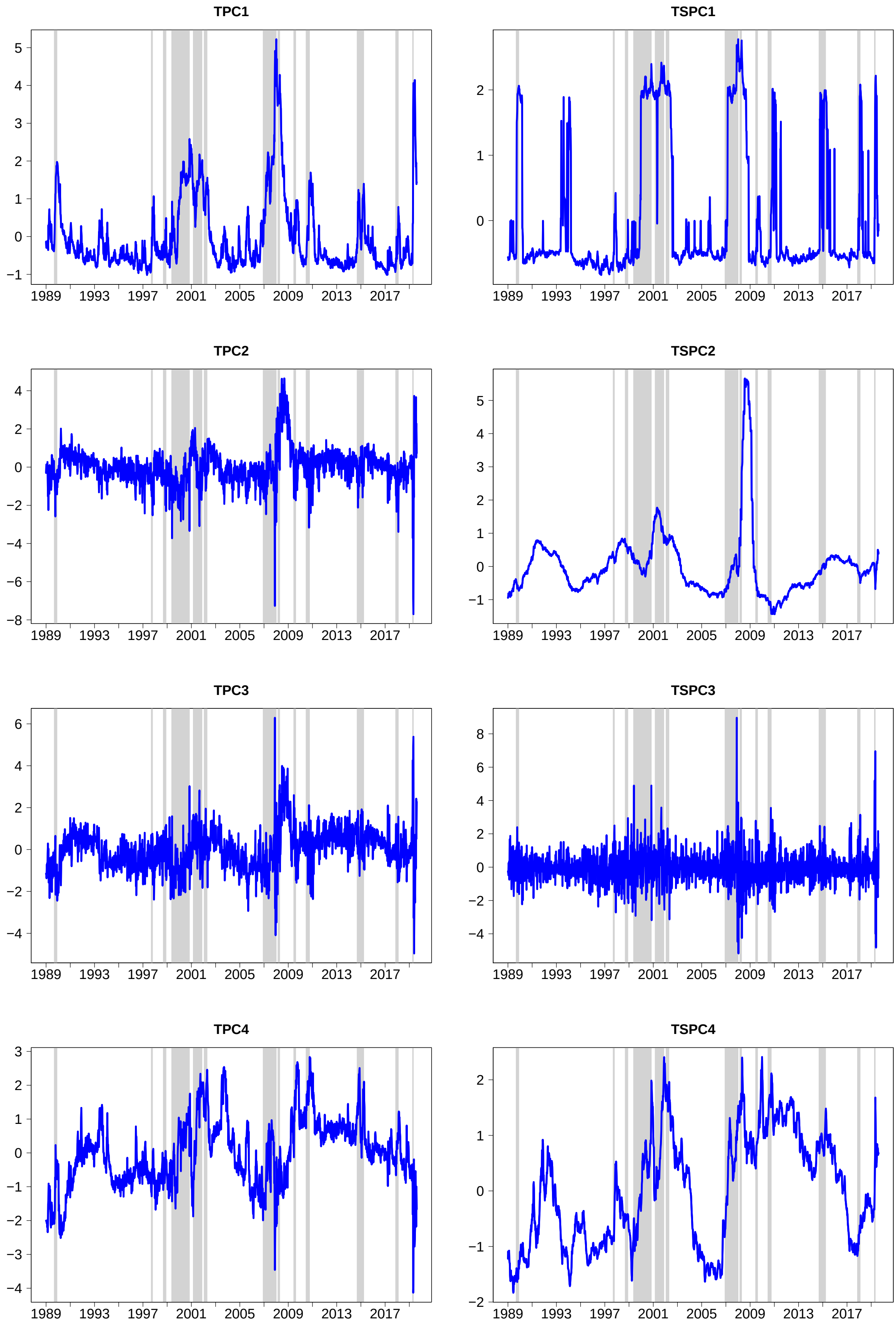


Figure B4: Factors of Targeted (Sparse) PCA Based on the Returns as Target



Appendix C: Evaluation Measures

To assess the out-of-sample performance, we distinguish alongside 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 on the interval between 0 (perfect accuracy) and 2 (worst possible accuracy). S_{t+1} corresponds to the ex post regime indicator 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 (Fawcett 2006):

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

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

Hence, the ROC curve visualizes the trade-off between the threshold choice and the benefits/costs of bear markets identification. For instance, a low threshold improves the accuracy of bear market predictions, but also causes an increase in false alarms. An attractive feature of the ROC curve is that it always lies within a unit square, that is, a classifier can directly be compared to a random guess represented by a diagonal line from (0,0) to (1,1). If the ROC curve is northwest of the diagonal, the model is able to extract valuable information from the data for classification purposes.

Although the ROC is a very flexible and robust visualization tool, it can be impractical when comparing different classifiers. In this context, a straightforward solution is 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 method to calculate confidence intervals for the AUC. To conserve space, we report the resulting intervals only in Figure 3.

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

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:

$$MSPE_i = \frac{1}{T_1}\sum_{t=T_0}^{T_1}(R_{t+1} - \hat{R}^i_{t+1})^2 \tag{C4}$$

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

$$RMSE_i = \sqrt{MSPE_i} \tag{C5}$$

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

$$R^2_{OS} = 1 - \frac{MSPE_i}{MSPE_0} \tag{C6}$$

$MSPE_0$ 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 corrected in the test statistic. $\hat{R}_{1,t+1}$ and $\hat{R}_{2,t+1}$ denote the one-step ahead forecasts from the restricted and the unrestricted model; $\hat{e}_{1,t+1}$ and $\hat{e}_{2,t+1}$ are the corresponding forecasting errors. Then, 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:

$$CW = \frac{\sqrt{T_1}\bar{f}}{\sqrt{\hat{V}}} \tag{C7}$$

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

Since the R^2_{OS} represents the relative performance only in an aggregated form and does not show the development of the error over time, it is also common to consider the cumulative difference in squared forecast error (CDSFE):

$$CDSFE_t = \sum_{t=T_0}^{T_1} (R_{t+1} - \bar{R}_{t+1})^2 - (R_{t+1} - \hat{R}_{t+1})^2 \tag{C8}$$

A positive (negative) slope indicates an outperformance (underperformance) of the strategy over the historical benchmark.

Finally, since point forecasts for returns are usually very difficult, we are additionally interested in the question to what extent we can at least forecast the 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^-).

C2: Economic Value

As Rapach and Zhou (2013) 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 the 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 Treasury Bonds 1Y-3Y ETF) 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:

$$CER = \hat{\mu}_i - \frac{1}{2}\gamma\hat{\sigma}_i^2 \tag{C9}$$

$\hat{\mu}_i$ ($\hat{\sigma}_i^2$) 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 (or another benchmark). The CER gain 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 $\gamma = 3$ (see, for example. Zhu and Zhu 2013) and set $\gamma = 1, 2, 5$ 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 follows:

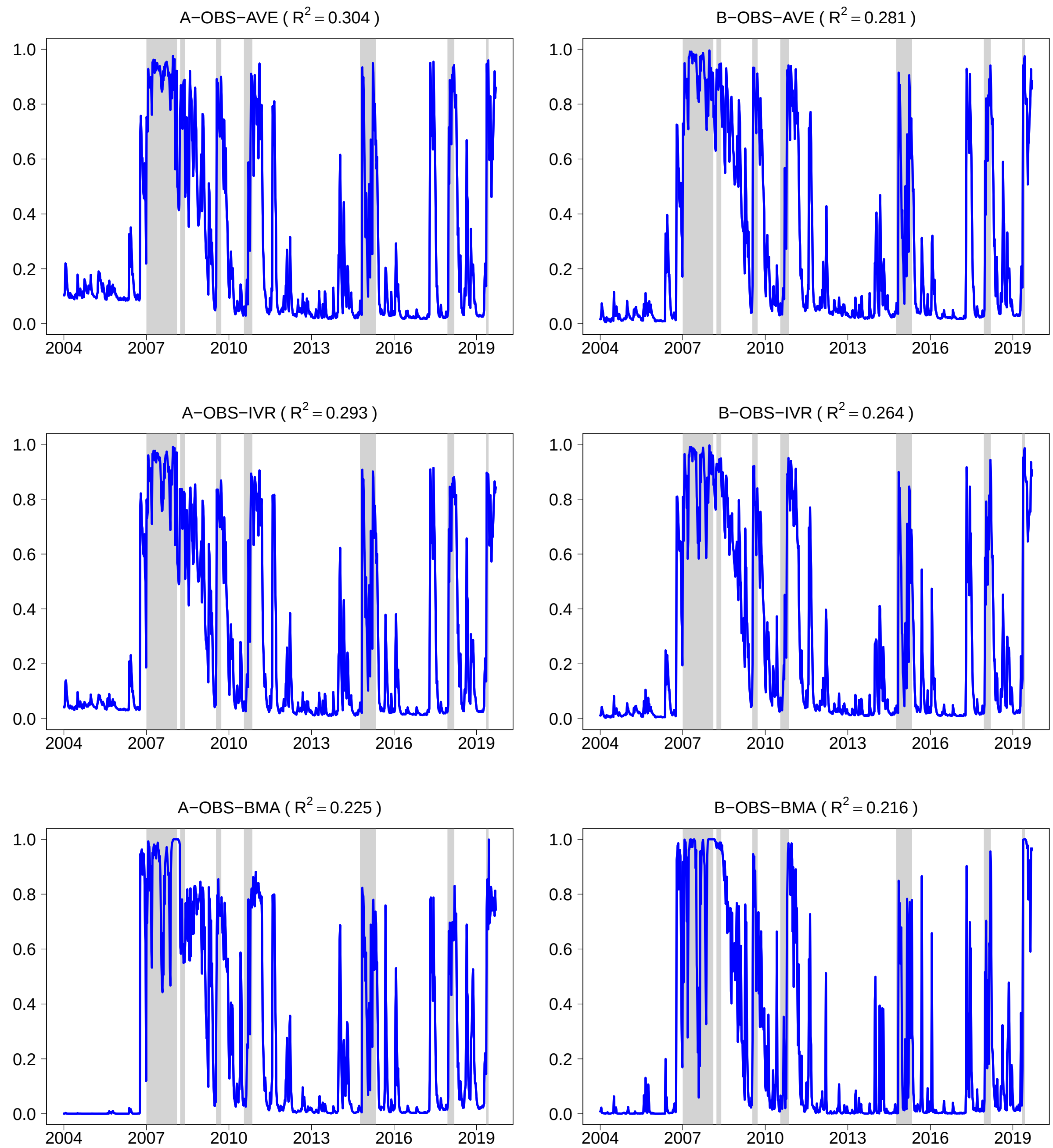
$$VaR_{1-\alpha}(R) = F_R^{-1}(1-\alpha) = \inf\{r \in \mathbb{R} : F_R^r \geq 1-\alpha\} \tag{C10}$$

Our calculation relies 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.

$$CVaR_{1-\alpha}(R) = E[r | r \leq F_R^{-1}(1-\alpha)] \tag{C11}$$

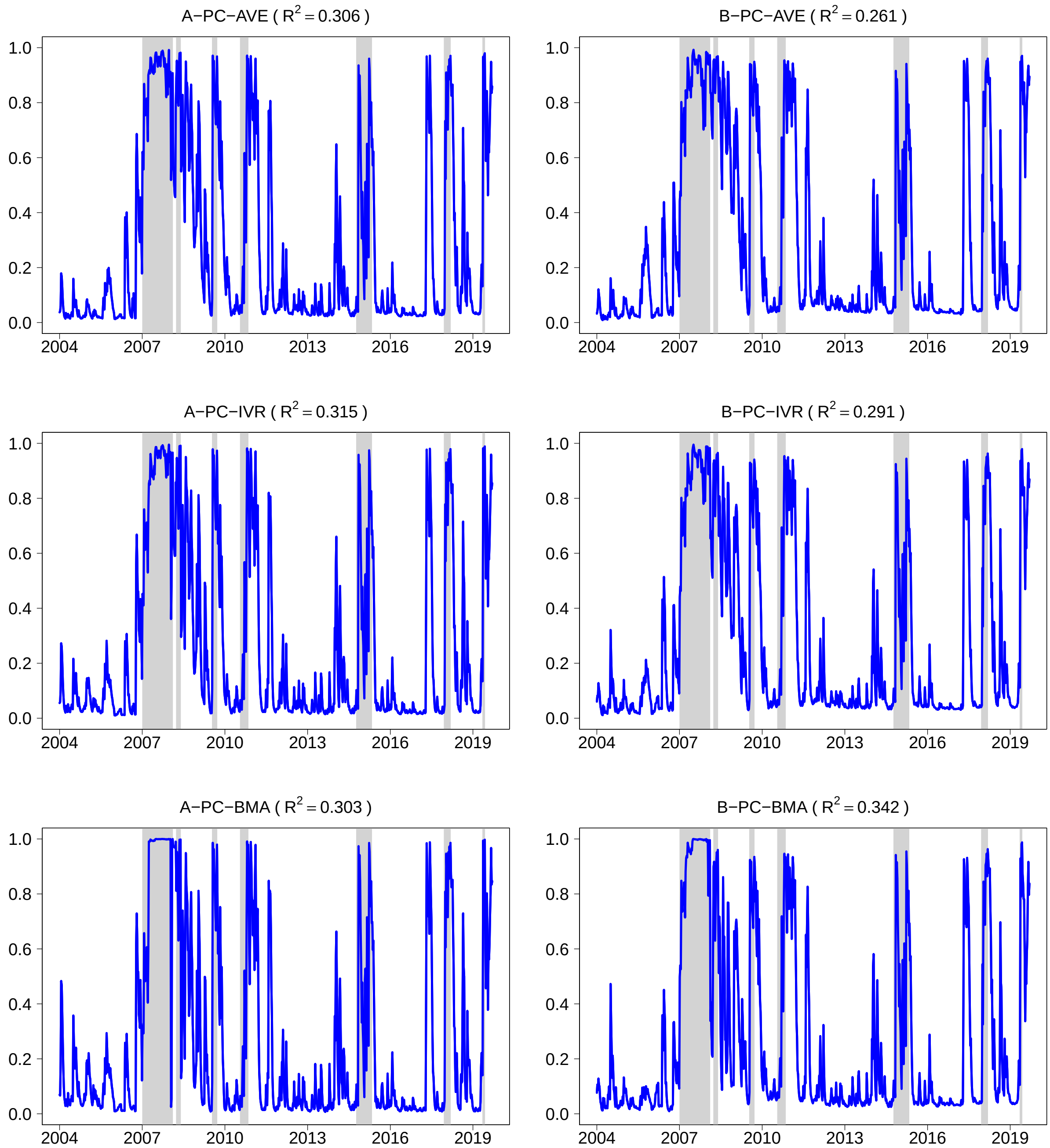
Appendix D: Additional Out-of-Sample Results

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



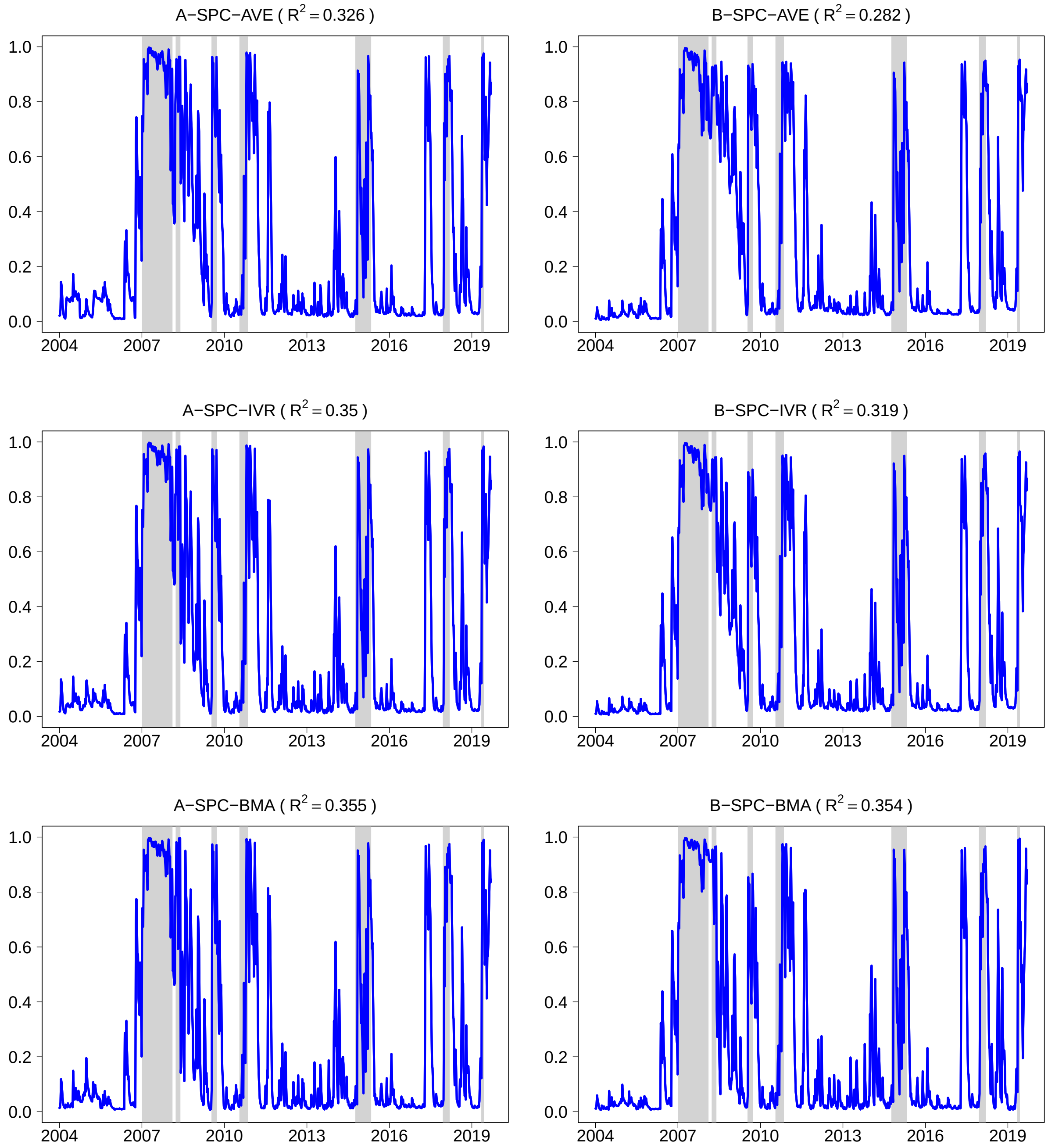
Notes: Figure shows the predicted out-of-sample probabilities of forecasting a bear market. Gray-shaded areas indicate actual bear market periods according to an ex post detection using the dating rule of Lunde and Timmermann (2004) (see also Section 4.1). The R^2 is obtained from a linear probability 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



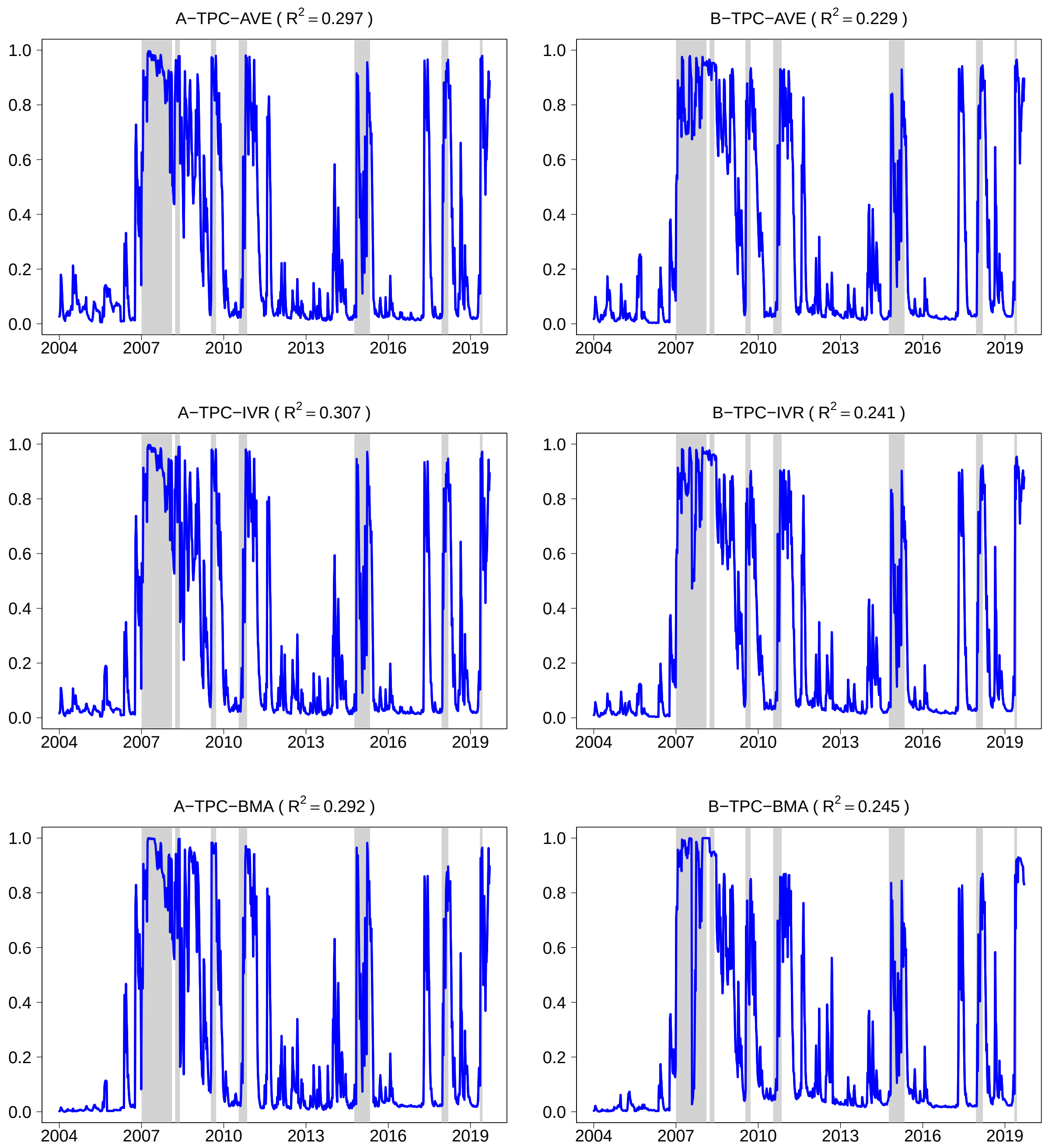
Notes: Figure shows the predicted out-of-sample probabilities of forecasting a bear market. Gray-shaded areas indicate actual bear market periods according to an ex post detection using the dating rule of Lunde and Timmermann (2004) (see also Section 4.1). The R^2 is obtained from a linear probability 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



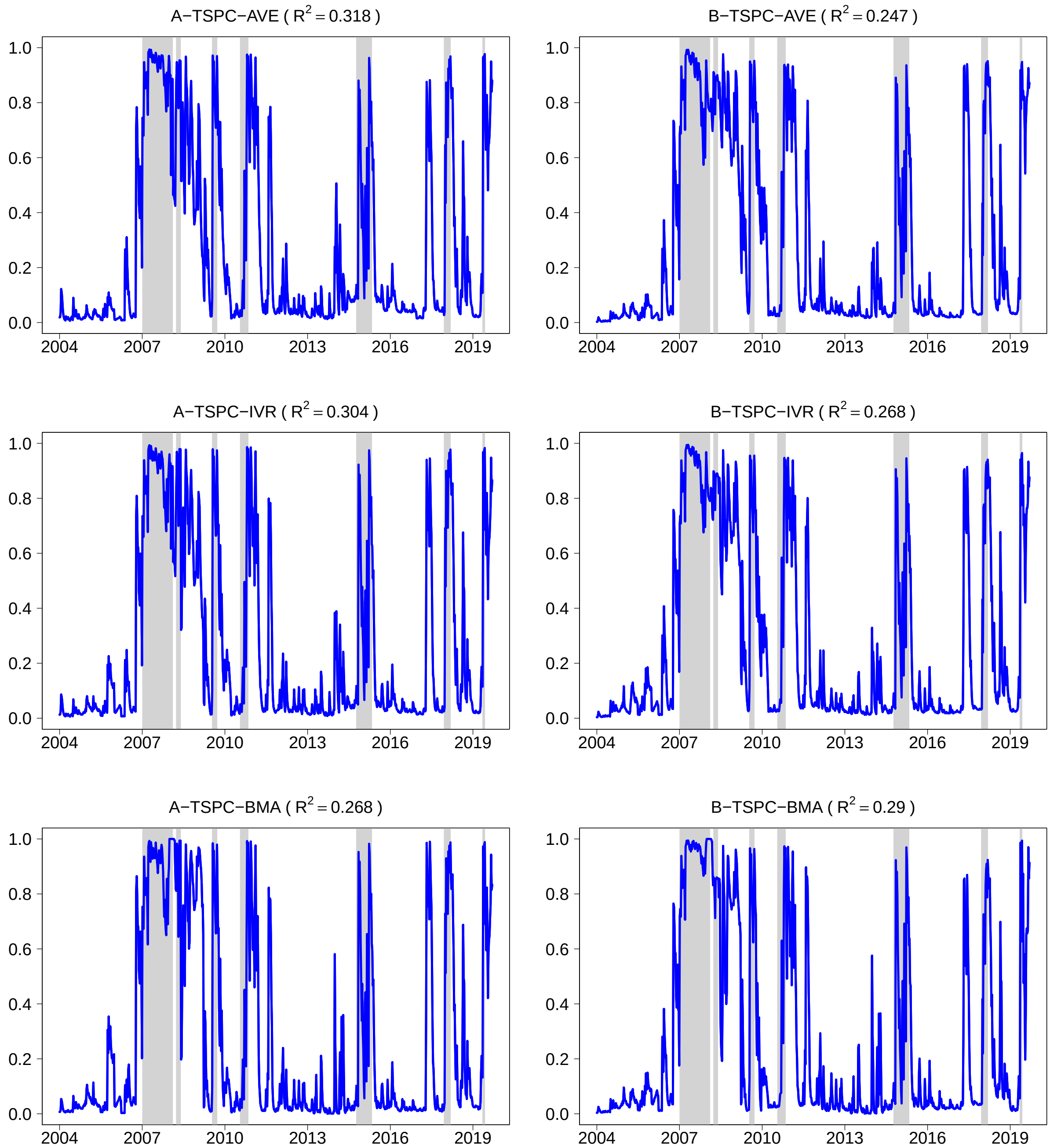
Notes: Figure shows the predicted out-of-sample probabilities of forecasting a bear market. Gray-shaded areas indicate actual bear market periods according to an ex post detection using the dating rule of Lunde and Timmermann (2004) (see also Section 4.1). The R^2 is obtained from a linear probability 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) (see also Section 4.1). 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) (see also Section 4.1). 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.

Table D1: 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)	+1	+4	+1	+1	+6	+6
Global Financial Crisis II (2009-01-09 to 2009-03-06)	0	0	0	+1	+44	+41
Flash Crash Aftermath (2010-04-30 to 2010-07-02)	+2	+2	+2	+3	+14	+14
Debt Crisis (2011-05-06 to 2011-08-19)	+9	+13	+9	+3	+23	+23
Chinese Market Crash (2015-07-24 to 2016-02-12)	+5	+5	+5	+1	+5	+6
Economic Slowdown Fear (2018-09-28 to 2018-12-21)	+3	+5	+5	+3	+7	+9
COVID-19 Crash (2020-02-21 to 2020-03-20)	+2	+2	+2	+2	+15*	+15*

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. 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 reference. * indicates that the delay still is ongoing at the end of the sample period for the COVID-19 crash.

Table D2: Number of Active Position Changes

	Regime Strategy			Return Strategy		
	AVE	IVR	BMA	AVE	IVR	DMSPE
A-OBS	52	46	50	486	518	495
A-PC	42	50	58	406	405	404
A-SPC	48	46	52	446	445	446
A-TPC	48	48	54	461	435	461
A-TSPC	52	40	40	478	474	475
B-OBS	36	40	68	102	106	100
B-PC	42	42	46	101	110	101
B-SPC	40	42	52	105	110	105
B-TPC	36	36	50	110	122	110
B-TSPC	40	40	44	110	101	110

Notes: Table shows the number of active position changes according to the different strategies. Model A contains predictors in the switching equation and the conditional mean equation, Model 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; IVR: inverse rank; BMA: Bayesian model averaging; DMSPE: discounted mean squared prediction error.

Appendix E: Robustness Tests

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

	QPS	Accuracy	Bear	Bull
MA_3M	0.4268	0.7720	0.7092	0.7850
MA_6M	0.3341	0.8171	0.7234	0.8365
MA_12M (D_t^{MA})	0.3439	0.8183	0.5177	0.8807
MA_18M	0.4049	0.7878	0.3050	0.8881
MA_36M	0.5195	0.7354	0.2695	0.8321

Notes: Table shows the statistical performance of regime forecasts when using different moving average 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.

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

	R^{cum}	$\bar{R}$	$\bar{\sigma}$	SR	CER	MaxDD	VaR	CVaR
MA_3M	2.10	4.81	10.84	0.24	3.53	19.23	–2.46	–3.93
MA_6M	2.58	6.20	10.77	0.37	4.86	20.12	–2.26	–3.87
MA_12M (D_t^{MA})	2.68	6.45	12.27	0.34	4.75	24.46	–2.66	–4.50
MA_18M	2.73	6.57	13.18	0.33	4.64	30.36	–3.07	–4.77
MA_36M	3.07	7.37	13.41	0.38	5.33	33.38	–2.87	–4.74

Notes: Table shows the economic value of regime forecasts when using different moving average 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; MaxDD: maximum drawdown; VaR: value-at-risk; CVaR: conditional value-at-risk. See Appendix C for a detailed description of the evaluation measures.

Table E3: Economic Value of Regime Forecasts: Threshold of 25%

	R^{cum}	$\bar{R}$	$\bar{\sigma}$	SR	CER	MaxDD	VaR	CVaR
S&P 500	3.86	8.94	18.00	0.37	5.35	54.61	–3.90	–6.16
Rf	1.39	2.10	1.20	0.00	2.06	1.90	–0.19	–0.31
50/50	2.51	6.02	9.12	0.42	5.02	28.76	–2.03	–3.09
MA_12M (D_t^{MA})	2.68	6.45	12.27	0.34	4.75	24.46	–2.66	–4.50
TCTP	3.07	7.37	9.29	0.55	6.25	12.25	–1.96	–3.20
A-OBS-AVE	3.25	7.76	9.44	0.58	6.59	13.79	–2.01	–3.24
A-PC-AVE	3.17	7.60	9.56	0.56	6.42	19.74	–2.02	–3.34
A-SPC-AVE	3.15	7.54	9.65	0.55	6.35	19.02	–2.05	–3.39
A-TPC-AVE	2.85	6.86	9.49	0.49	5.74	19.88	–2.02	–3.35
A-TSPC-AVE	3.06	7.36	9.65	0.53	6.18	20.12	–2.05	–3.39
A-OBS-IVR	2.29	5.39	9.12	0.35	4.42	15.47	–2.00	–3.24
A-PC-IVR	3.21	7.67	9.80	0.55	6.44	19.74	–2.06	–3.40
A-SPC-IVR	3.48	8.24	9.99	0.60	6.93	19.02	–2.09	–3.45
A-TPC-IVR	3.08	7.40	9.62	0.54	6.23	19.02	–2.02	–3.34
A-TSPC-IVR	3.15	7.55	9.66	0.55	6.35	20.58	–2.05	–3.39
A-OBS-BMA	2.04	4.61	9.14	0.27	3.68	29.96	–2.01	–3.31
A-PC-BMA	2.98	7.16	10.28	0.48	5.87	22.41	–2.11	–3.62
A-SPC-BMA	3.75	8.74	10.14	0.64	7.36	18.72	–2.11	–3.45
A-TPC-BMA	3.28	7.83	9.96	0.56	6.56	21.78	–2.06	–3.47
A-TSPC-BMA	3.59	8.45	9.92	0.62	7.14	19.13	–2.09	–3.43
B-OBS-AVE	2.86	6.88	9.30	0.50	5.80	14.15	–2.00	–3.25
B-PC-AVE	3.25	7.76	9.44	0.58	6.59	16.30	–2.01	–3.24
B-SPC-AVE	3.33	7.93	9.49	0.60	6.74	13.79	–2.01	–3.24
B-TPC-AVE	2.93	7.05	9.34	0.51	5.94	17.40	–2.00	–3.25
B-TSPC-AVE	2.73	6.58	9.40	0.46	5.50	13.79	–2.01	–3.31
B-OBS-IVR	2.12	4.89	9.26	0.29	3.92	24.75	–2.02	–3.39
B-PC-IVR	3.45	8.16	9.50	0.62	6.96	16.30	–2.01	–3.24
B-SPC-IVR	3.41	8.09	9.47	0.62	6.89	15.65	–2.01	–3.24
B-TPC-IVR	3.22	7.70	9.54	0.57	6.52	16.88	–2.01	–3.30
B-TSPC-IVR	3.40	8.08	9.55	0.61	6.87	13.91	–2.01	–3.30
B-OBS-BMA	1.76	3.66	9.63	0.16	2.68	22.73	–2.14	–3.61
B-PC-BMA	3.90	9.02	10.22	0.66	7.60	21.18	–2.06	–3.35
B-SPC-BMA	3.84	8.91	9.94	0.67	7.56	13.59	–2.09	–3.35
B-TPC-BMA	2.97	7.15	10.03	0.49	5.91	19.02	–2.13	–3.56
B-TSPC-BMA	3.82	8.87	9.79	0.67	7.55	12.25	–2.05	–3.34

Notes: Table shows the economic value of different investment strategies. S&P 500: stocks only; Rf: short-term government bonds only; 50/50: mixed strategy; D_t^{MA}: naïve 12-month moving average; $TCTP$: MS model with TCTP. Model A contains predictors in the switching equation and the conditional mean equation, Model 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; IVR: inverse rank; 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; MaxDD: maximum drawdown; VaR: value-at-risk; CVaR: conditional value-at-risk. See Appendix C for a detailed description of the evaluation measures.

Table E4: Economic Value of Regime Forecasts: Threshold of 75%

	R^{cum}	$\bar{R}$	$\bar{\sigma}$	SR	CER	MaxDD	VaR	CVaR
S&P 500	3.86	8.94	18.00	0.37	5.35	54.61	–3.90	–6.16
Rf	1.39	2.10	1.20	0.00	2.06	1.90	–0.19	–0.31
50/50	2.51	6.02	9.12	0.42	5.02	28.76	–2.03	–3.09
MA_12M (D_t^{MA})	2.68	6.45	12.27	0.34	4.75	24.46	–2.66	–4.50
TCTP	4.56	10.10	10.53	0.74	8.52	12.25	–2.14	–3.61
A-OBS-AVE	4.81	10.47	12.85	0.63	8.32	19.75	–2.64	–4.32
A-PC-AVE	5.40	11.29	12.79	0.70	9.07	19.08	–2.62	–4.26
A-SPC-AVE	6.02	12.06	12.71	0.76	9.78	14.93	–2.59	–4.17
A-TPC-AVE	5.38	11.26	12.65	0.70	9.08	18.50	–2.59	–4.23
A-TSPC-AVE	5.75	11.73	12.80	0.73	9.46	15.70	–2.60	–4.20
A-OBS-IVR	5.01	10.76	12.62	0.67	8.64	16.40	–2.62	–4.18
A-PC-IVR	5.10	10.89	12.96	0.66	8.67	18.77	–2.64	–4.31
A-SPC-IVR	6.24	12.31	12.85	0.77	9.97	16.46	–2.60	–4.22
A-TPC-IVR	6.12	12.17	12.53	0.78	9.93	16.40	–2.48	–4.04
A-TSPC-IVR	4.63	10.20	12.69	0.62	8.11	21.44	–2.66	–4.20
A-OBS-BMA	3.28	7.82	13.83	0.40	5.63	37.62	–2.90	–4.85
A-PC-BMA	5.15	10.95	12.96	0.66	8.72	20.58	–2.64	–4.39
A-SPC-BMA	6.06	12.11	12.88	0.76	9.78	18.27	–2.62	–4.24
A-TPC-BMA	3.48	8.24	13.51	0.44	6.12	37.13	–2.66	–4.66
A-TSPC-BMA	4.30	9.69	12.37	0.60	7.73	18.72	–2.62	–4.20
B-OBS-AVE	4.78	10.43	12.07	0.67	8.48	15.36	–2.48	–4.07
B-PC-AVE	4.67	10.26	11.60	0.68	8.44	16.12	–2.44	–3.86
B-SPC-AVE	4.17	9.48	12.00	0.60	7.63	17.70	–2.60	–4.10
B-TPC-AVE	3.81	8.85	12.08	0.54	7.03	24.22	–2.70	–4.10
B-TSPC-AVE	4.16	9.46	11.89	0.60	7.64	14.44	–2.48	–4.04
B-OBS-IVR	3.86	8.95	12.19	0.55	7.10	22.08	–2.64	–4.22
B-PC-IVR	5.49	11.41	11.99	0.76	9.38	15.00	–2.48	–3.98
B-SPC-IVR	6.12	12.18	12.28	0.80	10.00	14.93	–2.53	–4.13
B-TPC-IVR	3.22	7.71	12.21	0.45	5.94	30.34	–2.83	–4.25
B-TSPC-IVR	4.44	9.92	12.20	0.62	7.98	15.82	–2.64	–4.17
B-OBS-BMA	3.28	7.83	12.22	0.46	6.06	19.82	–2.70	–4.32
B-PC-BMA	7.29	13.42	12.30	0.90	11.10	17.36	–2.48	–4.02
B-SPC-BMA	5.30	11.15	12.83	0.69	8.94	18.61	–2.64	–4.28
B-TPC-BMA	3.14	7.54	12.77	0.41	5.65	20.41	–2.90	–4.57
B-TSPC-BMA	4.82	10.49	11.77	0.69	8.60	15.83	–2.44	–3.98

Notes: Table shows the economic value of different investment strategies. S&P 500: stocks only; Rf: short-term government bonds only; 50/50: mixed strategy; D_t^{MA}: naïve 12-month moving average; $TCTP$: MS model with TCTP. Model A contains predictors in the switching equation and the conditional mean equation, Model 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; IVR: inverse rank; 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; MaxDD: maximum drawdown; VaR: value-at-risk; CVaR: conditional value-at-risk. See Appendix C for a detailed description of the evaluation measures.

Table E5: Economic Performance of Return Forecasts with $\tau = 1$

	R^{cum}	$\bar{R}$	$\bar{\sigma}$	SR	CER	MaxDD	VaR	CVaR
S&P 500	3.86	8.94	18.00	0.37	8.59	54.61	–3.90	–6.16
Rf	1.39	2.10	1.20	0.00	2.07	1.90	–0.19	–0.31
50/50	2.51	6.02	9.12	0.42	5.85	28.76	–2.03	–3.09
Hist_Mean	3.85	8.93	18.00	0.37	8.58	54.61	–3.90	–6.16
TCTP	2.58	6.20	10.82	0.37	6.02	27.29	–2.42	–3.86
A-OBS-AVE	1.90	4.15	14.33	0.14	4.10	43.75	–2.64	–5.29
A-PC-AVE	2.77	6.68	14.32	0.31	6.50	35.24	–2.48	–5.13
A-SPC-AVE	2.64	6.35	14.21	0.29	6.18	31.77	–2.44	–5.01
A-TPC-AVE	2.50	5.98	15.07	0.25	5.83	38.96	–2.62	–5.28
A-TSPC-AVE	2.07	4.71	14.42	0.17	4.63	42.90	–2.59	–5.24
A-OBS-IVR	1.80	3.81	14.48	0.11	3.77	49.55	–2.66	–5.40
A-PC-IVR	2.62	6.30	14.27	0.28	6.14	32.73	–2.59	–5.15
A-SPC-IVR	2.06	4.68	14.17	0.17	4.61	40.60	–2.59	–5.19
A-TPC-IVR	2.13	4.90	14.12	0.19	4.82	37.65	–2.48	–5.18
A-TSPC-IVR	2.31	5.46	14.18	0.23	5.34	38.38	–2.48	–5.15
A-OBS-DMSPE	1.69	3.40	14.46	0.08	3.37	50.17	–2.64	–5.42
A-PC-DMSPE	2.67	6.42	14.30	0.29	6.25	35.27	–2.48	–5.13
A-SPC-DMSPE	2.58	6.19	14.13	0.28	6.04	31.71	–2.44	–5.01
A-TPC-DMSPE	2.50	5.98	14.86	0.25	5.83	38.96	–2.59	–5.23
A-TSPC-DMSPE	2.08	4.75	14.47	0.18	4.67	42.48	–2.53	–5.26
B-OBS-AVE	2.66	6.41	11.40	0.36	6.22	34.40	–2.48	–4.14
B-PC-AVE	2.77	6.67	11.02	0.40	6.46	29.64	–2.44	–3.90
B-SPC-AVE	2.77	6.67	10.97	0.40	6.47	30.09	–2.44	–3.89
B-TPC-AVE	2.65	6.38	11.08	0.37	6.19	31.20	–2.48	–3.94
B-TSPC-AVE	2.68	6.45	11.14	0.38	6.26	33.36	–2.48	–3.98
B-OBS-IVR	3.02	7.26	11.39	0.44	7.02	30.58	–2.48	–4.06
B-PC-IVR	2.80	6.76	10.99	0.41	6.55	28.93	–2.44	–3.87
B-SPC-IVR	2.73	6.58	11.21	0.39	6.39	32.35	–2.48	–4.01
B-TPC-IVR	2.75	6.62	11.04	0.40	6.42	29.16	–2.48	–3.92
B-TSPC-IVR	2.59	6.21	11.13	0.36	6.03	32.45	–2.48	–4.04
B-OBS-DMSPE	2.68	6.45	11.40	0.37	6.26	34.05	–2.48	–4.13
B-PC-DMSPE	2.77	6.67	11.01	0.40	6.47	29.60	–2.44	–3.90
B-SPC-DMSPE	2.78	6.69	10.97	0.40	6.49	29.99	–2.44	–3.89
B-TPC-DMSPE	2.65	6.38	11.08	0.37	6.20	31.14	–2.48	–3.94
B-TSPC-DMSPE	2.69	6.48	11.14	0.38	6.29	33.27	–2.44	–3.97

Notes: Table shows the economic value of different investment strategies. S&P 500: stocks only; Rf: short-term government bonds only; 50/50: mixed strategy; Hist_Mean: historical mean; $TCTP$: MS model with TCTP. Model A contains predictors in the switching equation and the conditional mean equation, Model 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; IVR: inverse rank; DMSPE: discounted mean squared prediction error; 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; MaxDD: maximum drawdown; VaR: value-at-risk; CVaR: conditional value-at-risk. See Appendix C for a detailed description of the evaluation measures.

Table E6: Economic Performance of Return Forecasts with $\tau = 2$

	R^{cum}	$\bar{R}$	$\bar{\sigma}$	SR	CER	MaxDD	VaR	CVaR
S&P 500	3.86	8.94	18.00	0.37	6.97	54.61	–3.90	–6.16
Rf	1.39	2.10	1.20	0.00	2.07	1.90	–0.19	–0.31
50/50	2.51	6.02	9.12	0.42	5.43	28.76	–2.03	–3.09
Hist_Mean	2.72	6.55	15.53	0.28	5.17	50.94	–3.08	–5.40
TCTP	2.55	6.12	10.33	0.38	5.42	21.81	–2.16	–3.64
A-OBS-AVE	1.40	2.15	13.81	–0.00	1.20	43.69	–2.49	–5.18
A-PC-AVE	2.02	4.54	13.83	0.17	3.52	33.50	–2.48	–5.04
A-SPC-AVE	1.94	4.28	13.62	0.15	3.29	31.72	–2.34	–4.90
A-TPC-AVE	1.83	3.92	14.44	0.12	2.83	39.04	–2.45	–5.12
A-TSPC-AVE	1.76	3.67	13.82	0.11	2.67	36.46	–2.40	–4.96
A-OBS-IVR	1.32	1.80	13.73	–0.03	0.87	46.00	–2.59	–5.12
A-PC-IVR	1.94	4.29	13.65	0.15	3.30	33.43	–2.41	–5.02
A-SPC-IVR	1.61	3.07	13.74	0.06	2.11	39.78	–2.32	–5.04
A-TPC-IVR	1.62	3.12	13.70	0.07	2.16	38.37	–2.44	–5.05
A-TSPC-IVR	1.76	3.63	13.64	0.11	2.67	38.12	–2.34	–4.98
A-OBS-DMSPE	1.25	1.45	13.93	–0.05	0.50	49.07	–2.58	–5.29
A-PC-DMSPE	1.98	4.43	13.82	0.16	3.41	33.50	–2.48	–5.03
A-SPC-DMSPE	1.88	4.10	13.59	0.14	3.12	31.70	–2.34	–4.89
A-TPC-DMSPE	1.80	3.78	14.30	0.11	2.72	39.47	–2.46	–5.13
A-TSPC-DMSPE	1.80	3.78	13.82	0.12	2.79	35.51	–2.40	–4.96
B-OBS-AVE	2.35	5.58	10.58	0.32	4.88	31.17	–2.34	–3.79
B-PC-AVE	2.58	6.20	10.52	0.38	5.47	24.07	–2.32	–3.71
B-SPC-AVE	2.63	6.32	10.33	0.40	5.60	24.18	–2.23	–3.62
B-TPC-AVE	2.27	5.33	10.59	0.29	4.64	32.21	–2.44	–3.82
B-TSPC-AVE	2.57	6.18	10.24	0.39	5.48	26.05	–2.17	–3.63
B-OBS-IVR	2.46	5.89	10.64	0.34	5.16	28.85	–2.31	–3.80
B-PC-IVR	2.57	6.17	10.44	0.38	5.45	24.53	–2.22	–3.68
B-SPC-IVR	2.61	6.26	10.35	0.39	5.55	24.78	–2.23	–3.63
B-TPC-IVR	2.28	5.37	10.63	0.30	4.67	31.05	–2.40	–3.82
B-TSPC-IVR	2.46	5.87	10.34	0.35	5.18	27.02	–2.22	–3.67
B-OBS-DMSPE	2.35	5.58	10.58	0.32	4.87	31.32	–2.35	–3.79
B-PC-DMSPE	2.60	6.24	10.51	0.38	5.51	24.04	–2.30	–3.70
B-SPC-DMSPE	2.63	6.33	10.34	0.40	5.61	24.11	–2.22	–3.62
B-TPC-DMSPE	2.28	5.37	10.60	0.30	4.67	32.14	–2.44	–3.82
B-TSPC-DMSPE	2.58	6.19	10.23	0.39	5.49	25.95	–2.18	–3.62

Notes: Table shows the economic value of different investment strategies. S&P 500: stocks only; Rf: short-term government bonds only; 50/50: mixed strategy; Hist_Mean: historical mean; $TCTP$: MS model with TCTP. Model A contains predictors in the switching equation and the conditional mean equation, Model 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; IVR: inverse rank; DMSPE: discounted mean squared prediction error; 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; MaxDD: maximum drawdown; VaR: value-at-risk; CVaR: conditional value-at-risk. See Appendix C for a detailed description of the evaluation measures.

Table E7: Economic Performance of Return Forecasts with $\tau = 5$

	R^{cum}	$\bar{R}$	$\bar{\sigma}$	SR	CER	MaxDD	VaR	CVaR
S&P 500	3.86	8.94	18.00	0.37	2.11	54.61	−3.90	−6.16
Rf	1.39	2.10	1.20	0.00	2.05	1.90	−0.19	−0.31
50/50	2.51	6.02	9.12	0.42	4.18	28.76	−2.03	−3.09
Hist_Mean	1.75	3.60	11.07	0.13	1.10	41.13	−2.07	−4.31
TCTP	2.07	4.71	8.52	0.30	3.16	14.64	−1.66	−3.09
A-OBS-AVE	0.85	−0.99	12.08	−0.25	−3.88	39.70	−2.09	−4.47
A-PC-AVE	1.11	0.68	12.13	−0.12	−2.23	35.52	−1.97	−4.46
A-SPC-AVE	1.04	0.23	11.91	−0.16	−2.58	35.47	−1.88	−4.34
A-TPC-AVE	1.09	0.55	12.30	−0.13	−2.45	36.97	−1.99	−4.44
A-TSPC-AVE	1.04	0.26	12.08	−0.15	−2.64	37.68	−1.88	−4.36
A-OBS-IVR	0.85	−1.03	12.09	−0.26	−3.93	42.46	−1.88	−4.55
A-PC-IVR	1.03	0.19	12.10	−0.16	−2.71	40.86	−2.15	−4.53
A-SPC-IVR	0.99	−0.04	12.09	−0.18	−2.94	42.47	−1.87	−4.43
A-TPC-IVR	1.03	0.21	12.09	−0.16	−2.69	39.28	−2.04	−4.38
A-TSPC-IVR	0.96	−0.23	11.92	−0.20	−3.04	46.28	−1.86	−4.37
A-OBS-DMSPE	0.83	−1.19	12.15	−0.27	−4.12	41.84	−2.09	−4.52
A-PC-DMSPE	1.10	0.63	12.12	−0.12	−2.28	35.99	−1.91	−4.46
A-SPC-DMSPE	1.03	0.19	11.90	−0.16	−2.61	35.83	−1.87	−4.34
A-TPC-DMSPE	1.07	0.41	12.23	−0.14	−2.56	36.99	−2.06	−4.44
A-TSPC-DMSPE	1.04	0.27	12.03	−0.15	−2.59	37.57	−1.88	−4.36
B-OBS-AVE	1.92	4.23	8.30	0.25	2.77	15.97	−1.71	−3.04
B-PC-AVE	1.91	4.20	8.41	0.24	2.71	15.75	−1.66	−3.10
B-SPC-AVE	2.04	4.62	8.45	0.29	3.09	15.30	−1.63	−3.09
B-TPC-AVE	1.84	3.94	8.49	0.21	2.43	20.72	−1.81	−3.09
B-TSPC-AVE	2.03	4.58	8.50	0.28	3.04	14.41	−1.67	−3.13
B-OBS-IVR	1.90	4.16	8.08	0.24	2.77	18.08	−1.59	−2.98
B-PC-IVR	1.81	3.83	8.37	0.20	2.36	16.32	−1.65	−3.14
B-SPC-IVR	2.07	4.73	8.43	0.30	3.21	14.53	−1.67	−3.05
B-TPC-IVR	1.78	3.71	8.46	0.18	2.22	18.49	−1.79	−3.14
B-TSPC-IVR	1.92	4.23	8.52	0.24	2.70	16.09	−1.69	−3.16
B-OBS-DMSPE	1.92	4.23	8.32	0.25	2.76	15.92	−1.71	−3.05
B-PC-DMSPE	1.91	4.20	8.41	0.24	2.71	15.72	−1.66	−3.10
B-SPC-DMSPE	2.04	4.62	8.45	0.29	3.10	15.29	−1.63	−3.09
B-TPC-DMSPE	1.83	3.92	8.50	0.21	2.41	20.53	−1.81	−3.11
B-TSPC-DMSPE	2.02	4.56	8.49	0.28	3.02	14.36	−1.68	−3.13

Notes: Table shows the economic value of different investment strategies. S&P 500: stocks only; Rf: short-term government bonds only; 50/50: mixed strategy; Hist_Mean: historical mean; $TCTP$: MS model with TCTP. Model A contains predictors in the switching equation and the conditional mean equation, Model 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; IVR: inverse rank; DMSPE: discounted mean squared prediction error; 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; MaxDD: maximum drawdown; VaR: value-at-risk; CVaR: conditional value-at-risk. See Appendix C for a detailed description of the evaluation measures.