Universität Trier

Quantitative Finance and Risk Analysis

Working Paper Series Nº 20 - 3:

Forecasting Stock Market Recessions in the US: Predictive Modeling using Different Identification Approaches

Felix Haase

Matthias Neuenkirch

Forecasting Stock Market Recessions in the US: Predictive Modeling using Different Identification Approaches

Felix Haase
University of Trier

Matthias Neuenkirch*
University of Trier and CESifo

First Draft: March 23, 2020

Abstract

Stock market recessions are often early warning signals for financial or economic crises. Hence, forecasting bear markets is important for investors, policymakers, and economic agents in general. In our two-step procedure, we first identify stock market regimes in the US using three different techniques (Markov-switching models, dating rules, and a naïve moving average). Second, we predict recessions in the S&P 500 with the help of several modeling approaches, utilizing the information of 92 macro-financial variables. Our results suggest that several variables are suitable for forecasting recessions in stock markets in-sample and out-of-sample. Our early warning models for the US equity market, in particular those using principal components to aggregate the information in the macro-financial variables, provide a statistical improvement over several benchmarks. In addition, these generate economic value by boosting returns, improving the sharp ratio and the omega, and substantially reducing drawdowns.

JEL Codes: C53; G11; G17.

Keywords: Dating Algorithms; Markov-Switching Models; Predictions; Principal Component Analysis; Specific-to-General Approach; Stock Market Recessions.

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

1 Introduction

The existence of different stock market regimes is widely accepted among academics and practitioners. Stock market cycles typically precede business cycles and are caused by time-varying expectations of future cash flows and discount rates. In bullish periods, prices rise and fluctuate only mildly, whereas in bearish periods, prices decrease and volatility increases. Hence, anticipating regime jumps and, in particular, recessions 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.

An early detection of warning signals is therefore crucial to reduce potential welfare losses and might initiate accommodative fiscal and monetary policies. From a regulator's perspective, the stock market's state helps anticipating instabilities in the financial sector triggered by a decrease in collaterals and an increase in default rates. Recession signals also influence the strategy of companies that seek to compensate projected revenue declines with cost-reducing measures and job cuts. Finally, the state of the stock market affects the consumption-savings decision of all economic agents. Motivated by all these reasons, this paper examines the extent to which macroeconomic and financial market variables are suitable for forecasting stock market recessions. In addition, we develop early warning models by using a large set of variables.

In this context, the first problem is how to identify the unobservable state of the stock market. Despite the common understanding of bull and bear markets, there is no consensus on how much stock prices must fall or which volatility change is required for a bear market. A further complication is that the current state of the stock market cycle can only be determined with a time lag (Nyberg 2013). We propose a two-step procedure to meet this challenge. In the first step, we apply Markov-switching (MS) models, dating rules, and the historical moving average to distinguish between bearish and bullish phases in the US stock market (for a similar approach, see Chen 2009). Thereafter, we treat the extracted recession probability or the recession signal as a dependent variable to forecast market regimes conditional on several predictors. To solve the problem of selecting appropriate variables out of many potential predictors, we apply two techniques: (i) we approximate latent factors for the underlying recession risk with a principal component analysis (PCA) and (ii) we employ a specific-to-general approach.

Our results suggest that several macro-financial variables are suitable for forecasting recessions in stock markets in-sample and out-of-sample. Hereby, risk spreads of bond markets, the VIX, the leading indicator of the Conference Board, and the business conditions of the Federal Reserve Bank of Philadelphia can be highlighted. In addition, the first and second moments of the forecasts by Consensus Economics help to predict the recession probability. Our early warning models for the US equity market, in particular those using principal components to aggregate the information in the macro-financial variables, provide a statistical improvement over several benchmarks. In addition, these generate economic value by boosting returns, improving the sharp ratio and the omega, and substantially reducing drawdowns. Finally, although dynamic model specifications are preferable from a statistical point of view, the market entry and exit around the GFC of 2007−2008 can be better anticipated with static models.

Our paper contributes to the literature on stock market forecasting. Previous papers have identified particularly relevant variables, for example, Lewellen (2004) advocates financial ratios or Rapach et al. (2005) recommend interest rates to forecast returns. However, other studies conclude that the predictive power of macroeconomic or financial variables is weak (see, among others, Welch and Goyal 2008). A major reason for the mixed results is the instability of predictors and expected returns with respect to the state of the business cycle (Hamilton and Lin 1996; Perez-Quiros and Timmermann 2000; Henkel et al. 2011) or the level of market volatility (Pesaran and Timmermann 1995). From a theoretical perspective, it is typically argued that changes in risk aversion cause time-varying expected returns. Campbell and Cochrane (1999) incorporate this behavior in a consumption-based capital asset pricing model and explain varying expected returns with macroeconomic cycles.

Taking these findings into account, forecasting regimes rather than returns has attracted more attention. For stock markets, three methods have been established in the literature. First, 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, MS models are used to estimate the regime probabilities and the corresponding first and second moments (Ang and Bekaert 2002; Kritzman et al. 2012). Hereby, the number of regimes is still subject to debate (see, for instance, Maheu et al. 2012; Hauptmann et al. 2014). Third, for dating rules, local extremes are defined by period lengths (Pagan and Sossounov 2003) or regime changes are marked by absolute price changes (Lunde and Timmermann 2004). The underlying algorithms need past and future prices for the dating of recession and, consequently, delayed signals may occur.

The remainder of this paper is organized as follows. Section 2 outlines the three different identification approaches. Section 3 introduces the dataset of macro-financial variables. Section 4 explains the different predictive models. Section 5 discusses the results of the identification and the predictions assuming knowledge of the full sample. Section 6 repeats this exercise in a real-time situation with recursive identification and recursive out-of-sample forecasts. Section 7 concludes.

2 Identification of Stock Market Recessions

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 recession 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 recession. 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).

According to Kole and Van Dijk (2017), differences between these two methods can be summarized as follows. Dating rules are based on more or less arbitrary definitions of price thresholds or durations and result in a binary indicator. In contrast, MS models directly estimate a certain regime probability. Non-parametric dating rule approaches focus on price changes and do not rely on distributional assumptions. Hence, these date bull and bear markets in a more transparent manner and are less prone to misspecification. However, these require future information to properly classify turning points, which leads to the risk that the current market regime can only be identified with a delay of several months. This can be a clear disadvantage in real-time applications as compared to MS models.

2.1 Markov-Switching Models

Since the pioneering work of Hamilton (1989), MS models became increasingly popular in economics. The underlying idea is that the latent regime variable S_t and the respective regime-dependent return r_t follow certain stochastic processes. Consequently, it is possible to estimate the individual state probabilities and to detect regimes. 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). In this context, the idea of mixture distributions is particularly helpful. For instance, a probability-weighted combination of two normal distributions with different means and variances might result in an i.i.d. mixed distribution that captures stylized facts of financial time series like fat tails, asymmetries, and volatility clustering (Ang and Timmermann 2012). Such properties help to account for time-varying risk premia and to separate the observed returns into different cycles.

Let r_t denote the log-return of the S&P 500 index in time t. A simple MS model can be used to describe regime-dependent returns:

$$r_t = \mu_{S_t} + \sigma_{S_t} u_t \tag{1}$$

with $u_t \sim i.i.d.\ N(0,1)$. The mean μ_{S_t} and the standard deviation σ_{S_t} depend on the current market regime S_t. S_t is an unobservable variable that follows a discrete, first-order Markov chain. Hence, the probability that the current regime S_t is in state j depends only on the most recent value of S_{t-1} and the transition probability p_{ij}.

Since S_t is a latent variable, the true number of regimes is unknown. An approximation with econometric tests is difficult because these tests do not follow standard distributions (Hansen 1992; Ang and Timmermann 2012). Therefore, a theoretically justified assumption is recommended. Our choice to consider two regimes refers to the detection of turning points in the business cycle literature (Hamilton 1989). Furthermore, a separation into two regimes is consistent with the number of regimes in the non-parametric approaches. In general, this assumption is an accepted and a frequently applied approach in the empirical literature (Chen 2009). Consequently, we rely on a two-regime model with $S_t = 1$ representing a stock market recession and $S_t = 0$ denoting an expansion. Typically, a bear market exhibits a negative mean and high volatility, while a bull market is characterized by a positive mean and lower volatility.[1]

We assume that the Markov chain is time-homogeneous.[2] Hence, the fixed transition probabilities are represented by the matrix $\mathbf{P}$:

$$\mathbf{P} = \begin{bmatrix} P(S_t = 0|S_{t-1} = 0) & P(S_t = 0|S_{t-1} = 1) \\ P(S_t = 1|S_{t-1} = 0) & P(S_t = 1|S_{t-1} = 1) \end{bmatrix} = \begin{bmatrix} p_{00} & p_{10} \\ p_{01} & p_{11} \end{bmatrix} = \begin{bmatrix} p_{00} & 1 - p_{11} \\ 1 - p_{00} & p_{11} \end{bmatrix} \tag{2}$$

In addition, we assume that there are only temporary changes in the time series structure (Hamilton 1994), which means that no absorbing state exists ($p_{00} < 1$ and $p_{11} < 1$). To complete the notation, let $\delta = P(S_1 = 0)$ denote the initial probability for the first regime. Then, the mean-variance MS model is completely described by the following vector of parameters:

$$\theta = (\mu_0, \mu_1, \sigma_0, \sigma_1, p_{00}, p_{11}, \delta)'$$

[1] We also tried to estimate regime-switching models with more than two regimes. However, the estimation of these models does not converge in a reliable manner due to a rather flat likelihood function, in particular in the out-of-sample exercises.

[2] It is also possible to assume time-varying transition probabilities. In that case, the probabilities are explicitly modeled via a discrete choice model conditional either on exogenous factors (Diebold et al. 1994) or on the time spent in a particular regime (Maheu and McCurdy 2000).

We estimate θ with maximum likelihood (ML) methods using the expectation maximization algorithm since the likelihood function depends on the unobservable market state S_t.[3]

In addition to the standard MS model in Eq. (1), we estimate a Markov-switching autoregressive (MSAR) model of order one. Accounting for short-run dependencies might improve the model's ability to distinguish between recessions and expansions. Furthermore, the question of whether returns correlate over time is of particular interest in the empirical financial literature (see, among others, Poterba and Summers 1988).

Finally, we have to make a decision about the threshold for the regime classification. In the case of two regimes, a threshold of 50% in the filtered probability seems straightforward. However, other thresholds are also conceivable since, for example, differences in the degree of risk aversion might be taken into account. Consequently, we also utilize a threshold of 25% as part of our robustness tests, thereby incorporating a higher degree of recession aversion.

2.2 Dating Algorithms

Dating rules also have their origin in the business cycle literature and go back to the BB algorithm by Bry and Boschan (1971). These serve as an identification method to detect turning points. Since the relationship to business cycles is very close (Hamilton and Lin 1996; Estrella and Mishkin 1998), it makes sense to use such methods in the context of stock market cycles, too.

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. The original BB algorithm defines minimum period lengths of recessions, expansions, and the entire cycle. Pagan and Sossounov (2003) adjust the BB dating rule to stock market characteristics. In particular, smoothing is not recommendable for the equity market where outliers count. Their algorithm (PS henceforth) can be described with the following steps:

1. Location of initial turning points:

a) Determine all local highs and lows in a price series. A local maximum (minimum) is identified when it has a higher (lower) price than all past and future prices available within an 8 month time window (τ_{window}) in both directions.

[3] 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, 2010).

b) Apply an alternating procedure that selects the highest peak and the lowest trough.

2. Censoring:

a) Eliminate all peaks and troughs of the first and last 6 month (τ_{censor}).

b) Eliminate market cycles that persist less than 16 months (τ_{cycle}).

c) Eliminate bull or bear market phases that last less than 4 months (τ_{phase}), unless the price rises or falls by a more than 20% (ω).

An alternative dating rule is provided by Lunde and Timmermann (2004). They filter local trends according to absolute price changes. Their identification procedure (LT henceforth) focuses on extreme events in the price series P_t and is as follows:

1. Given that the last observed extreme was a local maximum, referred to as P^{max}, the subsequent price series are 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 15% (λ_2) 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 are 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 20% (λ_1) or more.

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

Both methods produce a binary indicator that signals an expansion ($D_t^{PS,LT} = 0$) or a recession ($D_t^{PS,LT} = 1$). The key difference between both methods is that the PS rule is based on the specification of minimum lengths of phases and cycles, whereas the LT rule relies exclusively on price changes. As indicated at the beginning of Section 2, both procedures require past and future information for an accurate identification and might identify the current market regime with a time lag in real-time applications.

2.3 Naïve Approach

As third alternative, we use a simple moving average (MA) to detect local trends in the return series. The underlying assumption is that the current market state can be approximated by the average realized return over a certain rolling period. The MA is often applied as an indicator for market timing decisions (see, among others Brock et al. 1992). Hence, the MA rule can be seen as a naïve benchmark identification procedure. The moving average with length L is given by:

$$MA(L) = \frac{\sum_{l=1}^{L} r_{t-l}}{L} \tag{3}$$

To separate the smoothed performance into two regimes we define the binary variable D_t^{MA}:

$$
\begin{aligned}
D_t^{MA} &= 0 \quad \text{if} \quad MA(L) \geq 0 \quad \text{as an expansion phase} \\
D_t^{MA} &= 1 \quad \text{if} \quad MA(L) < 0 \quad \text{as a recession phase}
\end{aligned}
$$

To ensure smoothed cycles we set $L = 16$, which also corresponds to the recommended length in the PS dating rule. 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.

One advantage of the naïve approach is that we are able to classify the market regime with a short time lag only, that is, immediately after observing the returns of the previous month. However, stock returns must feature some persistence so that the regime detection based on this approach is not misleading. For the rest of this paper, we use the MA(16) as a benchmark (i) to assess the performance of the identification via MS models and dating rules and (ii) to signal bearish and bullish phases for a trading strategy in the out-of-sample forecasts.

3 Data

Our dataset consists of monthly 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 92 variables to predict regimes in the stock market. This involves bond market data, analysts' expectations, surveys of households and companies, and fundamental factors. To the best of our knowledge, this is the first paper that uses such a large number of variables to forecast bear markets. Other papers predicting market typically regimes rely on a small number of predictors (see, for instance, Chen 2009; Kole and Van Dijk 2017). However, as we outline below, a much larger number of variables might be justified from a theoretical point of view.

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 and over inflation-linked bonds. Since stock market recessions are often induced by an increase in risk aversion, credit spreads might also be useful in this context (Coudert and Gex 2008). Hence, we take the corporate bond spreads from Moody's and the TED spread into account. As additional predictors, we consider the volatility in the S&P 500 over the previous six months and the VIX, the latter of which reflects expected future volatility and a risk premium (Bekaert and Hoerova 2014). Further indicators that capture changes in risk perception are the gold price, the oil price, and the JPY/AUD exchange rate.[4]

Second, we utilize survey-based expectations as predictors. Consensus Economics asks analysts from banks or 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.[5] In addition, we employ sentiment measures, such as the surveys by the Conference Board and the University of Michigan. Following the idea of 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 Purchase Manager Index (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.

Third, the current valuation level is typically related to stock market turbulences (Lewellen 2004). Hence, we use the price-over dividend ratio (P/D), the price over earnings ratio (P/E), and the price over the cyclical adjusted 10-year earnings ratio (CAPE). Moreover, we incorporate the most recent S&P 500 returns and the six-month moving average of the returns in our predictor set. It might be argued that price "excesses" are a major cause of future recessions, which suggests that valuation ratios or historical returns correlate negatively with the risk of recession.

The length of our sample period varies according to the objective, that is, in-sample regime identification, in-sample regime prediction, and out-of-sample regime identification and prediction. The period January 1950–June 2019 is used to identify stock market regimes with full sample knowledge. For the purpose of predicting stock market regimes within our sample, we focus on the period October 1989–May 2019 due to data availability issues for some of the predictors. Finally, our out-of-sample real-time exercise is executed with the help of the most recent 175 months. The first training set for the identification of stock regimes and the forecast for the subsequent month ends in October 2004. Starting from this month, we employ a recursive scheme for regime identification and the forecasts of recession risk. In all cases, we rely on end-of-month

[4]The JPY is often assumed to be a safe haven and the AUD is substantially affected by commodity demand, which is increasing during economic expansions.

[5]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 monthly changes in stock market regimes. In the case of the macroeconomic forecasts, we calculate the weighted averages of the current-year and the next-year forecast to create a fixed-horizon forecast. For example, the weighted forecast for March is as follows: $Forecast_3^{FH} = 9/12 \times Forecast_3^{CY} + 3/12 \times Forecast_3^{NY}$.

data (if the data is available at a higher frequency) and all variables are shifted to their publication month to ensure a real-time perspective.

Table A1 in Appendix A lists all variables, alongside their definitions and sources. Table A2 shows descriptive statistics and indicates whether first differences have to be formed to ensure stationarity of the predictors.

4 Prediction Models

Next, we model the relationship between the future recession probability of S_t (or the stock market state D_t) and the predictors. In the case of the MS models, we explain the filtered probability $P(S_t = 1|\Omega_t;\hat{\theta})$ in a predictive linear regression. For the dating rules, we apply a probit approach to link the predictors to the binary recession indicator D_t. In both cases, we employ (i) static models and (ii) dynamic models to take the serial dependency of the recession probability or the state and the persistence in the market conditions into account. For the in-sample analysis, we assume knowledge of the complete data set, whereas for the out-of-sample forecasts, the recession probability or signal is sequentially updated for each month in real-time.

4.1 Linear Models for MS Identification

Following Chen (2009), we first consider a static predictive regression framework using the filtered probability of the recession state $Y_t = P(S_t = 1|\Omega_t;\hat{\theta})$:

$$Y_t = \beta_0 + \sum_{i=1}^{k} \beta_i \mathbf{x}_{i,t-1} + \epsilon_t \tag{4}$$

The number of predictors is indicated by k, which is either one in the case of the bivariate models or greater than one in the case of the multivariate models. β_i measures how the recession risk is related to the corresponding factor, β_0 denotes a constant term, and ϵ_t is the error term. Eq. (4) is estimated with ordinary least squares.

In a second step, we estimate dynamic ARMA(p,q) models and allow for persistence in the errors and in the lagged dependent variable (that is, the recession probability). Since the MA part is unobservable, the dynamic models are estimated using ML methods. In both, the static case and the dynamic case, we ensure that the predicted recession probabilities $\hat{Y}_t$ are within the interval [0,1] and censor values higher (lower) than 1 (0) after the fitting process.

4.2 Probit Models for Identification via Dating Algorithms

The dating rules approximate the market state either as bullish $D_t = 0$ or bearish $D_t = 1$. Discrete choice models help to translate the dummy variable D_t into a latent variable D_t^*. Under the assumption that the errors η_t follow a standard normal distribution, a static probit model can be written as follows:

$$D_t^* = \gamma_0 + \sum_{i=1}^{k} \gamma_i \mathbf{x}_{i,t-1} + \eta_t \tag{5}$$

The probabilities of the different outcomes of the underlying dummy variable can be obtained as follows (with Φ denoting the cumulative standard normal distribution):

$$\begin{aligned} Prob[D_t = 1|\mathbf{x}_{i,t-1}] &= \Phi\left(\mathbf{x}_{i,t-1}'\gamma\right) \\ Prob[D_t = 0|\mathbf{x}_{i,t-1}] &= 1 - \Phi\left(\mathbf{x}_{i,t-1}'\gamma\right) \end{aligned}$$

In a second step, we account for persistent states with a dynamic model. Since the dating rule approaches face the risk that a recession is identified in real-time only with a certain delay, we include the six-month lagged left hand-side variable as additional regressor in the out-of-sample exercises (Nyberg 2013). For the in-sample analysis, we include the previous period's state as additional covariate.

Eq. (5) and its dynamic extension are estimated using ML methods. The problem of complete or quasi-complete separation in the estimation is addressed with the reduced-bias estimator for generalized linear models (Kosmidis et al. 2018).

4.3 Selection of Multivariate Models

The number of 92 theoretically plausible predictors entails a selection problem in the empirical prediction of stock market recessions. On the one hand, including additional predictors improves the model's fit and mitigates a potential bias. On the other hand, each additional variable leads to a higher estimation variance due to multicollinearity between the regressors. Solving this trade-off between bias and variance implies finding a model that is as parsimonious as possible while, at the same time, capturing emerging recession signals as good as possible.

In this paper, we utilize two selection techniques: (i) dimension reduction via a principal component analysis (PCA) and (ii) a specific-to-general approach (StG) based on the Akaike information criterion (AIC). Employing regressors based on a PCA leads to parsimonious models by efficiently aggregating the information in the data, which might come at the expense of the economic interpretation of the aggregated components. In contrast, an StG approach typically leads to larger and more volatile models that, however, have a straightforward interpretation.

We utilize both selection methods for the in-sample predictions and the out-of-sample forecasts. In the latter case, we employ (i) a *fixed* selection procedure (that is, we select the components/predictors before the first forecast and keep these fixed in the recursive scheme) and (ii) a *continuously updated* selection procedure (that is, we update the selected components/predictors after each month).

4.3.1 Principal Component Analysis

Among others, Stock and Watson (2002) propose a PCA to approximate unobserved common factors to deal with a large number of (highly) correlated predictors in macroeconomic forecasting. In our paper, we approximate latent factors that are able to predict the underlying structure of the recession probability. These factors are linear combinations of the centered predictors and are ranked in descending order according to their explained variance.

Next, the number of considered indices in the predictive regression analysis is determined with a scree plot. To obtain our first forecast in the out-of-sample exercise, we use the latent risk factors and the corresponding model estimates based on the data available in October 2004 and predict the recession probability for November 2004. Using the data available in November 2004, we obtain the predicted recession probability for December 2004. This procedure is repeated for the entire test window, where the factor loadings are either fixed or continuously updated in the recursive scheme.

4.3.2 Specific-to-General Approach

As an alternative selection method, we propose a stepwise selection based on the AIC to deal with the trade-off between bias and variance. The AIC consists of the estimated log-likelihood (LL) and the number of parameters k:

$$AIC = -2LL + 2k$$

Although we do not know the true data generating process, the AIC typically finds a good approximation of the underlying process. Hence, it is a popular choice for model selection with the purpose of forecasting. Starting from a model that only contains a constant, one sequentially adds individual variables. The variable that most improves the model with respect to the AIC is selected. Through this process, the model is extended stepwise until no further reduction in the AIC can be achieved.[6]

[6]It has to be noted that there is no guarantee that the final model represents the best possible combination of predictors according to the AIC. The best subset method that tests all possible combinations to find the optimum model, however, is practically infeasible for large numbers of predictors (Hastie

Due to the low penalty for additional variables ($2k$), the AIC tends to choose a relatively large model, which might lead to additional noise in the forecast. Consequently, we first apply a pre-selection and calculate bivariate correlations between the predictors and the recession probability.[7] Second, we sort all variables in descending order according to their absolute correlation and select the variables with the 10% highest correlations in each recession period. Finally, we start the selection based on AIC from this candidate set. After the completion of the selection task, we obtain the one-step ahead predictions. In the out-of-sample exercises, the predictive model is re-estimated after each month and the one-step ahead forecasts are calculated using either a fixed selection or a continuously updated selection.

5 Full Sample Results

5.1 Identification

We start our discussion of the results with the identification of stock market recessions according to the three approaches presented in Section 2 and the full history of the adjusted S&P 500 price, beginning in January 1950 and ending in June 2019.

Table 1: Results of Markov-Switching Model for the Full Sample Period

	μ_0	μ_1	σ_0	σ_1	ϕ_0	ϕ_1	p_{00}	p_{11}	AIC
MS	1.06***	−0.92	3.20	6.14			0.96	0.85	−3025.1
	(0.16)	(0.88)							
MSAR	1.20***	−0.73	3.15	5.91	−0.088*	0.083	0.96	0.87	−3020.2
	(0.18)	(0.79)			(0.049)	(0.079)			

Notes: Table shows coefficients of the simple MS model and the MSAR model with standard errors in parentheses. *, **, and *** indicate significance at the 10%, 5%, and 1% level, respectively.

Table 1 suggests that the first two moments of the historical return distribution are indeed regime-dependent. The two-state MS models identifies bull and bear markets. Bullish periods have strong positive returns (μ_0) and lower volatilities (σ_0), whereas bearish periods are characterized by negative returns (μ_1) and high fluctuations (σ_1). In addition, returns in upward markets are characterized by a negative AR(1) coefficient (ϕ_0), while in downward markets returns are persistent (ϕ_1). In both, the MS and

et al. 2009). Consequently, the StG approach based on the AIC is often applied in empirical papers for the reason of its feasibility (see, for instance, Hauptmann et al. 2014).

[7] If the dependent variable is binary, we first estimate bivariate linear probability models and then select the predictors according to their R^2. For the out-of-sample case, we limit this pre-selection to the period up to September 2004. To capture emerging and declining signals, we are extending the recession periods by six (two) months in both directions for the out-of-sample predictions (in-sample case).

MSAR model the regimes are highly persistent (indicated by the large probabilities p_{00} and p_{11}), with bull markets lasting, on average, about three times as long as bear markets.[8] Finally, even though the estimated parameters of the two switching models are very similar, the simple MS fits the data slightly better than the MSAR as indicated by the lower AIC.

Figure 1 presents the identification results of the different methods.[9] At a first glance, many recession periods coincide in the five approaches. These include the market declines caused by the oil crisis at the beginning of the 1970s, the bursting of the dot-com bubble (2000–2002), and the GFC (2007–2008). Nevertheless, there are some discrepancies among the procedures in the exact timing. In particular, the MA(16) approach identifies substantially more turning points, which are often delayed or difficult to explain (see also Table B5 in Appendix B). Hence, the naïve indicator is inferior to the model and rule-based identification.

The switching models and the dating rules also differ in some important cases. For example, the major recovery in spring 2009 is still erroneously identified as a recession in both switching models, whereas the dating rules correctly identify the turning point. Another example is the turbulence at the end of 2018, where the model-based approaches estimate a recession that is still ongoing at the end of the sample period despite the recovery. The finding that dating rules detect turning points in the stock market more accurate than any other approach is not particularly surprising since these also use future information in the identification. Both switching models offer, despite their weaknesses, substantial value added over the naïve moving average approach.

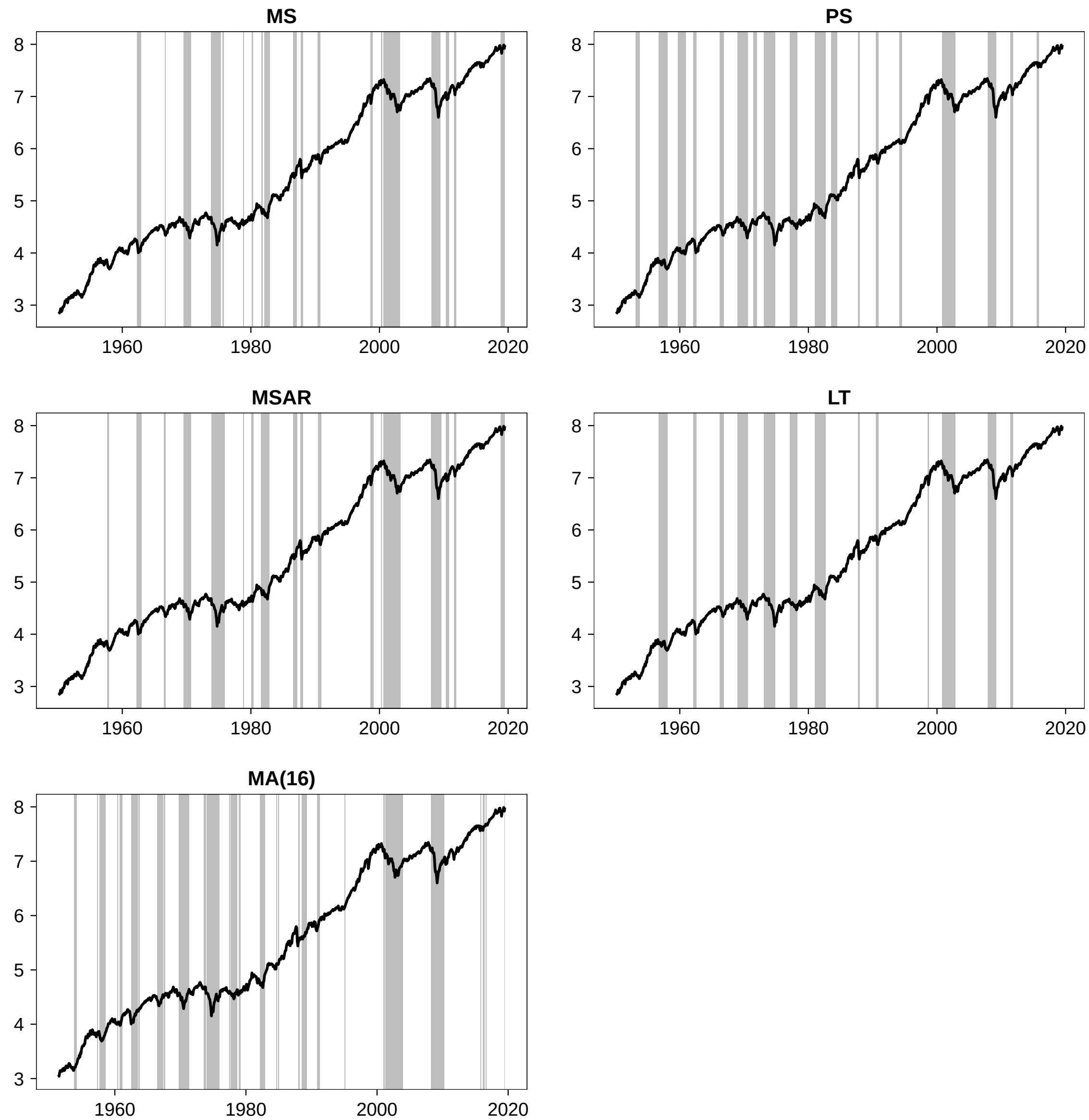
Table 2 shows some descriptive statistics of the identified market states. The average recession duration of model-based identification is 7.7 months (MS) and 9.9 months (MSAR) and is shorter than those of the rule-based approaches (PS: 12.3 months; LT: 12.6 months). A bear market is, on average, marked by a price decline of 24% (PS) and 29.4% (LT) in the rule-based approaches, whereas the average price decline for MS models is 11.3% (MS) and 10% (MSAR). These substantial differences result from the fact that the switching models consider the first and second moments of the stock market, whereas the dating rules focus on the first moment only.[10] When comparing both dating rules, it is noticeable that the PS algorithm recognizes more turning points than the LT filter. Consequently, the localization of extreme points is often based on price changes smaller than the LT thresholds of 20% and 15%.

[8] The average duration can be expressed as $1/(1-p_{ij})$.

[9] Tables B1–B5 in Appendix B provide the detailed timing of all bearish and bullish phases for the five approaches.

[10] One example of these differences is the period from July 1986 to February 1987, during which high volatility was accompanied by market growth of 20% (see Tables B1 and B2 in Appendix B).

Figure 1: Full-Sample Recession Identification



Notes: Figure shows the log S&P 500 index and the identified recession periods as gray-shaded areas. Top left: identification according to the simple MS model; top right: PS dating rule; middle left: MSAR model; middle right: PS dating rule; bottom left: naïve approach. The filtered recession probabilities of the MS model and the MSAR model can be found in Figure B1 in Appendix B.

Table 2: Summary Statistics of Bullish and Bearish Phases

	MS		**MSAR**		**PS**		**LT**		**Naïve**	
	Bull	Bear	Bull	Bear	Bull	Bear	Bull	Bear	Bull	Bear
Frequency	19	19	19	18	19	18	14	13	31	31
Duration (in Months)										
… Min	2	1	3	1	10	3	24	2	1	1
… Average	26.94	7.65	33.65	9.94	32.00	12.31	41.75	12.64	19.10	6.76
… Median	16	4	33	6	30	12.5	31.5	14	16	2
… Max	93	31	92	32	74	25	92	25	70	32
Amplitude (in %)										
… Min	0.15	−2.41	−2.93	−1.53	25.60	−7.75	41.55	−15.57	−11.39	−0.39
… Average	44.76	−11.31	53.10	−9.97	76.57	−23.95	96.24	−29.39	16.07	4.16
… Median	27.23	−8.73	33.88	−8.63	61.44	−19.02	69.38	−23.79	2.02	3.88
… Max	270.5	−41.17	273.0	−42.17	241.6	−52.56	273.0	−52.56	179.5	−19.94

5.2 Predictions

Next, we assess the relationship between different predictors and the recession risk assuming full-sample knowledge. We start with one-factor models for the filtered probabilities and for the binary recession indicators. Thereafter, we show the results of the multivariate models, using either PCA or an StG approach to select promising predictors. All prediction evaluation measures are explained in Appendix C.

5.2.1 Bivariate Models

In the case of the bivariate predictions, we only consider static models with Newey and West (1987) standard errors. We find that over a third of the covariates explain less than 1% of the recession probability. This includes supposedly promising candidates such as the 10-year treasury bonds or some term spreads. On the other hand, the t-test (or z-test), can be rejected for about half of the variables at a significance level of 10% for all four identification approaches. The ten models with the best (Pseudo-)R^2 are shown in Table 3.[11]

Irrespective of the identification approach, market-based variables explain the fluctuations of recession risk quite well. Particularly noteworthy are uncertainty measures such as the VIX and historical volatility as well as credit spreads between corporate and government bonds. The VIX accounts for more than 50% of the variability of the filtered recession probability (top panel of Table 3). Another interesting result is that, in contrast to the mean forecasts, the standard deviations of the consensus interest rate forecast provide substantial predictive power, in particular for the probit models (bottom panel of Table 3). Finally, three additional variables can be highlighted: (i)

[11] To conserve space, we do not show the results for the full list of covariates. All omitted results are available on request.

the leading indicator of the Conference Board, (ii) the survey of manufacturers by the Philadelphia Fed, and (iii) the Consensus unemployment forecast.

Table 3: Top-10 In-Sample Predictors According to Their (Pseudo) R^2

MS				**MSAR**			
Predictor	β_1	p-val.	R^2	Predictor	β_1	p-val.	R^2
VIX	0.03	0.00	0.538	VIX	0.03	0.00	0.555
Vola_6m	9.97	0.00	0.438	Vola_6m	10.71	0.00	0.473
SMA_6m	−9.32	0.00	0.376	SMA_6m	−9.45	0.00	0.362
RS_GOVBAA	0.19	0.00	0.293	RS_GOVBAA	0.21	0.00	0.325
RS_AAABAA	0.35	0.00	0.247	RS_AAABAA	0.38	0.00	0.280
Bus_Cond_FEDphil	−0.01	0.00	0.241	Bus_Cond_FEDphil	−0.01	0.00	0.274
unemp	1.07	0.00	0.221	unemp	1.14	0.00	0.232
RS_GOVAAA	0.25	0.00	0.205	RS_GOVAAA	0.27	0.00	0.224
Lead_Conf	−0.21	0.00	0.201	Lead_Conf	−0.22	0.00	0.211
stock_exp_lower	0.01	0.00	0.178	stock_exp_lower	0.02	0.00	0.210

PS				**LT**			
Predictor	β_1	p-val.	Ps. R^2	Predictor	β_1	p-val.	Ps. R^2
SMA_6m	−36.64	0.01	0.188	Lead_Conf	−1.10	0.00	0.177
i3m.3m.sd	6.61	0.00	0.148	SMA_6m	−34.44	0.01	0.168
Lead_Conf	−0.90	0.00	0.133	Bus_Cond_FEDphil	−0.04	0.00	0.165
Bus_Cond_FEDphil	−0.03	0.00	0.131	i3m.3m.sd	6.76	0.00	0.148
i3m.12m.sd	3.31	0.00	0.108	unemp	4.67	0.00	0.125
CAPE	−0.54	0.00	0.107	VIX	0.07	0.00	0.124
unemp	4.15	0.00	0.106	Int_exp_lower	0.06	0.00	0.121
Int_exp_lower	0.05	0.00	0.091	i3m.12m.sd	3.59	0.00	0.117
VIX	0.06	0.00	0.091	Int_exp_higher	−0.05	0.00	0.116
RS_GOVAAA	0.93	0.01	0.089	pmi	−0.11	0.00	0.115

Notes: Table shows selected results of in-sample bivariate predictive least squares (top panel) and probit (bottom panel) regressions with Newey and West (1987) standard errors.

To summarize, an increase of uncertainty (VIX, Vola_6m, and bond's risk spread) or forecast dispersion (i3m.3m.sd and i3m.12m.sd) as well as negative growth prospects (Lead_Conf, Bus_Cond_FEDphil, and unemp) signal a higher recession risk. To further illustrate how the four best variables for each identification approach behave over time, Figures D1 and D2 in Appendix D plot these against the recession periods. In some cases, the individual variables anticipate stock market recessions quite well, in other cases these perform poorly. Hence, a model-based approach combining the information of different variables might improve the predictive performance.

5.2.2 Multivariate Models

Principal Component Regressions. The first four components account for around 56% of the total variation.[12] Table D1 in Appendix D shows the top-10 loadings of all four components. The first component mainly represents the yield curve and covers 28% of the total variation. The second component is dominated by macroeconomic variables accounting for the current fundamental environment (for instance, expected GDP growth and its main constituents, the PMI, or business conditions). The third component is driven by term spreads that approximate future uncertainty. Finally, the fourth component covers the valuation level of the stock market and the consumer climate. From a theoretical perspective, it appears plausible that the approximated four factors drive the expected future cash flows and discount rates. Hence, these are good candidates to describe the state space of stock market dynamics.

Table 4: Results of Linear In-Sample PCR Models

	MS st.	MS dyn.	MSAR st.	MSAR dyn.
Constant	0.232***	0.234***	0.256***	0.258***
	(0.019)	(0.035)	(0.019)	(0.038)
AR(1)		0.775***		0.796***
		(0.042)		(0.041)
PC 1	−0.005	−0.003	−0.006*	−0.003
	(0.003)	(0.006)	(0.003)	(0.007)
PC 2	0.040***	0.018**	0.044***	0.017**
	(0.005)	(0.008)	(0.006)	(0.008)
PC 3	0.019**	0.018*	0.019**	0.017
	(0.008)	(0.010)	(0.008)	(0.011)
PC 4	0.031***	−0.003	0.028***	−0.005
	(0.009)	(0.005)	(0.010)	(0.005)
Adj. R^2	0.417	0.696	0.438	0.721
AIC	−101.26	−329.55	−91.12	−336.25
BIC	−78.04	−302.47	−67.90	−309.16

Notes: Table shows coefficients of least squares models with standard errors in parentheses. Static models are estimated with Newey and West (1987) standard errors. *, **, and *** indicate significance at the 10%, 5%, and 1% level, respectively.

Next, we estimate principal component regression (PCR) models to assess the performance of the approximated factors. Tables 4 and 5 show the results. PC 1 (representing the yield curve) does not exert much influence, since only one model (MSAR static in Table 4) indicates an estimate that is significantly different from zero. The third component (representing term spreads) has a significant positive correlation with

[12]The number of principal components is based on a visual inspection of the scree plot. Another popular decision criterion for determining number of components is the Kaiser criterion. However, since 17 components have an eigenvalue greater than one, we rely on the more parsimonious specification with four components.

the future recession probability only in the case of the linear models (see Table 4). Overall, PC 2 (representing fundamentals) and PC 4 (representing valuation and consumer climate) provide the most robust relationships. However, their impact decreases once we enhance our models with a dynamic component. In all models, the dynamic component itself has a strong impact on the recession risk, which is not surprising given the persistence of stock market regimes. Finally, the linear PCRs explain up to 72% (44%) of the future recession risk variability in the dynamic (static) case. The corresponding Pseudo R^2 figures of the probit PCR models are 66% (dynamic) and 26% (static).

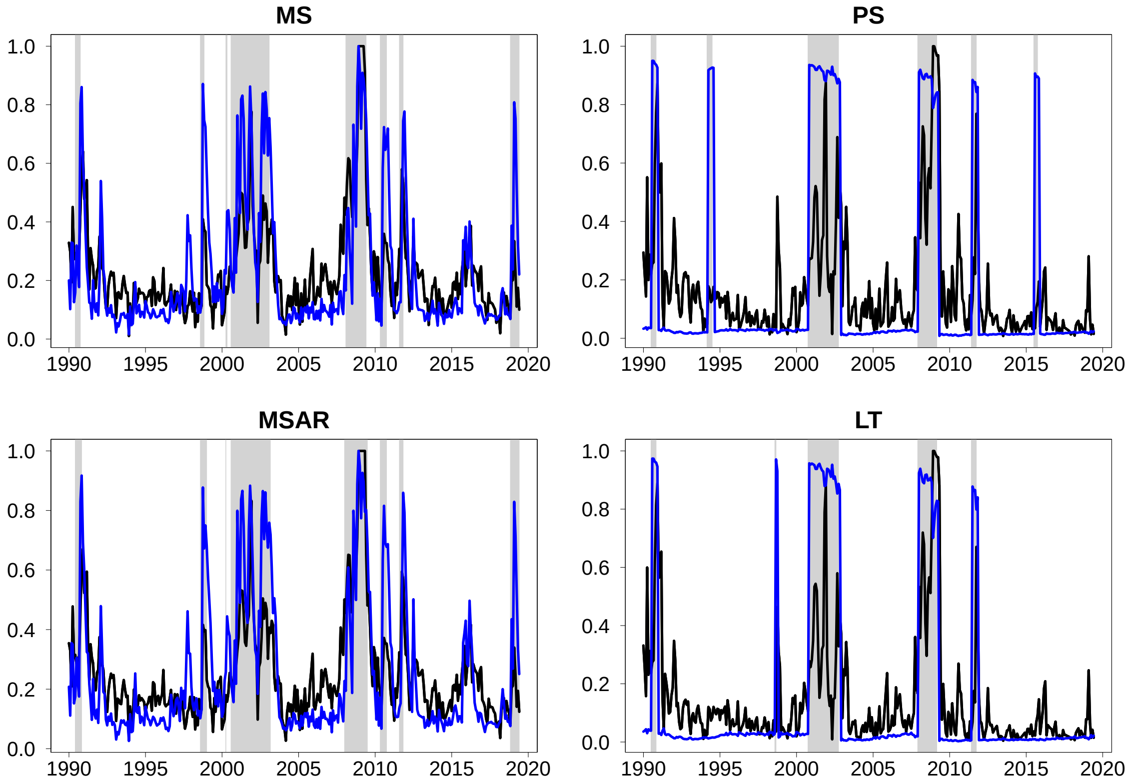
Table 5: Results of Probit In-Sample PCR Models

	PS st.	PS dyn.	LT st.	LT dyn.
Constant	-1.147^{***}	-2.082^{***}	-1.285^{***}	-2.216^{***}
	(0.160)	(0.175)	(0.167)	(0.196)
D_{t-1}		3.477^{***}		3.734^{***}
		(0.361)		(0.423)
PC 1	0.012	0.025	0.024	0.041
	(0.030)	(0.027)	(0.030)	(0.031)
PC 2	0.165^{***}	-0.013	0.181^{***}	-0.018
	(0.056)	(0.033)	(0.064)	(0.036)
PC 3	0.050	0.013	0.088	0.025
	(0.061)	(0.053)	(0.067)	(0.060)
PC 4	0.229^{***}	-0.031	0.206^{***}	-0.060
	(0.059)	(0.063)	(0.066)	(0.066)
Pseudo R^2	0.227	0.662	0.254	0.664
AIC	249.86	108.25	218.43	92.68
BIC	269.20	131.46	237.77	115.89

Note: Table shows coefficients of probit models with standard errors in parentheses. Static models are estimated with Newey and West (1987) standard errors. *, **, and *** indicate significance at the 10%, 5%, and 1% level, respectively.

Figure 2 shows the fitted recession probabilities of the static (black lines) and the dynamic model specifications (blue lines). The increase in recession probability typically occurs with considerable time lag in the case of the MS models and the static probit models. The dynamic component in the probit models ensures that recessions are detected at an earlier stage and that the probabilities are bipolar, providing a sharper distinction between bullish and bearish phases.

Figure 2: Fitted Recession Probabilities Based on PCA Regressions

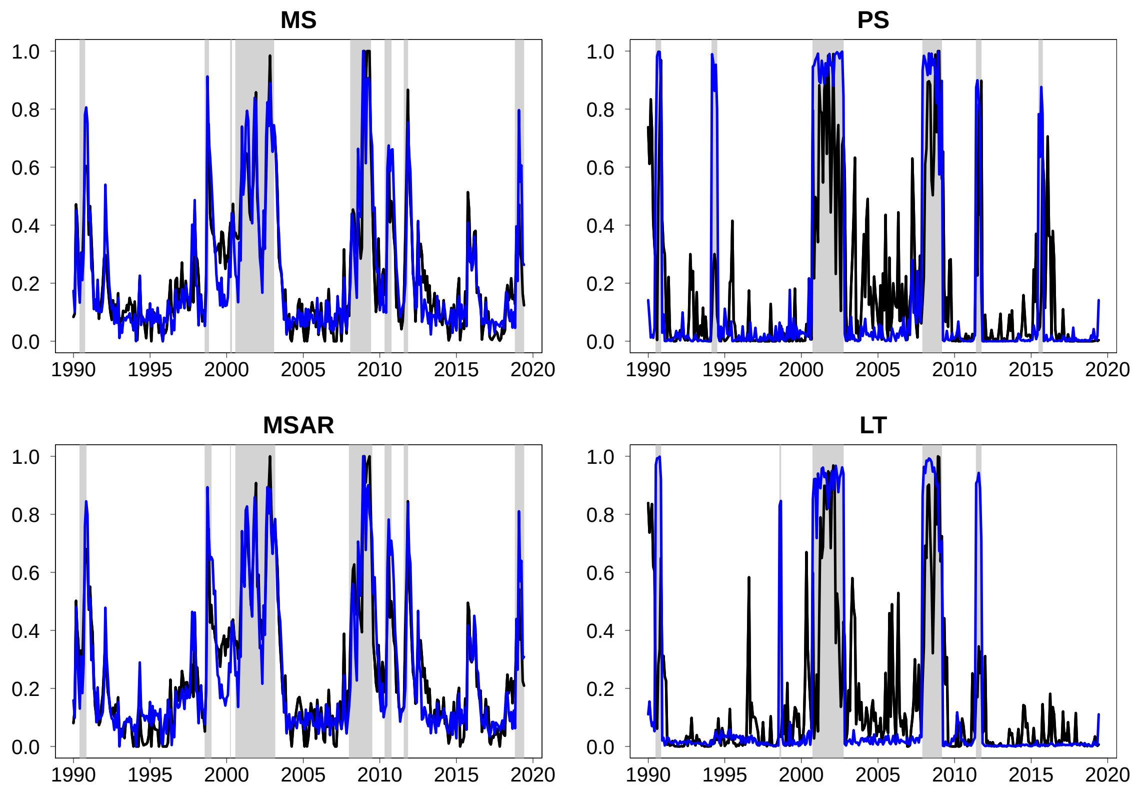


Notes: Recessions are highlighted by gray-shaded areas. Black lines represent static models and blue lines dynamic models.

Specific-to-General Approach. After searching for the variables with the highest correlation to the recession probability, the AIC recommends different models for each identification and modeling approach. Tables D2 and D3 in Appendix D show the results. The AIC typically selects more than ten variables in the static specifications and between four and eight variables in the dynamic specifications. The linear models have in common that these choose the VIX, the SMA_6m, the TED spread, and the business conditions of the Philadelphia Fed. All probit models select the i3m.12m.sd and the growth rate of the oil price. In general, the selection for the dynamic models is more parsimonious but also more heterogeneous. When comparing the goodness of fit, the models based on the MSAR outperform the simple MS models. In the case of the probit models, the PS-based models feature a higher *Pseudo* R^2, while the LT-based models are yield better information criteria.

Figure 3 shows the fitted recession probabilities. Here, the picture is very similar to the PCR case in Figure 2. The signals are often delayed and the static models feature considerable noise in the predictions, which might lead to false alarms (depending on the chosen threshold). The static probit models, whose recession probabilities fluctuate strongly between 2002 and 2008, are an example of this.

Figure 3: Fitted Recession Probabilities Based on StG Selection



Notes: Recessions are highlighted by gray-shaded areas. Black lines represent static models and blue lines dynamic models.

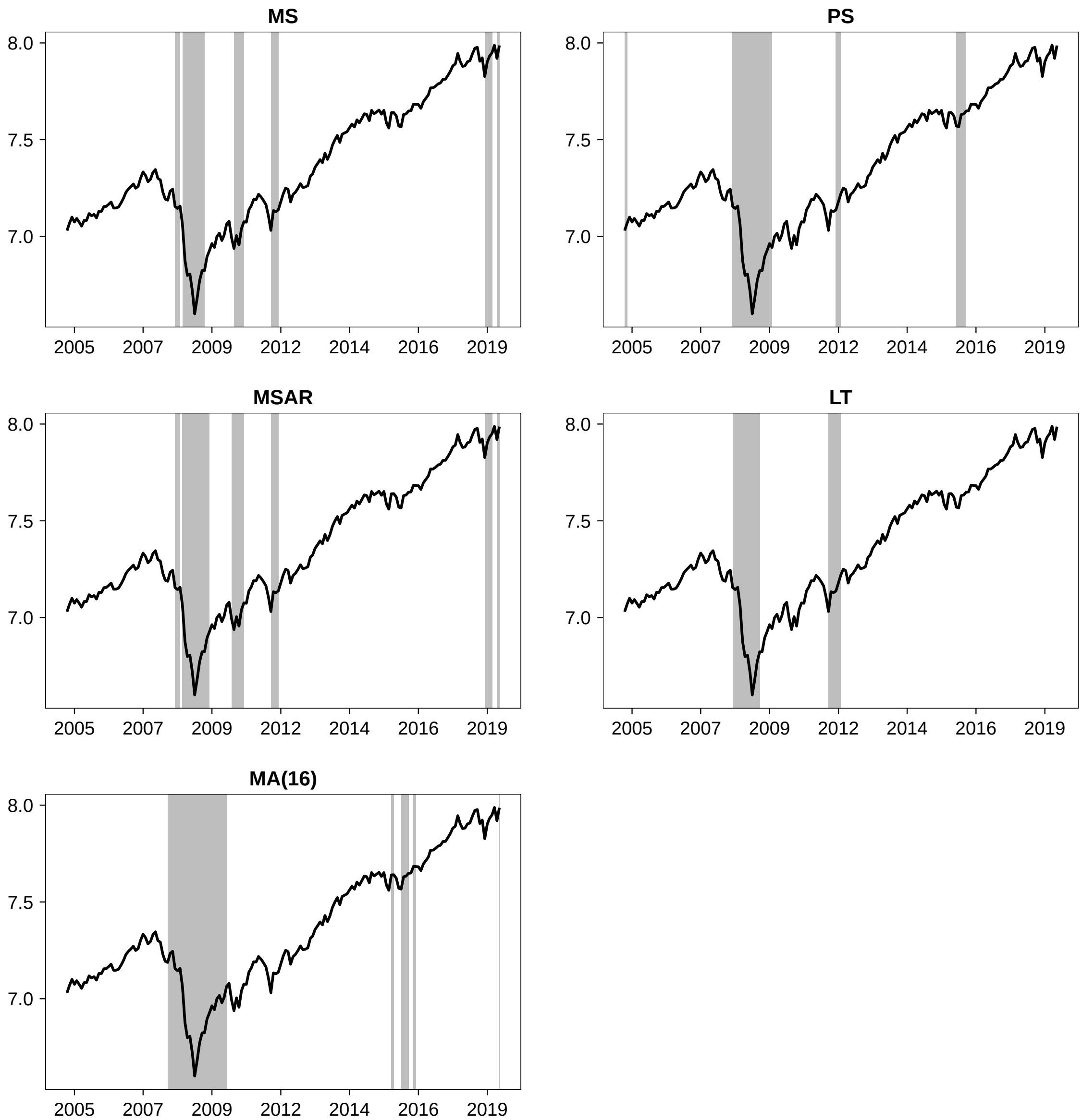
In comparison to the PCR models, the StG approach improves the goodness-of-fit measures substantially (R^2 and AIC). Only for the BIC, which penalizes additional variables to a larger extent, the StG approach cannot outperform the PCR models.

6 Out-of-Sample Results

6.1 Real-Time Identification

It is much more difficult to identify the current state of the market in real time since the future development is unknown at the time of the prediction. In our two-step approach, we first approximate the current market state recursively for each month during the period October 2004–June 2019. Since all approaches (with the exception of the naïve MA(16)) depend on the most recent observation, some substantial deviations from the full sample case might occur. In particular, the anticipating behavior of dating rules is not expected in a real time situation.

Figure 4: Real-Time Identification Results October 2004–June 2019



Notes: Figure shows the log S&P 500 index and the identified recession periods as gray-shaded areas. Top left: identification according to the simple MS model; top right: PS dating rule; middle left: MSAR model; middle right: PS dating rule; bottom left: naïve approach.

Figure 4 shows the results. The naïve approach identifies the downward trend in 2008 at the earliest stage, but identifies its end with the longest time lag. The PS dating rule identifies all recessions with a delay of seven months and does not classify the slumps of 2011 and 2015 as recessions. Worse, the subsequent recovery phases are wrongly marked as recessions and the bear market at the beginning of the out-of-sample period is wrongly classified, too. The LT approach signals the financial crisis at the same time as the PS rule, but recognizes the subsequent recovery much earlier. Both MS models also identify the turning points too late in a real-time situation.[13] Compared to the full-sample case, their time lag is, on average, three months. Nevertheless, these are able to recognize the GFC at the same time as the dating rules. In addition, the simple MS model anticipates the end of the GFC at an earlier stage than the MSAR and the PS approach. In general, the MS models are not very successful in identifying the recessions after the GFC. However, at least the MSAR model manages to anticipate a part of the price slump in 2010. To summarize, both MS models only achieve a moderate improvement in the real-time identification compared to the dating rules or the naïve approach.

6.2 Real-Time Predictions

As a final step, we address our main research question and examine the extent to which macroeconomic and financial market variables are suitable for forecasting stock market recessions in a real-time setting. In addition to the evaluation of the statistical performance of our model, we assess the economic value of a model-based investment strategy. We use a recursive forecasting procedure to capture stock market cycles from November 2004 to May 2019. Hereby, we face two additional uncertainty factors as compared to the case of complete knowledge: (i) the identification of stock market recessions takes place in real-time, which implies that the most recent observation may be misclassified (see also Section 6.1) and (ii) the common forecast uncertainty in one-step ahead out-of-sample forecasts.

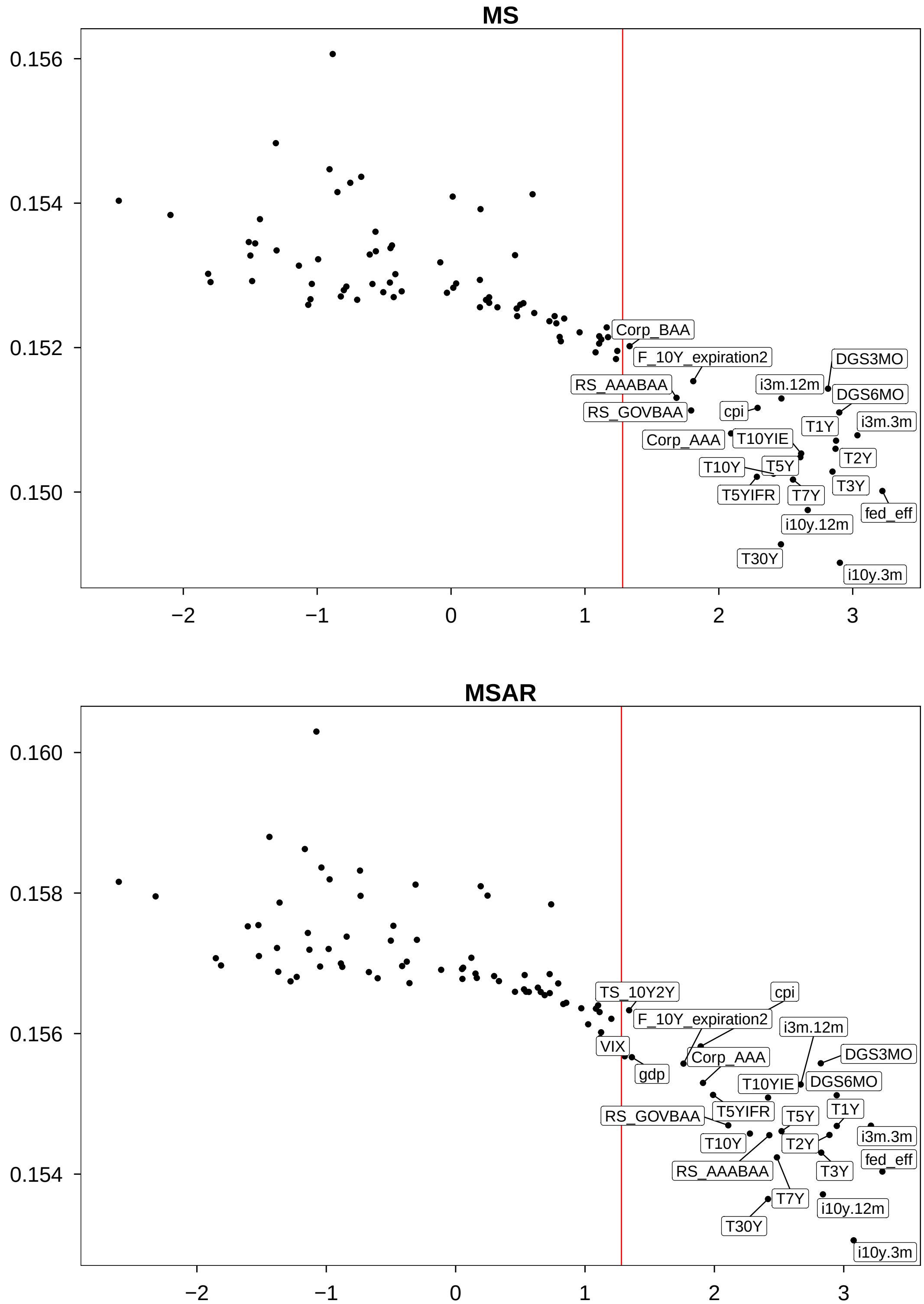
6.2.1 Bivariate Models

Statistical Performance. The out-of-sample results for the one-variable models confirm the in-sample findings that several variables help to predict stock market recessions.[14] For the linear forecast models, the inclusion of an AR(1) term always yields a better RMSE. In the probit case, we find the opposite and typically obtain worse QPS values in dynamic models. This is mainly due to misclassification or delays during the previous identification step, which are reflected in the dynamic component. The best prediction for the linear case is achieved with the three-months ahead forecast of the ten-year government bond (RMSE: 0.1490, MS dynamic). The leading indicator of the Conference Board represents the most successful predictor for probit models (QPS: 0.1503, LT static).

[13]Identification in the MS models is based on a threshold of 50% in the filtered probabilities. These do not differ much from the full-sample case and, hence, are not shown to conserve space.

[14]The detailed results are not shown to conserve space but are available on request.

Figure 5: Out-of-Sample Statistical Performance of Dynamic Linear Bivariate Models



Notes: The x-axis displays the CW statistics and the y-axis shows the RMSE. The vertical red line shows the critical value of the CW statistic at a significance level of 10%. All evaluation measures are explained in Appendix C.

Next, we are interested in whether the respective predictor can significantly improve the forecast compared to a nested benchmark model. For this purpose, we use the CW statistic where the benchmark model contains a constant (static models) or a constant and the lagged dependent variable (dynamic models). For the simple linear regression models, over two-thirds of the variables (65 of 92) beat the benchmark at a significance level of 10%. The results weaken somewhat if we add a dynamic component to the model. Figure 5 shows the statistical performance of dynamic MS models. The x-axis displays the CW statistics and the y-axis shows the RMSE. In the MS (MSAR) model, only 22 (24) variables achieve a better prediction accuracy than an AR(1) model of the filtered probability. In particular, bond market variables provide a significant improvement, thereby supporting previous empirical evidence (Chen 2009; Nyberg 2013). In addition, expected consumer prices, GDP expectations, and the VIX are helpful in this context.

We refrain from reporting the plots of the probit models since all predictors provide added value, regardless of whether the null model contains only a constant or additionally the lagged recession indicator. The corresponding CW statistics are highly significant in all cases.

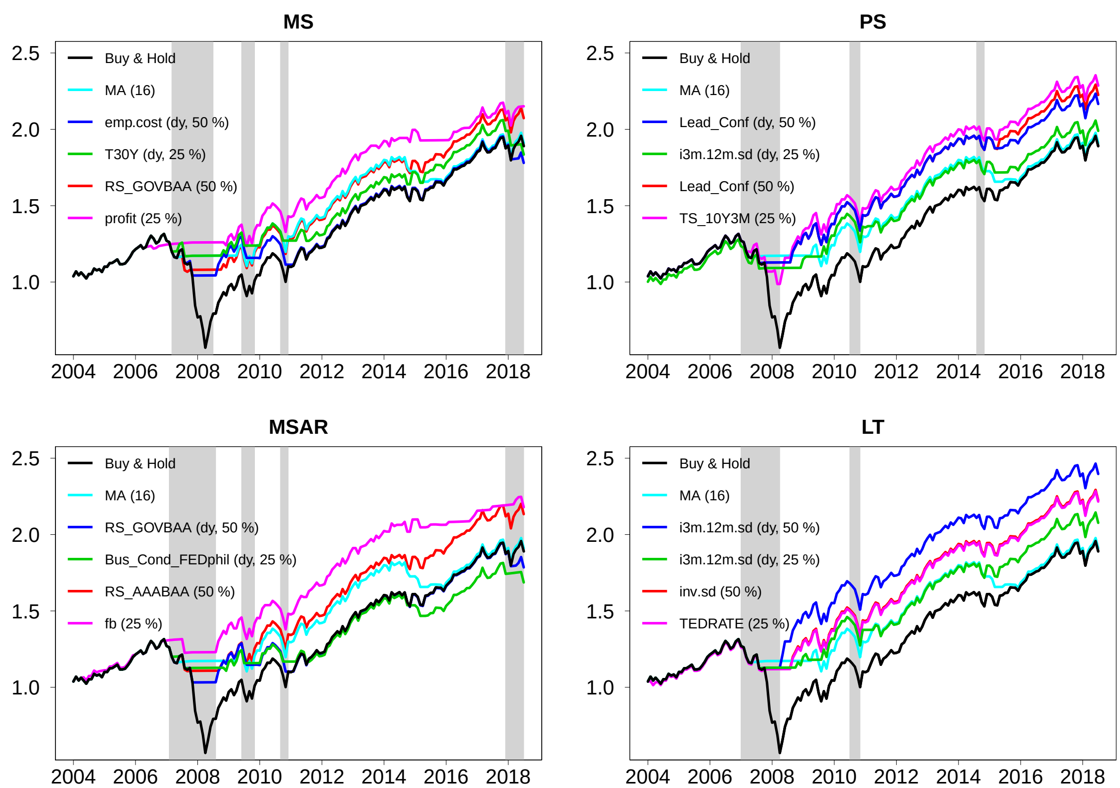
Economic Performance. Although a lot of predictors provide additional forecasting power in a statistical sense, there is no guarantee that these actually generate economic value. We apply the following simple trading strategy to check the economic performance of our forecasts. If the predicted recession probability exceeds a given threshold, we shift into the risk-free asset (that is, the one-month treasury bill). Otherwise, we keep the investment in the S&P 500. Naturally, such a strategy heavily depends on the recession threshold. Hence, we apply the natural candidate of 50% and for a more risk-averse agent we decrease the threshold to 25%. Lowering the threshold leads to an increase in the hit ratio of recessions but decreases the hit ratio for bull markets.

We use three benchmarks to examine the economic value. First, we use the buy-and-hold (BH) strategy. Hence, we pretend to have bought the S&P 500 in October 2004 and hold it until May 2019. This strategy yields an annualized Sharpe ratio (SR) of 0.3447 (return 6.10%, volatility 14.09%). Second, we use a mixed portfolio, which participates equally in the stock market and in the risk-free asset (50/50). This allocation achieves roughly the same SR than the BH strategy (return 3.67%, volatility 7.04%). Finally, in addition to these passive strategies, we are interested in the performance of the naïve strategy that acts according to the MA(16) identification rule. Since the historical average is always known, this strategy does not depend on an uncertain identification. The MA(16) outperforms the BH benchmark in all criteria and has an average annual SR of 0.4705.

In general, static linear regression models provide a higher economic benefit than dynamic ones, despite their inferior statistical performance.[15] In the case of the probit models, the evidence is mixed in that regard. The following discussion is based on the evaluation of eight different models (static and dynamic models for the four different identification strategies) against the benchmarks.

For a threshold of 50%, the spread between government bonds and BAA-rated corporate bonds (RS_GOVBAA) always exhibits a higher SR than the BH benchmark (average SR of 0.48). Four predictors (pmi, RS_AAABAA, fb, and fb.sd) beat the BH strategy in seven out of eight cases. In comparison to the MA(16), the leading indicator of the Conference Board outperforms this benchmark in six cases. The highest SR across all models (0.6974) is achieved with the standard deviation of the 12-month expectations for the short-term interest rate (LT identification and dynamic model).

Figure 6: Out-of-Sample: Cumulative Returns of the Best One-Factor Models



Notes: Recessions are marked by gray-shaded areas. Each figure shows the best model according to both thresholds (25% vs. 50%) and both model setups (static vs. dynamic).

For a threshold of 25%, the Bus_Cond_FEDphil shows the most robust results. This macroeconomic leading indicator outperforms the BH benchmark in all eight cases

[15]We do not consider transaction costs in our experiment, but since the position changes on average less than three times per year, the costs will not substantially affect the added value for an institutional investor.

(average SR of 0.43). Furthermore, a total of 18 variables are worse in only one case.[16] The unemployment forecast of Consensus Economics succeeds in six out of eight cases in achieving higher risk-adjusted performance than the naïve benchmark. Overall, the best strategy regarding the SR is the expected fiscal balance (fb), which reaches a value of 0.7068 (MSAR identification and static model). This is primarily due to an early detection of the financial crisis and a low participation rate since 2017.

Figure 6 summarizes the cumulative returns of the best trading strategies against the benchmark BH strategy and the MA(16) approach. It can be concluded that the outperformance is due to the avoidance of the large drawdown in 2008 combined with the accurate timing of the recovery phase in spring 2009.

6.2.2 Multivariate Models

In this subsection, we construct the corresponding multivariate models for the out-of-sample predictions. Again, we utilize a PCA and an StG approach for an efficient aggregation/utilization of information in the trade-off between fit and variance. We rely on one set of specifications where we obtain the principal components (regressors in the StG approach) once based the full training sample and keep the factor loadings (selected variables) fixed thereafter. Henceforth, we will refer to this as “fixed” PCR or “fixed” StG regression, respectively. In a second set, we update the principal components (regressors in the StG approach) every month, thereby using the incoming information in the recursive estimations as efficient as possible. Henceforth, we will refer to this as “updated” PCR or “updated” StG regression, respectively. In all four cases, we estimate a static prediction model (st) and a dynamic ARMA(p,q) prediction model (dy), which yields a total of eight models for each of the four identification methods.

Statistical Performance. Table 6 shows the RMSE and the QPS of all 32 prediction models. In addition, Tables E1 and E2 in Appendix E show the corresponding Diebold-Mariano test statistics. In the linear case, dynamic modeling achieves a substantially smaller RMSE and a better forecast performance according to the DM test. In particular, predictions based on dynamic PCR models dominate all other models in the case of recession identification using a simple MS model. If the filtered probability is extracted using an MSAR model, the forecasts of the two selection methods (PCA and StG) are more or less equivalent. In the case of the probit models, the static specification tends to generate better forecasts for both dating rules as indicated by lower QPS values and a superior DM test statistic. In addition, the PCR models outperform the StG models. Finally, it remains unclear whether continuous updates in the selection

[16] These variables are R, the unemployment rate, PD ratio, Lead_Conf, VIX, Int_exp lower, IP, TS_10Y1Y, TS_10Y3M, inv, profit, csales, unemp, fb, inv.sd, and i3m.3m.sd.

process over time yield some added value over a fixed selection of components and regressors.

Table 6: Out-of-Sample Prediction Accuracy

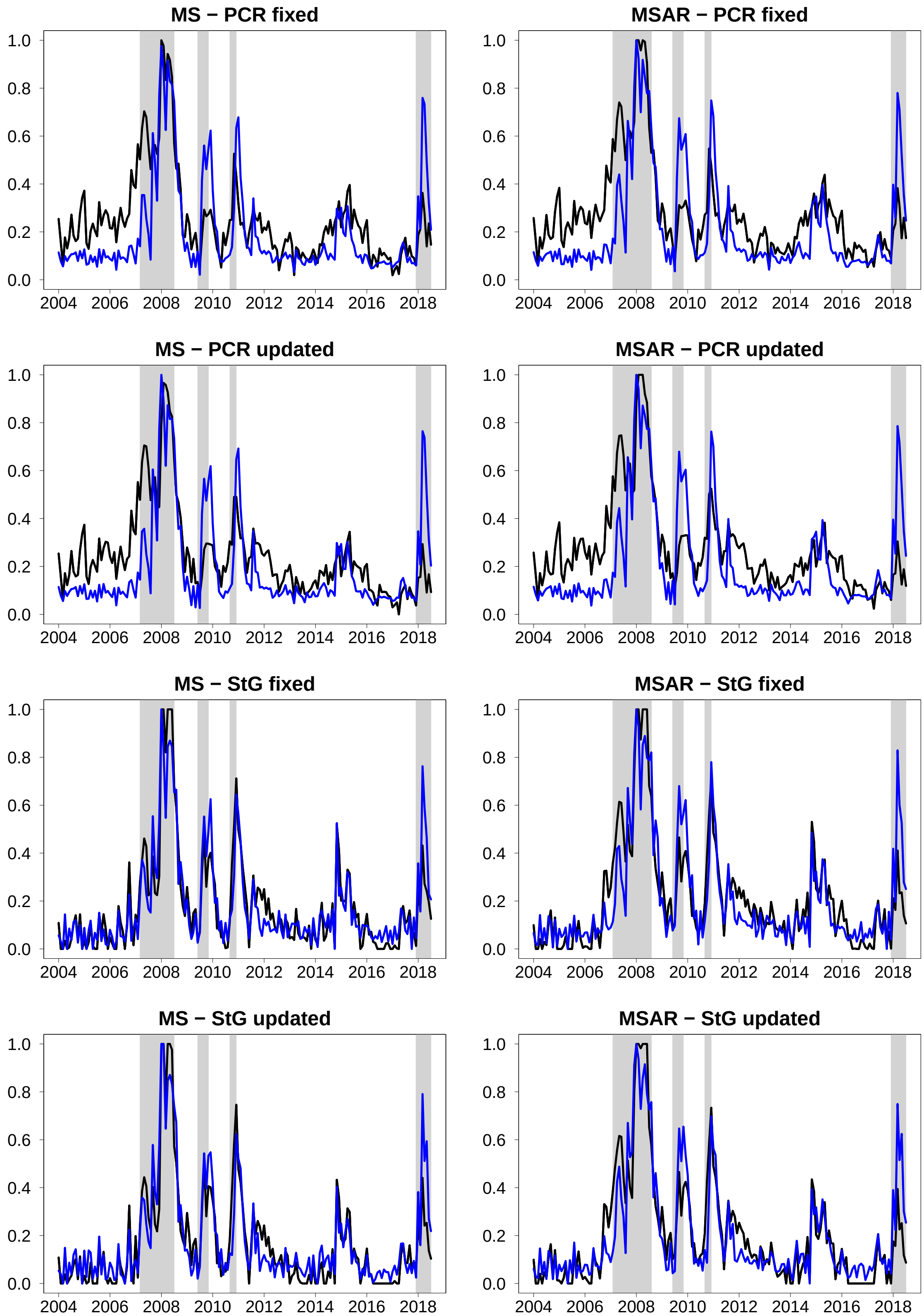
	RMSE MS	**RMSE MSAR**	**QPS PS**	**QPS LT**
PCR f/st	0.2089	0.2080	0.2149	0.1244
PCR f/dy	0.1516	0.1550	0.2376	0.1700
PCR u/st	0.2145	0.2131	0.2220	0.1253
PCR u/dy	0.1522	0.1558	0.2433	0.1755
StG f/st	0.1760	0.1750	0.3533	0.2643
StG f/dy	0.1601	0.1575	0.3906	0.3129
StG u/st	0.1737	0.1737	0.3344	0.2825
StG u/dy	0.1593	0.1557	0.3618	0.2761

Note: Table reports the RMSE and the QPS for one-step ahead forecasts. PCR: regression using principal components; StG: variable selection using a specific-to-general approach; f: components/variables are kept fixed during the test window; u: components/variables are updated every month during the test window; st: static prediction model; dy: dynamic prediction model.

Figures 7 and 8 show the predicted recession probabilities for the linear models and the probit models. A common problem of both switching approaches is that the predicted probabilities for dynamic models signal recessions with a slight delay (see Figure 7). This is mainly caused by the lagged real-time identification and the dominant influence of the AR coefficient in dynamic models (blue lines). As a result, bear markets cannot be anticipated in an early stage. The stock market contraction around the GFC is particularly well anticipated with the help of PCR models, whereas the StG models are more appropriate to forecast the recession in 2011. As Figure 8 points out, the PS dating rule is heavily affected by delayed real-time signals, which influence the regime classification accuracy negatively. This problem arises from the predetermined period lengths in the dating algorithm. Only static PCR models anticipate the recession around the GFC to some extent, which further underscores the problems with using dating rules in real time.

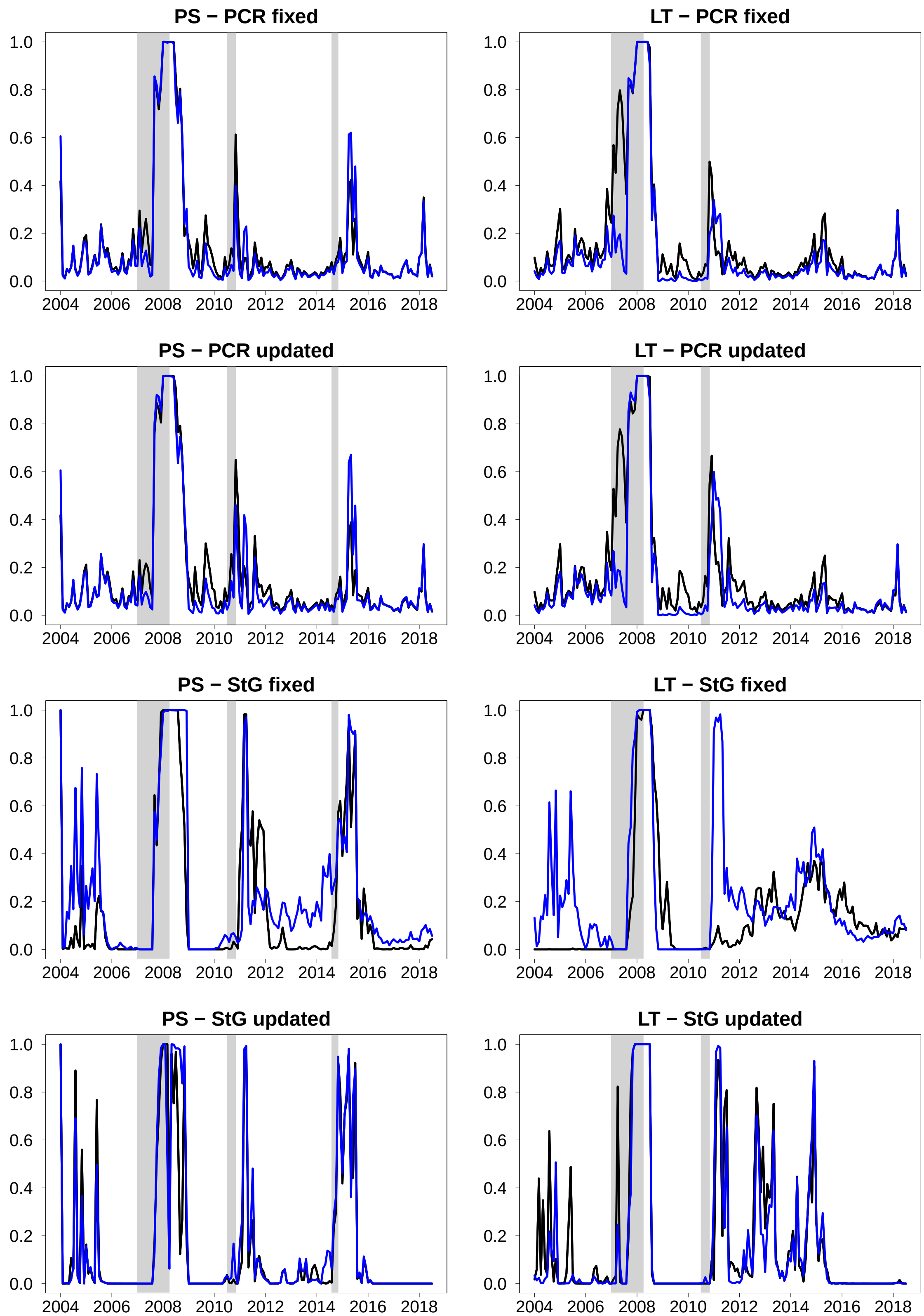
In terms of total accuracy rates (see Table E3 in Appendix E) PCR based models are particularly successful when applying a threshold of $\tau = 50\%$. With the help of MS models, up to 90% of the months can be predicted correctly (static MS model with fixed selection and static MSAR model with updated selection). When applying the LT dating rule, the hit ratio rises up to 94%. These large figures are particularly driven by the almost perfect prediction of expansions when applying the $\tau = 50\%$ threshold. Similar to Figure 8, the PS rule performs worse than the LT rule.

Figure 7: Predicted Recession Probability of Linear Regressions



Notes: Recessions are marked by gray-shaded areas. Black lines show static linear models and blue lines show dynamic linear models.

Figure 8: Predicted Recession Probability of Probit Models

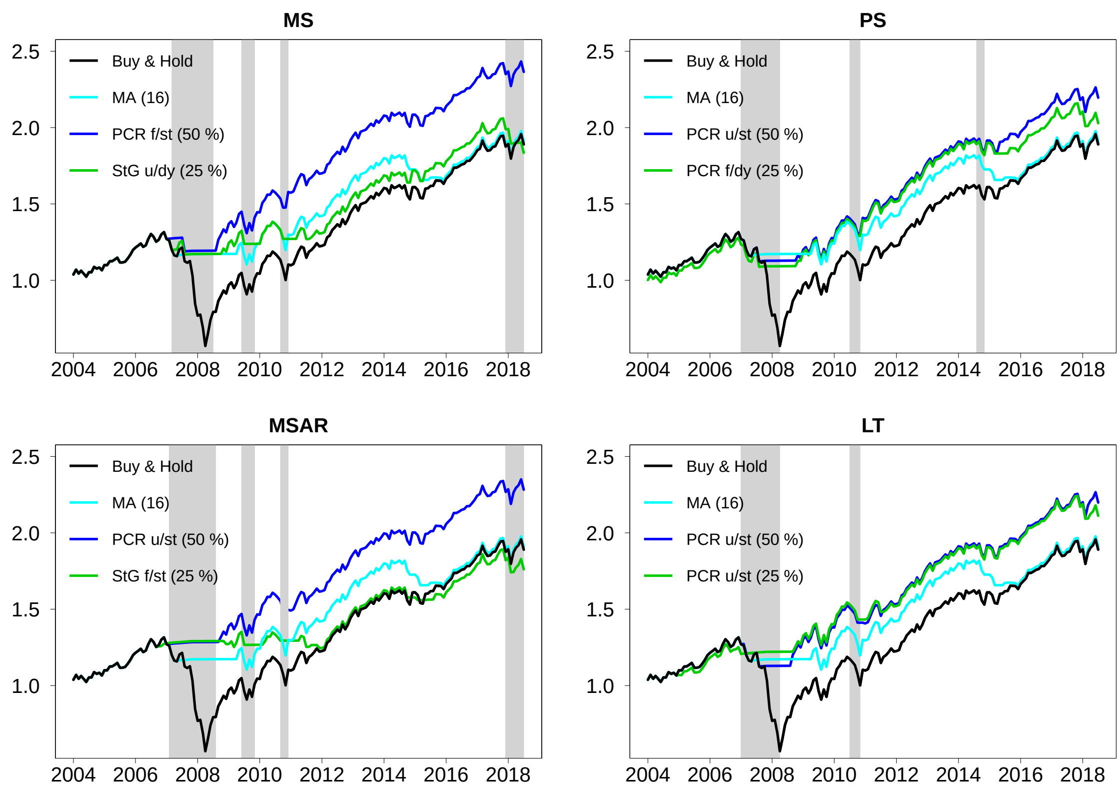


Notes: Recessions are marked by gray-shaded areas. Black lines show static probit models and blue lines show dynamic probit models.

However, we are particularly interested in predicting stock market *recessions* accurately and here the figures are a substantially lower than for expansions. Not surprisingly, the hit rations are higher for recessions when lowering the threshold to $\tau = 25\%$. Again, PCR models outperform StG models. Static LT models and static MSAR models with a fixed selection reach accuracy rates between 76% and 78%. However, achieving these figures for the prediction of recessions comes at some cost since the number of correctly predicted expansion months is (substantially) lower with a threshold of $\tau = 25\%$.

Economic Performance. As a final step, we are interested in whether the forecasting models offer higher economic value than our benchmarks. For this purpose, we apply the same trading strategy and the same benchmarks as in the previous subsection. Tables E4 and E5 in Appendix E show the detailed results for the predictive models' profitability and their tail risks. Figure 9 shows the cumulative returns of the best forecasting model for each recession identification method.

Figure 9: Cumulative Returns of Best Forecasting Model for Each Method



Note: Recessions are marked by gray-shaded areas. Each figure shows the best model according to the SR for both thresholds (25% vs. 50%).

Both tail risk measures (MaxDD and VaR) can be significantly reduced across all models compared to the BH benchmark for a threshold of 50% (see Table E4). In addition, the predictive models feature a larger omega, which means that our strategy generates more often positive returns. This also happens at the expense of yield losses. Finally, in more than half of the cases, the MA(16) strategy is dominated in a risk-adjusted manner.

The best model according to MS identification is the static regression with fixed PCA selection. On average, an annualized return of 9.36% can be achieved, whereby the maximum drawdown over the entire 15 years is less than 10%. From the top-left panel in Figure 9, we observe that this strategy recommends the risk-free assets during the GFC. In addition, the price decline in 2011 is smaller than for the BH strategy. None of the models that rely on MS identification performs substantially worse than the BH benchmark. At the same time, however, only one of these outperforms the naïve approach. It is also worth highlighting that the explicit consideration of a dynamic structure improves the forecast in a statistical sense (superior DM test results and lower RMSE), whereas static regressions are preferable from an economic perspective. This also in line with the bivariate findings from the previous subsection. Finally, if we compare the best bivariate model with the multivariate specifications, only the PCA with fixed weights and a static prediction model generates a value-added.

For MSAR identification, the static PCR model with an updated selection features impressive results. With an SR of 0.7250, the naïve approach can be outperformed by more than 50%. This early warning model has the lowest drawdown (−0.0959) and the highest omega (1.9266) across all models. All MSAR models are able to beat the BH benchmark in a risk-adjusted manner. Again, the economic value of static models is superior to dynamic ones. Finally, half of the multivariate models yield better risk-adjusted performance than the best bivariate model.

Although the forecasts of PS models are heavily affected by a time delay in the identification step, our investment strategy is able to outperform the naïve approach in five out of eight cases. In particular, PCR models are more profitable and achieve lower drawdowns. Similar to the linear models, the static specifications are superior. As indicated by Figure 9, both substantial price drops (at the beginning of 2008 and in 2011) can be absorbed to some extent. Finally, only the two static PCR models can outperform the best bivariate predictor in Figure 6 (Lead_Conf).

In the case of LT models, the best performer is the static probit model with an updated PCA selection. This model anticipates the large drawdown of the stock market during the GFC and is also able to limit losses in 2011. The LT models generate larger returns and a higher SR than all benchmarks. Compared to the bivariate case, minor improvements in terms of economic value can be achieved only in the case of static specifications.

As previously discussed, recessions can be predicted more precisely when applying a threshold of 25%. However, such a statistical success cannot be translated into higher economic value (see Table E5). The risk-adjusted performance (SR and omega) typically deteriorates when compared to the 50% threshold. The most notable exceptions are dynamic MS models and LT models with StG selection.

In general, probit models are less sensitive to the threshold reduction compared to their linear counterparts. In particular, probit models based on the LT dating rule appear to be the most robust ones. These are still able to exceed all benchmarks regarding SR and omega. Thereby, the static LT approach with an updated PCA achieves the best overall SR with 0.6272. From a risk perspective, the choice of a lower threshold leads in almost all cases to prolonged investments in the risk-free rate and, hence, reduces volatility and improves the VaR. Although the maximum drawdown can, on average, be reduced be applying the 25% threshold, a number of counterexamples should be highlighted (for instance, PCR f/st MS or PCR c/st MSA). Overall, the static MS model with a fixed StG selection shows the lowest maximum drawdown over the entire 15 years (−7.95%).

To summarize, dynamic modeling improves the forecasting accuracy of linear models, but this is accompanied by losses in economic value. Static models are more attractive from a return and risk perspective. For probit models, static specifications yield better results from both, a statistical and an economic point of view. Decisive for the outperformance of the benchmarks is the early anticipation of the GFC and the correct timing for re-entering the market at the end of the GFC. With respect to the model selection approach, the aggregation of predictors by PCA is preferred to the StG approach. The quality of the forecast does not particularly depend on whether the variable selection is carried out once or if is re-iterated for each month. Furthermore, the recession detection is more accurate when applying a threshold of 25%, but the risk-return trade-off is better managed by a threshold of 50%.

7 Conclusions

This paper offers a promising outlook on the predictability of stock market regimes. In our two-step procedure, we first identify stock market regimes in the US using three different techniques (Markov-switching models, dating rules, and a naïve moving average). Second, we predict recessions in the S&P 500 with the help of several modeling approaches, utilizing the information of 92 macro-financial variables.

Our results suggest that several variables are suitable for forecasting recessions in stock markets in-sample and out-of-sample. Risk spreads of bond markets, the VIX, the leading indicator of the Conference Board, and the business conditions of the Federal Reserve Bank of Philadelphia can be highlighted. In addition, the first and second moments of the forecasts by Consensus Economics help to predict the recession probability. Our early warning models for the US equity market, in particular those using principal components to aggregate the information in the macro-financial variables, provide a statistical improvement over several benchmarks. In addition, these generate economic value by boosting returns, improving the sharp ratio and the omega, and substantially reducing drawdowns. Finally, although dynamic model specifications are preferable from a statistical point of view, the market entry and exit around the GFC of 2007–2008 can be better anticipated with static models.

Our results offer a variety of starting points for future work. First, considering more than two regimes could be a promising extension. Since the two-regime model has problems in distinguishing recessions and recoveries, a three- or four-regime model might — if statistically feasible — help improve the modeling (see, for example, Maheu et al. 2012; Hauptmann et al. 2014). Second, allowing for time-varying transition probabilities that depend on exogenous predictors would ensure that market regimes can be estimated and predicted in a single step. This would provide the advantage that the filtered probability is not treated as an observable variable and hence its uncertainty is taken into account (Kole and Van Dijk 2017). Finally, the set of predictors could be extended by technical indicators and cross-sectional information of stock returns such as turbulence indices or factor portfolios.

References

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.

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

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

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.

Bry, G. and Boschan, C. (1971). *Cyclical Analysis of Time Series: Selected Procedures and Computer Programs*. Columbia University Press, New York.

Campbell, J. Y. and Cochrane, J. H. (1999). By force of habit: A consumption-based explanation of aggregate stock market behavior. *Journal of Political Economy*, 107(2):205–251.

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.

Diebold, F. X., Lee, J.-H., and Weinbach, G. C. (1994). Regime switching with time-varying transition probabilities. In Hargreaves (1994), pages 283–302.

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

Durlauf, S. N. and Blume, L. E., editors (2010). *Macroeconometrics and Time Series Analysis*. Palgrave Macmillan, London.

Estrella, A. (1998). A new measure of fit for equations with dichotomous dependent variables. *Journal of Business and Economic Statistics*, 16(2):198–205.

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.

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.

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*. 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. (2010). Regime switching models. In Durlauf and Blume (2010), pages 202–209.

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. (1992). The likelihood test under non-standard conditions: Testing the markov trend model of GNP. *Journal of Applied Econometrics*, 7(S1):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.

Hargreaves, C., editor (1994). *Nonstationary Time Series Analysis and Cointegration*. Oxford University Press, Oxford.

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

Hastie, T., Tibshirani, R., and Friedman, J. (2009). *The Elements of Statistical Learning: Data Mining, Inference, and Prediction*. Springer, Berlin.

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.

Keating, C. and Shadwick, W. F. (2002). A universal performance measure. *Journal of Performance Measurement*, 6(3):59–84.

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

Kosmidis, I., Pagui, E. C. K., and Sartori, N. (2018). Mean and median bias reduction in generalized linear models. *Statistics and Computing*, 30:1–17.

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

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.

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.

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.

Newey, W. K. and West, K. D. (1987). A simple, positive semi-definite, heteroskedasticity and autocorrelationconsistent covariance matrix. *Econometrica*, 55(3):703–708.

Nyberg, H. (2013). Predicting bear and bull stock markets with dynamic binary time series models. *Journal of Banking and Finance*, 37(9):3351–3363.

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.

Perez-Quiros, G. and Timmermann, A. (2000). Firm size and cyclical variations in stock returns. *Journal of Finance*, 55(3):1229–1262.

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

Poterba, J. M. and Summers, L. H. (1988). Mean reversion in stock prices: Evidence and implications. *Journal of Financial Economics*, 22(1):27–59.

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

Stock, J. H. and Watson, M. W. (2002). Forecasting using principal components from a large number of predictors. *Journal of the American Statistical Association*, 97(460):1167–1179.

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.

Appendix A: Data Description

Table A1: Data Description and Sources

Variable Description	Abbreviation	Source
Effective Federal Funds Rate	fed_eff	Federal Reserve System
3 M Treasury Bill	DGS3MO	Federal Reserve System
6 M Treasury Bill	DGS6MO	Federal Reserve System
1 Y Treasury Bonds	T1Y	Federal Reserve System
2 Y Treasury Bonds	T2Y	Federal Reserve System
3 Y Treasury Bonds	T3Y	Federal Reserve System
5 Y Treasury Bonds	T5Y	Federal Reserve System
7 Y Treasury Bonds	T7Y	Federal Reserve System
10 Y Treasury Bonds	T10Y	Federal Reserve System
30 Y Treasury Bonds	T30Y	Federal Reserve System
Future Nearest Expiration 10 Y US Bond	F_10Y_expiration1	Chicago Board of Trade
Future 2nd-Nearest Expiration 10 Y US Bond	F_10Y_expiration2	Chicago Board of Trade
Term Spread 30 Y and 10 Y	TS_30Y10Y	Federal Reserve System
Term Spread 10 Y and 2 Y	TS_10Y2Y	Federal Reserve System
Term Spread 10 Y and 1 Y	TS_10Y1Y	Federal Reserve System
Term Spread 10 Y and 3 M	TS_10Y3M	Federal Reserve System
Term Spread 5 Y and 3 M	TS_5Y3M	Federal Reserve System
5 Y 5 Y Forward Inflation Expectation	T5YIFR	Federal Reserve System
10 Y Break-Even Inflation Rate	T10YIE	Federal Reserve System
Corporate Bonds Yield AAA	Corp_AAA	Moody's
Corp. Bonds Yield BAA	Corp_BAA	Moody's
Average Corp. Bonds Yield	Corp_Average	Moody's
Risk Spread BAA Corp. and AAA Corp.	RS_AAABAA	Moody's
Risk Spread AAA Corp. Bond and Gov. Bond	RS_GOVAAA	Moody's
Risk Spread BAA Corp. Bond and Gov. Bond	RS_GOVBAA	Moody's
3 M USD LIBOR and 3 M T-Bill Spread	TEDRATE	Federal Reserve System
Rolling 6 Month Volatility of S&P 500 Returns	Vola_6m	Yahoo Finance
Implied Volatility Index for S&P 500 Options	VIX	CBOE
Return Gold Price	Gold	LBMA
Return WTI Oil Price	Oil	NY Mercantile Exchange
Exchange Rate JPY / AUD	JPYAUD	European Central Bank
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
10 Y Int. Rate (in 3 Months) Mean Forecast	i10y.3m	Consensus Economics
10 Y Int. Rate (in 12 Months) Mean Forecast	i10y.12m	Consensus Economics
3 M Int. Rate (in 3 Months) Mean Forecast	i3m.3m	Consensus Economics
3 M Int. Rate (in 12 Months) Mean Forecast	i3m.12m	Consensus Economics

Table A1: Data Description and Sources (Continued)

Variable Description	Abbreviation	Source
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
3 M Int. Rate (in 3 Months) Forecast Std. Dev.	i3m.3m.sd	Consensus Economics
3 M Int. Rate (in 12 Months) Forecast Std. Dev.	i3m.12m.sd	Consensus Economics
10 Y Int. Rate (in 3 Months) Forecast Std. Dev.	i10y.3m.sd	Consensus Economics
10 Y Int. Rate (in 12 Months) Forecast Std. Dev.	i10y.12m.sd	Consensus Economics
12 M Expectation Fed Rate Increase	Int_exp_higher	The Conference Board
12 M Expectation Fed Rate Decrease	Int_exp_lower	The Conference Board
12 M Expectation Stock Price Increase	stock_exp_higher	The Conference Board
12 M Expectation Stock Price Decrease	stock_exp_lower	The Conference Board
Inflation Expectation Survey TCB	Inf_exp_Conf	The Conference Board
Consumer Climate Survey TCB	Cons_clim_Conf	The Conference Board
Consumer Situation Survey TCB	Cons_sit_Conf	The Conference Board
Consumer Expectation Survey TCB	Cons_exp_Conf	The Conference Board
Consumer Climate Survey Univ. Michigan	Cons_clim_Mich	University Michigan
Leading Economic Index for the US	Lead_Conf	The Conference Board
Purchasing Manager Index	pmi	ISM
Business Condition Philadelphia Fed	Bus_Cond_FEDphil	Philadelphia Fed
Trade-Weighted USD Real Exchange Rate	FX_TW	Federal Reserve System
M1 Growth Rate	M1	Federal Reserve System
M2 Growth Rate	M2	Federal Reserve System
CPI Inflation Rate	CPI	Bureau of Labor Statistics
Core CPI Inflation Rate	Core_CPI	Bureau of Labor Statistics
Industrial Production Growth	IP	Federal Reserve System
Unemployment Rate	Unem	Bureau of Labor Statistics
House Prices Growth Rate	house_market_nahb	NAHB
Price / Dividends (S&P 500)	PD	Robert Shiller Data
Price / Earnings (S&P 500)	PE	Robert Shiller Data
Price / Cyclical Adjusted Earnings (S&P 500)	CAPE	Robert Shiller Data
Log. Returns of S&P 500 (Adj. Close Price)	R	Yahoo Finance
6-Month Moving Average of S&P 500 Returns	SMA_6m	Yahoo Finance

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, T5YIFR, T10YIE, and VIX), we face the problem of missing values. We solve this by substituting the respective entries with 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% (and a risk premium of 0%), we can replace the TIPS yields with a simple proxy for market-based expectations before 2003.[17] 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 the realized twelve-month rolling volatility of the S&P 500.

Table A2: Descriptive Statistics

Variable	Max.	Min.	Mean	Std. Dev.	$I(d)$
fed_eff	8.5000	0.2500	2.9859	2.3572	I(0)
DGS3MO	8.0700	0.0000	2.8190	2.3257	I(0)
DGS6MO	8.4400	0.0300	2.9519	2.3544	I(0)
T1Y	8.5800	0.0900	3.0857	2.3502	I(0)
T2Y	8.9600	0.2000	3.3910	2.3584	I(0)
T3Y	9.0500	0.3000	3.6134	2.2985	I(0)
T5Y	9.0400	0.5900	4.0193	2.1566	I(0)
T7Y	9.0600	0.9800	4.3298	2.0485	I(0)
T10Y	9.0400	1.4600	4.5684	1.9255	I(0)
T30Y	9.0300	2.2300	5.1182	1.7112	I(0)
F_10Y_expiration1	134.9100	93.2500	114.3807	10.0353	I(0)
F_10Y_expiration2	133.9400	93.1600	113.5908	10.0430	I(0)
TS_30Y10Y	1.4600	−0.3800	0.5498	0.3801	I(0)
TS_10Y2Y	2.8400	−0.4700	1.1774	0.8782	I(0)
TS_10Y1Y	0.8600	−0.6300	0.0002	0.2095	I(1)
TS_10Y3M	1.0300	−0.8400	0.0006	0.2826	I(1)
TS_5Y3M	3.1200	−0.9000	1.2003	0.8317	I(0)
T5YIFR	7.2212	0.4800	3.3706	1.4350	I(0)
T10YIE	7.0400	0.1100	3.0872	1.4006	I(0)
Corp_AAA	9.5600	3.2800	5.9882	1.6631	I(0)
Corp_BAA	10.7400	4.2200	6.9411	1.6318	I(0)
Corp_Average	10.0300	3.6300	6.4207	1.6605	I(0)
RS_AAABAA	3.3800	0.5500	0.9529	0.3886	I(0)
RS_GOVAAA	3.1600	0.4200	1.4198	0.4852	I(0)
RS_GOVBAA	6.2300	1.2200	2.3727	0.7624	I(0)
TEDRATE	3.1500	0.1200	0.4920	0.3581	I(0)
Vola_6m	0.0947	0.0056	0.0364	0.0181	I(0)
VIX	59.8900	9.5100	19.2216	7.4146	I(0)
Gold	17.3500	−17.3800	0.4466	4.4666	I(0)
Oil	45.1600	−32.6200	0.7616	9.3555	I(0)
JPYAUD	12.5478	−23.4541	−0.0151	4.0488	I(0)
gdp	4.5355	−1.6845	2.5114	0.9157	I(0)
cons	4.3061	−1.2124	2.5112	0.8248	I(0)
inv	10.9000	−12.3832	4.8686	4.0110	I(0)
profit	15.8500	−10.6345	5.2290	4.6465	I(0)
prod	5.2450	−6.3727	2.6394	1.7647	I(0)
cpi	5.1540	−0.4485	2.4597	0.8389	I(0)
ppi	4.8701	−2.9446	1.9364	1.1242	I(0)
emp.cost	0.3344	−0.2833	−0.0045	0.0849	I(1)

[17]The assumption of a real interest rate of 2% is very simplistic, but, on average, valid for the US between 1989 and 2002 (Neely and Rapach 2008).

Table A2: Descriptive Statistics (Continued)

Variable	Max.	Min.	Mean	Std. Dev.	$I(d)$
csales	5.1002	−0.9914	0.0194	0.3054	I(1)
housep	0.0586	−0.1150	−0.0005	0.0218	I(1)
unemp	0.7859	−0.4114	−0.0050	0.1192	I(1)
ca	92.2855	−46.5704	−1.3141	15.3252	I(1)
fb	95.5325	−201.5910	−2.1874	31.8330	I(1)
term.spread.exp.3m	3.7009	−0.3382	1.7991	1.0946	I(0)
term.spread.exp.12m	3.6817	−0.0927	1.7396	1.0250	I(0)
i3m.3m	7.9300	0.0591	2.9209	2.2814	I(0)
i3m.12m	7.7920	0.1164	3.2805	2.1322	I(0)
i10y.3m	8.7714	1.7004	4.7200	1.8410	I(0)
i10y.12m	8.5840	2.1996	5.0201	1.6759	I(0)
gdp.sd	7.7533	1.4732	3.3799	1.1881	I(0)
cons.sd	0.9747	0.1033	0.3224	0.1059	I(0)
inv.sd	2.7172	0.2535	0.7176	0.2870	I(0)
profit.sd	0.7533	0.1061	0.3273	0.1172	I(0)
prod.sd	0.9468	0.1072	0.3703	0.1060	I(0)
cpi.sd	0.3881	0.0643	0.2001	0.0573	I(0)
ppi.sd	114.7762	6.7674	42.9085	23.2209	I(0)
emp.cost.sd	0.3682	−0.4399	−0.0009	0.0636	I(1)
csales.sd	0.6016	−0.2931	−0.0006	0.0640	I(1)
housep.sd	1.7821	−0.9602	−0.0030	0.2658	I(1)
unemp.sd	0.5258	−0.9136	−0.0016	0.1498	I(1)
ca.sd	0.0558	−0.0564	−0.0001	0.0110	I(1)
fb.sd	98.6003	−111.3816	0.1412	15.2484	I(0)
i3m.3m.sd	0.6377	0.0296	0.1821	0.0916	I(0)
i3m.12m.sd	0.8399	0.0665	0.3888	0.1603	I(0)
i10y.3m.sd	0.5690	0.1094	0.2370	0.0737	I(0)
i10y.12m.sd	0.6801	0.1860	0.4047	0.0966	I(0)
Int_exp_higher	79.2000	23.4000	54.9381	11.8206	I(0)
Int_exp_lower	45.8000	5.2000	14.8398	8.8214	I(0)
stock_exp_higher	51.0000	18.1000	35.4415	6.2067	I(0)
stock_exp_lower	54.9000	15.3000	27.7412	7.6464	I(0)
Inf_exp_Conf	1.7500	−1.0500	−0.0003	0.2701	I(1)
Cons_clim_Conf	21.7000	−23.0000	0.0345	6.0866	I(1)
Cons_sit_Conf	186.8000	20.2000	101.0025	47.9325	I(0)
Cons_exp_Conf	37.1000	−25.8000	−0.0105	7.6965	I(1)
Cons_clim_Mich	17.3000	−12.7000	0.0093	3.9767	I(1)
Lead_Conf	1.5500	−2.8400	0.1305	0.5860	I(1)
pmi	61.4000	34.5000	52.2989	4.8459	I(0)
Bus_Cond_FEDphil	41.4000	−48.2000	7.3133	15.7862	I(0)
FX_TW	6.4500	−3.5500	0.0303	1.2235	I(0)
M1	5.7800	−3.3200	0.4456	0.8710	I(0)
M2	2.3000	−0.4600	0.4382	0.3457	I(0)
CPI	6.3800	−1.9600	2.4838	1.2863	I(0)
Core_CPI	5.6500	0.6000	2.4319	0.9202	I(0)
IP	8.5400	−15.3300	1.9329	3.9092	I(0)
Unem	0.5000	−0.5000	−0.0042	0.1539	I(1)
house_market_nahb	10.0000	−10.0000	0.0424	2.9278	I(1)
PD	6.8897	−9.2904	0.0594	1.9138	I(1)
PE	87.9912	13.5250	24.1853	9.3205	I(0)
CAPE	3.5313	−4.1392	0.0381	0.9144	I(1)
R	0.1058	−0.1856	0.0061	0.0413	I(0)
SMA_6m	0.0547	−0.0928	0.0059	0.0179	I(0)

Notes: The order of integration $I(d)$ is determined using an Augmented Dickey-Fuller test.

Appendix B: Full Sample Identification Results

Table B1: Expansion and Recession Periods According to MS Model

Bull Markets Date	Durat. (Months)	Ampl. (in %)	**Bear Markets** Date	Durat. (Months)	Ampl. (in %)
Jan 1950 – Feb 1962	146	310	Mar 1962 – Oct 1962	8	−19
Nov 1962 – Jul 1966	45	34	Aug 1966 – Aug 1966	1	0
Sep 1966 – May 1969	33	35	Jun 1969 – Jul 1970	14	−20
Aug 1970 – Sep 1973	38	33	Oct 1973 – Mar 1975	18	−23
Apr 1975 – Jun 1975	3	9	Jul 1975 – Aug 1975	2	−2
Sep 1975 – Sep 1978	37	22	Oct 1978 – Oct 1978	1	0
Nov 1978 – Jan 1980	15	21	Feb 1980 – Mar 1980	2	−10
Apr 1980 – Jul 1981	16	23	Aug 1981 – Sep 1981	2	−5
Oct 1981 – Dec 1981	3	1	Jan 1982 – Oct 1982	10	11
Nov 1982 – Jun 1986	44	81	Jul 1986 – Jan 1987	7	16
Feb 1987 – Aug 1987	7	16	Sep 1987 – Dec 1987	4	−23
Jan 1988 – Apr 1990	28	29	May 1990 – Sep 1990	5	−15
Oct 1990 – Jun 1998	93	273	Jul 1998 – Oct 1998	4	−2
Nov 1998 – Feb 2000	16	17	Mar 2000 – Apr 2000	2	−3
May 2000 – Jun 2000	2	2	Jul 2000 – Jan 2003	31	−40
Feb 2003 – Dec 2007	59	75	Jan 2008 – May 2009	17	−33
Jun 2009 – Mar 2010	10	27	Apr 2010 – Sep 2010	6	−4
Oct 2010 – Jun 2011	9	12	Jul 2011 – Oct 2011	4	−3
Nov 2011 – Sep 2018	83	134	Oct 2018 – Jun 2019	9	8

Table B2: Expansion and Recession Periods According to MSAR Model

Bull Markets Date	Durat. (Months)	Ampl. (in %)	**Bear Markets** Date	Durat. (Months)	Ampl. (in %)
Jan 1950 – Jul 1957	91	181	Aug 1957 – Oct 1957	3	−9
Nov 1957 – Feb 1962	52	68	Mar 1962 – Nov 1962	9	−10
Dec 1962 – May 1966	42	36	Jun 1966 – Aug 1966	3	−9
Sep 1966 – May 1969	33	35	Jun 1969 – Jul 1970	14	−20
Aug 1970 – Sep 1973	38	33	Oct 1973 – Oct 1975	25	−18
Nov 1975 – Sep 1978	35	12	Oct 1978 – Oct 1978	1	0
Nov 1978 – Dec 1979	14	14	Jan 1980 – Apr 1980	4	−7
May 1980 – Jun 1981	14	18	Jul 1981 – Oct 1982	16	2
Nov 1982 – Jun 1986	44	81	Jul 1986 – Feb 1987	8	20
Mar 1987 – Jul 1987	5	9	Aug 1987 – Dec 1987	5	−25
Jan 1988 – Apr 1990	28	29	May 1990 – Oct 1990	6	−16
Nov 1990 – Jun 1998	92	252	Jul 1998 – Dec 1998	6	10
Jan 1999 – Feb 2000	14	7	Mar 2000 – Mar 2000	1	0
Apr 2000 – Jun 2000	3	0	Jul 2000 – Feb 2003	32	−41
Mar 2003 – Nov 2007	57	75	Dec 2007 – Jun 2009	19	−37
Jul 2009 – Mar 2010	9	18	Apr 2010 – Sep 2010	6	−4
Oct 2010 – Jun 2011	9	12	Jul 2011 – Oct 2011	4	−3
Nov 2011 – Sep 2018	83	134	Oct 2018 – May 2019	8	1
Jun 2019 – Jun 2019	1	0			

Table B3: Expansion and Recession Periods According to PS Dating Rule

Bull Markets Date	Durat. (Months)	Ampl. (in %)	**Bear Markets** Date	Durat. (Months)	Ampl. (in %)
Jan 1950 – Dec 1952	36	56	Jan 1953 – Aug 1953	8	−12
Sep 1953 – Jul 1956	35	112	Aug 1956 – Dec 1957	17	−16
Jan 1958 – Jul 1959	19	45	Aug 1959 – Oct 1960	15	−10
Nov 1960 – Dec 1961	14	29	Jan 1962 – Jun 1962	6	−20
Jul 1962 – Jan 1966	43	60	Feb 1966 – Sep 1966	8	−16
Oct 1966 – Nov 1968	26	35	Dec 1968 – Jun 1970	19	−30
Jul 1970 – Apr 1971	10	33	May 1971 – Nov 1971	7	−6
Dec 1971 – Dec 1972	13	16	Jan 1973 – Sep 1974	21	−45
Oct 1974 – Dec 1976	27	45	Jan 1977 – Feb 1978	14	−15
Mar 1978 – Nov 1980	33	58	Dec 1980 – Jul 1982	20	−21
Aug 1982 – Jun 1983	11	40	Jul 1983 – May 1984	11	−7
Jun 1984 – Aug 1987	39	115	Sep 1987 – Nov 1987	3	−28
Dec 1987 – May 1990	30	46	Jun 1990 – Oct 1990	5	−15
Nov 1990 – Jan 1994	39	49	Feb 1994 – Jun 1994	5	−5
Jul 1994 – Aug 2000	74	231	Sep 2000 – Sep 2002	25	−43
Oct 2002 – Oct 2007	61	75	Nov 2007 – Feb 2009	16	−50
Mar 2009 – Apr 2011	26	71	May 2011 – Sep 2011	5	−16
Oct 2011 – May 2015	44	68	Jun 2015 – Sep 2015	4	−7
Oct 2015 – Jun 2019	45	41			

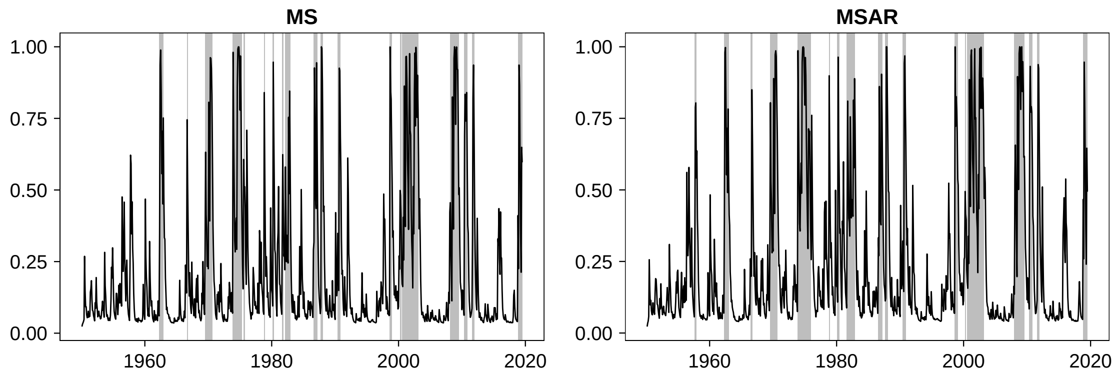
Table B4: Expansion and Recession Periods According to LT Dating Rule

Bull Markets Date	Durat. (Months)	Ampl. (in %)	**Bear Markets** Date	Durat. (Months)	Ampl. (in %)
Jan 1950 – Jul 1956	79	190	Aug 1956 – Dec 1957	17	−16
Jan 1958 – Dec 1961	48	72	Jan 1962 – Jun 1962	6	−20
Jul 1962 – Jan 1966	43	60	Feb 1966 – Sep 1966	8	−16
Oct 1966 – Nov 1968	26	35	Dec 1968 – Jun 1970	19	−30
Jul 1970 – Dec 1972	30	51	Jan 1973 – Sep 1974	21	−45
Oct 1974 – Dec 1976	27	45	Jan 1977 – Feb 1978	14	−15
Mar 1978 – Nov 1980	33	58	Dec 1980 – Jul 1982	20	−21
Aug 1982 – Aug 1987	61	176	Sep 1987 – Nov 1987	3	−28
Dec 1987 – May 1990	30	46	Jun 1990 – Oct 1990	5	−15
Nov 1990 – Jun 1998	92	252	Jul 1998 – Aug 1998	2	−15
Sep 1998 – Aug 2000	24	49	Sep 2000 – Sep 2002	25	−43
Oct 2002 – Oct 2007	61	75	Nov 2007 – Feb 2009	16	−50
Mar 2009 – Apr 2011	26	71	May 2011 – Sep 2011	5	−16
Oct 2011 – Jun 2019	93	135			

Table B5: Expansion and Recession Periods According to Naïve Approach

Bull Markets Date	Durat. (Months)	Ampl. (in %)	**Bear Markets** Date	Durat. (Months)	Ampl. (in %)
May 1951 – Aug 1953	28	8	Sep 1953 – Jan 1954	5	12
Feb 1954 – Mar 1957	38	69	Apr 1957 – Apr 1957	1	0
May 1957 – Jul 1957	3	1	Aug 1957 – Jun 1958	11	0
Jul 1958 – Apr 1960	22	15	May 1960 – May 1960	1	0
Jun 1960 – Aug 1960	3	0	Sep 1960 – Jan 1961	5	15
Feb 1961 – May 1962	16	−6	Jun 1962 – May 1963	12	29
Jun 1963 – Jun 1963	1	0	Jul 1963 – Aug 1963	2	5
Sep 1963 – May 1966	33	20	Jun 1966 – Apr 1967	11	11
May 1967 – May 1967	1	0	Jun 1967 – Jul 1967	2	5
Aug 1967 – Aug 1969	25	2	Sep 1969 – Jan 1971	17	3
Feb 1971 – Feb 1971	1	0	Mar 1971 – Mar 1971	1	0
Apr 1971 – Jun 1973	27	0	Jul 1973 – Jul 1973	1	0
Aug 1973 – Aug 1973	1	0	Sep 1973 – Oct 1973	2	−0
Nov 1973 – Nov 1973	1	0	Dec 1973 – Oct 1975	23	−9
Nov 1975 – May 1977	19	5	Jun 1977 – Jun 1977	1	0
Jul 1977 – Jul 1977	1	0	Aug 1977 – Jul 1978	12	4
Aug 1978 – Oct 1978	3	−10	Nov 1978 – Jan 1979	3	6
Feb 1979 – Jan 1982	36	25	Feb 1982 – Oct 1982	9	18
Nov 1982 – Jul 1984	21	9	Aug 1984 – Aug 1984	1	0
Sep 1984 – Oct 1984	2	−0	Nov 1984 – Nov 1984	1	0
Dec 1984 – Nov 1987	36	38	Dec 1987 – Jan 1988	2	4
Feb 1988 – May 1988	4	−2	Jun 1988 – Feb 1989	9	6
Mar 1989 – Sep 1990	19	4	Oct 1990 – Feb 1991	5	21
Mar 1991 – Dec 1994	46	22	Jan 1995 – Jan 1995	1	0
Feb 1995 – Nov 2000	70	170	Dec 2000 – Jan 2001	2	3
Feb 2001 – Feb 2001	1	0	Mar 2001 – Oct 2003	32	−9
Nov 2003 – Feb 2008	52	26	Mar 2008 – Feb 2010	24	−16
Mar 2010 – Sep 2015	67	64	Oct 2015 – Oct 2015	1	0
Nov 2015 – Jan 2016	3	−7	Feb 2016 – Apr 2016	3	7
May 2016 – Jun 2016	2	0	Jul 2016 – Jul 2016	1	0
Aug 2016 – May 2019	34	27	Jun 2019 – Jun 2019	1	0

Figure B1: Filtered Recession Probability of MS Models



Note: Left panel shows the filtered recession probability of the simple MS model and right panel shows the filtered recession probability of the MSAR model. Recession periods are indicated by gray-shaded areas.

Appendix C: Evaluation Measures

To assess the **in-sample** predictive power of every variable, a t-test is used for linear regressions and a z-test is applied for probit models. The goodness of fit is measured in three ways. In addition to the standard information criteria AIC and BIC, we consider the R^2. For probit models, we rely on the Pseudo-R^2 proposed by Estrella (1998):

$$\text{Pseudo } R^2 = 1 - \left(\frac{\log L_u}{\log L_c}\right)^{-(2/T)\log L_c}$$

L_u represents the log-likelihood of the predictive model and L_c denotes the log-likelihood of the null model with only a constant. For the multivariate models, we penalize both R^2 measures for the reduction in the degrees of freedom by using the adjusted R^2:

$$\text{adj. } R^2 = 1 - (1 - R^2)\frac{n-1}{n-k-1}$$

In the **out-of-sample** predictions, we distinguish between statistical and economic performance. Statistical accuracy is measured in three ways. First, we use the root mean squared error (RMSE) to evaluate the predictive performance in linear models:

$$RMSE = \sqrt{\sum_{t=1}^{T} \frac{(Y_{t+1} - \hat{Y}_{t+1})^2}{T}}$$

In the case of the probit models, we rely on the quadratic probability score (QPS):

$$QPS = \frac{1}{T}\sum_{t=1}^{T} 2[\hat{p}_{t+1} - D_{t+1}]^2$$

The RMSE depends on the scale of the response and takes only positive numbers. The QPS is defined on the interval between 0 (perfect accuracy) and 2 (worst possible accuracy). The term T reflects the number of predictions.

Second, we use the adjusted mean squared prediction error (MSPE) by Clark and West (2007) to compare nested predictions. This measures tests whether the null hypothesis "equal prediction power" or the alternative "the larger model has a smaller adjusted MSPE" is more likely. Since a larger model produces additional noise in the prediction, the ordinary MSPE is corrected in the test statistic. $\hat{Q}_{1,t+1}$ and $\hat{Q}_{2,t+1}$ 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 $\hat{f}_{t+1} = \hat{e}_{1,t+1} - [\hat{e}_{2,t+1} - (\hat{Q}_{1,t+1} - \hat{Q}_{2,t+1})^2]$. Using the sample average $\bar{f} = 1/T \sum_{t=1}^{T} \hat{f}_{t+1}$ and the sample variance $\hat{V} = 1/(T-1)(\hat{f}_{t+1} - \bar{f})$, the CW test statistic is as follows:

$$CW = \frac{\sqrt{T}\bar{f}}{\sqrt{\hat{V}}}$$

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

To evaluate forecasts from non-nested models, we compare them pairwise with the test by Diebold and Mariano (1995). e_{it} and e_{jt} denote the forecasts errors of models i and j; $g(e_{it})$ and $g(e_{jt})$ are the corresponding quadratic loss functions. The loss differential $d = g(e_{it}) - g(e_{jt})$ is assumed to have an expected value of zero, to be covariance stationary, and to be asymptotically normally distributed. Then, the resulting DM test statistic is as follows:

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

$\bar{d}$ denotes the average loss differential and the denominator represents the standard deviation of d. $\hat{f}_d(0)$ indicates an estimate of the spectral density at frequency zero.[18] The DM test statistic relies on the null hypothesis of equal prediction accuracy with the alternative that the forecast of model j is more accurate.

As a final measure of statistical performance, we are interested in the hit ratios, that is, how often can we correctly classify bullish and bearish months depending on the chosen identification approach. Since the aim of this paper is to predict market *recessions*, this measure is of particular interest.

The **economic value** of model-based strategies is examined with commonly used performance measures. In addition to the annualized average return R and the volatility σ, we calculate two risk-adjusted performance measures: the Sharpe ratio (SR) and omega. The SR expresses the ratio between the excess return of a strategy over the risk-free interest rate (one-month treasury bill) and the strategy's volatility. Since the SR considers only the first two moments of the return distribution, it is only optimal for normally distributed returns. Since this assumption is often questionable for stock returns, Keating and Shadwick (2002) propose a universal performance measure (omega) that considers all moments of the strategy's return distribution $F(x)$. Using a threshold of 0, omega can be expressed as follows:

$$omega = \frac{\int_0^b 1 - F(x)dx}{\int_a^0 F(x)dx}$$

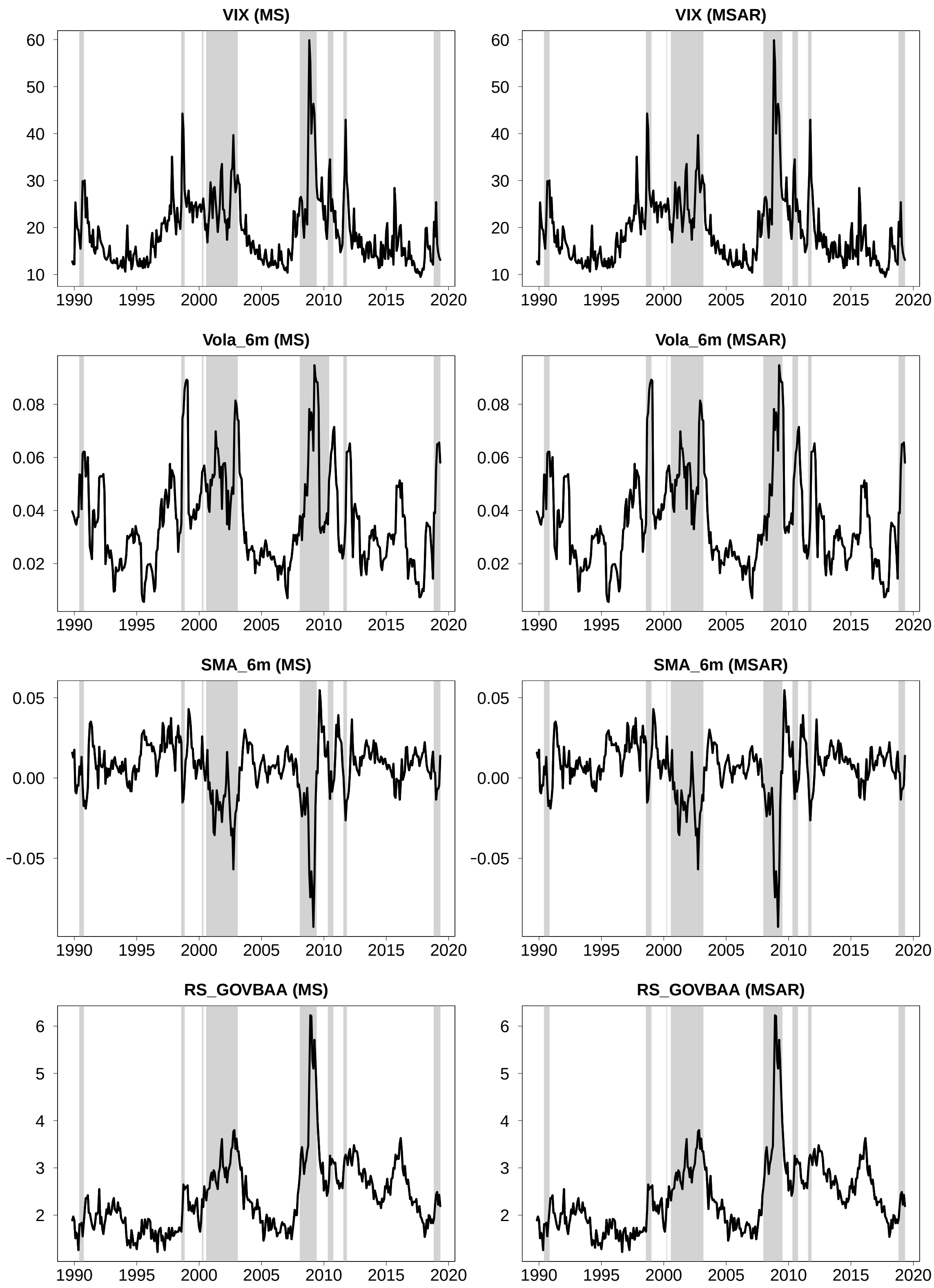
[18] See Diebold and Mariano (1995) for details. To account for small sample properties, we use the corrected version of Harvey et al. (1997) in this paper.

a and b are the minimum and maximum values of the return distribution. A value larger (smaller) than one implies that the probability mass for positive returns is larger (smaller) than for negative returns.

Finally, we consider downside-risk measures to assess whether the strategy protects from significant losses. For this purpose, we calculate the maximum drawdown (MaxDD) and the value-at-risk (VaR). The VaR indicates the maximum loss that will not be exceeded with a certain probability and over a certain time horizon. Our calculation relies a confidence level of 95% and a one-month time horizon. We utilize the historical return distribution to calculate the VaR.

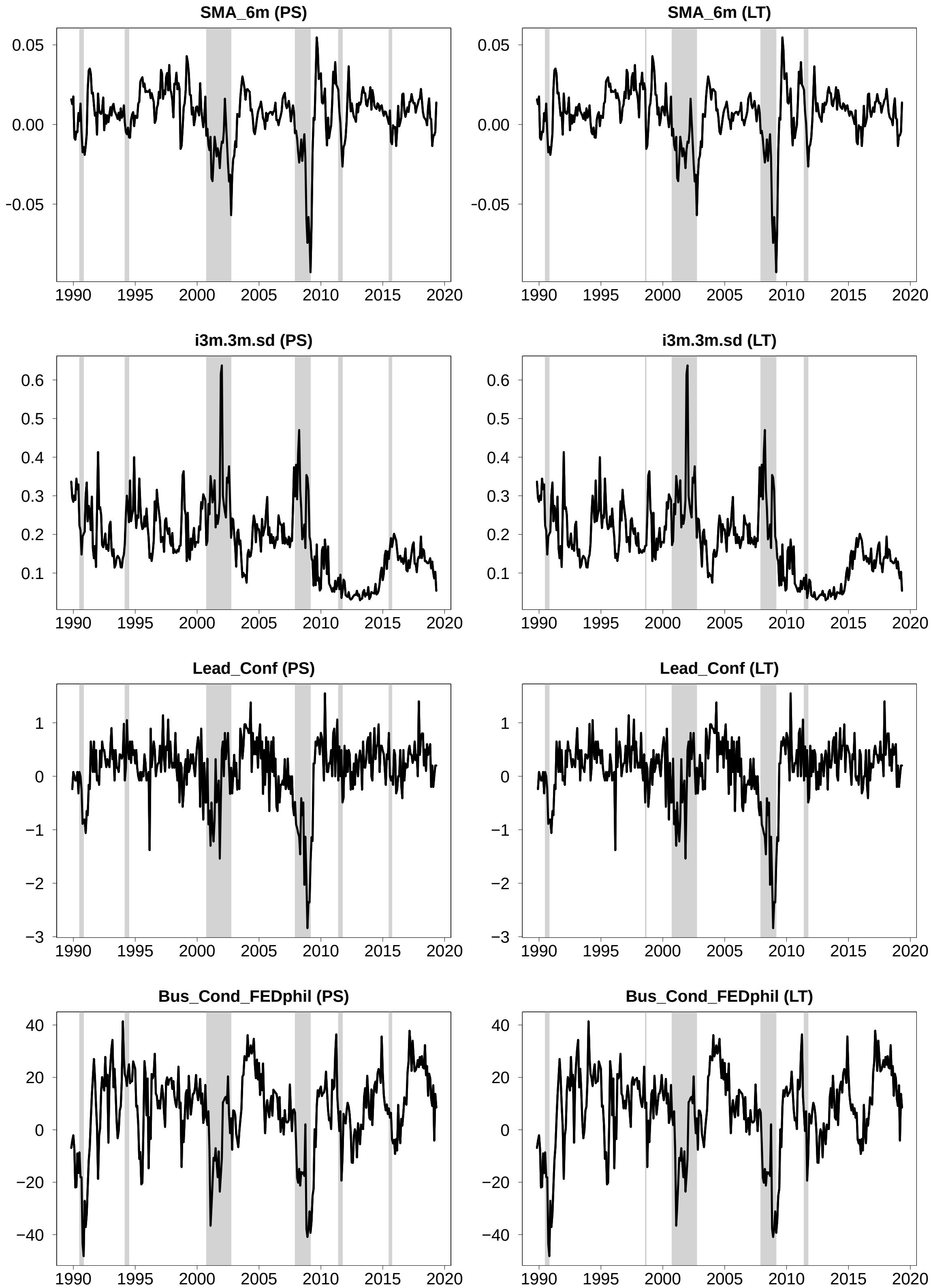
Appendix D: In-Sample Results

Figure D1: Top-4 Variables According to MS and MSAR Identification



Notes: Recessions are highlighted by gray-shaded areas.

Figure D2: Top-4 Variables According to PS and LT Identification



Notes: Recessions are highlighted by gray-shaded areas.

Table D1: Top-10 PCA Loadings

PC 1		**PC 2**	
T3Y	2.84	gdp	2.77
T2Y	2.83	cons	2.75
T5Y	2.83	prod	2.70
i3m.12m	2.80	RS_AAABAA	2.69
T1Y	2.80	pmi	2.68
T7Y	2.80	inv	2.63
DGS6MO	2.78	Bus_Cond_FEDphil	2.62
DGS3MO	2.77	unemp	2.56
i3m.3m	2.77	Int_exp_lower	2.53
T10Y	2.75	cons.sd	2.50

PC 3		**PC 4**	
term.spread.exp.3m	4.32	CAPE	4.73
TS_5Y3M	4.26	PD	4.63
term.spread.exp.12m	4.06	Cons_exp_Conf	3.91
TS_10Y2Y	3.70	R	3.80
Cons_sit_Conf	3.58	Cons_clim_Conf	3.72
profit	2.92	Cons_clim_Mich	3.22
Lead_Conf	2.35	JPYAUD	2.93
TEDRATE	2.27	SMA_6m	2.77
gdp.sd	2.26	profit	2.61
TS_30Y10Y	2.01	prod	2.59

Notes: Table show the relative shares on each principal component (in %).

Table D2: Specific-to-General Linear Models

	MS st.	MS dyn.	MSAR st.	MSAR dyn.
ar1		0.062		0.017
ar2		0.391*		0.506*
ar3		0.154***		0.158
ma1		0.570*		0.629*
ma2				−0.002
VIX	0.021***	0.011***	0.021***	0.010***
SMA_6m	−2.521***	0.229	−1.006	0.541
TEDRATE	−0.186***	−0.064	−0.186***	−0.065
CPI		0.018		
Cons_sit_Conf			0.002***	0.001
Int_exp_lower			0.003	
Bus_Cond_FEDphil	−0.002*	−0.002**	−0.002*	−0.003***
inv.sd	−0.138**		−0.108*	
RS_GOVAAA			−0.133*	
RS_GOVBAA	0.177***		0.248***	
TS_30Y10Y	−0.169***			
IP	0.010**		0.013*	
T2Y			−0.023	
cpi	0.048**		0.070**	
cons		0.024		
housep.sd	0.056*	0.038		
pmi			0.004	
Oil	0.001		0.001	
stock_exp_higher			−0.004*	
F_10Y_expiration2			−0.006**	
inv		−0.012	−0.007	
Constant	−0.413***	0.031	0.004	0.053
adj. R^2	0.648	0.707	0.679	0.729
AIC	−273.06	−336.03	−276.44	−341.75
BIC	−222.76	−281.86	−202.92	−295.32

Notes: Table shows coefficients of least squares models. Standard errors are not shown to conserve space. Static models are estimated with Newey and West (1987) standard errors. *, **, and *** indicate significance at the 10%, 5%, and 1% level, respectively.

Table D3: Specific-to-General Probit Models

	PS st.	PS dyn.	LT st.	LT dyn.
D_{t-1}		3.616***		3.535***
Bus_Cond_FEDphil	−0.044***			
i3m.12m.sd	7.140***	3.843***	3.374**	1.852*
i10y.12m.sd	−7.869***	−5.459***		
i3m.12m	−0.435***			
profit			−0.075	
profit.sd	6.318***		4.118***	
PE	−0.093***		−0.069***	
CPI	0.480**		1.088***	
Unem		1.719*		
house_market_nahb		−0.114*		
cpi			−0.387	
cons			0.980***	
inv			−0.273**	
Oil	0.034***	0.030*	0.025*	0.027
JPYAUD	−0.072**		−0.074**	
Core_CPI	−0.389		−1.216***	
IP	−0.197***			
gdp	0.698**			
fb.sd	0.016**		0.017**	
ppi.sd			−0.014	
fed_eff			−0.144	
ppi				0.165
Constant	−1.052	−1.718***	−1.168	−3.315***
adj. *Pseudo* R^2	0.421	0.718	0.366	0.684
AIC	197.72	92.59	196.22	85.48
BIC	251.89	119.66	254.26	104.81

Notes: Table shows coefficients of probit models. Standard errors are not shown to conserve space. Static models are estimated with Newey and West (1987) standard errors. *, **, and *** indicate significance at the 10%, 5%, and 1% level, respectively.

Appendix E: Out-of-Sample Results

Table E1: Diebold-Mariano Tests for Linear Models

MS	PCR f/st	PCR f/dy	PCR u/st	PCR u/dy	StG f/st	StG f/dy	StG u/st	StG u/dy
PCR f/st		0.0000	0.9407	0.0001	0.0029	0.0002	0.0019	0.0001
PCR f/dy	1.0000		1.0000	0.8196	0.9827	0.9569	0.9722	0.9601
PCR u/st	0.0593	0.0000		0.0000	0.0002	0.0000	0.0001	0.0000
PCR u/dy	0.9999	0.1804	1.0000		0.9804	0.9479	0.9690	0.9486
StG f/st	0.9971	0.0173	0.9998	0.0196		0.0316	0.0901	0.0326
StG f/dy	0.9998	0.0431	1.0000	0.0521	0.9684		0.9429	0.3843
StG u/st	0.9981	0.0278	0.9999	0.0310	0.9099	0.0571		0.0573
StG u/dy	0.9999	0.0399	1.0000	0.0514	0.9674	0.6157	0.9427	
MSA	**PCR f/st**	**PCR f/dy**	**PCR u/st**	**PCR u/dy**	**StG f/st**	**StG f/dy**	**StG u/st**	**StG u/dy**
PCR f/st		0.0001	0.9317	0.0002	0.0000	0.0002	0.0000	0.0000
PCR f/dy	0.9999		1.0000	0.8676	0.9658	0.7612	0.9534	0.5698
PCR u/st	0.0683	0.0000		0.0001	0.0000	0.0000	0.0000	0.0000
PCR u/dy	0.9998	0.1324	0.9999		0.9592	0.6858	0.9456	0.4873
StG f/st	1.0000	0.0342	1.0000	0.0408		0.0422	0.1963	0.0320
StG f/dy	0.9998	0.2388	1.0000	0.3142	0.9578		0.9428	0.3209
StG u/st	1.0000	0.0466	1.0000	0.0544	0.8037	0.0572		0.0453
StG u/dy	1.0000	0.4302	1.0000	0.5127	0.9680	0.6791	0.9547	

Note: Table reports p-values according to a one-step ahead forecast with a quadratic loss function. The Diebold-Mariano test compares the forecasting performance of model A (rows) with model B (columns). The null hypothesis is that the forecasts of A and B have the same accuracy. The alternative hypothesis is that the forecasts of B are more accurate. More details can be found in Appendix C. PCR: regression using principal components; StG: variable selection using a specific-to-general approach; f: components/variables are kept fixed during the test window; u: components/variables are updated every month during the test window; st: static prediction model; dy: dynamic prediction model.

Table E2: Diebold-Mariano Tests for Probit Models

PS	PCR f/st	PCR f/dy	PCR u/st	PCR u/dy	StG f/st	StG f/dy	StG u/st	StG u/dy
PCR f/st		0.9989	0.9174	0.9996	1.0000	1.0000	0.9995	1.0000
PCR f/dy	0.0011		0.0618	0.9254	1.0000	1.0000	0.9981	0.9999
PCR u/st	0.0826	0.9382		0.9875	1.0000	1.0000	0.9990	0.9999
PCR u/dy	0.0004	0.0746	0.0125		1.0000	1.0000	0.9974	0.9998
StG f/st	0.0000	0.0000	0.0000	0.0000		0.9480	0.2033	0.6673
StG f/dy	0.0000	0.0000	0.0000	0.0000	0.0520		0.0191	0.0944
StG u/st	0.0005	0.0019	0.0010	0.0026	0.7967	0.9809		0.9331
StG u/dy	0.0000	0.0001	0.0001	0.0002	0.3327	0.9056	0.0669	

LT	PCR f/st	PCR f/dy	PCR u/st	PCR u/dy	StG f/st	StG f/dy	StG u/st	StG u/dy
PCR f/st		0.9926	0.5770	0.9963	1.0000	1.0000	1.0000	1.0000
PCR f/dy	0.0074		0.0118	0.8813	1.0000	1.0000	1.0000	1.0000
PCR u/st	0.4230	0.9882		0.9949	1.0000	1.0000	1.0000	1.0000
PCR u/dy	0.0037	0.1187	0.0051		1.0000	1.0000	1.0000	1.0000
StG f/st	0.0000	0.0000	0.0000	0.0000		0.9611	0.7252	0.6612
StG f/dy	0.0000	0.0000	0.0000	0.0000	0.0389		0.1277	0.0347
StG u/st	0.0000	0.0000	0.0000	0.0000	0.2748	0.8723		0.3314
StG u/dy	0.0000	0.0000	0.0000	0.0000	0.3388	0.9653	0.6686	

Note: See Table E1.

Table E3: Accuracy of Regime Prediction

	Selection Approach	Recession Identification	Model	Selection Fixed $\tau = 25\%$	Updated $\tau = 25\%$	Fixed $\tau = 50\%$	Updated $\tau = 50\%$
Total	StG	MS	Static	82.3	83.4	84.0	85.1
			Dynamic	85.1	86.9	85.1	86.9
		MSAR	Static	80.0	81.7	86.9	87.4
			Dynamic	85.1	85.7	86.9	88.0
		PS	Static	74.3	78.3	78.3	80.6
			Dynamic	72.6	77.1	79.4	80.0
		LT	Static	77.1	77.1	87.4	84.6
			Dynamic	73.7	80.6	84.6	84.0
	PCR	MS	Static	74.9	73.1	**89.1**	86.9
			Dynamic	86.9	87.4	86.3	86.9
		MSAR	Static	66.3	68.0	89.1	**90.3**
			Dynamic	85.7	85.7	88.0	88.0
		PS	Static	84.6	82.9	**87.4**	**87.4**
			Dynamic	83.4	81.1	85.1	85.1
		LT	Static	90.3	90.9	**93.7**	**93.7**
			Dynamic	88.0	88.0	90.9	90.3
Expansion	StG	MS	Static	87.1	88.6	98.6	**99.3**
			Dynamic	87.1	91.4	98.6	97.9
		MSAR	Static	81.9	83.3	99.3	**100.0**
			Dynamic	88.4	88.4	97.1	97.8
		PS	Static	80.7	86.0	86.0	89.3
			Dynamic	77.3	84.0	87.3	88.7
		LT	Static	84.4	82.5	96.1	90.9
			Dynamic	78.6	86.4	91.6	91.6
	PCR	MS	Static	75.7	73.6	**99.3**	**99.3**
			Dynamic	90.7	91.4	97.9	97.9
		MSAR	Static	63.0	66.7	99.3	99.3
			Dynamic	89.1	89.1	98.6	98.6
		PS	Static	91.3	90.0	**96.0**	**96.0**
			Dynamic	91.3	88.7	94.0	94.0
		LT	Static	92.2	92.9	**98.1**	97.4
			Dynamic	94.2	93.5	**98.1**	97.4
Recession	StG	MS	Static	62.9	62.9	25.7	28.6
			Dynamic	65.7	68.6	37.1	42.9
		MSAR	Static	73.0	75.7	40.5	40.5
			Dynamic	73.0	75.7	48.6	51.4
		PS	Static	36.0	32.0	32.0	28.0
			Dynamic	**44.0**	36.0	32.0	28.0
		LT	Static	23.8	38.1	23.8	38.1
			Dynamic	38.1	38.1	33.3	28.6
	PCR	MS	Static	**71.4**	**71.4**	48.6	37.1
			Dynamic	**71.4**	**71.4**	40.0	42.9
		MSAR	Static	**78.4**	73.0	51.4	56.8
			Dynamic	73.0	73.0	48.6	48.6
		PS	Static	**44.0**	40.0	36.0	36.0
			Dynamic	36.0	36.0	32.0	32.0
		LT	Static	**76.2**	**76.2**	61.9	66.7
			Dynamic	42.9	47.6	38.1	38.1

Notes: Best accuracy rates for each identification approach are in bold. Total months: 175. Recession months: 35 (MS), 37 (MSAR), 25 (PS), and 21 (LT).

Table E4: Economic Performance with Threshold τ=50%

	R p.a.	σ p.a.	SR p.a.	Omega	VaR (95%)	Max. DD
PCR f/st MS	**0.0936**	0.1113	**0.7291**	**1.9261**	**−0.0531**	**−0.0972**
PCR f/dy MS	0.0518	**0.1112**	0.3537	1.4314	−0.0636	−0.2074
PCR u/st MS	0.0647	0.1282	0.4079	1.4990	−0.0646	−0.3562
PCR u/dy MS	0.0507	0.1112	0.3442	1.4224	−0.0636	−0.2074
StG f/st MS	0.0646	0.1247	0.4182	1.5090	−0.0636	−0.3570
StG f/dy MS	0.0594	0.1129	0.4158	1.5022	−0.0636	−0.2074
StG u/st MS	0.0582	0.1239	0.3692	1.4521	−0.0636	−0.3570
StG u/dy MS	0.0507	**0.1112**	0.3442	1.4224	−0.0636	−0.2074
PCR f/st MSAR	**0.0887**	0.1100	0.6929	1.8776	−0.0531	−0.1023
PCR f/dy MSAR	0.0533	0.1103	0.3705	1.4567	−0.0636	−0.2155
PCR u/st MSAR	0.0879	**0.1041**	**0.7250**	**1.9266**	**−0.0496**	**−0.0959**
PCR u/dy MSAR	0.0533	0.1103	0.3705	1.4567	−0.0636	−0.2155
StG f/st MSAR	0.0780	0.1129	0.5803	1.7049	−0.0599	−0.2084
StG f/dy MSAR	0.0505	0.1101	0.3456	1.4324	−0.0636	−0.2155
StG u/st MSAR	0.0762	0.1133	0.5624	1.6770	−0.0599	−0.2084
StG u/dy MSAR	0.0593	0.1071	0.4375	1.5382	−0.0599	−0.1452
PCR f/st PS	**0.0820**	0.1125	**0.6180**	**1.7533**	**−0.0563**	**−0.1452**
PCR f/dy PS	0.0703	0.1131	0.5113	1.6187	−0.0599	−0.1493
PCR u/st PS	**0.0820**	0.1125	**0.6180**	**1.7533**	**−0.0563**	**−0.1452**
PCR u/dy PS	0.0703	0.1131	0.5113	1.6187	−0.0599	−0.1493
StG f/st PS	0.0639	**0.1072**	0.4794	1.6224	**−0.0563**	−0.1570
StG f/dy PS	0.0589	0.1099	0.4223	1.5330	−0.0599	−0.1505
StG u/st PS	0.0645	0.1161	0.4482	1.5575	−0.0636	−0.2480
StG u/dy PS	0.0580	0.1146	0.3977	1.5016	−0.0636	−0.2647
PCR f/st LT	0.0841	0.1148	0.6237	1.7608	−0.0599	−0.1447
PCR f/dy LT	0.0840	0.1159	0.6174	1.7375	−0.0599	−0.1452
PCR u/st LT	0.0822	**0.1100**	**0.6335**	**1.7796**	**−0.0563**	−0.1447
PCR u/dy LT	**0.0844**	0.1159	0.6206	1.7429	−0.0599	−0.1452
StG f/st LT	0.0703	0.1176	0.4915	1.5800	−0.0636	−0.2159
StG f/dy LT	0.0691	0.1135	0.4993	1.6050	−0.0599	−0.1539
StG u/st LT	0.0747	0.1102	0.5644	1.7108	**−0.0563**	**−0.1243**
StG u/dy LT	0.0767	0.1109	0.5791	1.7224	**−0.0563**	−0.1532
MA(16)	0.0624	0.1061	0.4705	1.6048	−0.0563	−0.1608
50/50	0.0367	0.0704	0.3447	1.4911	−0.0357	−0.2990
BH	0.0610	0.1409	0.3447	1.3980	−0.0727	−0.5668

Notes: Best performing models according to each metric are in bold. PCR: regression using principal components; StG: variable selection using a specific-to-general approach; f: components/variables are kept fixed during the test window; u: components/variables are updated every month during the test window; st: static prediction model; dy: dynamic prediction model.

Table E5: Economic Performance with Threshold τ=25%

	R p.a.	σ p.a.	SR p.a.	Omega	VaR (95%)	Max. DD
PCR f/st MS	0.0501	0.0851	0.4425	**1.6798**	−**0.0367**	−0.1164
PCR f/dy MS	0.0486	0.0973	0.3717	1.5055	−0.0540	−0.1462
PCR u/st MS	0.0374	**0.0836**	0.2980	1.4916	−0.0367	−0.1211
PCR u/dy MS	0.0541	0.0993	0.4193	1.5623	−0.0540	−0.1462
StG f/st MS	0.0387	0.0943	0.2780	1.4098	−0.0472	−0.1634
StG f/dy MS	0.0530	0.0988	0.4099	1.5504	−0.0540	−0.1233
StG u/st MS	0.0394	0.0969	0.2783	1.3993	−0.0554	−0.2168
StG u/dy MS	**0.0572**	0.0999	**0.4480**	1.5947	−0.0540	−**0.1117**
PCR f/st MSAR	0.0437	**0.0771**	0.4053	**1.7240**	−**0.0316**	−0.1200
PCR f/dy MSAR	0.0443	0.0907	0.3507	1.4940	−0.0472	−0.1579
PCR u/st MSAR	0.0357	0.0804	0.2889	1.5154	−0.0336	−0.1421
PCR u/dy MSAR	0.0443	0.0907	0.3507	1.4940	−0.0472	−0.1579
StG f/st MSAR	**0.0522**	0.0843	**0.4713**	1.7006	−0.0336	−**0.0795**
StG f/dy MSAR	0.0408	0.0924	0.3068	1.4457	−0.0472	−0.1610
StG u/st MSAR	0.0477	0.0855	0.4124	1.6114	−0.0367	−0.0877
StG u/dy MSAR	0.0453	0.0914	0.3595	1.5149	−0.0413	−0.1647
PCR f/st PS	0.0590	0.1053	0.4418	1.5487	−0.0563	−0.1619
PCR f/dy PS	**0.0705**	0.1094	**0.5306**	**1.6581**	−0.0563	−**0.1493**
PCR u/st PS	0.0555	0.1046	0.4115	1.5184	−0.0563	−0.2027
PCR u/dy PS	0.0627	0.1084	0.4639	1.5854	−0.0563	−0.1494
StG f/st PS	0.0622	0.1067	0.4658	1.6200	−0.0563	−0.1660
StG f/dy PS	0.0595	**0.1048**	0.4491	1.6172	−0.0563	−0.1539
StG u/st PS	0.0665	0.1119	0.4832	1.6229	−0.0563	−0.2480
StG u/dy PS	0.0611	0.1103	0.4408	1.5793	−0.0563	−0.2516
PCR f/st LT	0.0737	0.1009	0.6068	1.7786	−**0.0496**	−0.1011
PCR f/dy LT	0.0695	0.1124	0.5077	1.6153	−0.0599	−0.1369
PCR u/st LT	0.0762	**0.1016**	**0.6272**	**1.8061**	−**0.0496**	−**0.1003**
PCR u/dy LT	0.0670	0.1075	0.5078	1.6214	−0.0563	−0.1369
StG f/st LT	**0.0767**	0.1107	0.5796	1.7592	−0.0599	−0.2302
StG f/dy LT	0.0718	0.1045	0.5684	1.7725	−0.0506	−0.1456
StG u/st LT	0.0757	0.1057	0.5986	1.8214	−0.0531	−0.1206
StG u/dy LT	0.0743	0.1073	0.5759	1.7584	−0.0531	−0.1457
MA(16)	0.0624	0.1061	0.4705	1.6048	−0.0563	−0.1608
50/50	0.0367	0.0704	0.3447	1.4911	−0.0357	−0.2990
BH	0.0610	0.1409	0.3447	1.3980	−0.0727	−0.5668

Notes: See Table E4.