

How to Identify and Forecast Bull and Bear Markets?*

Erik Kole[†]
Dick van Dijk

Econometric Institute, Erasmus School of Economics, Erasmus University Rotterdam

October, 2013

Abstract

The state of the equity market, often referred to as a bull or a bear market, is of key importance for financial decisions and economic analyses. Its latent nature has led to several methods to identify past and current states of the market and forecast future states. These methods encompass semi-parametric rule-based methods and parametric regime-switching models. We compare these methods by new statistical and economic measures that take into account the latent nature of the market state. The statistical measure is based directly on the predictions, while the economic measure is based on the utility that results when a risk-averse agent uses the predictions in an investment decision. Our application of this framework to the S&P500 shows that rule-based methods are preferable for (in-sample) identification of the market state, but regime-switching models for (out-of-sample) forecasting. In-sample only the direction of the market matters, but for forecasting both means and volatilities of returns are important. Both the statistical and the economic measures indicate that these differences are significant.

Key words: forecast evaluation, economic comparison, stock market, regime switching

JEL classification: C52, C53, G11, G17

*We thank Christophe Boucher, Thijs Markwat and seminar participants at the 66th Meeting of the Econometric Society in Málaga, the 18th Conference on CEF in Prague 2012, the 20th Meeting of the SNDE in Istanbul 2012, the (EC)² in Florence 2011, Robeco Quantitative Strategies, the 8th International Paris Finance Meeting 2010, Inquire's UK Autumn Seminar 2010 and Erasmus University for helpful comments and discussions. We thank Anne Opschoor for skillful research assistance and Inquire UK for financial support. Kole thanks the Netherlands Scientific Organisation (NWO) for financial support.

[†]Corresponding author. Address: Burg. Oudlaan 50, Room H11-13, P.O. Box 1738, 3000DR Rotterdam, The Netherlands, Tel. +31 10 408 12 58. E-mail addresses kole@ese.eur.nl (Kole) and djvandijk@ese.eur.nl (Van Dijk).

1 Introduction

The state of the equity market, a bull or a bear market, is an important variable for financial decision-making and economic analysis. Unfortunately, this variable is latent, which has led to several methods to identify and forecast the state of the equity market. In this paper we compare the performance of these methods on identification and on forecasting. By identification, we refer to the dating of past bull and bear markets. Comparing the identification shows how the properties of bull and bear markets, such as the mean and volatility of returns and the duration of phases, vary across methods, and which method works best in-sample. Comparing probability forecasts of bull and bear markets and moment forecasts of the returns can show the value of the different methods for decision making, and which method performs best out-of-sample.

Investors and economists need good information on the past, current and future state of the equity market. Investors require it for investment strategies and risk management, because of the different behavior of returns depending on market sentiment (see Perez-Quiros and Timmermann, 2000). Several authors find evidence for a relation between the state of the equity market and time-variation in risk premia (see, for example, Veronesi, 1999; Gordon and St-Amour, 2000; Ang et al., 2006), making the market state relevant for asset pricing. Regulators monitor the state of the equity market because of its relation to the credit supply (see Rigobon and Sack, 2003; Bohl et al., 2007). Economists use the state of the equity market to forecast macro-economic variables. Harvey (1989); Estrella and Mishkin (1998); Chauvet (1999) and Stock and Watson (2003) report evidence that the state of the equity market helps predicting the business cycle (see also Marcellino, 2006, for an overview).

We group the various methods to identify and forecast bull and bear markets into two fundamentally different types, semi-parametric methods and parametric ones. The semi-parametric methods treat identification and forecasting as two separate steps. In the first step, they use rules to determine the past states of the market. This step transforms a stock index series into a binary time series, based on its peaks and troughs. Different sets of rules can be used to determine which peaks and troughs mark switches between bull and bear markets. We consider Pagan and Sossounov (2003), whose rules combine

requirements for duration and price changes, and Lunde and Timmermann (2004), whose rules consists of requirements for price increases and decreases.¹ In the second step, these methods use models for binary time series to make forecasts. We use a dynamic logit model in this step, whose parameters vary according to the most recent state of the market. We consider several macro-finance variables as predictors.

As parametric methods we consider Markov regime-switching models in the style of Hamilton (1989, 1990).² In these models the return distribution depends on a latent state that follows a Markov chain, which naturally corresponds with bull and bear markets. Whereas the rule-based methods determine bull and bear markets mostly based on mean returns and regime duration, regime-switching models also take the volatility of returns into account. Moreover, the transition probabilities of the Markov chain can be a function of predictor variables (see Diebold et al., 1994). As a consequence, identification and forecasting result from a single step. Regime-switching models can accommodate more than two regimes (see Guidolin and Timmermann, 2006a,b, 2007), or specific features of financial markets such as crashes (see Kole et al., 2006) or bull market rallies (see Maheu et al., 2012).

Both sets of methods have their advantages and disadvantages. The main advantage of the semi-parametric methods is the transparency of the rules, though they require some arbitrary settings. As a disadvantage, forecasts follow from a two-step approach. Predictor variables that are relevant for forecasting are not used for the identification in the first step. To the contrary, the regime switching models specify one data-generating process that is used for both identification and forecasting. This specification gives the advantage of a complete density of returns at each point in time, which can be directly used for all kinds of analysis like the evaluation of (expected) utility functions, or the calculation of Value-at-Risk. Rule-based approaches require additional assumptions. As a disadvantage, specifying

¹See Edwards et al. (2003); Candelon et al. (2008); Chen (2009) and Kaminsky and Schmukler (2008) for applications of the approach proposed by Pagan and Sossounov (2003), and Chiang et al. (2009) for the method proposed by Lunde and Timmermann (2004).

²See for instance Hamilton and Lin (1996); Maheu and McCurdy (2000); Chauvet and Potter (2000); Ang and Bekaert (2002); Guidolin and Timmermann (2008) and Chen (2009) for applications, and the surveys by Ang and Timmermann (2011) and Guidolin (2011a,b).

a data-generating process makes parametric approaches vulnerable to misspecification, to which rule-based approaches may be more robust.

We propose two new measures, one statistical and the other economic, to compare the identification and forecasts of the different methods that can take into account the latent nature of bull and bear markets. Existing tests of equal predictive ability do not suffice, because they assume that the variable being predicted can eventually be observed. A generally accepted chronology of bull and bear markets like the NBER chronology of the business cycle could be used as observations of the true market state, but does not exist. As argued by for example Pesaran and Skouras (2002), in-sample and out-of-sample predictions can be evaluated by statistical and economic measures of predictive accuracy. Statistical measures are general and relevant in many circumstances, while economic measures may better reflect actual costs and benefits of applying methods.

The statistical measure, which we call the Integrated Absolute Difference (IAD), is based on the absolute difference between the probabilities for the different states. It quantifies how different two methods identify or forecast bull and bear markets. While closely related to the Integrated Squared Difference of Pagan and Ullah (1999) and Sarno and Valente (2004), the IAD is easier to interpret as a difference in probability. The measure can be applied to bivariate and multivariate state distributions, and can handle binary realizations, i.e. zero or one.

As economic measure we propose the maximum fee or transaction cost that a risk-averse investor would pay to switch from one method to another as the base of an investment strategy. The investment strategy uses the predictions of the methods to take a position in futures contracts. This setting yields a realistic comparison of the various methods, and shows which method leads to the best performance. While fees have been proposed before to measure economic impact³, we adopt it as a statistic and address its uncertainty (McCracken and Valente, 2012, consider fees in a similar way).

We conduct an empirical comparison of the various methods for the US equity market proxied by the S&P500 over the period 1955–2010. We consider several macroeconomic and financial variables to predict switches between bull and bear markets. The comparison of

³See for example West et al. (1993); Fleming et al. (2001); Ang and Bekaert (2002); Das and Uppal (2004) and Marquering and Verbeek (2004).

the methods for their in-sample identification is based on the whole sample period, whereas we use the second half to compare the forecasts.

Our in-sample results indicate a preference for the rule-based methods. The IADs indicate that they lead to identifications of bull and bear markets that differ significantly from those by the parametric methods. On the other hand, the difference between the methods of Pagan and Sossounov (2003) and Lunde and Timmermann (2004) is negligible. Since identification is done *ex post*, separating periods with increasing prices from those with decreasing prices leads to the best decisions and hence a positive fee. Including volatility, as regime switching models do, does not improve identification.

For forecasting, our results show that regime-switching models work best. The forecasts of the states by regime-switching models differ significantly from those by the semi-parametric methods. Because regime-switching models take volatility into account, investments based on its forecasts are more prudent and produce a higher utility. An investor is willing to pay a maximum transaction cost or fee of 30 basis points to adopt a regime-switching model instead of a semi-parametric method. Differences between the different regime switching models are small, but the forecasts by the semi-parametric methods differ considerably from each other. This difference comes from the different way in which these methods determine the most recent peak or trough.

We conclude that incorporating the volatility of the market improves forecasts, but reduces the quality of the dating of past bull and bear markets. This conclusion relates directly to the debate between Harding and Pagan (2003a,b) and Hamilton (2003) on the best method to date business cycle regimes. Harding and Pagan advocate simple classification rules, while Hamilton proposes regime-switching models. In the business cycle context, both methods produce comparable results, because regimes in the GDP growth process differ mostly in their means. For the state of the equity market, volatility is a distinguishing characteristic, too. This explains why we find considerable differences between rule-based, which focus on the trend, and model-based methods, which use both means and volatilities.

Both types of methods can be extended to include predictor variables. We consider several macro-financial variables, but find that they lead to marginal improvements at best. The IADs with respect to the models without predictor variables are significant,

indicating that the variables have a non-negligible effect. The maximum fees on the other hand are small and insignificant, which shows that the differences have little value in financial decision-making. The small effect of these macro-financial variables is in line with the results of Welch and Goyal (2008) for direct forecasts of stock returns. We find less predictive ability of macro-financial variables than Chen (2009). He also considers regime-switching models, and links the inference probabilities to predictor variables. His two-step analysis may overstate the effect of predictor variables, because it treats those probabilities as observed, ignoring that they are inferred.

2 Different methods and their comparison

The state of the equity market is a latent process. Both semi-parametric and parametric methods exist to identify and forecast it, based on observations of a market index and possibly a set of predictor variables. We develop a unifying framework for all these methods. This framework forms the basis for the measures to compare the identifications and forecasts. We first discuss this framework, then consider the two types of methods and their link to predictor variables, and finally introduce the measures to compare them.

2.1 A general framework

We use S_t to denote the state of the equity market at time t . This state can be one out of a finite set of possible states \mathcal{S}_m . The subscript m indicates that the set of states is method-specific, because methods can differ in the number and nature of the states they distinguish. S_t is a binary variable when only bull and bear markets are considered. When more states are distinguished, S_t is a categorical variable. Additional states can capture sudden booms and crashes as in Guidolin and Timmermann (2006a, 2007) and Kole et al. (2006) or bear market rallies and bull market corrections as in Maheu et al. (2012).

A method m to identify or forecast the state of the equity market can be seen as function F_m that returns a (conditional) probability for each possible state $s \in \mathcal{S}_m$ at time t ,

$$\Pr[S_t = s | \Omega_{t_0, T_0}] = F_m(s, t; \Omega_{t_0, T_0}, \boldsymbol{\theta}_m). \quad (1)$$

This function requires an information set Ω_{t_0, T_0} and parameters gathered in a vector θ_m . The information set contains a time-series of returns and a set of explanatory variables \mathbf{z}_t which can be empty, $\Omega_{t_0, T_0} = \{(r_\tau, \mathbf{z}_\tau)\}_{\tau=t_0}^{T_0}$.⁴ The information set starts at t_0 and ends at T_0 . The time t to which the probability pertains can be before or after the final date T_0 in the information set. If $t \leq T_0$, the resulting probability is an inference probability. These probabilities give the identification at time t , which is thus typically based on past, current and future information. In the analysis of regime-switching models, these probabilities are therefore called smoothed inference probabilities (see Hamilton, 1994, Ch. 22). If $t > T_0$, F_m returns a probability forecast. The parameters in θ_m can be specified by the user or estimated from the information set Ω_{t_0, T_0} or a subset of it.

We can compare different methods by their resulting probabilities for the different states. For a statistical comparison, we propose a measure based on the absolute difference between them. This measure indicates how different two methods are. To establish the preference of one method over another, we propose a measure based on economic decision-making. We derive the maximum fee that would make an investor indifferent in utility terms between two methods, when she applies them for an investment strategy. As pointed out by Pesaran and Skouras (2002), both comparisons have their merits. The statistical measure is more general, and can be relevant in various circumstances. In contrast, the economic measure relates to a specific decision problem, and can take costs and benefits of predictions into account. Wrong decisions may be more harmful than good decisions are beneficial. Because the economic measure is founded on utility functions, it has a clear link to economic theory, as advocated by West et al. (1993) and Granger and Pesaran (2000).

While many tests of comparable predictive ability exist⁵, the latent nature of the state of the equity market prevents us from applying these tests. These tests typically define a loss function over the forecast error, which is the difference between the prediction and the observed realization of a certain variable. Because the variable that is predicted remains

⁴Some non-parametric methods actually work with price index series to define peaks and troughs. These index series can be reconstructed from a return series up to a multiplication factor, which does not influence the results of these methods. Because a stationary information set is advantageous from a statistical point of view, a set of returns is included in Ω_{t_0, T_0} .

⁵See for example Diebold and Mariano (1995); West (1996); White (2000); Corradi and Swanson (2007).

latent in our case, we propose new measures.

We propose these measures primarily for the comparison of different methods, that is, different functions F_m and F_n that use the same information set Ω_{t_0, T_0} . Our measures can also be used to compare the results of the same method with different information sets. We can compare forecasts $F^m(s, t, \Omega_{t_0, T_0}; \boldsymbol{\theta}_m)$, $t > T_0$ with identifications $F^m(s, t, \Omega_{t_0, T_1}; \boldsymbol{\theta}_m)$, $t \leq T_1$. To determine the influence of a predictive variable, we can consider an information set Ω_{t_0, T_0} that includes this particular variable, and a second information set $\tilde{\Omega}_{t_0, T_0} \subset \Omega_{t_0, T_0}$ without it. The influence of the length of the information set can be investigated by varying t_0 or T_0 .

2.2 Identification and forecasting based on rules

The first type of methods for identification and forecasting of bull and bear markets are semi-parametric methods. These methods treat identification and forecasting separately. In the identification step, they apply a set of rules to a price index series, transforming it into a binary series. In the second step, they construct forecasts based on models for binary time series. These models can combine past values of the constructed binary series with other observed variables to make forecasts. Because the methods consist of a rule-based first step and a model-based second step, they are semi-parametric.

2.2.1 Identification by rules

We consider the rule-based algorithms proposed by Lunde and Timmermann (2004) and by Pagan and Sossounov (2003). Both methods obtain a classification of bull and bear markets using peaks and troughs in a market index P_t . They differ in the criteria used for the selection of peaks and troughs that constitute the actual switch points between bull and bear markets.

Lunde and Timmermann (2004, LT henceforward) select peaks and troughs as switch points based on the change in the level of a stock index. Their method requires two parameters λ_1 and λ_2 indicating the minimum increase or decrease since the last trough or peak that the index should show to label a switch. They mostly set $\lambda_1 = 0.20$ and $\lambda_2 = 0.15$, so an increase of 20% since the last trough signals a switch to a bull market,

and a decrease of 15% since the last peak a switch to a bear market. The identification of peaks and troughs in a price series uses an iterative search procedure (see Appendix A for more details).

Pagan and Sossounov (2003, PS henceforward) base their approach on the identification of business cycles in macroeconomic data (see also Harding and Pagan, 2002). They also use peaks and troughs to mark the switches between bull and bear markets. An observation at time t is a maximum (minimum) if all observations between $t - \tau_{\text{window}}$ and $t + \tau_{\text{window}}$ are lower (higher). To determine which peaks and troughs constitute switch points, PS put restrictions on the minimum duration of individual phases and complete cycles (see Appendix A for details). These criteria form the main difference with the LT algorithm. Because the selection of maxima and minima is less reliable at the beginning and end of the sample, peaks and troughs during the first and last τ_{censor} observations are censored. For the information set Ω_{t_0, T_0} censoring means that the last switch can have occurred the latest at $T_0 - \tau_{\text{censor}}$. PS use $\tau_{\text{window}} = 32$ weeks and $\tau_{\text{censor}} = 13$ weeks.

2.2.2 Forecasting by models

In the second step, the semi-parametric methods need a model to make forecasts. Such a model should take into account that the state variable S_t is binary. We use a dynamic logit model to make forecasts for two reasons. First, a logit model is one of the standard models to relate binary dependent variables to predictor variables. Second, in a dynamic logit model the effect of a predictor variable can depend on the current state of the market. We thus allow for the possibility that some variables may have a different (or no) effect on the probability of a bull-bear switch than a bear-bull switch.

We model the probability of state s to occur at time $t + 1$, conditional on the current state q and predictor variables \mathbf{z}_t as

$$\pi_{qs,t+1}^m \equiv \Pr[S_{t+1} = s | S_t = q, \Omega_t] = \Lambda(\boldsymbol{\beta}_q^m \mathbf{z}_t), s, q \in \mathcal{S}_m \quad m = \text{LT, PS}, \quad (2)$$

where $\Lambda(x)$ denotes the logistic function, and $\boldsymbol{\beta}_q^m$ is the coefficient vector on the \mathbf{z}_t variables (including a constant). The coefficient $\boldsymbol{\beta}_q^m$ depends on the current state of the market q . If $\boldsymbol{\beta}_q^m$ is the same for all q , a static logit model results. If \mathbf{z}_t consists of a constant only, the market state follows a Markov chain of order one.

Our specification of the dynamic logit model is close to the dynamic probit model of Kauppi and Saikkonen (2008). Their specification is more general, as they allow for a dependence of $\pi_{qs,t+1}^m$ on lagged predictions $\pi_{qs,t-j}^m, j \geq 0$, and more lagged realizations S_{t-j} . However they only consider one predictor variable (the yield spread). Our specification incorporates the advice of Harding and Pagan (2011) to use Markov chain models for constructed binary time series. The order of the Markov chain can reflect duration restrictions that have been applied to the construction. Because the LT method does not have such restrictions, we use a Markov chain of order one.

To form a forecast for S_{t+1} , the prevailing state at time t is needed, which may not be available. In the LT-approach, only if P_t equals the last observed maximum (minimum), and is at least a fraction λ_1 above (λ_2 below) the prior minimum (maximum) is the market surely in a bull (bear) state. The PS-approach suffers from this drawback too, since extremes in the last τ_{censor} periods are censored. So, the market may already have switched, but this will only become clear later. In these cases, the state of the market is known until the period $t^* < t$ with the last extreme value. We construct the forecast for S_{t+1} , starting from the period t^* by the recursion

$$\Pr[S_{\tau+1} = s | \mathbf{z}_\tau] = \sum_{q \in \mathcal{S}_m} \Pr[S_{\tau+1} = s | S_\tau = q, \mathbf{z}_\tau] \Pr[S_\tau = q | \mathbf{z}_{\tau-1}] \quad t^* < \tau \leq t + 1. \quad (3)$$

Starting with the known state at t^* , we construct predictions for $t^* + 1$, which we use for the predictions of $t^* + 2$ and so on. The recursion stops at $t + 1$. The recursion essentially calculates the forecast probability for state s over the tree formed by all possible paths from t^* onwards.

2.3 Identification and forecasting by regime-switching models

The parametric methods we consider are Markov regime-switching models in the style of Hamilton (1989, 1990). In this approach, the state of the stock market follows a first order Markov chain with a specified number of regimes. The number of regimes can be set equal to two (a bullish and bearish regime) or larger.

Using such a model-based approach has several advantages. First, it offers more insight into the process. The model can be evaluated by its theoretical properties, like autocorre-

lation, or the expected duration of a phase. Second, we can extend the number of states in the model and test whether such extensions yield significant improvements. A third advantage is the ease with which we can compare results for different markets and different time periods. Models can typically be summarized by their coefficients, whereas a simple characterization of the rule-based results is less straightforward. These advantages come at the cost of misspecification risk. Misspecification of the number of regimes or structural breaks in the data generating process can severely impact the results. Because rule-based approaches do not make assumptions on distributions or on the absence or presence of variation over time, they may be more robust.

We consider Markov chain regime-switching models with k regimes (used as first suffix) and having either constant or time-varying transition probabilities (suffix C or L). For example, the label RS3L indicates a regime-switching model with three states and time-varying transition probabilities. Based on the results of Guidolin and Timmermann (2006a, 2007, 2008), we consider $k = \{2, 3\}$.

We assume that the index return r_t follows a normal distribution in each regime s with regime specific means and variances,

$$r_t \sim N(\mu_s^m, \omega_s^m), \quad s \in \mathcal{S}_m. \quad (4)$$

We order the regimes based on the estimated means. While asymmetric or fat-tailed distributions can be used for the regimes, Timmermann (2000) shows that a mixture of normal distributions can accommodate these features too.

The state of the market follows a first order Markov chain with transition matrix \mathbf{P}_t^m that can vary over time. This matrix contains transition probabilities

$$\pi_{qst}^m \equiv \Pr[S_t^m = s | S_{t-1}^m = q, \mathbf{z}_{t-1}], \quad s, q \in \mathcal{S}_m, \quad (5)$$

where the restriction $\forall q, \forall t : \sum_{s \in \mathcal{S}_m} \pi_{sqt}^m = 1$ applies. When the transition probabilities are constant, this restriction leaves $k(k-1)$ free parameters. If the probabilities are time-varying, we use a multinomial logit specification

$$\pi_{qst}^m = \frac{e^{\beta_{qs}^{m'} \mathbf{z}_{t-1}}}{\sum_{\zeta \in \mathcal{S}} e^{\beta_{q\zeta}^{m'} \mathbf{z}_{t-1}}}, \quad s, q \in \mathcal{S}_m, \quad (6)$$

with $\forall q \exists s \in \mathcal{S} : \beta_{qs} = \mathbf{0}$ to ensure identification. If the number of regimes equals two, this multinomial specification reduces to the standard logit specification

$$\pi_{qt}^m \equiv \Pr[S_t^m = u | S_{t-1}^m = q, \mathbf{z}_{t-1}] = \Lambda(\beta_q^{m'} \mathbf{z}_{t-1}), \quad m = \text{RS2L}. \quad (7)$$

This specification is mathematically similar to the dynamic logit models for the semi-parametric methods in Equation (2), though it is an integral part of the regime-switching model.

We finish by introducing parameters for the probability that the process starts in a specific state, $\xi_s^m \equiv \Pr[S_1^m = s]$. Again the restriction $\sum_{s \in \mathcal{S}_m} \xi_s^m = 1$ should be satisfied. We treat the remaining parameters as free, and estimate them.

We estimate the parameters of the resulting regime-switching model by means of the EM-algorithm of Dempster et al. (1977) (see also Hamilton, 1994, Ch. 24). In appendix B we extend the method of Diebold et al. (1994) to estimate the parameters of the multinomial logit specification.

Our approach differs from Maheu et al. (2012). These authors allow for bear markets that can exhibit short rallies and bull markets that can show brief corrections. They ensure identification by imposing that the expected return during bear markets including rallies is negative, while it is positive during bull markets including corrections. This setup can improve the dating of bull and bear markets, though the added value for forecasting is less obvious. The difference in the forecasted return distributions between a bull market and a bear market rally is likely to be small.

2.4 The statistical measure of predictive ability

The statistical distance measure uses the absolute difference between the probabilities p_{st}^m and p_{st}^n that result from two different methods m and n for a given state s . Because the true state is latent, we integrate over the different states,

$$d(\mathbf{p}_t^n, \mathbf{p}_t^m) \equiv \sum_{s \in \mathcal{S}_m} \phi_{st} |p_{st}^m - p_{st}^n|, \quad (8)$$

where $\phi_{st} \equiv \Pr[S_t = s]$ is the probability that state s prevails under the true probability measure. The statistic can only compare two methods if their set of states coincide,

$\mathcal{S}_m = \mathcal{S}_n$. If one set is larger, two or more states can be aggregated. The weight of the difference between p_s^m and p_s^n increases when state s is more likely to occur. A general measure for the closeness of two methods m and n is then

$$d_{m,n} \equiv E^\phi[d(\mathbf{p}_t^n, \mathbf{p}_t^m)], \quad (9)$$

where E^ϕ denotes the expectation with respect to the true distribution of the states. The expression can be interpreted as the Integrated Absolute Difference (IAD), similar to the integrated square difference (ISD) in Pagan and Ullah (1999) and Sarno and Valente (2004).

Generally, to evaluate Equation (9) an assumption on the true probability distribution of the states ϕ_{st} is needed. One can for example assume that method m gives the true distribution and set $\phi_t = \mathbf{p}_t^m$. Then $E^{p^m}[d(\mathbf{p}_t^n, \mathbf{p}_t^m)]$ measures the divergence of method n from method m , similar to the Kullback-Leibler Information Criterion (KLIC) (see Kullback and Leibler, 1951).⁶ As for KLIC, the measure $d_{m,n}$ is asymmetric, so the divergence between methods n and m may be different when m is assumed true than when n is assumed true.

The binomial case presents an exception to the required assumption on the true probability distribution. When S_t is binary with states 1 and 2, Equation (8) simplifies to $d(\mathbf{p}_t^n, \mathbf{p}_t^m) = \phi_{1t}|p_{1t}^m - p_{1t}^n| + \phi_{2t}|p_{2t}^m - p_{2t}^n| = |p_{1t}^m - p_{1t}^n|$. This means that we do not need to know the true distribution of bull and bear markets to compare two methods, and we can directly work with the absolute difference between the probabilities for either bull or bear markets.

We estimate $d_{m,n}$ by the sample mean

$$\hat{d}_{m,n} \equiv \frac{1}{T - R + 1} \sum_{t=R}^T \sum_{s \in \mathcal{S}} \phi_{s,t} |p_{s,t}^m - p_{s,t}^n|, \quad (10)$$

where R is the first period for which we compare the probabilities. When we compare identifications, we can use the full sample and typically have $R = t_0$. For predictions, $R > 1$ and the observations before R are used as in-sample period to estimate model

⁶We cannot measure the difference between p_s^m and p_s^n by the ratio of their logarithms as proposed by Kullback and Leibler (1951) since either probability can equal zero or one, when we consider identification in the rule-based approaches.

parameters. We propose a bootstrap to derive the variance of $\hat{d}_{m,n}$, which we discuss in Section 2.6.

2.5 The economic measure of predictive ability

To construct an economic measure for the difference between two methods, we need a decision based on the state probabilities, and an evaluation of this decision. We consider an investor who uses the predictions to invest in the equity market. If she predicts a bull (bear) market, she takes a long (short) position in one-period futures contracts on the stock market index. Because such futures contracts are highly liquid, this investment strategy can be easily implemented in practice and hardly involves transaction costs.

We assume that the investor has a quadratic utility function defined over returns,

$$U(r) = r - \frac{1}{2}\gamma r^2, \quad (11)$$

where γ is her coefficient of risk aversion. At each point in time t she uses method m to make predictions for the first and second moments of the equity return, and calculates the proportion of wealth w_t that maximizes the expected utility $E_t[U(w_t r_{t+1})]$. The optimal proportion of wealth w_t^m for method m satisfies

$$w_t^m = \mu_t^m / (\gamma \psi_t^m), \quad (12)$$

where μ_t^m is the predicted mean and ψ_t^m the predicted raw second moment of r_{t+1} . Both depend on the forecast probability for each state $p_{s,t}^m$,

$$\mu_t^m \equiv \sum_{s \in \mathcal{S}_m} p_{s,t}^m \mu_s^m \quad (13)$$

$$\psi_t^m \equiv \sum_{s \in \mathcal{S}_m} p_{s,t}^m \psi_s^m = \sum_{s \in \mathcal{S}_m} p_{s,t}^m ((\mu_s^m)^2 + \omega_s^m), \quad (14)$$

where μ_s^m , ψ_s^m and ω_s^m are the state-specific mean, raw second moment and variance for model m .

We choose a quadratic utility function instead of other fully specified utility functions for three reasons. First, the rule-based approaches do not provide predictions on the complete distribution of r , but just predict its sign. Quadratic utility only requires an estimate for

the mean and the variance. For the rules based methods, we sort the returns into two groups based on the identification and then estimate the means and variances for both groups of returns. Second, a quadratic utility function fits well with the regime-switching models that have estimates for both the mean and variance. Third, quadratic utility can be seen as a second-order Taylor approximation to more general utility functions.

We evaluate the optimal weight w_t^m by calculating the value function with respect to the unconditional distribution G_r of r_{t+1} ,

$$V(w_t^m) = E_{G_r} \left[w_t^m r_{t+1} - \frac{1}{2} \gamma (w_t^m r_{t+1})^2 \right]. \quad (15)$$

If method m is superior to method n the value of $V(w_t^m)$ will be higher. The maximum fee $\eta_{m,n}$ that an investor would be willing to pay to use method m instead of n reduces $V(w_t^m)$ to $V(w_t^n)$. This fee acts as a (the difference in) proportional transaction cost on method m . It is expressed as a return, and solves

$$E_{G_r} \left[w_t^m r_{t+1} - \eta_{m,n} - \frac{1}{2} \gamma (w_t^m r_{t+1} - \eta_{m,n})^2 \right] = E_{G_r} \left[w_t^n r_{t+1} - \frac{1}{2} \gamma (w_t^n r_{t+1})^2 \right]. \quad (16)$$

If $\eta_{m,n} > 0$, method m is superior to method n , and the investor is willing to pay up to $\eta_{m,n}$. A negative $\eta_{m,n}$ can be interpreted as a compensation that the investor requires for adopting an inferior method m instead of n . Because of the utility functions, risk aversion is taken into account. The measure is not symmetric, so exchanging methods m and n does not produce the negative of the original fee, $\eta_{m,m} \neq -\eta_{m,n}$ for $m \neq n$. Based on series for $\{w_t^m\}_{t=R-1}^{T-1}$, $\{w_t^n\}_{t=R-1}^{T-1}$ and $\{r_t\}_{t=R}^T$, we can estimate $\hat{\eta}_{m,n}$. We advocate a bootstrap to construct its distribution.

While a fee and the closely related concept of a certainty equivalent return have been used before to compare the economic values of investment strategies, we consider it as a statistic and take estimation uncertainty into account. Our approach is similar to Fleming et al. (2001) and Marquering and Verbeek (2004), who calculate the fees that equate the realized quadratic utility of a dynamic investment strategy to that of a static buy-and-hold strategy. As in our definition, the fees are subtracted from the returns of the dynamic strategies. West et al. (1993) interpret the relative increase in initial wealth that would be needed to equate the realized utility levels of two strategies also as a fee. Their

approach slightly differs from the former methods because the amount available for investment changes. The certainty equivalent return (CER) that is calculated by Das and Uppal (2003) and Ang and Bekaert (2002) is closely related to fees. They calculate the relative increase in initial wealth that is required to equate the expected utility of a suboptimal investment to that of the optimal investment. The CER differs from fees, because it is calculated based on the expected utility under the model from which the optimal investment is derived, whereas fees are calculated based on the actually realized returns of different strategies.

In a recent paper, McCracken and Valente (2012) also consider fees as a measure for the preference of one method over another. They use asymptotic theory to derive its statistical properties. Their approach can test whether the fee is equal to zero versus the alternative that it is strictly larger than zero. This setting corresponds naturally with a setting of nested models, where the null hypothesis is that the unrestricted model does not improve a restricted version. In our setting, methods are generally non-nested which leads to hypotheses of negativity versus positivity. Therefore, we use bootstraps to construct confidence intervals. While this approach does not offer us a formal test, it can show how large the fee is when estimation uncertainty is taken into account.

2.6 Bootstrap

We use a bootstrap to determine the distribution of $\hat{d}_{m,n}$ or $\hat{\eta}_{m,n}$ for two reasons. First, a bootstrap can accommodate various degrees of autocorrelation in the series of \mathbf{p}_t^m and \mathbf{p}_t^n . Considering just a limited number of autocovariances as proposed by Diebold and Mariano (1995) may not be sufficient. Second, a bootstrap can deal with the lower bound of zero of the distribution of $\hat{d}_{m,n}$. We implement the bootstrap in two ways, depending on whether we compare the identification or predictions of different methods. In both cases, we construct bootstrapped samples from the original information set $\Omega_{t_0,T}$, so including returns and predictor variables. To account for autocorrelation in these series, we apply the stationary bootstrap of Politis and Romano (1994).

When we compare the different methods for identification, we use the bootstrapped sample $\Omega_{t_0,T}^j$ to calculate new estimates $\hat{\boldsymbol{\theta}}_m^j$ and to construct new series of inference prob-

abilities. In turn, these lead to bootstrapped estimates $\hat{d}_{m,n}^j$ and $\hat{\eta}_{m,n}^j$. These set of bootstrapped estimates converges to the true distributions and can be used for testing and the construction of confidence intervals.

In the case of comparing predictions, we follow the approach of White (2000). We resample from the second part of the information, starting from the first prediction R . So, in this case the first part of the information set $\Omega_{t_0, R-1}^j = \Omega_{t_0, R-1}$ for each bootstrap j . We create new series $\{\mathbf{p}_t^{m,j}\}_{t=R}^T$ and $\{\mathbf{p}_t^{n,j}\}_{t=R}^T$, but use the parameter estimates $\hat{\boldsymbol{\theta}}_m$ based on the original series. With these series, we calculate bootstrap estimates $\hat{d}_{m,n}^j$ and $\hat{\eta}_{m,n}^j$ and construct their distributions.

3 Data and implementation

3.1 Stock market data

We determine the state of the stock market against the benchmark of a riskless investment. Financial theory states that in the long run the stock market should yield the risk-free rate plus a positive risk premium. Accordingly, a riskless investment is a natural benchmark to determine the attractiveness of the stock market. If the stock market yields positive returns below the risk-free rate, investors do not earn the risk premium, which makes the stock market unattractive.

A riskless bank account B_t earns the risk-free interest rate r_τ^f over period τ . Starting with $B_0 = 1$, the value of this bank account obeys

$$B_t \equiv \prod_{\tau=0}^{t-1} (1 + r_\tau^f). \quad (17)$$

For a stock market index P_t the relevant series to determine the state is thus given by

$$\tilde{P}_t = P_t/B_t. \quad (18)$$

The return on this index is the market return in excess of the risk-free rate. It also corresponds with the return on a long position in a one-period futures contract on the stock market index.

We consider the US stock market, proxied by the S&P500 price index at a weekly frequency. We splice together a time-series for the S&P500 by combining the data of Schwert (1990) with the S&P500 series that the Federal Reserve Bank of St. Louis has made available on FRED.⁷ Schwert’s data set covers the period from February 17, 1885 until July 2, 1962, whereas the FRED series starts on January 4, 1957 and is kept up-to-date. For the risk-free rate we use the three-month T-Bill rate, also from FRED. This series starts on January 8, 1954. Because of the availability of the predictor variables, the data sample that we analyze starts on January 7, 1955.

We use weekly observations because of their good trade-off between precision and data availability. Higher frequencies lead to more precise estimates of the switches between bull and bear markets. On the other hand, data of predictor variables at a lower frequency is available for a longer time span. Weekly data does not cut back too much on the time span, and gives a satisfactory precision.

Figure 1 shows the excess stock price index for the US. The index has been set to 100 on 1/7/1955. The graph shows a clear alternation of periods of rise and decline in the 1950s and 1960s, the prolonged slump during the late 1970s and early 1980s, the crash of 1987, the dramatic rise during the IT-bubble of the late 1990s and the subsequent bust in 2000-2002, and finally the fall during the recent credit crisis.

[Figure 1 about here.]

3.2 Predictor variables

Macro-economic and financial variables can help predict whether the next period will be bullish or bearish. Our choice of variables is motivated by prior studies that have reported the success of several variables for predicting the direction of the stock market.⁸ We gather data for inflation, industrial production, unemployment, the T-Bill rate, the term and credit spread, and the dividend-to-price ratio. To ensure stationarity, we transform some

⁷We kindly thank Bill Schwert for sharing his data with us.

⁸The literature on return predictability is vast. Welch and Goyal (2008) provide a critical review of the performance of many variables. For the combination of regime-switching models with predictor variables we refer to Ang and Bekaert (2002) and Chen (2009).

of the predictor variables. To ease the interpretation of coefficients on these variables, we standardize each series. As a consequence, coefficients all relate to a one-standard deviation change and the economic impact of the different variables can directly be compared. In Appendix C we provide more information on the predictor variables.

The observations for inflation, industrial production and unemployment are available at a monthly frequency. We lag them by one month, and assume that the series are constant within a month. Weekly observations of the term and credit spreads become available from January, 1962. Before that date, we use monthly observations. The T-Bill rate and the D/P-ratio are available at a weekly frequency for the complete sample period. We lag weekly observations by one week. Because the predictor variables have a mixed frequency, we combine the stationary bootstrap of Politis and Romano (1994) discussed in Section 2.6 with a block-bootstrap. We apply the stationary bootstrap at a monthly frequency. If a certain month is drawn, we draw all four or five corresponding weekly observations.

We consider in total seven predictor variables. Not all these variables might be helpful in predicting specific transitions. Therefore, we use a specific-to-general procedure for variable selection. For each dynamic logit model and regime switching model we start with a specification with only constants included. Next, we calculate for each variable and transition combination the improvement its inclusion would yield in the likelihood function. We test whether the variable-transition combination with the largest improvement is significant with a likelihood ratio test. If so, we add the variable for that specific transition and repeat the search procedure for the remaining combinations of predictor variables and transitions. The procedure stops when no further significant improvement is found.

This approach differs from the general-to-specific approach, which would include all variables first and then consider removing the variables with insignificant coefficients. For the RS3L-model, we would need to estimate a model with $3 \cdot 2 \cdot 7 = 42$ transition coefficients, which is typically infeasible. For the same reason, we do not follow Pesaran and Timmermann (1995), who compare all different variable combinations based on general model selection criteria such as AIC, BIC and R^2 .

4 Results

4.1 Full Sample Identification

We first compare the identification of the different methods for the full sample. This in-sample comparison shows in detail for the largest available information set how and why the outcomes of the various methods differ. We first consider the actual dating of bull and bear markets. We use the IADs to measure the differences between the dating of the different methods and use graphs to analyze it in more detail. Next, we consider the return distributions during the different states and calculate switching fees to determine which method is preferable.

The IADs in Table 1 indicate how different the methods date bull and bear markets. The difference is rather small for the LT- and PS-algorithms with an IAD of 0.068. This value indicates that these two methods lead to a different identification for 6.8% of the observations, because the compared probabilities are either zero or one. Nevertheless, the 90%-confidence interval does not contain zero, which means that for samples similar to the S&P500 the LT- and PS-algorithms generally lead to small differences.

[Table 1 about here.]

Figures 2a and 2b show that the LT- and PS-methods identify many periods identically. Because both methods are based on peaks and troughs, all switches take place at maxima and minima. The differences between the methods are concentrated in bull and bear markets that are relatively short or show only subdued price changes. For example, the PS-method dates two bear markets between 1991 and 1995, but the price decreases are too small to be picked up by the LT-method. On the other hand, the LT-method dates a bear market in 1998 that does not meet the duration requirements of the PS-method.

[Figure 2 about here.]

The differences between the two-state regime-switching models and the LT- and PS-methods are much larger. The IADs in Table 1 range from 0.320 to 0.368. The confidence intervals show considerable sampling variation in these IADs, in particular on the upside.

The lower bounds of 0.263 to 0.301 imply that these differences are not a coincidence. A two-state regime-switching model leads to substantially different results than a non-parametric method. The effect of time-variation in the transition probabilities is negligible with an IAD of only 0.032 and small confidence intervals around it.

The dating by the RS2-models in Figures 2c and 2d is based on rounding the smoothed inference probabilities (drawn by a thin black line). The result hardly resembles the LT or PS dating. Because the probabilities for a bull market are close to zero or one, the large IADs are not due to a comparison of a binary identification with a probabilistic one. The RS2-models do not produce a clear alternation of bull and bear markets, but instead periods of bull markets that are interrupted by brief bear markets (e.g., 1955–1975 and 1983–1995) and periods of bear markets that are interrupted by brief bull markets (e.g., 1979–1983 and 1997–2003). Maheu et al. (2012) explicitly accommodate rallies during bear markets and corrections during bull markets in their regime-switching models.

To see why the rule-based methods and the RS2-models date bull and bear markets differently, we report the means and volatilities of the regimes in Table 2. Bull markets have a positive average return and low volatility, whereas bear markets exhibit negative returns and high volatility. The LT and PS-methods concentrate the differences between bull and bear markets in the average return, which is about a full 1% lower for bear market. The volatility of bear markets is about 1.30 times higher than bull market volatility. Regime-switching models pay more attention to volatility. In the RS2-models the volatility ratio is around 2.25, while the difference in average returns is only around 0.44%. As a consequence, the RS2-models identify volatile periods as bearish, even if prices eventually increase, and tranquil periods as bullish, even if prices fall. That is why the periods with gradually falling prices in the 1950s are seen as bull markets, and why the volatile price increases from 1997 to 2000 are qualified as bearish. Of course, a risk-averse investor may indeed judge volatile price increases as unattractive. We consider such a comparison when we calculate the switching fees.

[Table 2 about here.]

We consider an extension of the regime-switching models to three states. Table 2 shows that the bullish regimes of the RS3-models and RS2-models share the same return

characteristics. The other two regimes split the bearish regime of the RS2-models into a mildly bearish regime with an average return just below zero and a lower volatility, and a strongly bearish regime with a large negative average return and very high volatility. To calculate the IADs we have aggregated the mild and strongly bearish regimes.

Table 1 shows that the dating of the RS3-models differs as much from the LT and PS-methods as the RS2 models do. The differences in dating are largest between the non-parametric rule-based methods on the one hand and the parametric RS-models on the other. Differences between the RS3-models and RS2-models are smaller, though still considerable with IADs between 0.154 and 0.182. The confidence intervals show that the IADs exhibit substantial sample variation, with upper bounds exceeding 0.5. RS3-models can date states completely different from all two-state methods. The effect of time-varying transition probabilities is again small, judged by the IAD for the RS3C and RS3L models of 0.032.

The comparison of Figures 2e and 2f with Figures 2c and 2d illustrates the differences between the two-state and three-state regime-switching models. Instead of long bull markets with short bearish spells, and bear markets with short bullish spells, the RS3-models produce identifications that alternate more regularly. The strongly bearish regime captures large price drops, notably the big declines by the end of 1974, the crash of October 1987, and the big drops during the credit crisis in 2008. It is mostly preceded and succeeded by the mildly bearish regime, which justifies their aggregation. The comparison of Figures 2e and 2f with Figures 2a and 2b confirms the result from the IADs that the RS-identifications have more in common with each other than with the LT and PS-identifications.

The analysis so far cannot establish a preference from one method over another. The maximum fee that an investor would be willing to pay to switch from one method to another fills this gap. Because dating takes place in-sample, the fees result from portfolios that are constructed and evaluated in-sample, too.

In line with their small IAD, the LT and PS-methods lead to almost the same performance in an investment strategy. Average returns and volatilities in Table 3a are high for both methods at 0.93% and 4.2% per week, and exceed the ones reported in Table 2 because the optimal weights exceed 1. Because the investment corresponds with a futures contract, leverage is not an issue here. The utility of the PS-method is slightly higher than

that of the LT-method. As a consequence, the investor is willing to pay at most 0.25 bps to switch from the LT-method to the PS-method. A risk-averse investor thus prefers the PS method over the LT method. However, the confidence interval includes zero, which indicates that this preference is not significant.

[Table 3 about here.]

The performance of the regime-switching models is considerably worse. The average returns vary between 0.14% and 0.15% per week and volatilities between 1.47% and 1.57%. Sharpe ratios are halved, and utilities are much lower. The maximum fees indicate a preference for the LT and PS methods over all regime-switching models. An investor would be willing to pay a maximum of 36.1 bps to switch from the RS2C model to the PS model. The 90%-confidence interval does not contain zero, indicating a significant preference. The fee is smaller than the difference between the average returns of 75 bps, which reflects the effect of the higher volatility of the PS-based investment strategy. The maximum fees for switches from the other RS-models are of similar magnitude and their confidence intervals do mostly not contain zero.

The maximum fees for a switch from a three-state to a two-state regime-switching model are positive, indicating that RS2-models are preferred to RS3-models. The average returns for the RS2 models exceed those of the RS3 models. Because the volatilities are almost the same, the maximum fees reflect mainly the difference in average returns. The confidence intervals contain zero, so the preference varies. Also, the confidence intervals are wide, indicating that the sampling error for the maximum fees is large. In some bootstraps the RS3-models command a high fee when switching from the RS2-models. The fees for a switch from constant transition probabilities to time-varying ones are close to zero, indicating the small added value of predictor variables.

We conclude that the rule-based approaches date bull and bear markets differently than regime-switching models. Dating by rules is definitely preferable. The driving force behind these differences is volatility, which is important for identification by regime-switching but completely ignored by the rule-based methods. The maximum fees indicate that for dating only peaks and troughs should matter. In the next subsection we examine whether this result also holds for forecasting.

4.2 Forecasts

Because the state of the stock market is a comprehensive economic indicator, forecasting is more interesting than identification. An accurate forecast tells investors whether to expect a good or bad investment performance. More generally, it signals whether the economic outlook is good or bad, and whether risk premia are low or high. Therefore, we compare the forecasts of the different methods. The in-sample preference for rule-based methods need not carry over to forecasting. Importantly, the rule-based methods indicate switches between bull and bear states with delay. For example, the LT-method needs a price increase of 20% to establish a switch to a bull market. The PS-method excludes the last 13 weeks from identification. If it is unknown which state currently prevails, forecasts may be false, in particular when a switch has just occurred. Regime-switching models may need less time to pick up changes in market sentiment.

To compare the forecasting quality of the methods, we conduct the following out-of-sample experiment. An investor uses the first part of the sample period until June 24, 1983 for identification and the estimation of model parameters, and uses these to make forecasts for the second part of the sample period. At every point in time t , she updates her information set to construct a forecast for the next week $t + 1$. When 52 weeks have passed, the investor expands the estimation window and reestimates the parameters of the different models. Given the duration of bull and bear markets, 52 weeks offers an acceptable trade-off between estimation speed and accuracy. We compare the resulting forecasts again in terms of IADs and maximum fees.

The investor applies the methods in the same way as in the previous section. When she uses the LT or PS-method, she first identifies bull and bear markets in the in-sample period, and next estimates the Markov-logit model in Equation (2) with or without predictor variables. When transition probabilities (either in the Markov-logit or the regime-switching models) can be time-varying, the specific-to-general approach is used to select the relevant predictor variables. This approach implies that the set of selected variables varies over time. Details on the evolution of the parameter estimates are provided in Appendix E.

At each point in time t , the investor combines the most recent parameter estimates with the current information set to forecast the state of the market at $t + 1$. In the LT-method,

she knows the state of the market until the last extremum $t^* \leq t$. She constructs the forecast by the recursion in Equation (3), starting at t^* . In the PS-method, predictions start at $t - 13$, as switches in the last 13 weeks are ignored. For the regime-switching models, she applies the filter to infer the state of the market at time t , and subsequently calculates the forecast probabilities. For the weekly predictor variables, the values of time t are used, and for the monthly variables the values of the month ending before week t .

We start our analysis by comparing the rule-based methods without predictor variables. The IAD of 0.25 in Table 4 shows that the LT and PS methods lead to quite different forecasts. This large values contrast with the small IAD of 0.068 for identification in Table 1. The lower bound of the confidence interval is considerably higher than zero, indicating structural differences. Even when methods lead to almost the same identification, they can lead to quite different forecasts.

[Table 4 about here.]

The differences between the forecasts by the LT and PS methods are illustrated by Figures 3a and 3b. In both methods the forecasts for $t + 1$ depend on the inferred state at time t , which is determined with a delay. In the PS method this delay is always equal to the censoring window of 13 weeks, whereas it varies in the LT method, depending on the last extremum. If the last extremum occurred longer ago, the forecasts of the LT method converge to the steady state probabilities. This effect is absent in the PS method. In particular correct forecasts of bear markets take long when the LT method is used. Sometimes more than half of the bear market has already passed.

[Figure 3 about here.]

[Figure 3 (continued) about here.]

The introduction of predictor variables does not lead to largely different forecasts. The IADs between the LT or PS methods with and without predictor variables are small, 0.060 or 0.082. Comparing Figures 3a and 3c or Figures 3b and 3d shows that the most recent state of the market remains the main driver of the forecasts. The predictor variables have

some impact on the forecasts, but not enough to change the pattern of the forecasts without predictor variables.

The forecasts given by the two-state regime-switching model without predictor variables are quite different from the rule-based methods. The IADs vary between 0.239 and 0.331, indicating that the RS2C forecasts differ as much from the LT and PS method as the forecasts of those methods from each other. While there are still considerable differences between the different methods, the forecasts seem to be closer than the identifications. The IADs and their confidence intervals in Table 4 are smaller than in Table 1. Figure 3e shows that the RS2C-model picks up switches in the state of the market more quickly than the rule-based methods.

The results for the two-state regime-switching model with predictor variables (RS2L) are quite similar to those from the RS2C model. The IAD is small with a value of 0.061, and the confidence intervals indicate that the estimation uncertainty is limited. The evolution of the forecasts in Figures 3e and 3f look very much alike. We can conclude that predictor variables do not improve performance, and may even lead to overfitting.

The forecasts by the RS3C-model differ considerably from the rule-based methods, but they are rather close to the RS2C model. To make the forecasts comparable, we aggregate the forecast probabilities for states with a positive mean.⁹ IADs with the LT and PS methods vary from 0.239 to 0.327. The IAD between the forecasts of the RS2C and RS3C model is only 0.090. The small confidence intervals indicate that this small value is not a coincidence. The extension of the RS2C- to the RS3C-model leads to just marginally different predictions, which is also obvious comparing Figures 3e and 3g.

The addition of predictor variables has again a limited effect on the forecasts. The IAD of the forecasts of the RS3C and RS3L models is very small, only 0.034, and also has a small confidence interval. IADs to the other methods are in line with the results for the RS3C: close to the RS2-models and quite different from the rule-based methods.

To quantify the preference of one method over another we determine the maximum fee or cost that an investor would be willing to pay to switch between methods. The fees correspond with a setting where the investor combines the forecast probabilities with the

⁹We show in Figure E.1(e-f) in Appendix E that before 2009 two bullish states result from the estimation, contrary to the single bullish and two bearish regime from the full-sample estimation.

state-dependent means and volatilities to construct the optimal portfolio in Equation (12). The state-dependent means and volatilities are reestimated together with the other parameters every 52 weeks.

We report the performance of the investment strategies based on the different methods in Table 5a. For reference we have included the performance of a strategy that always takes a long futures position in the stock market. All strategies yield a positive excess return. The LT method without predictor variables leads to the highest average returns, also higher than the market futures: 0.168% versus 0.066% per week (8.74% versus 3.43% per year). Part of these higher returns results from taking more risk, as the average absolute weight equals 1.82. Consequently, volatility is almost twice as high as for the market futures strategy. As a consequence of this higher volatility, the resulting utility of the LTC method is negative and lower than the utility of holding market futures. The Sharpe ratio is higher, but this measure is insensitive to the weight that is invested.

[Table 5 about here.]

The PS method without predictor variables performs considerably worse than the LTC method. The average return is much lower (0.028% per week or 1.46% per year). This is only partially compensated by lower volatility (3.88% vs. 4.50% per week). Consequently, utility and the Sharpe ratio are lower for the PSC method. The fee to switch from the PSC method to the LTC method is therefore positive with a value of 1.0 bps. The confidence interval contains zero, indicating that the LTC method is not significantly better. While the IAD between the two methods point at considerable differences, the fees do not indicate a clear preference for one over the other.

Combining the LT or PS methods with predictor variables leads to a worse performance. In case of the LT-method realized utility goes down from -3.38×10^{-3} to -3.54×10^{-3} ; for the PS-method it decreases from -3.48×10^{-3} to -3.91×10^{-3} . Correspondingly, fees to switch from methods without to methods with predictor variables are negative. The maximum fee to switch from LTC to LTL (from PSC to PSL) equals -1.57 (-4.32) bps. The confidence intervals indicate there is quite some uncertainty around these fees. Predictor variables do not much affect forecasts when combined with non-parametric methods.

The regime-switching models lead to quite different investment strategies and performances than the rule-based methods. Portfolio weights are more restrained with average absolute weights of about 0.63. Average excess returns are generally smaller than for market futures or the rule-based methods, but this is more than offset by the reduction in volatility. The realized utilities exceed those of a long futures strategy as well as those of the LT and PS methods. The highest utility results from the RS2 model with predictor variables. The maximum fees to switch from any of the rule-based methods to any of the RS-models are all positive with values around 30 bps. The confidence intervals are wide, pointing at considerable estimation uncertainty, but they do not include zero. We conclude that the forecasts by the RS-model are preferred to the rule-based forecasts. This preference stems from larger prudence. Forecasts by the RS-models better reflect the state of the market as indicated by its mean and volatility.

The differences in performance between the different regime-switching models are small. The RS3C model performs worse than the RS2 model. We already concluded based on the IADs that an extension to three states does not lead to considerably different forecasts. The realized utility of the resulting investment is lower, and fees for a switch from the RS2C to RS3C model are consequently negative with a value of -2.5 bps.

Including predictor variables leads to small improvements. Realized utilities go up, and fees are positive. However, the fees are quite small: 0.5 (1.5) bps for a switch from the RS2C to the RS2L (RS3C to the RS3L) model. All confidence intervals include zero, indicating that differences are not significant. The IADs showed that the forecasts of the different regime-switching models are close. The fees indicate that a clear preference cannot be established.

We can draw two conclusions from this analysis. First, regime-switching models perform best from a utility perspective. Since they take both means and volatilities into account, they lead to more accurate forecasts, less extreme positions and higher utilities. A risk averse investor is willing to pay a significant fee to use a regime-switching model, mainly representing the lower volatility that results. Also for other economic decisions where both the predicted direction and volatility of the stock market are important, regime-switching models can be used best.

Second, the forecast from the rule-based methods and the regime-switching models

differ considerably. As expected, this difference is obvious for the rule-based methods on the one hand, and the regime-switching models on the other hand. Regime-switching models take both means and volatilities into account, whereas the rule-based method just look at the trend of the market. However, we find difference between the LT and PS-methods that are just as large. Similarity in identification does not imply similarity in forecasts. Restrictions on price changes lead to quite different results compared to restrictions on duration.

5 Robustness checks

Our results and conclusions may be sensitive to some arbitrary choices. The rule-based methods require parameters to determine which peaks and troughs signal switches between bull and bear markets. In the calculation of the fee, the coefficient of risk aversion is crucial. In this section we analyze the sensitivity of our results to these settings.

5.1 Thresholds in the LT-method

The parameters λ_1 and λ_2 are the thresholds for switches to bull and bear markets in the LT method. The values of 20% and 15% we have used so far have been argued by LT to be most conventional. They also consider lower threshold combinations of (0.20, 0.10), (0.15, 0.15) and (0.15, 0.10). We consider these values as well. Lower thresholds may be particularly interesting, because our results show that the speed with which the current regime is identified is crucial for good forecasts. Lower thresholds make it easier to identify a switch, but may also lead to more false alarms. We compare the performance of the LT-methods with different thresholds and discuss the main results here. The full results and a detailed discussion are available in Appendix F.1.

The IADs based on the full sample identification are low, and indicate that the identification is different in less than 5% of the weeks. The thresholds do not much affect the pattern of bull and bear markets in Figure 2a. Even though the changes are statistically small, the economic comparison indicates improvements. The fees that an investor is willing to pay to switch to (0.15, 0.10)-thresholds are positive and significant. These

results strengthen our conclusion that the LT-method works better for identification than the regime-switching methods.

For forecasting the results are more mixed. The IADs between the forecasts of the LT-methods with different thresholds vary between 0.07 and 0.16. While larger than the differences for identification, they are smaller than the IADs in Table 4. These difference do not mean improvements. Lower thresholds lead to investments that produce higher average returns, but also higher volatilities. Sharpe ratios still increase, but utility is sometimes lower. Only a lower threshold for λ_2 of 0.10 leads to an increase in utility and positive, though insignificant, fees. The utility of the (0.20,0.10)-thresholds is still substantially lower than the utility for the RS2-models in Table 5. We conclude that lower thresholds do not improve the economic quality enough to beat the RS2-models.

5.2 Parameters in the PS-method

The PS-method uses minimum constraints on the length of complete cycles and individual phases to select the peaks and troughs that indicate switches between bull and bear markets, and censors a set of first and last observations. The settings for these constraints are based on the algorithm of Bry and Boschan (1971) for business cycle identification and on common market lore. As PS do not consider robustness checks themselves, we consider some changes in the parameters based on our results so far. The LT-method performs better with relaxed restrictions, so we mainly investigate relaxations. We consider a lower minimum on cycle duration of 52 weeks (instead of 70 weeks). For the minimum on phase duration we consider 12 and 20 weeks (instead of 16 weeks). We relax the minimum price change to overrule the phase constraint to 15%. We also investigate the consequences of censoring more (26 weeks) and less (7 weeks). We discuss the main results here. For the full results and a discussion, we refer to Appendix F.2.

For identification, changes are negligible. Identification does not change at all when we change the constraints on phase duration or price change. Censoring 7 weeks leads to an extra bear market at the end of the sample period, but censoring 26 weeks has no effect. Lowering the cycle constraint to 52 weeks leads to one extra cycle. Together, it means that the conclusion that the PS-method works well for identification is robust to these

parameter changes.

Neither does a relaxation of the constraints on minimum length or minimum price change largely impact forecasts. The IADs between the forecasts for different settings and the forecasts by the basic setting are not significantly different from zero. The fees for switching are economically small and insignificant.

Censoring shows a larger impact, in particular for the economic difference measures. The IADs between the forecasts of the standard approach and approaches with more and less censoring are quite substantial. The economic comparison shows that more censoring leads to less extreme portfolio weights, higher means, lower volatilities, higher Sharpe ratios and higher utility, while censoring less has the opposite effect. These effects can all be explained by the effect that censoring has on a forecast. When more observations at the end are censored, an investor essentially makes a forecast for a longer horizon, applying a longer recursion of Equation (3). If the forecast horizon becomes longer, the forecast converges to the long-term average (see e.g., Table D.1b) and typically become less extreme. Less extreme forecasts lead to less extreme investments, and to a better performance. When 26 weeks are censored, utility is close to the utility of long position in the market and only slightly below the utility of the two-state regime-switching models.

Overall, we find that our conclusions are robust to changes in the design of the PS-method. Constraints on the duration of cycles or phases or on price changes only have a small effect. The effects for censoring show that the PS-method relies too much on the past direction of the market when making forecasts. Paying more attention to volatility as regime-switching models do leads to better performance. Shrinking forecasts towards their long-term average also leads to less extreme investments and a better performance.

5.3 Risk Aversion

The economic measure for comparison in Section 2.5 depends on the coefficient of risk aversion γ . Throughout our analysis so far, we have set $\gamma = 5$, which is conservative though still in the generally accepted range for this parameter. This subsection shows the sensitivity of our results to the choice for γ .

The optimal portfolio weight w^m in Equation (12) is proportional to the inverse of γ .

Consequently, the average return of an investment strategy is proportional to $1/\gamma$ as well, and the variance is proportional to $1/\gamma^2$. Both effects exactly cancel out in the Sharpe ratio. The expected utility being the sum of the expected return and the second moment weighted by $-\gamma/2$ is also inversely related to γ ,

$$\begin{aligned} V(w^m) &= \mathbb{E}_{G_r} \left[w^m r - \frac{1}{2} \gamma (w^m r)^2 \right] = \mathbb{E}_{G_r} \left[\frac{\mu^m}{\gamma \psi^m} r - \frac{1}{2} \gamma \left(\frac{\mu^m}{\gamma \psi^m} r \right)^2 \right] \\ &= \frac{1}{\gamma} \mathbb{E}_{G_r} \left[\frac{\mu^m}{\psi^m} r - \frac{1}{2} \left(\frac{\mu^m}{\psi^m} r \right)^2 \right]. \end{aligned} \quad (19)$$

Finally, the switching fee is proportional to $1/\gamma$. To derive this result, we rewrite the implicit definition of $\eta_{m,n}$ in Equation (16) as

$$V(w^m) - V(w^n) - (1 - \gamma \mathbb{E}_{G_r} [w_t^m r_{t+1}]) \eta_{m,n} - \frac{1}{2} \gamma \eta_{m,n}^2 = 0. \quad (20)$$

We make the effect of γ explicit, where we use Equation (12) and write $V(\mu^m/\psi^m)$ for the expected utility for $\gamma = 1$ corresponding with model m

$$\frac{1}{\gamma} \left(V \left(\frac{\mu^m}{\psi^m} \right) - V \left(\frac{\mu^n}{\psi^n} \right) \right) - \left(1 - \mathbb{E}_{G_r} \left[\frac{\mu_t^m}{\psi_t^m} r_{t+1} \right] \right) \eta_{m,n} - \frac{1}{2} \gamma \eta_{m,n}^2 = 0. \quad (21)$$

Solving for $\eta_{m,n}$ produces

$$\begin{aligned} \eta_{m,n} &= -\frac{1}{\gamma} \left(1 - \mathbb{E}_{G_r} \left[\frac{\mu_t^m}{\psi_t^m} r_{t+1} \right] \right) \pm \\ &\quad \sqrt{\left(1 - \mathbb{E}_{G_r} \left[\frac{\mu_t^m}{\psi_t^m} r_{t+1} \right] \right)^2 + 2 \left(V \left(\frac{\mu^m}{\psi^m} \right) - V \left(\frac{\mu^n}{\psi^n} \right) \right)} \end{aligned} \quad (22)$$

An increase (reduction) in risk-aversion will lead to less (more) extreme results in Tables 3a and 5a. Means, volatilities, utilities and switching will all decrease (increase) by the same relative amount. However, the ordering of the strategies will not be affected and our conclusions regarding the preferred method are robust. The confidence intervals around the fees are also proportional to $1/\gamma$, as the proportionality applies similarly to simulated fees. So, fees that are significantly different from zero remain significant, independently of the choice of γ .

6 Conclusion

In this paper, we compare several methods to date past bull and bear markets and to forecast the future state of the equity market. These methods can be split into semi-parametric, rule-based ones, and parametric regime switching models. For the first type, we consider the methods by Pagan and Sossounov (2003) and Lunde and Timmermann (2004). While both types of methods have their pros and cons, the main difference between the two methods is the information they take into account. The semi-parametric methods use the mean return to distinguish bull and bear markets, while regime-switching models also use volatility.

We propose a statistical and an economic measure to compare how the various methods date past bull and bear markets and forecast future ones. Because the true regime remains latent, existing tests for equal predictive ability cannot be used. The statistical measure, the Integrated Absolute Difference, uses the absolute differences between the probabilities for each state. The economic measure is the maximum fee that a risk-averse investor would be willing to pay to switch from one method to another for an investment strategy. The IAD indicates how different two methods are, whereas the maximum fee quantifies the preference for one method over another. Both measures can be used in-sample and out-of-sample.

Our results for dating past bull and bear markets show that the semi-parametric methods are significantly different from and better than parametric methods. The rule-based methods use peaks and troughs to mark switches between bull and bear markets, and consequently only pay attention to the trend in prices. For dating, so *ex post*, only the trend is required. Regime-switching models also pay attention to volatilities, without adding value. Differences between the different rule-based methods, or between the regime-switching models are small, with no clear preference for a specific method.

For forecasting, regime-switching models work best. Their forecasts differ from those produced by semi-parametric methods and command a significant fee. Because regime-switching models pay attention to means and volatilities, they lead to more prudent investments and higher utility. Regime-switching models are also quicker to pick up a switch between states. Rule-based methods need quite some time before they can indicate a

switch. Because the current state is the most important predictor of the future state, this delay hampers their performance.

Harding and Pagan (2003a,b) and Hamilton (2003) have investigated the difference between rule-based and model-based methods for dating business cycles. As the resulting identifications are largely similar, the main differences are the larger transparency for the rule-based approaches versus the deeper insight into the data generating process for the regime-switching models. For financial time series, differences are larger because volatility is an important characteristic. For dating, taking volatility into accounts leads to worse results, but for forecasting it is beneficial.

References

- Ang, A. and Bekaert, G. (2002). International asset allocation with regime shifts. *Review of Financial Studies*, 15(4):1137–1187.
- Ang, A., Chen, J., and Xing, Y. (2006). Downside risk. *Review of Financial Studies*, 19(4):1191–1239.
- Ang, A. and Timmermann, A. G. (2011). Regime changes and financial markets. Working paper, Columbia University.
- Bohl, M. T., Siklos, P. L., and Werner, T. (2007). Do central banks react to the stock market? the case of the Bundesbank. *Journal of Banking & Finance*, 31(3):719–733.
- Bry, G. and Boschan, C. (1971). *Cyclical Analysis of Time Series: Selected Procedures and Computer Programs*. NBER, New York, NY, USA.
- Campbell, J. Y. (1991). A variance decomposition for stock returns. *Economic Journal*, 101(405):157–179.
- Candelon, B., Piplack, J., and Straetmans, S. (2008). On measuring synchronization of bulls and bears: The case of East Asia. *Journal of Banking & Finance*, 32(6):1022–1035.
- Chauvet, M. (1999). Stock market fluctuations and the business cycle. *Journal of Economic and Social Measurement*, 25(3/4):235–258.
- 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 & Finance*, 33:211–223.
- Chiang, M., Lin, T., and Yu, C. J. (2009). Liquidity provision of limit order trading in the futures market under bull and bear markets. *Journal of Business Finance & Accounting*, 36(7-8):1007–1038.
- Corradi, V. and Swanson, N. R. (2007). Nonparametric bootstrap procedures for predictive inference based on recursive estimation schemes. *International Economic Review*, 48(1):67–109.
- Das, S. R. and Uppal, R. (2003). Systemic risk and international portfolio choice. Working paper, Santa Clara University, Santa Clara, CA, USA.
- Das, S. R. and Uppal, R. (2004). Systemic risk and international portfolio choice. *Journal of Finance*, 59(6):2809–2834.
- Dempster, A. P., Laird, N. M., and Rubin, D. B. (1977). Maximum likelihood from incomplete data via the EM algorithm. *Journal of the Royal Statistical Society: Series B*, 39(1):1–38.

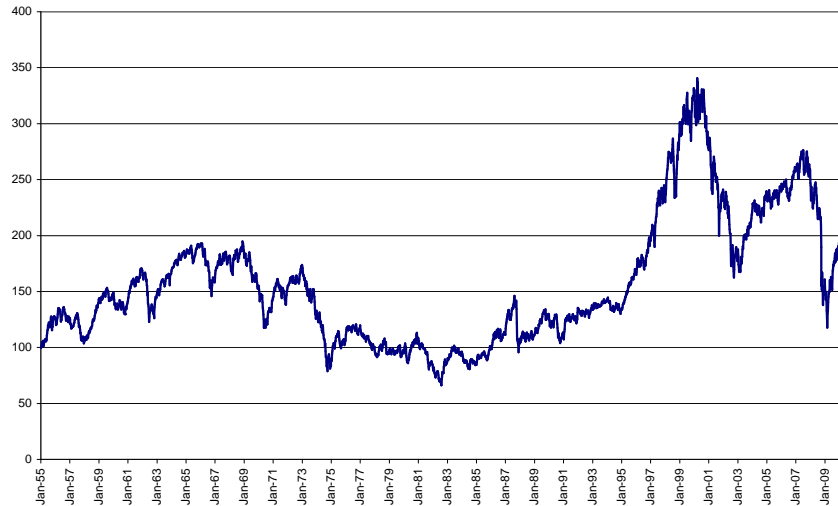
- Diebold, F. X., Lee, J.-H., and Weinbach, G. C. (1994). Regime switching with time-varying transition probabilities. In Hargreaves, C., editor, *Non-Stationary Time Series Analysis and Cointegration*, pages 283–302. Oxford University Press, Oxford, UK.
- Diebold, F. X. and Mariano, R. S. (1995). Comparing predictive accuracy. *Journal of Business & Economic Statistics*, 13(3):134–144.
- Edwards, S., Gómez Biscarri, J., and Pérez de Gracia, F. (2003). Stock market cycles, financial liberalization and volatility. *Journal of International Money and Finance*, 22(7):925–955.
- Estrella, A. and Mishkin, F. S. (1998). Predicting US recessions: Financial variables as leading indicators. *Review of Economics and Statistics*, 80(1):45–61.
- Fleming, J., Kirby, C., and Ostdiek, B. (2001). The economic value of volatility timing. *Journal of Finance*, 56(1):329–352.
- Gordon, S. and St-Amour, P. (2000). A preference regime model of bull and bear markets. *American Economic Review*, 90(4):1019–1033.
- Granger, C. W. J. and Pesaran, M. H. (2000). Economic and statistical measures of forecast accuracy. *Journal of Forecasting*, 19(7):537–560.
- Guidolin, M. (2011a). Markov switching in portfolio choice and asset pricing models: A survey. In Drukker, D. M., editor, *Missing Data Methods: Time-Series Methods and Applications*, volume 27 of *Advances in Econometrics*, pages 87–178. Emerald Group Publishing Limited.
- Guidolin, M. (2011b). Markov switching models in empirical finance. In Drukker, D. M., editor, *Missing Data Methods: Time-Series Methods and Applications*, volume 27 of *Advances in Econometrics*, pages 1–86. Emerald Group Publishing Limited.
- Guidolin, M. and Timmermann, A. (2006a). An econometric model of nonlinear dynamics in the joint distribution of stock and bond returns. *Journal of Applied Econometrics*, 21(1):1–22.
- Guidolin, M. and Timmermann, A. (2006b). Term structure of risk under alternative econometric specifications. *Journal of Econometrics*, 131:285–308.
- Guidolin, M. and Timmermann, A. (2007). Asset allocation under multivariate regime switching. *Journal of Economics Dynamics & Control*, 31(11):3503–3544.
- Guidolin, M. and Timmermann, A. (2008). International asset allocation under regime switching, skew and kurtosis preference. *Review of Financial Studies*, 21(2):889–935.
- Hamilton, J. D. (1989). A new approach to the economic analysis of nonstationary time series and the business cycle. *Econometrica*, 57:357–384.
- 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, USA.

- Hamilton, J. D. (2003). Comment on “a comparison of two business cycle dating methods”. *Journal of Economic Dynamics and Control*, 27(9):1691–1693.
- Hamilton, J. D. and Lin, G. (1996). Stock market volatility and the business cycle. *Journal of Applied Econometrics*, 11:573–593.
- Harding, D. and Pagan, A. (2002). Dissecting the cycle: A methodological investigation. *Journal of Monetary Economics*, 49(2):365–381.
- Harding, D. and Pagan, A. (2003a). A comparison of two business cycle dating methods. *Journal of Economic Dynamics and Control*, 27(9):1681–1690.
- Harding, D. and Pagan, A. (2003b). Rejoinder to james hamilton. *Journal of Economic Dynamics and Control*, 27(9):1695–1698.
- Harding, D. and Pagan, A. (2011). An econometric analysis of some models for constructed binary time series. *Journal of Business and Economic Statistics*, 29(1):86–95.
- Harvey, C. R. (1989). Forecasts of economic growth from the bond and stock markets. *Financial Analysts Journal*, 45(5):38–45.
- Kaminsky, G. L. and Schmukler, S. L. (2008). Short-run pain, long-run gain: Financial liberalization and stock market cycles. *Review of Finance*, 12(2):253–292.
- Kauppi, H. and Saikkonen, P. (2008). Predicting U.S. recessions with dynamic binary response models. *Review of Economics and Statistics*, 90(4):777–791.
- Kim, C.-J. (1994). Dynamic linear models with markov-switching. *Journal of Econometrics*, 60(1):1–22.
- Kole, E., Koedijk, K., and Verbeek, M. (2006). Portfolio implications of systemic crises. *Journal of Banking & Finance*, 30(8):2347–2369.
- Kullback, S. and Leibler, R. A. (1951). On information and sufficiency. *Annals of Mathematical Statistics*, 22(1):79–86.
- Lunde, A. and Timmermann, A. (2004). Duration dependence in stock prices: An analysis of bull and bear markets. *Journal of Business & 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 & 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 & Economic Statistics*, 30(3):391–403.
- Marcellino, M. (2006). Leading indicators: What have we learned? In Elliot, G., Granger, C. W., and Timmermann, A., editors, *Handbook of Economic Forecasting*, pages 879–960. Elsevier, Amsterdam, Netherlands.

- Marquering, W. and Verbeek, M. (2004). The economic value of predicting stock index returns and volatility. *Journal of Financial and Quantitative Analysis*, 39(2):407–429.
- McCracken, M. W. and Valente, G. (2012). Testing the economic value of asset return predictability. Working Paper 2012-049, Federal Reserve Bank of St. Louis, St. Louis, MO, USA.
- Newey, W. K. and West, K. D. (1987). A simple, positive semi-definite, heteroskedasticity and autocorrelation consistent covariance matrix. *Econometrica*, 55(3):703–708.
- Pagan, A. and Ullah, A. (1999). *Nonparametric Econometrics*. Cambridge University Press, Cambridge, UK.
- 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 Skouras, S. (2002). Decision-based methods for forecast evaluations. In Clements, M. P. and Hendry, D. F., editors, *A Companion to Economic Forecasting*, chapter 11, pages 241–267. Basil Blackwell, Oxford, UK.
- Pesaran, M. H. and Timmermann, A. (1995). Predictability of stock returns: Robustness and economic significance. *Journal of Finance*, 50(4):1201–1228.
- Politis, D. N. and Romano, J. P. (1994). The stationary bootstrap. *Journal of the American Statistical Association*, 89(428):1303–1313.
- 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.
- Rigobon, R. and Sack, B. (2003). Measuring the reaction of monetary policy to the stock market. *Quarterly Journal of Economics*, 118(2):639–669.
- Sarno, L. and Valente, G. (2004). Comparing the accuracy of density forecasts from competing models. *Journal of Forecasting*, 23(8):541–557.
- Schwert, G. W. (1990). Indexes of u.s. stock prices from 1802 to 1987. *Journal of Business*, 63(3):399–426.
- Shiller, R. J. (2000). *Irrational Exuberance*. Princeton University Press, Princeton NJ, USA.
- Stock, J. H. and Watson, M. W. (2003). How did leading indicator forecasts do during the 2001 recession? *Federal Reserve Bank of Richmond Economic Quarterly*, 89:71–90.
- Timmermann, A. (2000). Moments of markov switching models. *Journal of Econometrics*, 96(1):75–111.
- Veronesi, P. (1999). Stock market overreaction to bad news in good times: A rational expectations equilibrium model. *Review of Financial Studies*, 12(5):975–1007.

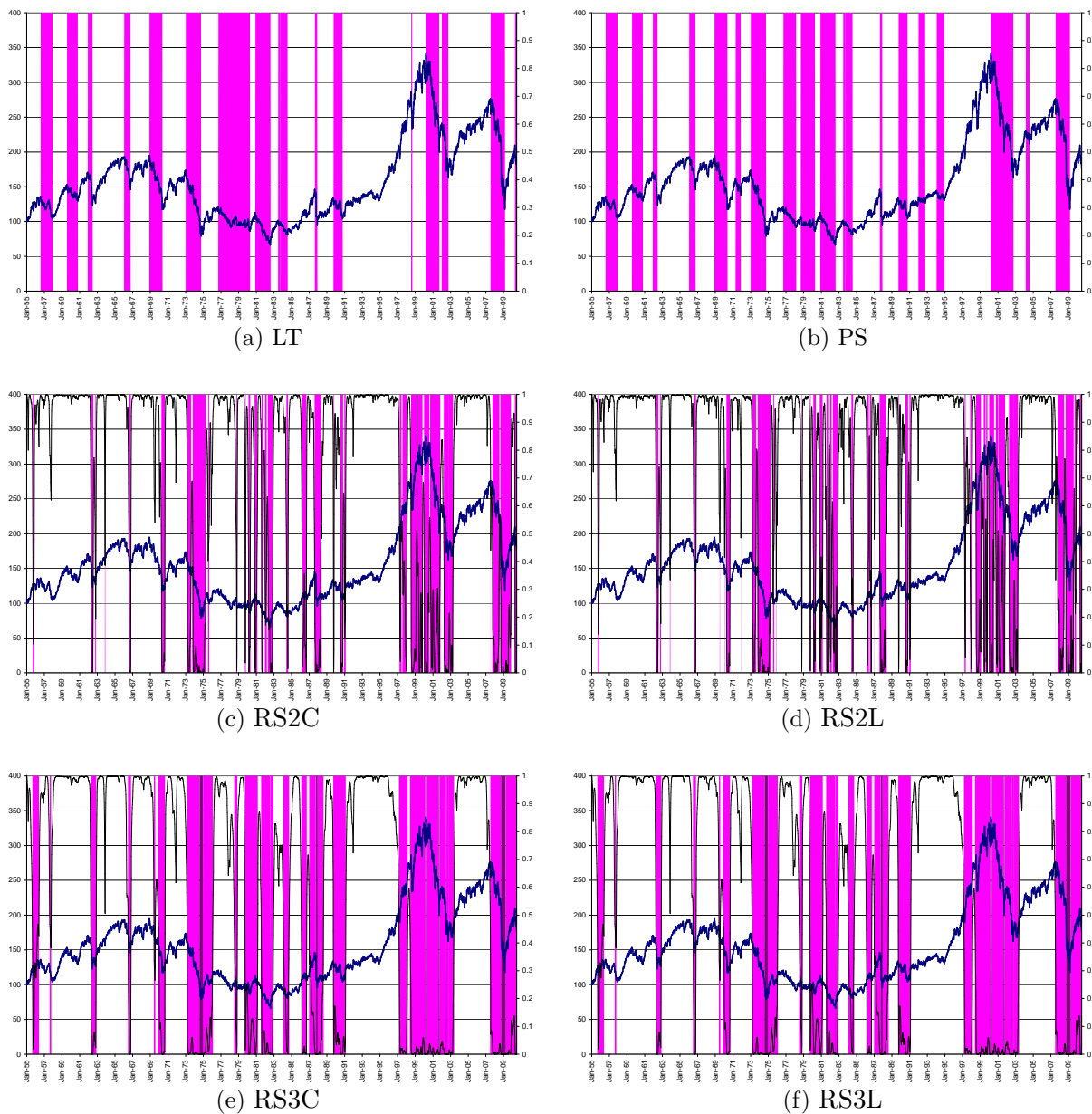
- Welch, I. and Goyal, A. (2008). A comprehensive look at the empirical performance of equity premium prediction. *Review of Financial Studies*, 21(4):1455–1508.
- West, K. D. (1996). Asymptotic inference about predictive ability. *Econometrica*, 64(5):1067–1084.
- West, K. D., Edison, H. J., and Cho, D. (1993). A utility-based comparison of some models of exchange rate volatility. *Journal of International Economics*, 35:23–45.
- White, H. (2000). A reality check for data snooping. *Econometrica*, 68(5):1097–1126.

Figure 1: Performance US Market



This figure shows weekly observations of the US stock market in excess of the risk-free rate over the period January 7, 1955 until July 2, 2010 ($1/7/1955 = 100$). The excess stock market index is calculated as the ratio P_t/B_t , where P_t is the value of the stock market index and B_t is the cumulation of a riskless bank account, $B_t \equiv \prod_{\tau=0}^{t-1} (1 + r_\tau^f)$. For the stock market index, we use the S&P500. The risk-free rates is the three-month T-Bill rate. Data have been taken from FRED at the Federal Reserve Bank of St. Louis and Schwert (1990).

Figure 2: Identification of Bull and Bear Markets



This figure shows the identification of bull and bear periods for the US, based on the different methods. The thick blue line plots the excess stock market index (left y -axis). In panel (a), bull and bear markets are identified by the LT-algorithm, and in panel (b) by the PS-algorithm. Panels (c–d) and (e–f) reflect identification by two-state and three-state regime-switching models. These models have constant or time-varying transition probabilities (panels c and e, and panels d and f). The thin black line indicates the smoothed inference probability of a bull market (right y -axis). Bullish (bearish) regimes are indicated with white (pink) areas. A bull (bear) market prevails in the RS-models when the smoothed inference probability for regime 1 (2) exceeds 0.5. The strong bear regime prevails when the smoothed inference probability for regime 3 exceeds 0.5, and is indicated in purple (only for panels e–f).

Figure 3: Forecasts of Bull and Bear Markets

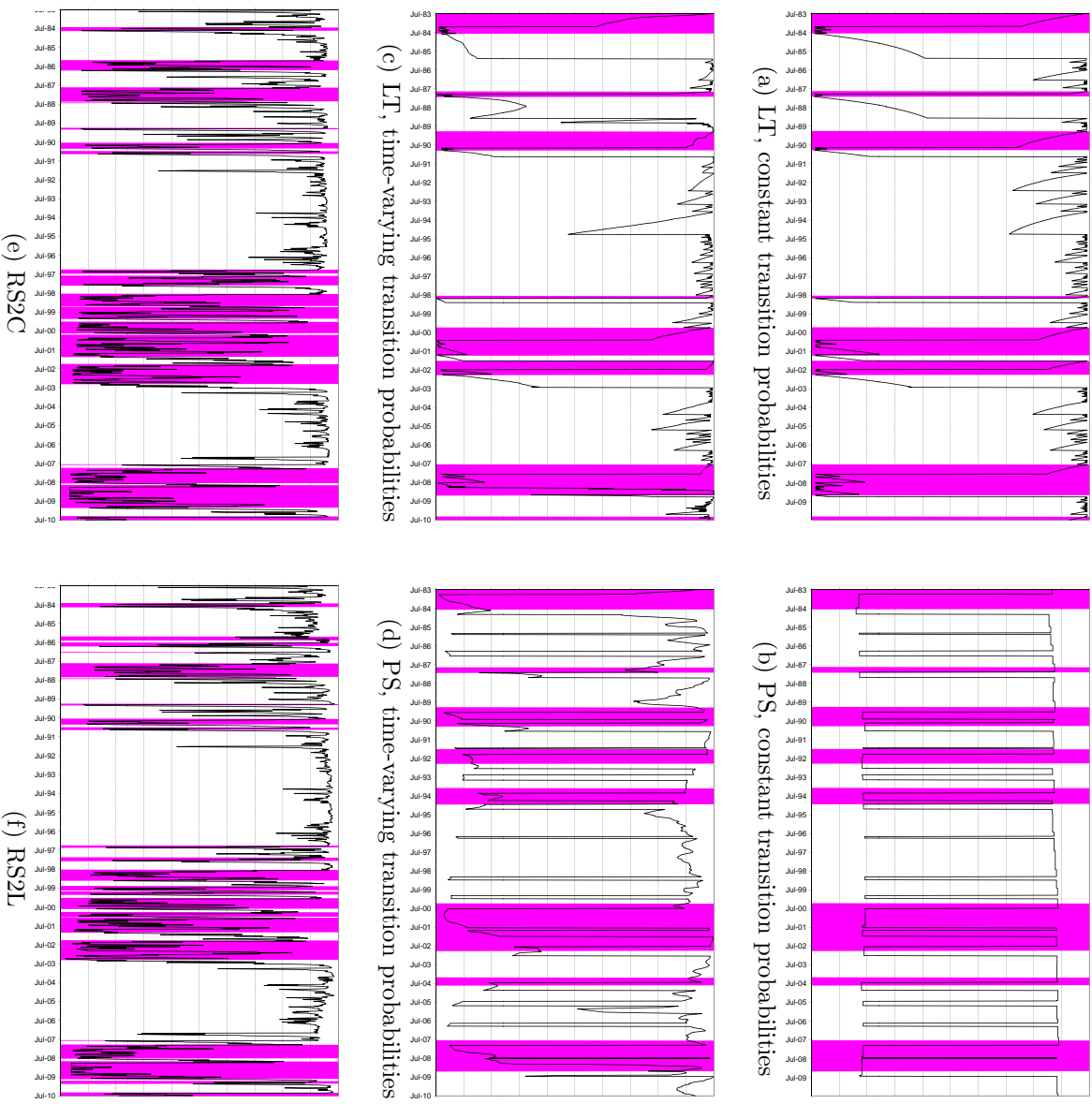
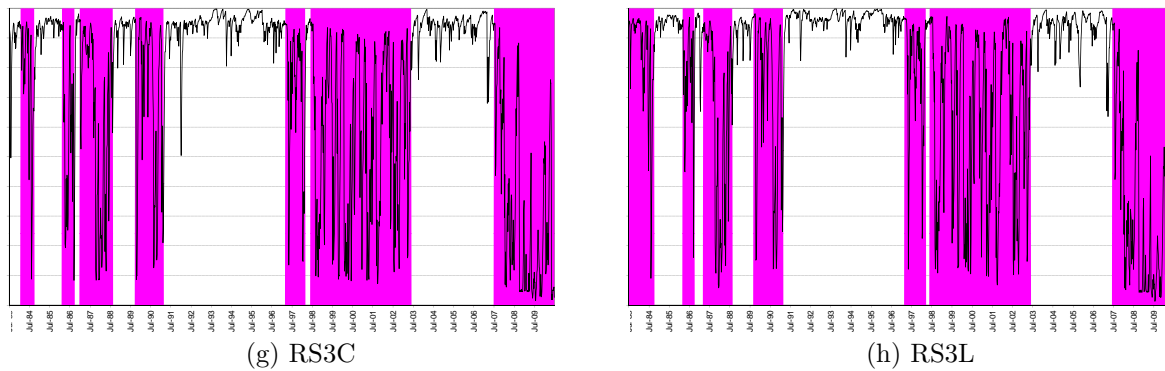


Figure continues on next page.

Figure 3: Forecasts of Bull and Bear Markets – *continued*



This figure shows the forecasts for each method over time. The first forecast is made for July 1, 1983 and the last for July 2, 2010, giving a total of 1410 predictions. To make a forecast for week $t + 1$, the LT-method follows the recursion in Equation (3) from the last extremum onwards. The PS-method starts this recursion 13 weeks prior to week t . The regime-switching models apply the standard filter technique with the last available model parameters. All approaches use weekly explanatory variables up to week t and monthly explanatory variables up to the last month prior to week t . The model parameters are updated every 52 weeks. The lines give the forecast probabilities of a bull market. For the LT and PS methods, the black solid (green dashed) line gives the forecast without (with) predictor variables. For the RS3-models, we calculate the probability of a bull market as the sum of the predictions for regimes with positive means. The pink areas indicate the periods of bear markets as identified based on the full sample as in Figure 2.

Table 1: Integrated Absolute Differences of Identification

	PS	RS2C	RS2L	RS3C	RS3L
LT	0.068 [0.035, 0.142]	0.325 [0.263, 0.434]	0.320 [0.268, 0.433]	0.346 [0.296, 0.538]	0.342 [0.303, 0.558]
PS		0.368 [0.301, 0.449]	0.363 [0.301, 0.448]	0.389 [0.331, 0.512]	0.383 [0.327, 0.530]
RS2C			0.032 [0.015, 0.067]	0.154 [0.081, 0.543]	0.167 [0.079, 0.595]
RS2L				0.168 [0.081, 0.554]	0.182 [0.073, 0.589]
RS3C					0.032 [0.028, 0.166]

This table reports the integrated absolute difference between the identification of the different approaches, based on two states as in Equation (10). When the RS3-models are part of the comparison, we aggregate the mild and strong bear regimes. We report 5% and 95% percentiles between brackets for each statistic based on 200 bootstrap samples of both the return and the explanatory variable series following Politis and Romano (1994). The probability p of randomly drawing a new observation from the sample is equal to 0.05.

Table 2: Return Characteristics of Bull and Bear Markets

regime		LT		PS		RS2C		RS2L		RS3C		RS3L	
bull	mean	0.38	(0.04)	0.40	(0.04)	0.16	(0.04)	0.16	(0.04)	0.15	(0.04)	0.16	(0.04)
	vol.	1.82	(0.06)	1.86	(0.07)	1.47	(0.04)	1.47	(0.04)	1.38	(0.04)	1.39	(0.04)
(mild) bear	mean	-0.60	(0.08)	-0.54	(0.07)	-0.27	(0.13)	-0.30	(0.14)	-0.04	(0.08)	-0.06	(0.08)
	vol.	2.46	(0.16)	2.36	(0.15)	3.28	(0.24)	3.34	(0.25)	2.43	(0.16)	2.36	(0.15)
strong bear	mean									-0.95	(0.73)	-0.74	(0.62)
	vol.									5.65	(1.53)	5.50	(1.30)

This table shows the mean and the volatility (in % per week) for the different regimes under the different approaches. In the approach of LT and PS identification is based on peaks and troughs in the price series. Conditioning on the regimes, we estimate the means and volatilities of the returns distributions. In the regime-switching approach, we estimate the model for the return distributions with two regimes (bull and bear, columns RS2C and RS2L) and with three regimes (bull, mild bear, and strong bear, columns RS3C and RS3L). We report standard errors in parentheses. For the LT- and PS-methods we calculate HAC consistent standard errors of Newey and West (1987) in a GMM setting. For the regime-switching models, we use the Fisher information matrix to compute standard errors.

Table 3: Performance Measures and Fees Based on Full-sample Identification

(a) Performance Measures						
	LT	PS	RS2C	RS2L	RS3C	RS3L
Mean	0.93	0.93	0.18	0.19	0.14	0.15
Volatility	4.20	4.21	1.57	1.57	1.47	1.57
Sharpe ratio	0.22	0.22	0.11	0.12	0.10	0.10
Utility ($\times 1,000$)	4.63	4.65	1.18	1.27	0.87	0.91

(b) Fees (in bps)						
	LT	PS	RS2C	RS2L	RS3C	RS3L
LT	0	-0.25	35.8	34.8	39.0	38.5
		[-10.3, 9.5]	[19.8, 46.3]	[19.0, 46.2]	[5.6, 43.6]	[-4.3, 43.0]
PS	0.25	0	36.1	35.1	39.3	38.8
	[-9.5, 10.4]		[19.9, 46.2]	[19.3, 44.9]	[5.1, 42.0]	[0.7, 41.3]
RS2C	-35.1	-35.4	0	-1.0	3.1	2.6
	[-45.1, -19.5]	[-45.1, -19.6]		[-3.8, 2.1]	[-24.2, 4.5]	[-32.9, 4.2]
RS2L	-34.1	-34.4	1.0	0	4.1	3.6
	[-45.1, -18.7]	[-43.8, -18.9]	[-2.1, 3.8]		[-23.1, 5.0]	[-32.6, 3.7]
RS3C	-38.2	-38.5	-3.1	-4.1	0	-0.5
	[-42.6, -5.5]	[-41.0, -5.1]	[-4.5, 24.3]	[-5.0, 23.2]		[-22.7, 2.5]
RS3L	-37.8	-38.0	-2.6	-3.6	0.5	0
	[-42.1, 4.2]	[-40.4, -0.6]	[-4.2, 33.0]	[-3.7, 32.8]	[-2.5, 22.8]	

This table reports performance measures of investment strategies that use the different methods as input for asset allocation and the fees an agent would be willing to pay to exchange two methods. The optimal portfolio for method m at time t equals $w_t^m = \mu_t^m / (\gamma \psi_t^m)$, with risk aversion parameter $\gamma = 5$. The values for the mean μ_{t+1}^m and second moment ψ_t^m are calculated as in Equations (13) and (14). The state-dependent mean μ_s^m and second moment ψ_s^m are based on the full-sample and reported in Table 2. The state-dependent probabilities are taken as the identification at time t , which are binary for the rule-based methods and smoothed inference probabilities for the regime-switching approaches. Based on the realized returns of the asset allocations, we calculate the mean and volatility (in % per week), the Sharpe ratio, and the realized utility as in Equation (15) (multiplied by 1,000). The fee (in basis points) $\eta_{m,n}$ to switch from strategy m (in rows) to strategy n (in columns) solves Equation (16). We report 5% and 95% percentiles between brackets for each statistic based on 200 bootstrap samples of both the return and the explanatory variable series following Politis and Romano (1994). The probability p of randomly drawing a new observation from the sample is equal to 0.05.

Table 4: Integrated Absolute Differences of Forecasts

	LTL	PSC	PSL	RS2C	RS2L	RS3C	RS3L
LTC	0.060 [0.057, 0.110]	0.250 [0.210, 0.288]	0.276 [0.271, 0.367]	0.239 [0.203, 0.356]	0.213 [0.193, 0.332]	0.239 [0.189, 0.338]	0.228 [0.178, 0.324]
LTL		0.279 [0.216, 0.303]	0.275 [0.237, 0.345]	0.272 [0.223, 0.374]	0.244 [0.212, 0.360]	0.269 [0.201, 0.364]	0.255 [0.194, 0.352]
PSC			0.082 [0.060, 0.145]	0.284 [0.253, 0.332]	0.262 [0.236, 0.320]	0.284 [0.239, 0.332]	0.281 [0.230, 0.328]
PSL				0.331 [0.294, 0.386]	0.305 [0.282, 0.379]	0.327 [0.278, 0.384]	0.318 [0.276, 0.374]
RS2C					0.061 [0.048, 0.077]	0.090 [0.059, 0.103]	0.116 [0.074, 0.126]
RS2L						0.086 [0.072, 0.114]	0.095 [0.071, 0.118]
RS3C							0.034 [0.029, 0.054]

This table reports the integrated absolute difference between the forecasts of the different approaches, based on two states. The forecasts are constructed as in Figure 3. When the RS3-models are part of the comparison, we concentrate on the predictions of bullish regimes. We calculate the probability of a bull market as the sum of the forecasts for regimes with positive means. We report 5% and 95% percentiles between brackets for each statistic based on 200 bootstrap samples of both the return and the predictive variable series following Politis and Romano (1994).

Table 5: Performance Measures and Fees Based on Forecasts
(a) Performance Measures

	market	LTC	LTL	PSC	PSL	RS2C	RS2L	RS3C	RS3L
Av. Abs. Weight	-	1.82	1.90	1.70	1.87	0.62	0.66	0.61	0.64
Mean	0.066	0.168	0.137	0.028	0.043	0.019	0.029	0.006	0.023
Mean Bull		0.172	0.198	0.019	0.108	0.070	0.089	0.083	0.128
Mean Bear	-	0.154	-0.052	0.049	-0.094	-0.081	-0.106	-0.072	-0.076
Volatility	2.32	4.50	4.43	3.88	4.17	1.31	1.40	1.50	1.51
Sharpe Ratio	0.029	0.037	0.031	0.007	0.010	0.014	0.021	0.004	0.015
Utility ($\times 1,000$)	-0.68	-3.38	-3.54	-3.48	-3.91	-0.25	-0.19	-0.49	-0.35

(b) Fees (in bps)

	LTC	LTL	PSC	PSL	RS2C	RS2L	RS3C	RS3L
LTC	0	1.6	1.0	5.3	-31.9	-32.4	-29.4	-30.9
		[-10.0, 3.6]	[-15.0, 15.7]	[-16.6, 18.5]	[-49.4, -16.4]	[-50.5, -17.1]	[-47.3, -14.8]	[-49.4, -13.8]
LTL	-1.6	0	-0.6	3.7	-33.5	-34.0	-30.9	-32.4
	[-3.6, 9.9]		[-11.8, 23.8]	[-13.1, 22.1]	[-44.8, -12.8]	[-46.2, -14.3]	[-43.4, -10.9]	[-44.8, -10.6]
PSC	-1.0	0.6	0	4.3	-32.7	-33.2	-30.1	-31.6
	[-15.7, 14.9]	[-23.8, 11.8]		[-10.9, 11.8]	[-50.7, -16.0]	[-50.6, -18.2]	[-48.3, -15.6]	[-48.2, -15.5]
PSL	-5.3	-3.7	-4.3	0	-37.1	-37.6	-34.5	-36.0
	[-18.4, 16.4]	[-22.1, 13.1]	[-11.8, 10.9]		[-50.2, -17.2]	[-52.1, -19.9]	[-49.5, -15.4]	[-49.8, -17.2]
RS2C	31.2	32.7	32.1	36.4	0	-0.5	2.5	1.0
	[16.0, 47.9]	[12.5, 43.8]	[15.8, 49.7]	[17.0, 49.0]		[-3.5, 1.2]	[-2.2, 4.7]	[-3.2, 5.7]
RS2L	31.7	33.2	32.6	36.9	0.5	0	3.0	1.5
	[16.7, 48.8]	[14.0, 45.0]	[17.9, 49.7]	[19.5, 51.3]	[-1.2, 3.5]		[-1.9, 6.5]	[-2.8, 6.5]
RS3C	28.7	30.2	29.6	33.9	-2.5	-3.0	0	-1.5
	[14.4, 45.9]	[10.8, 42.2]	[15.4, 47.6]	[15.2, 48.3]	[-4.7, 2.2]	[-6.5, 1.9]		[-3.5, 4.3]
RS3L	30.2	31.7	31.1	35.4	-1.0	-1.5	1.5	0
	[13.6, 48.0]	[10.5, 43.8]	[15.2, 47.1]	[16.9, 48.7]	[-5.7, 3.2]	[-6.5, 2.8]	[-4.3, 3.5]	

This table reports performance measures of investment strategies that use the different methods as input for asset allocation and the fees an agent would be willing to pay to exchange two methods. The optimal portfolio for method m at time t equals $w_t^m = \mu_t^m / (\gamma \psi_t^m)$, with risk aversion parameter $\gamma = 5$. The values for the mean μ_{t+1}^m and second moment ψ_t^m are calculated as in Equations (13) and (14). The state-dependent mean μ_s^m and second moment ψ_s^m are estimated as discussed in Figure 3. Based on the realized returns of the asset allocations, we calculate the mean and volatility (in % per week), the Sharpe ratio, and the realized utility as in Equation (15) (multiplied by 1,000). Mean Bull (Mean Bear) is the annualized mean during the subperiods identified *ex post* as bull (bear) markets. For the purpose of comparison, we report the statistics of a long position in the market. The fee $\eta_{m,n}$ (in basis points) to switch to strategy m (in rows) from strategy n (in columns) solves Equation (16). We report 5% and 95% percentiles between brackets for each statistic based on 200 bootstrap samples of both the return and the explanatory variable series following Politis and Romano (1994). The probability p of randomly drawing a new observation from the sample is equal to 0.05.

A Rules for identifying bull and bear markets

A.1 Lunde and Timmermann (2004)

The identification rules in Lunde and Timmermann (2004) for an (adjusted) index P_t can be summarized as follows:

1. Suppose the last observed extreme value before period t was a peak with index value P^{\max} .
 - (a) If the index value P_t exceeds P^{\max} , the maximum is updated, $P^{\max} = P_t$.
 - (b) If the index value P_t is more than a fraction λ_2 below P^{\max} , a trough has been found, and $P^{\min} = P_t$.
 - (c) If neither of these conditions is satisfied, no update takes place.

The agent now moves to $t + 1$.

2. Suppose the last observed extreme value before period t was a trough with index value P^{\min} .
 - (a) If the index value P_t is below P^{\min} , the minimum is updated, $P^{\min} = P_t$.
 - (b) If the index exceeds P^{\min} by more than a fraction λ_1 , a peak has been found, and $P^{\max} = P_t$.
 - (c) If neither of these conditions is satisfied, no update takes place.

The agent now moves to $t + 1$.

To commence the procedure we have to determine whether the market is initially bullish or bearish. For this purpose we count the number of times the maximum and minimum of the index have to be adjusted since the first observation. If the maximum has to be adjusted three times first, the market starts bullish, otherwise it starts bearish.

A.2 Pagan and Sossounov (2003)

The algorithm of Pagan and Sossounov (2003) consists of five steps:

1. Identify all local maxima and minima in a price series. A local maximum (minimum) is higher (lower) than all prices in the past and future τ_{window} periods.
2. Construct an alternating sequence of peaks and troughs by selecting the highest maxima and lowest minima, in case two (or more) extrema of the same type occur consecutively.
3. Censor peaks and troughs in the first and last τ_{censor} periods.
4. Eliminate cycles of bull and bear markets that last less than τ_{cycle} periods.
5. Eliminate bull market or bear markets that lasts less than τ_{phase} periods, unless the absolute price change exceeds a fraction ζ .

We mostly follow PS for the values of these parameters, adjusted for the weekly frequency of our data. We set $\tau_{\text{window}} = 32$, $\tau_{\text{cycle}} = 70$, $\tau_{\text{phase}} = 16$ and $\zeta = 0.20$ (see also PS, Appendix B). We censor switches in the first and last 13 weeks, opposite to the 26 weeks taken by PS. Censoring for 26 weeks would mean that only after half a year an investor can be sure whether a bear or a bull market prevails, which we consider a very long delay. Since we will use this information in making forecasts, we use a shorter period of 13 weeks to establish the initial and the ultimate state of the market.

B Multinomial logit transitions

In Section 2.3 we propose a regime-switching model with time-varying transition probabilities. The probability of a transition from regime q to regime s at time t is linked to predictor variables \mathbf{z}_{t-1} by a multinomial logit transformation

$$\pi_{qst} \equiv \pi_{qs}(\mathbf{z}_{t-1}) \equiv \Pr[S_t = s | S_{t-1} = q, \mathbf{z}_{t-1}] = \frac{e^{\beta'_{qs} \mathbf{z}_{t-1}}}{\sum_{s \in \mathcal{S}} e^{\beta'_{qs} \mathbf{z}_{t-1}}}, \quad s, q \in \mathcal{S}, \quad (\text{B.1})$$

with $\exists s \in \mathcal{S} : \beta_{qs} = \mathbf{0}$ to ensure identification. We have dropped the model-superscript m for notational convenience.

B.1 Estimation

To estimate the parameters β_{qs} , we extend the approach of Diebold et al. (1994), based on the EM-algorithm by Dempster et al. (1977). Diebold et al. (1994) consider estimation when the transition probabilities are linked via a standard (binomial) logit transformation. This extension maintains the attractive feature of the EM-algorithm that the expectation of the complete-data log likelihood can be split in terms related to only a subset of the parameter space. The transition part of the expectation of the likelihood function, which is only related to the parameters β_{qs} , is given by

$$\ell(B) = \sum_{t=1}^T \sum_{s \in \mathcal{S}} \sum_{q \in \mathcal{S}} \xi_{qst} \log \pi_{qst}, \quad (\text{B.2})$$

where $B = \{\beta_{qs} : s, q \in \mathcal{S}\}$ is the set of all parameters β_{qs} and $\xi_{qst} \equiv \Pr[S_t = s | S_{t-1} = q, \Omega_T]$ is a smoothed inference probability. These probabilities are based on the complete data set of returns and predictor variables Ω_T , and are calculated with the method of Kim (1994).

In the expectation step the set of smoothed inference probabilities is determined. In the maximization step new parameters values are calculated that maximize the expected likelihood function. The first order conditions that apply to β_{qs} result from differentiating Eq. (B.2)

$$\frac{\partial \ell(B)}{\partial \beta_{qs}} = \sum_{t=1}^T \sum_{\varsigma \in \mathcal{S}} \xi_{q\varsigma t} \frac{1}{\pi_{q\varsigma}} \frac{\partial \pi_{q\varsigma}}{\partial \beta_{qs}}.$$

Based on Eq. (B.1) we find

$$\frac{\partial \pi_{q\varsigma}}{\partial \beta_{qs}} = \begin{cases} \pi_{qs}(1 - \pi_{qs})\mathbf{z}_{t-1} & \text{if } \varsigma = s \\ -\pi_{q\varsigma}\pi_{qs}\mathbf{z}_{t-1} & \text{if } \varsigma \neq s \end{cases}.$$

Combining these two expressions yields the first order condition

$$\sum_{t=1}^T (\xi_{qst} - \xi_{q,t-1}\pi_{qst}) \mathbf{z}_{t-1} = \mathbf{0} \quad \forall q, s \in \mathcal{S} \quad (\text{B.3})$$

where $\xi_{q,t} = \Pr[S_t = q | \Omega_T]$. For each departure state q the set of the first order conditions for the different $s \in \mathcal{S}$ comprise a system that determines the set $B_q = \{\beta_{qs} : s \in \mathcal{S}\}$. Numerical techniques can be used to find parameters β_{qs} that solve this system.

B.2 Marginal Effects

Because the multinomial logit transformation is non-linear, the coefficients on the explanatory variables cannot be interpreted in a straightforward way. To solve this problem, we calculate the marginal effect of the change in one variable z_i , evaluated at specific values for all variables $\bar{\mathbf{z}}$. The marginal effect is given by the first derivative of (6) with respect to z_i :

$$\left. \frac{\partial \pi_{qs}(\mathbf{z})}{\partial z_i} \right|_{\mathbf{z}=\bar{\mathbf{z}}} = \pi_{qs}(\bar{\mathbf{z}}) \left(\beta_{qsi} - \sum_{\varsigma \in \mathcal{S}} \pi_{q\varsigma}(\bar{\mathbf{z}}) \beta_{q\varsigma i} \right), \quad (\text{B.4})$$

where β_{qsi} denotes the coefficient on z_i . It is easy to verify that the sum of this expression over the destination states s is equal to zero. Since the probabilities for the destination states should add up to one, a marginal increase in one probability should be accompanied by decreases in the other probabilities. When only two regimes are available, the above expression reduces to the familiar expression for marginal effects in logit models, $\pi_{qs}(\bar{\mathbf{z}})(1 - \pi_{qs}(\bar{\mathbf{z}}))\beta_{qsi}$.

C Predictor variables

The predictor variables are based on data from the Federal Reserve Bank of St. Louis available via FRED¹⁰, except the unemployment rate (from the Bureau of Labor Statistics) and the dividend-to-price ratio. If a variable has a unit root, we transform the data to create a stationary time series. The results are reported in Table C.1, together with the mean and standard deviation of the (transformed) series.

We construct the inflation rate as the monthly relative change in the consumer price index. We reject a unit root for the inflation rate at conventional confidence levels, and find an average inflation rate of 3.78% per year with a standard deviation 1.10%. We construct the growth rate of industrial production as the yearly relative change in industrial production. It does not show evidence of a unit root, and has an average of 3.06% with a standard deviation of 5.27%. Because the unemployment rate has a high AR(1)-coefficient and a p -value for the ADF-statistic close to 0.05, we transform the series by taking a

¹⁰See <https://research.stlouisfed.org/fred2/>

yearly change. The average is close to zero with a standard deviation of 1.14%. These three variables are constructed at the monthly frequency and are seasonally adjusted.

[Table C.1 about here.]

We transform the three-month T-Bill rate by taking the difference with respect to the yearly moving average as in Campbell (1991); Rapach et al. (2005), because a unit root is not rejected. The resulting difference is on average zero and shows a standard deviation of 1.04%. We construct the term spread as the yield on a ten-year government bond minus the three-month T-Bill rate. A weekly series for the ten-year yield becomes available from 1962 onwards. The AR(1)-coefficient and ADF-test are based on this smaller subsample. They indicate stationarity, so we do not transform this series. We use the monthly series for the ten-year government bond and three-month T-Bill rate to construct the observations from 1955 to 1962, and assume that the difference stays constant within a month. The average term spread is 1.42% with a volatility of 1.23%. We determine the credit spread as the difference between the yield on Baa-rated and Aaa-rated corporate bonds (as calculated by Moody's, seasonally adjusted). As weekly data become available starting in 1962 again, we follow the same procedure as for the term spread. We reject the null-hypothesis of a unit root for the subsample with weekly data, and find an average spread of 0.99% and a volatility of 0.46% for the whole sample.

The dividend-to-price ratio is constructed from several series. We use the dividends series for the S&P500 as in Shiller (2000), which are available on Robert Shiller's homepage.¹¹ The series consists of monthly observations of the moving total dividends over the past twelve months. To calculate the D/P-ratio for time t (in weeks), we divide the dividends over the last 12 months prior to the month corresponding with time t by the current price level. As the resulting series shows evidence of a unit root, we take the difference with respect to the moving yearly average. This difference is on average zero and has a volatility of 0.34%.

¹¹See <http://www.econ.yale.edu/~shiller/data.htm> for more information.

Table C.1: Characteristics of Predictor Variables

	ID	Frequency	AR(1)	ADF	p-value	Transformation	Mean	Std. Dev.
Inflation	CPIAUCSL	monthly	0.623	-3.26	0.017		3.78	1.10
Ind. Prod., yearly growth rate	INDPRO	monthly	0.966	-5.20	< 0.001		3.06	5.27
Unemployment	LNS14000000	monthly	0.997	-3.05	0.031	yearly change	0.08	1.14
Three-month Tbill rate	WTB3MS	weekly	0.998	-2.36	0.152	change to yearly average	-0.01	1.04
Term spread	WGS10YR	weekly	0.991	-4.28	< 0.001		1.42	1.23
Credit spread	WAAA, WBAA	weekly	0.993	-3.81	0.003		0.99	0.46
D/P ratio		weekly	0.998	-1.78	0.390	change to yearly average	-0.02	0.34

This table shows the set of predictor variables with their ID and frequency. For each variable we conduct an Augmented Dickey-Fuller (ADF) test. We report the first order autocorrelation coefficient, the ADF test statistic and the p -value for the null hypothesis of the presence of a unit root. If this hypothesis is not rejected, the next column shows the transformation that is applied to the variable. The last two columns show the mean and standard deviation of the (transformed) variables in %. The monthly series run from December 1954 to May 2010; the weekly series from January 7, 1955 until June 25, 2010.

D Additional Results on Identification

D.1 Constant versus Time-Varying Transition Probabilities

The results in Section 4.1 indicate that the difference between constant and time-varying transition probabilities in the regime-switching models are minor. Here we investigate in more detail how different the results are.

Table D.1 reports the transition probabilities under the assumption that they are constant over time. Bull and bear markets are quite persistent with probabilities of around 0.95 or higher that the current state prevails for another week. Bull markets tend to be slightly more persistent than bear markets. Only the strongly bearish regime is somewhat less persistent, though its probability of continuation for another week is still 0.89.

[Table D.1 about here.]

As an alternative to constant transition probabilities, we link the probabilities to predictor variables, which makes them time-varying. According to Table D.2, time-variation can be related to a few economic variables. The D/P ratio is selected for all models. A rise in the D/P-ratio increases the likelihood of a bull market in the next period in the LT, PS and RS3L-models. This result comes as no surprise, since an increase in the D/P ratio can point at higher future expected returns. In the RS2L-model the D/P-ratio decreases the likelihood of a bull market, which is more puzzling. It is similarly puzzling than an increase in the D/P-ratio strengthens the persistence of the strongly bearish regime in the RS3L-model.

A rise in the inflation rate negatively affects the persistence of bull markets in the LT, PS and RS3L-models, but has no effect when the market is bearish. Other variables show up less consistently. An increase in the unemployment rate during a bull market leads to a higher probability of continuation in the PS and RS3L-model. In the PS-approach, a higher term spread or a lower credit spread increases the probability of a switch from a bear to a bull market, while in the RS2L-model they increase the probability that a bull market continues. A higher T-bill rate decreases the probability of a bull market in the LT-model.

To determine by how much the transition probabilities change when the predictor variables change, we calculate the marginal effect that a one-standard deviation change in a predictor variable has on a reference probability π . As a reference point we use the average forecast probability

$$\bar{\pi}_{qs} = \frac{\sum_{t=1}^T \Pr[S_{t+1} = s | S_t = q, \mathbf{z}_t] \Pr[S_t = q | \Omega_t]}{\sum_{t=1}^T \Pr[S_t = q | \Omega_t]}, \quad (\text{D.1})$$

where Ω_t denotes the information set up to time t . In this expression, each forecast probability $\Pr[S_{t+1} = s | S_t = q, \mathbf{z}_t]$ of a switch from state q to state s is weighted by the likelihood of an occurrence of state q at time t , $\Pr[S_t = q | \Omega_t]$. In the rules based approaches, the weights are either zero or one. In the regime-switching approaches the weights are the so-called inference probabilities. We report these probabilities for the different regimes and models in the last row of Table D.2. They confirm the strong persistence reported in Table D.1.

The marginal effects are calculated from the reference probability π and the coefficient β . When logit transformations, the marginal effect of variable z_i with coefficient β_i is given by $\pi(1 - \pi)\beta_i$. For the multinomial logit transformation we derive the marginal effects in Appendix B. Overall, marginal effects are small, around 0.01–0.02. However, the probability of a switch from a bear to a bull market can still double. For instance, in the PS-model, the marginal effect of a one-standard deviation increase in the D/P-ratio increases the probability a bear-bull switch from 0.02 to 0.04.

[Table D.2 about here.]

Combining Tables D.1 and D.2, we conclude that all methods identify persistent bull and bear markets. The evidence for time-variation in the transition probabilities is limited. However, if the sentiment of the stock market is a good predictor of other economic processes, it is not surprising that other economic variables fail to predict the stock market sentiment well. The D/P-ratio, which is closely related to expected returns in the stock market, is most consistently selected.

D.2 Model choice

To judge the quality of the different models, we report log likelihood values in Table D.3. For all models, the improvements by introducing predictor variables in the transition probabilities are significant by construction. Also, the Akaike Information Criterion favors the models with time-variation in the transition probabilities. However, the improvements of the likelihood are not enough to improve the Bayesian Information Criterion, which puts a heavier penalty on additional parameters. We conclude that the evidence favoring time-varying transition probabilities is at best marginal.

[Table D.3 about here.]

For the regime-switching models we can also evaluate the added value of an extra regime. For both cases (constant and time-varying transition probabilities), an additional regime leads to large improvements in the likelihood values. Both information criteria prefer a model with three regimes over one with only two. When transition probabilities are constant, the model with two regimes is nested in the model with three regimes, and we could conduct a likelihood ratio test (the LR-statistic equals 105.50). However, due to presence of unidentified nuisance parameters under the null hypothesis of two regimes, the statistic does not follow a standard χ^2 distribution and simulations are needed. As our interest is not in selecting the best statistical model we do not conduct this test here. Given the magnitude of the LR-statistic and the values of the information criteria, the evidence so far supports a model with three regimes.

Table D.1: Constant Transition Probabilities
(a) Probability Estimates

from	to	LT	PS	RS2C	RS3C
bull	bull	0.992	0.990	0.981	0.986
	bear	0.008	0.010	0.019	0.014
	crash				< 0.001
bear	bull	0.015	0.016	0.052	0.019
	bear	0.985	0.984	0.948	0.972
	crash				0.009
crash	bull				< 0.001
	bear				0.106
	crash				0.894

(b) Unconditional Regime Probabilities

	LT	PS	RS2C	RS3C
bull	0.652	0.615	0.730	0.570
(mild) bear	0.348	0.385	0.27	0.396
strong bear				0.034

This table shows the transition probabilities between the different regimes under the different approaches and the resulting unconditional regime probabilities. We assume that the probabilities are constant over time. In the approaches of LT and PS, we first apply their algorithms to identify the sequences of bull and bear markets. As a second step we estimate the probabilities. For the regime-switching models the probabilities result directly from the estimation. The regime-switching model can either have 2 regimes (RS2C) or 3 regimes (RS3C). The unconditional probabilities $\bar{\pi}$ satisfy $\bar{\pi}^m \mathbf{P}^m = \bar{\pi}^m$.

Table D.2: Time-varying transition probabilities, (multinomial) logit models

model from to	LT		PS		RS2L		RS3L					
	bull bull	bear bull	bull bull	bear bull	bull bull	bear bull	bull bull	mild bear mild bear	strong bear bull	mild bear mild bear	strong bear mild bear	
constant	5.16	-5.32	5.23	-5.35	3.62	-2.33	88.58	83.35	-0.29	3.72	-63.27	0.18
inflation	-0.34 [-0.003]	-	-0.85 [-0.008]	-	-	-	-1.04 [-0.013]	-	-	-	-	-
prod. growth	-	-	-	-	-	-	-	-	-	-	-	-
unempl.	-	-	0.73 [0.007]	-	-	-	-	-0.95 [-0.012]	-	-	-	-
t-bill rate	-0.77 [-0.006]	-	-	-	-	-	-	-	-	-	-	-
term spread	-	-	-	0.70 [0.011]	0.32 [0.009]	-	-	-	-	-	-	-
credit spread	-	-	-	-0.40 [-0.006]	-0.68 [-0.018]	-	-	-	-	-	-	-
D/P ratio	0.76 [0.006]	1.01 [0.015]	1.05 [0.010]	1.25 [0.020]	-	-0.43 [-0.031]	1.14 [0.014]	-	-	-	-	-1.14 [-0.207]
$\bar{\pi}_{qs}$	0.99	0.02	0.99	0.02	0.97	0.08	0.99	0.01	0.02	0.96	< 0.01	0.24

This table shows the estimated coefficients and marginal effects of the predictor variables, when they are linked to the transition probabilities by (multinomial) logit models. The predictor variables have been standardized by subtracting their full-sample mean and dividing by their full-sample standard deviation. In the approaches of LT and PS, we first apply the algorithms to identify bullish and bearish periods. In the second step we estimate a Markov-logit model as in Equation (2). In the two-state regime-switching model, RS2L, the logistic transformation in Equation (7) is used to link the predictor variables to the transition probabilities. For the three-state regime-switching model, RS3L, the multinomial logistic transformation in Equation (6) is used. In that case the coefficients for a switch to the strong bear regime have been fixed at zero. The variable-transition combinations that subsequently produce the biggest increase in the likelihood function are included in the models. The procedure stops when the remaining combinations of predictor variables and transitions fail to produce an increase in the likelihood function that is significant on the 10%-level. For the RS3L-model we allow a maximum of four combinations to be included. The marginal effects in brackets are calculated for the average forecast probability $\bar{\pi}_{qs}$ reported in the last row of the table. The average forecast probability is calculated as in Equation (D.1). For the two-state approaches, the marginal effect of predictor variable i is calculated as $\bar{\pi}_{qs}(1 - \bar{\pi}_{qs})\beta_{qi}$. For the three-state approaches, the marginal effect is given by Equation (6).

Table D.3: Log likelihood values and information criteria of different model specifications

		LT	PS	RS2	RS3
constant	# parameters	2	2	7	14
	$\log L$	-170.12	-192.61	-5979.75	-5927.01
	AIC	0.119	0.134	4.136	4.104
	BIC	0.123	0.139	4.150	4.133
time-varying	# parameters	6	8	11	18
	$\log L$	-153.82	-168.63	-5970.87	-5916.96
	AIC	0.110	0.122	4.133	4.100
	BIC	0.123	0.139	4.155	4.137
LR		32.60	47.96	17.78	20.09
df		4	6	4	4
Pr(0.01)		13.28	16.81	13.28	13.28

This table shows the log likelihood values of the different models. For the rules-based approaches LT and PS, we report the log likelihood values of the Markov-logit models as in (2). For the regime-switching models with two or three states, we report the likelihood of the complete model. The transition probabilities can be constant (corresponding with Table D.1) or time-varying (corresponding with Table D.2). Based on the log likelihood values we calculate the Akaike and Bayesian Information Criterion (AIC and BIC). In the row labeled “LR” we report the likelihood ratio statistic for time-varying vs. constant transition probabilities, which has a χ^2 distribution with degrees of freedom listed in the row below. The last row reports 1% critical values of the corresponding χ^2 distribution.

E Additional Results on Forecasting

When comparing the forecasts of the various methods, we reestimate the parameters of the different models every 52 weeks. Here we show how the different parameters evolve. Besides aiding our understanding of the forecasts, it also shows how robust the parameters and characteristics are when more information becomes available.

Figure E.1 shows the evolution of the means and volatilities of the different regimes. Generally, these are stable. As we concluded in the full-sample analysis, the difference between the mean for the bull and for the bear regime is more extreme for the rules-based approaches, while the difference between the volatilities for these two regimes is more extreme for the RS2-models.

[Figure E.1 about here.]

The evolution of the means and volatilities of the regimes in the RS3-models in Figures E.1e and E.1f shows the influence of the credit crisis. When data until May 2008 is used, two bullish and one bearish regime are identified. One regime has strongly bullish characteristics with a high mean and a low volatility, while the other is more mild, with a mean slightly above zero, and a higher volatility. The characteristics of the bearish regime match quite well with those of the bearish regime in the RS2-models. After 2008, one bullish and two bearish regimes show up. One bearish regime has mild characteristics, while the other has much stronger bearish features. It stresses the exceptional behavior of the US stock market during the credit crisis.

The evolution of the transition probabilities when assumed constant within an estimation window are in Figure E.2. The methods with two states produce transition probabilities that are also quite stable over time. Persistence is high for both regimes. The probability of remaining in a bull state never falls below 0.95. For the rules-based approaches, the same applies to the bear regime. In the RS2C model, a bear market seems slightly less persistent, but the probability of continuation almost always exceeds 0.90.

[Figure E.2 about here.]

The transition probabilities in the RS3C-model vary a bit more than in the two-state models, in particular the probabilities for a switch to another regime. For example, the

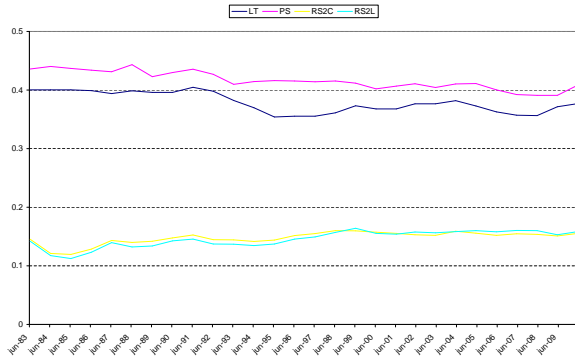
probability of a switch from a (strongly) bearish to a mildly bullish/bearish regime ranges from 0.04 to 0.11. The probabilities that a specific regime continues are high (around 0.90 or more) and quite stable.

The parameter dynamics of the models with time-varying transition probabilities in Figure E.3 indicate whether the parameters are stable, but also whether the variable selection is stable. In the rules-based method, the selection and the parameters are quite stable. The D/P-ratio is consistently present for all switches in both the LT and PS-methods. The T-bill rate is selected for forecasts from the bullish regime in both models. In the PS-method, the inflation rate, the unemployment rate and the term spread also show up consistently. In the LT-approach, these are sometimes selected. In the RS-models the selection shows a more haphazard pattern. However, given that a variable is selected, the corresponding coefficient is stable.

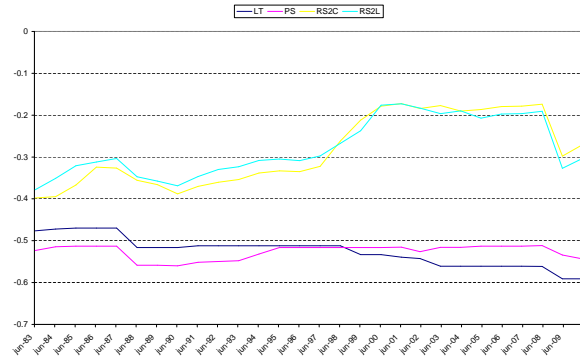
[Figure E.3 about here.]

[Figure E.3 (continued) about here.]

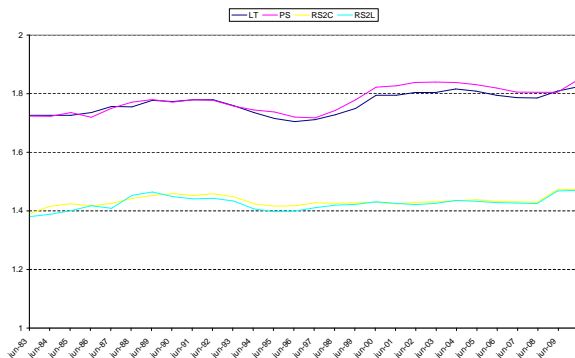
Figure E.1: Evolution of means and volatilities per regime



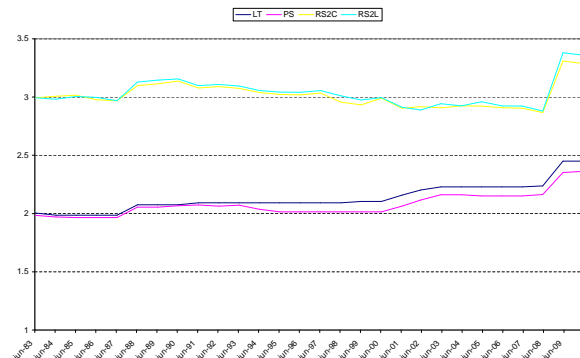
(a) mean of bull regimes in two-state models



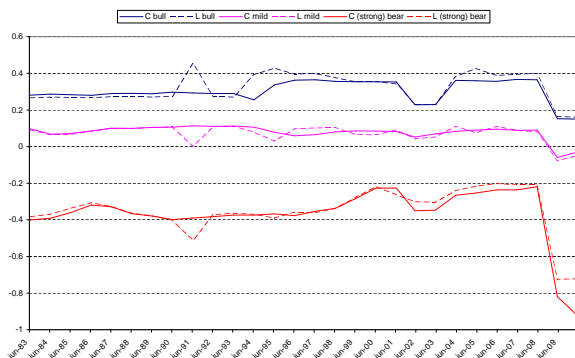
(b) mean of bear regimes in two-state models



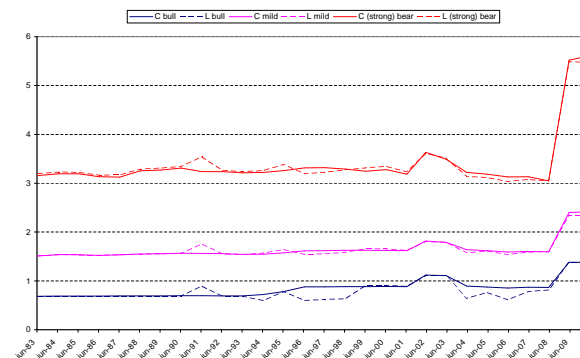
(c) volatility of bull regimes in two-state models



(d) volatility of bear regimes in two-state models



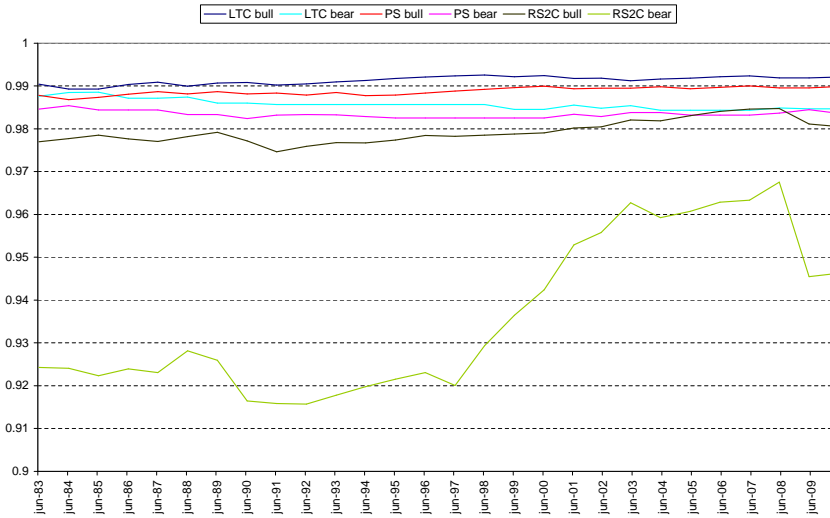
(e) mean of regimes in RS3-models



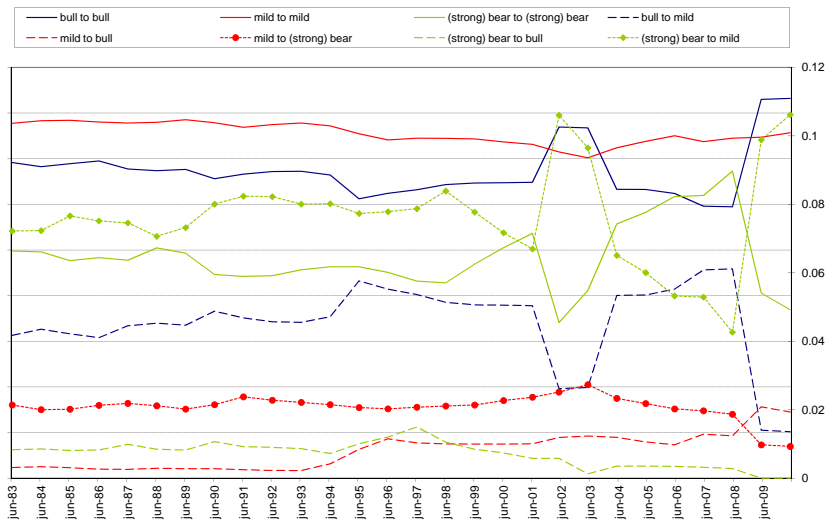
(f) volatility of regimes in RS3-models

This figure shows the evolution of the means and volatilities when estimated with an expanding window (end date on the x -axis). The first window comprises the period January 7, 1955 – June 24, 1983 (1485 observations), and is continuously expanded with 52 weeks until we reach the end of the sample (July 2, 2010). In the approaches of LT and PS, we first apply their algorithms to identify the sequences of bull and bear markets for each estimation window. As a second step we calculate means and volatilities per regime. The regime-switching models can either have two regimes or three regimes, and constant or time-varying transition probabilities. The means and volatilities of the regimes follow directly from the estimation.

Figure E.2: Evolution of constant transition probabilities



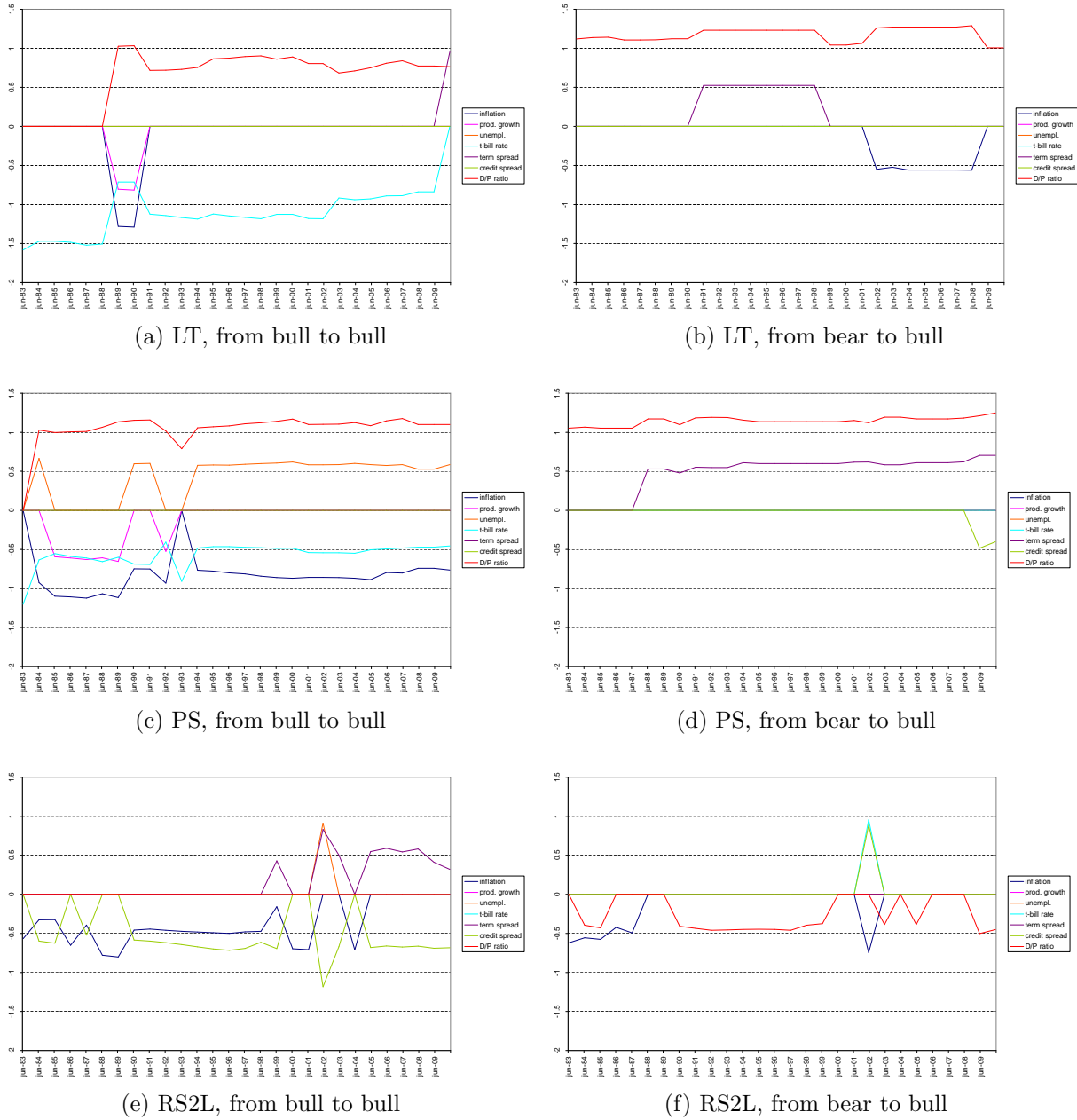
(a) 2-state models



(b) RS3C

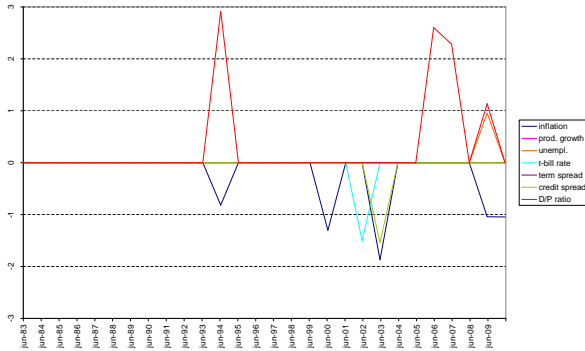
This figure shows the evolution of the transition probabilities when estimated with an expanding window (end date on the x -axis). We assume that the probabilities are constant within each estimation window. The first window comprises the period January 7, 1955 – June 24, 1983 (1485 observations), and is continuously expanded with 52 weeks until we reach the end of the sample (July 2, 2010). In the approaches of LT and PS, we first apply their algorithms to identify the sequences of bull and bear markets for each estimation window. As a second step we estimate the probabilities. For the regime-switching models the probabilities result directly from the estimation. The regime-switching models can either have two regimes (RS2C) or three regimes (RS3C). For the methods with two states, we plot the probabilities of a bull-bull and a bear-bear switch in Panel (a). For the RS3C-model in Panel (b) we indicate the transition in the legend above the subfigure. Dashed lines correspond with the secondary y -axis. We do not show transition probabilities that never exceeded 0.001.

Figure E.3: Evolution of parameters in logit models

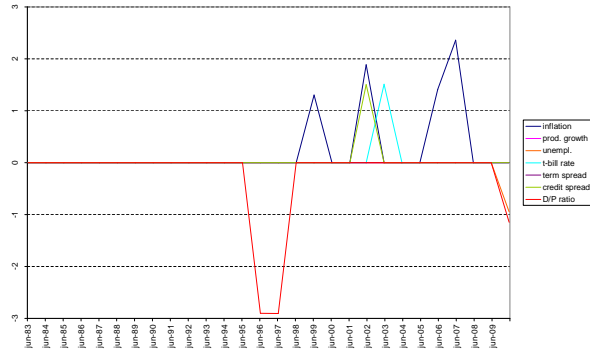


This figure plots the evolution of the coefficients in the (multinomial) logit transitions for the predictor variables in Table C.1, when estimated with an expanding window (end date on the x -axis). The first window comprises the period January 7, 1955 – June 24, 1983 (1485 observations), and is continuously expanded with 52 weeks until we reach the end of the sample (July 2, 2010). The predictor variables have been standardized by subtracting their full-sample mean and dividing by their full-sample standard deviation. In the approaches of LT and PS, we first apply the algorithms to identify bullish and bearish periods in the subperiod under consideration. In the second step we estimate a Markov-logit model as in Equation (2), where the coefficients depend on the departure state. In the RS2L-model the logistic transformation in Equation (7) is used to link the predictor variables to the transition probabilities. *Figure note continues on next page.*

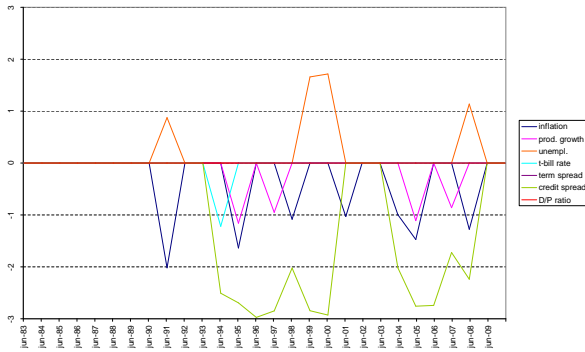
Figure E.3: Evolution of parameters in logit models – *continued*



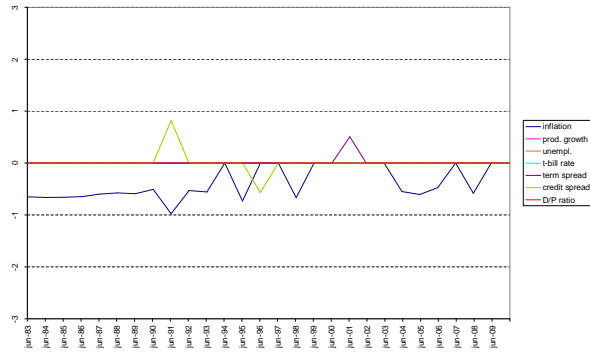
(g) RS3L, from bull to bull



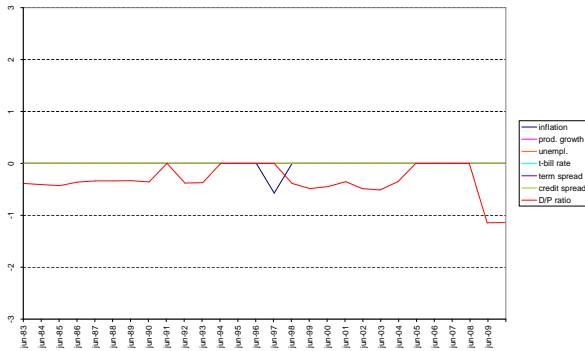
(h) RS3L, from bull to mild bear



(i) RS3L, from mild bear to bull



(j) RS3L, from mild bear to mild bear



(k) RS3L, from strong bear to mild bear

Continued from previous page. In the RS3L, the multinomial logistic transformation in Equation (6) is used. For identification, all coefficients for a switch to the (strong) bear regime have been fixed at zero. The variable-transition combinations that subsequently produce the biggest increase in the likelihood function are included in the models. The procedure stops when the remaining variable-transition combinations fail to produce an increase in the likelihood function that is significant on the 10%-level. In each subfigure, we only plot the variables that have been selected at least once.

F Result of Robustness Checks

F.1 Robustness of the LT-method

Lunde and Timmermann (2004) consider four combinations for the values of the thresholds λ_1 and λ_2 to identify switches between bull and bear markets. They argue that a value of 20% for λ_1 is conventionally used. A lower value of 15% for λ_2 subsequently accounts for the positive drift of the stock market. Other combinations they consider for (λ_1, λ_2) are $(0.20, 0.10)$, $(0.15, 0.15)$ and $(0.15, 0.10)$. Since we conclude that a quick identification of the current state is important when making forecasts, we follow LT and also consider these combinations of lower thresholds.

Lower thresholds lead to a more rapid alternation of bull and bear markets, which as a consequence last briefer. In the identification based on the full sample, the number of cycles, reported in Table F.1, increases from 16 in the original $(0.20, 0.15)$ setting to 24 for the lowest thresholds $(0.15, 0.10)$. Lowering both thresholds has a much stronger effect than lowering only one of them. Average and median durations decrease for lower thresholds, though the result is less pronounced than in Lunde and Timmermann (2004). This is due the fact that the longest bull market is unaffected by the choice of thresholds.

[Table F.1 about here.]

We use the IAD and maximum fee to determine how different the identifications and forecasts are that results from the LT-method with different thresholds. Table F.2a shows that the integrated absolute difference between these series is quite small. Their average difference varies between 0.011 and 0.059, while the upper bound of the 90% confidence intervals is around 0.10. Comparing these difference with those in Table 1 shows that they are slightly smaller than the difference between the standard LT and the PS identifications, and much smaller than the differences between the standard LT-method and the RS-models. So from a statistical point of view, the type of method matters more than the specific thresholds.

[Table F.2 about here.]

The economic comparison in Tables F.2b and F.2c shows more substantial differences. A faster identification of bull and bear markets leads to a considerably higher mean return, rising from 0.93% to 1.27% per week, at the cost of a slight increase in the volatility from 4.20% to 4.89%. As a consequence, both the Sharpe ratio and the realized utility increase. An investor would be willing to pay large fees up to 18.6 bps to switch to lower thresholds. Since fees to switch from the standard (0.20, 0.15) are all positive, the fees in Table 3b may underestimate the preference of the standard LT-method over the regime-switching methods. The fees to switch from (0.20, 0.15) to (0.15, 0.10) are significant, as confidence intervals are substantially bounded from zero. Only the joint lowering of thresholds commands a significant fee. However, these fees correspond with identification *ex post*, and it is not obvious that working with lower thresholds leads to better forecasts or allocations *ex ante*.

The IAD-measures in Table F.3 indicate that the differences between forecasts based on the different threshold combinations are larger than those between identification. Logically, IADs are larger when both threshold values are different, with maxima around 0.16, meaning that 16% of the forecasts are different. Confidence intervals indicate that the differences are significant. A comparison with Table 4 shows that the forecasts by LT-method with different thresholds still have more in common than the standard LT-method with the other methods, including the PS method. In line with earlier results, the use of predictor variables only leads to marginal differences in the forecasts.

[Table F.3 about here.]

Finally, we compare the performance of investment strategies that are based on the LT-methods with different thresholds in Table F.4. Lower thresholds lead to higher means, which rise from around 1.9% to 2.22%, but also to higher volatilities. Part of the increases come from a larger position in futures contracts. Judged by Sharpe ratio and realized utility, the threshold combination (0.20, 0.15) works best, while a lowering of the λ_1 threshold actually leads to lower utility because of the increase in volatility. The highest utility in Table F.4a is still lower than the utilities for the regime-switching models in Table 5a.

[Table F.4 about here.]

The fees in Table F.4b indicate that the economic improvement of a strategy by one threshold combination compared to another combination is still relatively small and mostly insignificant. Changing from the standard (0.20, 0.15) combination to the (0.20, 0.10) commands a fee of 7.7–11.1 bps, but the confidence intervals are quite wide and stretch beyond zero. The only significant fees correspond with a switch from the (0.15, 0.10) combination with constant transition probabilities. The magnitude of the fees is small compared to the fees reported in Table 5. Lowering the thresholds leads to better identification but also to more risk-taking, which in the end makes a slower, more cautious strategy more attractive.

Overall, we conclude that lowering the thresholds improves identification and forecasts. Lower thresholds lead to better identification, both in a statistical and an economic performance. We already concluded that the rules-based methods are better suited for identification. The results for lower thresholds indicate an even larger difference. For forecasts, the results are more mixed. If hits and false alarms are equally important, lower thresholds improve the performance of the LT-methods to a level comparable to the two-state regime-switching models. However, the economic comparison shows that the false alarms actually become costlier, and the overall balance in terms of economic performance is negative. Nonetheless, lowering the threshold λ_2 makes sense as a switch to this setting commands a positive fee and leads to higher utility. Since utility is still considerably lower than the level produced by the regime-switching models, the latter models lead to forecasts that are economically more valuable.

F.2 Robustness of the PS-method

Pagan and Sossounov (2003) base their algorithm on the algorithm for business cycle identification of Bry and Boschan (1971), and adjust some of the original settings based on early literature on bull and bear markets, the so-called Dow theory after Charles Dow. They do not investigate how robust their findings are to changes in these settings. We conduct a modest robustness analysis, where we change one parameter at a time. In comparison with the LT-method, the PS-method identifies a few more cycles, so we consider only a relaxation of the constraint on cycle length, which we lower to 52 weeks instead of 70. For the constraint on the length of a phase, we consider a lower value of 12 weeks and

a higher value of 20 weeks. A higher value may lead to less false alarms. The bound on price change that must be crossed to overrule the phase constraint is currently at 20%. As in the robustness checks for the LT-method, we consider a value of 15%. Censoring (currently 13 weeks) will not influence the identification much, but may have considerable impact on forecasts. We consider a lower value of 7 weeks and a higher value of 26 weeks, which corresponds with the setting of PS.

The identification is quite robust to these changes. In the basic setting, 18 bull and 18 bear markets result. The results do not change for a change the phase constraint or the price change bound. Censoring 7 weeks at the beginning and end leads to one extra switch to a bear market at the end of of the sample. Censoring 26 weeks does not lead to different results. Lowering the cycle to 52 weeks leads to 19 bull and 19 bear markets. Because results change at most slightly, the further analyses based on identification will also change only slightly. Our conclusion that the PS-method works well for identification remains unaffected.

The forecasts and allocations based on the PS-method change more when the parameters are changed. A forecast for week $t + 1$ based on the PS-method uses only information up to time t . The identification over the subsample until week t may differ from the full-sample identification and may be more sensitive to parameter choices. This larger sensitivity then carries over to the forecasts.

Table F.5 shows a further analysis of the impact of parameter changes. Changes in the degree of censoring leads to forecasts that are quite different from the forecasts in the basic setting, with IADs of around 0.17 when 7 weeks are censored and 0.22 when 26 weeks are censored. Less censoring leads to a more aggressive allocation with weights larger than two, while more censoring leads to a less aggressive allocation with weights close to one. This pattern is caused by forecasts being closer to the steady state distribution of bull and bear markets, when more observations are censored. Effectively, more censoring necessitates a forecast for more steps ahead. More censoring and less aggressive investments lead to a better performance: the mean return increases while the volatility decreases. Less censoring shows the opposite effect. It means that the forecasts are too extreme, and that shrinking them towards their long-term average makes sense. The effect of more censoring is quite impressive with a Sharpe ratio that increases to 0.051 when predictor variables are

used. This ratio is similar in magnitude to the best LT-method in the previous subsection. Moreover, utility increases compared to the basis strategies. The resulting utility is only slightly lower than the utility the results from using two-state regime-switching models. Because more censoring leads to higher utility, an investor is willing to pay significant fees of around 27 basis points to switch from the basic PS-method.

[Table F.5 about here.]

All other parameter choices do not matter much. Changing the phase constraint does not give different results. When we put the constraint at 12 or 20 weeks, the forecasts are exactly the same as in the basic setting. Therefore we have not include the phase constraint in further analysis. A relaxation of the constraints on the duration of a cycle or the price change to trigger a switch do not lead to large statistical or economic differences in Table F.5. The characteristics of the investment strategies are largely similar. Consequently, fees for switching differ only a little bit from zero. The confidence intervals for the IADs and the fees encompass zero, indicating that the differences are not significant.

Table F.1: Number and Duration of Market Regimes for Different Thresholds in the LT-method

		λ_1	λ_2	0.20	0.15
		0.20	0.10	0.15	0.10
bull	number	16	18	18	24
	avg. duration	119	98	108	77
	med. duration	90	59	84	50
	std. dev. duration	97	97	97	90
	max. duration	405	405	405	405
	min. duration	15	15	6	5
bear	number	16	18	18	24
	avg. duration	62	63	53	44
	med. duration	60	55	60	39
	std. dev. duration	44	49	29	30
	max. duration	187	187	91	91
	min. duration	7	7	7	4

This table shows the number of spells of the different market regimes, their average and median duration, the standard deviation of the duration, and its maximum and minimum for different thresholds λ_1 and λ_2 .

Table F.2: Statistical and Economic Comparison of the Identification by the LT-method for Different Thresholds

(a) Integrated Absolute Differences				
(λ_1, λ_2)	(0.20, 0.10)	(0.15, 0.15)	(0.15, 0.10)	
(0.20, 0.15)	0.047 [0.017, 0.102]	0.011 [0.000, 0.059]	0.045 [0.034, 0.112]	
(0.20, 0.10)		0.059 [0.031, 0.126]	0.038 [0.014, 0.087]	
(0.15, 0.15)			0.033 [0.017, 0.089]	

(b) Performance Measures				
(λ_1, λ_2)	(0.20, 0.15)	(0.20, 0.10)	(0.15, 0.15)	(0.15, 0.10)
Mean	0.93	1.00	1.02	1.27
Volatility	4.20	4.35	4.39	4.89
Sharpe ratio	0.22	0.23	0.23	0.26
Utility ($\times 1,000$)	4.63	4.98	5.08	6.37

(c) Fees (in bps)				
(λ_1, λ_2)	(0.20, 0.15)	(0.20, 0.10)	(0.15, 0.15)	(0.15, 0.10)
(λ_1, λ_2)	(0.20, 0.15)	(0.20, 0.10)	(0.15, 0.15)	(0.15, 0.10)
(0.20, 0.15)	0	-3.7 [-14.6, -0.8]	-4.8 [-12.9, 0.0]	-18.4 [-32.4, -11.3]
(0.20, 0.10)	3.7 [0.8, 14.7]	0	-1.0 [-8.6, 9.6]	-14.7 [-24.3, -7.3]
(0.15, 0.15)	4.8 [0.0, 13.0]	1.0 [-9.5, 8.6]	0	-13.7 [-26.3, -6.5]
(0.15, 0.10)	18.6 [11.4, 32.9]	14.8 [7.3, 24.5]	13.8 [6.5, 26.6]	0

This table shows the statistical and economic comparison of the identification that results from the LT-method when different thresholds λ_1 and λ_2 are used. The statistics are calculated as in Tables 1 and 3.

Table F.3: Integrated Absolute Difference of Forecasts based on the LT-method with Different Settings

	(0.20, 0.15, C)	(0.20, 0.15, L)	(0.20, 0.10, C)	(0.20, 0.10, L)	(0.15, 0.15, C)	(0.15, 0.15, L)	(0.15, 0.10, C)
(0.20, 0.15, L)	0.060 [0.057, 0.110]	0.067 [0.035, 0.136]	0.113 [0.096, 0.206]	0.067 [0.032, 0.099]	0.106 [0.097, 0.166]	0.127 [0.092, 0.180]	0.150 [0.139, 0.239]
(0.20, 0.10, C)		0.116 [0.096, 0.185]	0.098 [0.073, 0.181]	0.113 [0.083, 0.162]	0.077 [0.038, 0.119]	0.165 [0.122, 0.207]	0.162 [0.119, 0.217]
(0.20, 0.10, L)			0.063 [0.041, 0.105]	0.128 [0.094, 0.204]	0.161 [0.142, 0.242]	0.080 [0.052, 0.116]	0.116 [0.101, 0.172]
(0.15, 0.15, C)				0.163 [0.121, 0.247]	0.155 [0.108, 0.213]	0.121 [0.089, 0.161]	0.101 [0.065, 0.135]
(0.15, 0.15, L)					0.052 [0.056, 0.104]	0.082 [0.055, 0.134]	0.108 [0.104, 0.192]
(0.15, 0.10, C)						0.118 [0.107, 0.181]	0.109 [0.086, 0.165]
(0.15, 0.10, L)							0.054 [0.055, 0.093]

This table reports the integrated absolute difference between the forecasts of the LT-method constructed with different thresholds and with constant or time-varying transition probabilities. The settings are reported in the headings in parentheses, where the first two elements indicate the values for λ_1 and λ_2 and the third element indicates constant (C) or time-varying (L) transition probabilities. See Table 4 for further explanation.

Table F.4: Performance Measures and Fees when forecasts are based on the LT-method with Different Settings

(a) Performance Measures

(λ_1, λ_2)	(0.20, 0.15)		(0.20, 0.10)		(0.15, 0.15)		(0.15, 0.10)	
Transition	C	L	C	L	C	L	C	L
Av. Abs. Weight	1.82	1.90	1.91	2.04	1.91	1.96	2.15	2.26
Mean	0.17	0.14	0.26	0.27	0.18	0.20	0.23	0.28
Mean Bull	0.17	0.20	0.24	0.29	0.20	0.27	0.19	0.27
Mean Bear	0.15	-0.05	0.31	0.21	0.11	-0.02	0.35	0.29
Volatility	4.50	4.43	4.55	4.53	4.68	4.62	5.22	5.10
Sharpe Ratio	0.037	0.031	0.056	0.060	0.038	0.043	0.044	0.054
Utility ($\times 1,000$)	-3.38	-3.54	-2.62	-2.44	-3.72	-3.37	-4.52	-3.75

(b) Fees (in bps)

(λ_1, λ_2)	(0.20, 0.15)		(0.20, 0.10)		(0.15, 0.15)		(0.15, 0.10)	
Transition	C	L	C	L	C	L	C	L
(0.20, 0.15, C)	0	1.6	-7.7	-9.5	3.4	-0.2	11.4	3.7
		[-10.0, 3.6]	[-11.3, 6.2]	[-16.8, 4.7]	[-3.3, 15.3]	[-11.0, 7.3]	[3.2, 30.8]	[-7.2, 21.0]
(0.20, 0.15, L)	-1.6	0	-9.2	-11.1	1.8	-1.8	9.8	2.1
	[-3.6, 9.9]		[-11.2, 13.2]	[-13.2, 7.5]	[-4.0, 22.7]	[-5.2, 10.3]	[3.5, 39.7]	[-3.9, 26.3]
(0.20, 0.10, C)	7.7	9.2	0	-1.8	11.1	7.5	19.1	11.4
	[-6.2, 11.3]	[-13.3, 11.2]		[-10.7, 1.7]	[-5.5, 21.1]	[-12.0, 13.7]	[5.4, 31.4]	[-3.2, 22.1]
(0.20, 0.10, L)	9.5	11.1	1.8	0	12.9	9.3	20.9	13.2
	[-4.7, 16.8]	[-7.5, 13.3]	[-1.7, 10.7]		[-4.6, 28.4]	[-8.2, 16.8]	[8.6, 37.0]	[2.8, 26.0]
(0.15, 0.15, C)	-3.4	-1.8	-11.1	-12.9	0	-3.6	8.0	0.3
	[-15.3, 3.3]	[-22.9, 4.0]	[-21.2, 5.5]	[-28.6, 4.6]		[-17.0, 2.0]	[-1.4, 25.3]	[-17.2, 18.0]
(0.15, 0.15, L)	0.2	1.8	-7.5	-9.3	3.6	0	11.6	3.9
	[-7.3, 11.0]	[-10.4, 5.2]	[-13.7, 11.9]	[-16.9, 8.2]	[-2.0, 16.9]		[3.8, 35.9]	[-6.0, 23.4]
(0.15, 0.10, C)	-11.5	-9.9	-19.2	-21.1	-8.1	-11.7	0	-7.8
	[-31.2, -3.2]	[-40.4, -3.5]	[-31.8, -5.5]	[-37.7, -8.7]	[-25.5, 1.4]	[-36.4, -3.8]		[-21.4, -0.7]
(0.15, 0.10, L)	-3.7	-2.2	-11.5	-13.3	-0.3	-3.9	7.7	0
	[-21.1, 7.3]	[-26.5, 3.9]	[-22.3, 3.2]	[-26.2, -2.8]	[-18.1, 17.2]	[-23.6, 6.0]	[0.7, 21.3]	

See Table 5 for explanation. The different settings in the LT-method are indicated by the values for λ_1 and λ_2 and the choice for transition probabilities, which can be constant (C) or time-varying (L).

Table F.5: Statistical and Economic Comparison of Forecasts based on the PS-method with Different Settings

Setting	IAD	Av. Abs. Weight	Mean Mean	Mean Bull	Mean Bear	Vol.	Sharpe Ratio	Utility	Fee
Basic, C	-	1.70	0.028	0.019	0.049	3.88	0.007	-3.48	-
Basic, L	-	1.87	0.043	0.108	-0.094	4.17	0.010	-3.91	-
Censoring: 7, C	0.175 [0.158, 0.188]	2.01	-0.102	-0.187	0.072	4.56	-0.022	-6.23	-27.5 [-43.3, -7.5]
Censoring: 7, L	0.169 [0.153, 0.185]	2.13	-0.074	-0.109	-0.001	4.77	-0.015	-6.44	-25.4 [-42.8, -3.8]
Censoring: 26, C	0.212 [0.186, 0.228]	1.20	0.103	0.147	0.006	2.73	0.038	-0.84	26.4 [12.0, 37.6]
Censoring: 26, L	0.222 [0.181, 0.243]	1.42	0.174	0.308	-0.114	3.41	0.051	-1.18	27.3 [3.7, 30.8]
Cycle: 52, C	0.009	1.71	0.030	0.027	0.038	3.89	0.008	-3.48	0.04
Cycle: 52, L	0.028	1.81	0.009	0.078	-0.139	4.12	0.002	-4.15	-2.4
Change: 15%, C	0.001 [0.000, 0.023]	1.70	0.040	0.036	0.049	3.88	0.010	-3.36	1.2 [-0.3, 3.9]
Change: 15%, L	0.001 [0.000, 0.029]	1.87	0.059	0.131	-0.094	4.17	0.014	-3.75	1.6 [-0.4, 5.1]

The column IAD gives the Integrated Absolute Difference between the forecasts based on a specific setting and the basic setting for the PS-method. The next columns subsequently report performance measures of investment strategies that use the different forecasts as input for an asset allocation as discussed in Section 2.5. We report the average absolute weight of each strategy. Based on the realized returns of the asset allocations, we calculate the mean and volatility (in % per week), the Sharpe ratio, and the annualized realized utility as in Equation (15). Mean Bull (Mean Bear) is the annualized mean during the subperiods identified *ex post* as bull (bear) markets. The column fee reports the maximum fee (in basis points) that an investor is willing to pay to switch from a basic setting to another setting in the PS-method. When calculating the IAD or the fee, the setting for the transition probabilities (constant or time-varying) in the basic strategy corresponds with the setting in the alternative method. We report 5% and 95% percentiles between brackets for the IAD and the fee based on 200 bootstrap samples of both the return and the explanatory variable series following Politis and Romano (1994). We do not report confidence intervals for the different cycle restrictions, as changing cycle minima cannot be combined with a stationary bootstrap.

ERIM Report Series <i>Research in Management</i>	
ERIM Report Series reference number	ERS-2013-016-F&A
Date of publication	2013-10-14
Version	14-10-2013
Number of pages	77
Persistent URL for paper	http://hdl.handle.net/1765/41558
Email address corresponding author	kole@ese.eur.nl
Address	Erasmus Research Institute of Management (ERIM) RSM Erasmus University / Erasmus School of Economics Erasmus University Rotterdam PO Box 1738 3000 DR Rotterdam, The Netherlands Phone: +31104081182 Fax: +31104089640 Email: info@erim.eur.nl Internet: http://www.erim.eur.nl
Availability	The ERIM Report Series is distributed through the following platforms: RePub, the EUR institutional repository Social Science Research Network (SSRN) Research Papers in Economics (RePEc)
Classifications	The electronic versions of the papers in the ERIM Report Series contain bibliographic metadata from the following classification systems: Library of Congress Classification (LCC) Journal of Economic Literature (JEL) ACM Computing Classification System Inspec Classification Scheme (ICS)