How to Identify and Predict Bull and Bear Markets?*

Erik Kole[†]

Dick J.C. van Dijk

Econometric Institute, Erasmus School of Economics, Erasmus University Rotterdam

September 30, 2010

Abstract

Characterizing financial markets as bullish or bearish comprehensively describes the behavior of a market. However, because these terms lack a unique definition, several fundamentally different methods exist to identify and predict bull and bear markets. We compare methods based on rules with methods based on econometric models, in particular Markov regime-switching models. The rules-based methods purely reflect the direction of the market, while the regime-switching models take both signs and volatility of returns into account, and can also accommodate booms and crashes. The out-of-sample predictions of the regime-switching models score highest on statistical accuracy. To the contrary, the investment performance of the algorithm of Lunde and Timmermann [Lunde A. and A. Timmermann, 2004, Duration Dependence in Stock Prices: An Analysis of Bull and Bear Markets, *Journal of Business & Economic Statistics*, 22(3):253–273] is best. With a yearly excess return of 10.5% and Sharpe ratio of 0.60, it outperforms the other methods and a buy-and-hold strategy.

1 Introduction

Bull and bear markets are key elements in analyzing and predicting financial markets. Investors who actively manage their portfolios seek to invest in assets with bullish prospects and stay away from assets with bearish prospects, or even to go short in those. To successfully implement such a strategy, they require accurate identification and prediction of bullish and bearish periods. The academic literature does not offer a single preferred method for this purpose. An important reason for this lack of consensus is the absence of a clear definition of bull and bear markets. Bull markets are commonly understood as prolonged periods of gradually rising prices, while bear markets are characterized by falling prices and higher volatility than during bull markets. How large price increases or decreases should be, or how long rising or falling tendencies should last is not uniquely specified.

In this paper we conduct an extensive empirical analysis of the two main types of methods that have been put forward for the identification and prediction of bullish and bearish periods. One type concerns methods based on a set of rules, while the other type makes use of more fully specified econometric models. We compare the two types of methods along several dimensions. First, we examine their identification of bullish and bearish periods in the US stock market. Then we investigate which predictive variables have a significant effect on forecasting switches between bull and bear markets. We consider macro variables related to the business cycle, and financial variables such as the short rate and the dividend yield. Finally, we determine which method works best for an investor who bases her allocation on bull and bear markets. We pay attention to both the statistical accuracy of the predictions and the economic value in terms of the performance of the investment strategy.

From the methods that use a set of rules for identification, we consider the algorithmic methods of Pagan and Sossounov (2003) and Lunde and Timmermann (2004). These methods first determine local peaks and troughs in a time series of asset prices, and then apply certain rules to select those peaks and troughs that constitute genuine turning points between bull and bear markets. They are based on the algorithms used to date recessions and expansions in business cycle research (see Bry and Boschan, 1971, among others), and

have been adapted in different ways for application in financial markets. The main rule in the approach of Pagan and Sossounov (2003) (PS henceforward) is the requirement of a minimum length of bull and bear periods.[1] By contrast, Lunde and Timmermann (2004) (LT from now) impose a minimum on the price change since the last peak or trough for a new peak or trough to qualify as a turning point.[2]

As an alternative to a rules-based approach, we analyze Markov regime-switching models. They belong to the category of methods that are based on a specific model for the data generating process underlying asset prices. To accommodate bullish and bearish periods, these models contain two or more regimes. Within this class, Markov regime switching models pioneered by Hamilton (1989, 1990) are most popular. The regime process is latent and follows a first order Markov chain. Empirical applications typically distinguish two regimes with different means and variances and normally distributed innovations.[3] The bull (bear) market regime exhibits a high (low or negative) average return and low (high) volatility. The number of regimes can easily be increased to improve the fit of the model (see Guidolin and Timmermann, 2006a,b, 2007) or to model specific features of financial markets such as crashes (see Kole et al., 2006) or bull market rallies (see Maheu et al., 2009). Other regime switching models such as threshold autoregressive models can be applied as well (see, e.g., Coakley and Fuertes, 2006).

The difference between these two categories is fundamental. The rules-based approaches are typically more transparent than the model-based methods. The identification based on the best statistical fit can be more difficult to grasp than that based on a set of rules. On the other hand, a full-blown statistical model offers more insight into the process under scrutiny and its drivers. It shows directly what constitutes a bull or a bear market. As a second difference, the rules-based methods require some arbitrary or subjective settings that possibly affect the outcomes. The regime switching models let the data decide, or offer statistical techniques to evaluate settings as, for example, the number of regimes. As a final

---

[1]See Edwards et al. (2003); Gómez Biscarri and Pérez de Gracia (2004); Candelon et al. (2008); Chen (2009) and Kaminsky and Schmukler (2008) for applications.

[2]Chiang et al. (2009) adopt this method.

[3]See 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.

difference, the regime switching models can treat identification and prediction in one go, while making predictions with the rules-based methods always follows as a separate second step. Jointly handling identification and prediction offers gains in statistical efficiency.

We devise a new statistical technique, the Integrated Absolute Difference (IAD), to compare the identification and predictions that result from the different methods. It is suitable for the binary identification of the rules-based methods, but also works for the probabilities with which the regime switching models identify regimes. For predictions, all methods produce probabilities for each state. The IAD is closely related to the Integrated Square Difference of Pagan and Ullah (1999) and Sarno and Valente (2004), but is easier to interpret as a difference in probability. We show how this technique can handle the complication of the true sequence of bull and bear markets being hidden.

A comparison of the identification resulting from the different methods for the period January 1980 – July 2009 shows that the two rules-based approaches are largely similar with IADs close to zero, and purely reflect the recent direction of the stock market. To the contrary, regime switching models take a risk-return trade-off into account. High expected returns and low volatility characterize bullish periods, while low means and high volatilities typify bear markets. Consequently, some periods that are considered bullish by the rules-based approaches as the market goes up, may be identified as bearish by the regime switching approach because the volatility is high. Regime switching models with four regimes show the added value of explicitly including crash and boom states. Compared with the two-state case, this model can better accommodate brief crashes during bull markets, or booms during bear markets.

When it comes to predicting bullish and bearish periods, differences between the methods are larger. We evaluate several investment strategies, using means, variances or sign forecasts. The performance of the LT-method stands out, whereas the differences between the others methods are smaller. Over the period July 1994 – June 2009, all strategies based on the LT-method beat the benchmark of a buy-and-hold strategy. The former yield excess returns of 6.6% up to 15.1% per year, and Sharpe ratios ranging from 0.38 to 0.6, compared to an average excess return of 2.4% per year and a Sharpe ratio of 0.14 for the benchmark. These dynamic strategies produce substantial economic value, since an investor would be willing to pay fees ranging from 4.1% to 12.3% per year to switch to them from the buy-

4

and-hold strategy. The highest Sharpe ratio and fee for the PS-method equals 0.26 and 3.1%, for the regime-switching models with two and four states they equal 0.21 and 1.2%. However, for some investment strategies the PS and regime-switching methods perform worse than the benchmark, and command negative fees.

Our results show that quickly picking up bull-bear changes is crucial for successfully predicting bull and bear markets. Bullish and bearish periods are highly persistent, so the sooner a switch is identified, the larger the gains. All methods identify switches with some delay, but the regime switching models are fastest in signalling switches. However, they do not warn against small negative returns, which is why they do not outperform the benchmark. The LT-method identifies a bull-bear (bear-bull) switch only after a decrease (increase) of 15% (20%) in the stock index. Though this may take some time (several weeks up to half a year), it is still fast enough to make a profit. The PS-method rapidly picks up switches, but produces many false alarms.

The use of financial and macro variables has mixed effects on the quality of the predictions. We use a specific-to-general selection procedure to include predictive variables. For the rules-based approaches their use consistently lowers performance, whereas performance improves when predictive variables are included in the transition probabilities of the regime-switching models (see Diebold et al., 1994). This result indicates that directly including predictive variables in a model, which preserves the latent nature of the bull-bear process, is preferable to treating the bull-bear process as observed.

Our research 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 dating rules to classify months as a recession or expansion, while Hamilton proposes regime switching models. In the dating of recessions and expansions, both methods base their identification mainly on the sign of GDP growth and produce comparable results. For dating bull and bear periods in the stock market by regime switching models, the volatility of recent returns seems at least as important (if not more) than their sign. Consequently, their identification differs substantially from the rules-based approaches. Since price increases are necessary for a profitable active management strategy, focussing purely on the recent tendency leads to better results than combining it with the volatility of returns.

We also add to the discussion on predictability in financial markets. We extend the analysis of Chen (2009) in several ways. First, we consider the dynamic combination of more predictive variables. Second, we include predictive variables directly in the regime switching models and do not need Chen (2009)'s two-step procedure. He treats the smoothed inference probabilities as observed dependent variables in a linear regression, which does not take their probabilistic nature into account. Our results for the rules-based approaches show that the in-sample added value of the predictive variables is not met with out-of-sample quality. Strategies with predictive variables perform worse than those without. For the regime switching models, we find quite some variation in the selected variables and their coefficients. Taken together, these results are in line with those documented by Welch and Goyal (2008) for direct predictions of stock returns. The added value of predictive variables in the regime switching model fits in with the discussion of sign predictability and volatility persistence in Christoffersen and Diebold (2006).

This article is structured as follows. In Section 2 we introduce the data. Section 3 discusses the different methods for identifying and predicting bull and bear markets. In Section 4 we propose distance measures to determine the differences between the competing methods. Section 5 analyses the identification results from applying the different approaches to the full sample. In Section 6 we assess the performance of the different methods when their out-of-sample predictions are used in an investment strategy. Section 7 concludes.

2 Data

2.1 Stock market data

When an investor speculates on the direction of the stock market, her natural benchmark is a riskless investment. This implies that bullish or bearish periods should be determined with respect to a riskless bank account B_t, which earns the continuously compounded risk-free interest rate r_τ^{f} over period τ. Starting with $B_0 = 1$, the value of this bank account

obeys

$$B_t = \exp \sum_{\tau=0}^{t-1} r_\tau^{\mathrm{f}}. \tag{1}$$

The investor considers a stock market index P_t relative to this benchmark and focuses on the series

$$\tilde{P}_t = P_t/B_t. \tag{2}$$

The log return on this index gives the return on an investment in the stock market 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. Futures contracts are the natural asset to speculate on the direction of the stock market, as they are cheap and easily available. Studying the excess market index $\tilde{P}_t$ thus corresponds directly with the return on an investment opportunity.

Our analysis considers the US stock market, proxied by the MSCI price index on a weekly frequency. For the risk-free rate we use the Financial Times / ICAP 1-Week Euro rate. Our data series start on December 26, 1979 and end on July 1, 2009. All data series are obtained from Thompson Datastream.

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 predicting 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 26/12/1979. The graph exhibits the familiar financial landmarks of the last 30 years, i.e., the slump during 1981-1982, the crash of 1987 and the IT boom and bust around 2000, and the big drop during the credit crunch in 2008. At first sight, the periods December 1980–August 1982, March 2000–October 2002 and January 2008–March 2009 qualify as bear markets. A more detailed inspection shows several other shorter periods with sustained declines in stock prices, i.e., July 1983–July 1984, June 1990–January 1991 and July 1998–October 1998. In the next section we examine how the different methods handle these periods.

[Figure 1 about here.]

2.2 Predicting variables

We consider macro-economic and financial variables to 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. Hamilton and Lin (1996), Avramov and Wermers (2006) and Beltratti and Morana (2006) use business cycle variables like industrial production. Ang and Bekaert (2002) show the added value of the short term interest rate. Avramov and Chordia (2006) provide evidence favoring the term spread and the dividend yield. Chen (2009) considers a wide range of variables with the term spread, the inflation rate, industrial production and change in unemployment being the most successful.

We join this literature and gather data accordingly. We construct monthly inflation rates based on the seasonally adjusted consumer price index from the FRED database of the Federal Reserve Bank of St. Louis.[4] From the same database, we take the three-month Treasury Bill rate, the 10-year government bond yield and Moody's AAA and BAA corporate bond yields. We construct the yield spread as the difference between the government bond yield and the treasury bill rate. The difference between the BAA and AAA yields produces the credit spread. The trade weighted exchange rate is also taken from FRED. From the International Financial Statistics Database (IFS) of the IMF, we use the unemployment rate (code USI67R), and industrial production (volume based, not seasonally adjusted, code USI66..IG). The dividend yield has been provided by Thompson DataStream.

To ensure stationarity, we transform some of the predictive variables. The T-Bill rate and the dividend yield exhibit a unit root and show a downward sloping pattern over most of our sample period. We construct a stationary series by subtracting the prior one-year average from each observation, used more often in forecasting (see e.g., Campbell, 1991; Rapach et al., 2005). We apply the same transformation to the trade weighted exchange rate. For the unemployment rate we construct yearly differences. We transform

---

[4]Series ID CPIAUCSL, see `https://research.stlouisfed.org/fred2/` for more information.

the industrial production series to yearly growth rates. We do not transform the inflation, the yield spread or the credit spread series. 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 be compared directly. We provide summary statistics on the predictive variables in Appendix B.1.

3 Rules or regime switching: theory

Both the rules-based approaches and regime switching models aim at identifying the state of the equity market. At each point in time, the market is in a specific state S_t^m. We add a superscript m to indicate that the states can depend on the selected method. When rules are applied, the number of states typically equals two, i.e. a bull state and a bear state. In a regime switching approach, more states can be introduced, for example to capture sudden booms and crashes as in Guidolin and Timmermann (2006b, 2007) and Kole et al. (2006) or bear market rallies and bull market corrections as in Maheu et al. (2009). The rules-based and regime switching methods propose different ways to make inferences on the prevailing regime at time t. In both methods we relate the occurrence of a specific state to a set of predictive variables z_{t-1}, which are lagged one period to enable prediction. This section discusses the methods, their similarities and differences.

3.1 Identification and prediction based on rules

We consider two sets of rules that have been put forward in the literature to identify bull and bear markets. Our first set has been proposed by Lunde and Timmermann (2004, LT henceforward). In their approach, investors use peaks and troughs in the stock market index to define bullish (between a trough and the subsequent peak) and bearish (between a peak and the subsequent trough) periods. A bull market occurred if the index has increased by at least a fraction λ_1 since the last trough. A bear market occurred if the index has decreased by at least a fraction λ_2 since the last peak. To identify peaks and troughs in a time series, the investor uses an iterative search procedure that starts with a peak or

trough. The identification rules can be summarized as follows:

1. The last observed extreme value was a peak with index value P^{max}. The investor considers the subsequent period.

   (a) If the index has exceeded the last maximum, the maximum is updated.

   (b) If the index has dropped with a fraction λ_2, a trough has been found.

   (c) If neither of the conditions is satisfied, no update takes place.

2. The last observed extreme value was a trough with index value P^{min}. The investor considers the subsequent period.

   (a) If the index has dropped below the last minimum, the minimum is updated.

   (b) If the index has increased with a fraction λ_1, a peak has been found.

   (c) If neither of the conditions in satisfied, no update takes place.

After these decision rules the investor considers the next period.

We follow LT by setting $\lambda_1 = 0.20$ and $\lambda_2 = 0.15$. This implies that an increase of 20% over the last trough signifies a bull market, and that a decrease of 15% since the last peak indicates a bear market. To commence the search procedure we determine whether the market is initially bullish of bearish. 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.

The second approach we investigate has been put forward by Pagan and Sossounov (2003, PS henceforward). Their approach is based 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. However, their identification is quite different from the approach taken by Lunde and Timmermann (2004). As the main difference, PS do not impose requirements on the magnitude of the change of the index during bull or bear markets, but instead put restrictions on the minimum duration of phases and cycles. In the first step, all local maxima and minima are located. A price constitutes a local maximum (minimum) if it is higher (lower) than all prices in the past

and future τ_{window} periods. This step can produce a series of subsequent peaks or troughs. In the second step, an alternating sequence of peaks and troughs is constructed, consisting of the highest maxima and lowest minima. Next, peaks and troughs in the first and last τ_{censor} periods are censored. Fourth, cycles of bull and bear markets that last less than τ_{cycle} periods are eliminated. Fifth, a bull market or bear market that lasts less than τ_{phase} periods is eliminated, 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 have $\tau_{\text{window}} = 32$, $\tau_{\text{cycle}} = 70$, $\tau_{\text{phase}} = 16$ 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 time. Since we will use this information in making predictions, we use a shorter period of 13 weeks to establish the initial and the ultimate state of the market.

The next step is to relate the resulting series of bull and bear states to a set of explanatory variables, $\boldsymbol{z}_{t-1}$. We code bull markets as $S_t^m = \text{u}$ and bear markets as $S_t^m = \text{d}$. Since the dependent variable is binary, a logit or probit model can be used. We opt for a logit model, as this model can be easily extended to a multinomial logit model when more states are present. We adjust the standard logit model such that the effect of an explanatory variable on the probability of a future state can depend on the current state. Some macro-finance variables may help predicting a switch from a bear market, but not from a bull market, or may have a different effect on the probability. The probability for a bull state to occur at time t is modeled as

$$\pi_{qt}^m \equiv \Pr[S_t^m = \text{u}|S_{t-1}^m = q, \boldsymbol{z}_{t-1}] = \Lambda(\boldsymbol{\beta}_q^{m\prime}\boldsymbol{z}_{t-1}), \ m = \text{LT, PS} \tag{3}$$

where $\Lambda(x) \equiv 1/(1 + \text{e}^{-x})$ denotes the logistic function, and $\boldsymbol{\beta}_q^m$ is the coefficient vector on the $\boldsymbol{z}_{t-1}$ variables, which depends on the previous state of the market q. For notational convenience, we assume the first variable in $\boldsymbol{z}_{t-1}$ is a constant to capture the intercept term. We call this model a Markovian logit model, as it combines a logit model with the Markovian property that the probability distribution of the future state S_{t+1}^m is (partly) determined by S_t^m. If the coefficient $\boldsymbol{\beta}_q^m$ does not depend on q, a normal logit model results. If only a constant is used, the market state process is a standard stochastic process with

11

the Markov property.

To form the one-period ahead prediction for π_{T+1}^m, the prevailing state at time T is needed. For the rules-based approaches, this information may not be available. In the LT-approach, only if P_T equals the last observed maximum (minimum), and is a fraction λ_1 above (λ_2 below) the prior minimum is the market surely in a bull (bear) state. The PS-alogirthm suffers from this problem too, since only the state up to the last 13 weeks is known. So, the market may already have switched, but this will only become obvious later. In that case, the state of the market is known until the period of the last extreme value, which we denote with $T^* < T$. We construct the one-period ahead prediction in a recursive way

$$
\begin{aligned}
\Pr[S_{t+1}^m = s | \boldsymbol{z}_t] = \Pr[S_{t+1}^m = s | S_t^m = \mathrm{u}, \boldsymbol{z}_t] \Pr[S_t^m = \mathrm{u} | \boldsymbol{z}_{t-1}] + \\
\Pr[S_{t+1}^m = s | S_t^m = \mathrm{d}, \boldsymbol{z}_t] \Pr[S_t^m = \mathrm{d} | \boldsymbol{z}_{t-1}], \quad T^* < t \leq T+1. \quad (4)
\end{aligned}
$$

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. This iteration stops at $T + 1$.

3.2 Identification and prediction by regime-switching models

We also consider a method for identifying and predicting bull and bear markets that is fundamentally different from the algorithms considered in the previous section. Instead of applying a set of rules to a given series, we now first write down a model that can be the data generating process of a stock market index that allows for prolonged bullish and bearish periods. Estimating such a model produces probabilistic inferences on periods of bull and bear markets in a certain index.

Using such a model-based approach has several advantages. First, it offers more insight into the process under study. We can derive theoretical properties of the model and see whether it yields desirable features. Second, we can easily extend the number of states in the model. We can test whether such extensions imply 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 may not be straightforward. The advantages come

at the cost of misspecification risk. In particular (missed or misspecified) changes in the data generating process can have severe impact on the results. As rule-based approaches do not make strict assumptions on distributions or on the absence or presence of variation over time, they may be more robust.

We consider four Markov chain regime switching models for the stock market, having either two or four regimes (suffix 2 or 4) and having either constant or time-varying probabilities (suffix C or L). For example, the label RS4L means a Markov regime switching model with four states and time-varying transition probabilities.

In the two-regime case, the set of states comprises a bull and a bear state (again denoted by u and d). In both states the excess index return r_t obeys a normal distribution, with mean and variance that depend on the nature of the state:

$$r_t = \begin{cases} r_{\mathrm{u}}, \ r_{\mathrm{u}}^m \sim N(\mu_{\mathrm{u}}^m, \omega_{\mathrm{u}}^m) & \text{if } S_t^m = \mathrm{u} \\ r_{\mathrm{d}}, \ r_{\mathrm{d}}^m \sim N(\mu_{\mathrm{d}}^m, \omega_{\mathrm{d}}^m) & \text{if } S_t^m = \mathrm{d}, \end{cases} \qquad m = \text{RS2C, RS2L}. \tag{5}$$

For the regime switching models with four states, we extend the two-states models with a boom (denoted by b) and a crash state (denoted by c). The full specification of the excess return on the market index reads

$$r_t = \begin{cases} r_{\mathrm{u}}^m, \ r_{\mathrm{u}}^m \sim N(\mu_{\mathrm{u}}^m, \omega_{\mathrm{u}}^m) & \text{if } S_t^m = \mathrm{u} \\ r_{\mathrm{d}}^m, \ r_{\mathrm{d}}^m \sim N(\mu_{\mathrm{d}}^m, \omega_{\mathrm{d}}^m) & \text{if } S_t^m = \mathrm{d} \\ r_{\mathrm{b}}^m = l_{\mathrm{b}} + e^{x_{\mathrm{b}}}, \ x_{\mathrm{b}} \sim N(\mu_{\mathrm{b}}^m, \omega_{\mathrm{b}}^m) & \text{if } S_t^m = \mathrm{b} \\ r_{\mathrm{c}}^m = u_{\mathrm{c}} - e^{x_{\mathrm{c}}}, \ x_{\mathrm{c}} \sim N(\mu_{\mathrm{c}}^m, \omega_{\mathrm{c}}^m) & \text{if } S_t^m = \mathrm{c}, \end{cases} \qquad m = \text{RS4C, RS4L}. \tag{6}$$

If a bull or bear state prevails, any return on the real line can be realized. A return during a boom state should be big and positive, and therefore we use shifted lognormal distribution with lower bound $l_{\mathrm{b}} > 0$. Since crashes constitute by definition big negative returns, we model crash returns by a mirrored and shifted log-normal distribution, with upper bound u_{c}.

Our approach explicitly constructs the boom (crash) state as return distribution with a specific lower (upper) bound. We deviate from from Guidolin and Timmermann (2006b, 2007), who allow a maximum of six regimes with each a different normal distribution,

and interpret the regimes based on the estimated means and variances. We impose slightly more structure to ensure that a crash regime can only mean losses, and a boom regime only implies gains. Moreover, we use this structure later on to model the transition probabilities.

Our four-states specification also differs from Maheu et al. (2009). These authors allow for bear markets that can exhibit short rallies and bull markets that can show brief corrections. They enable 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 identification, though the added value for prediction is less obvious. The difference in the predicted return distributions between a bull market and a bear market rally is likely to be less than this difference between a bull market and a boom state as in our specification.

Since the actual state of the market is not directly observable, we treat it as a latent variable that follows a first order Markov chain with transition matrix $\boldsymbol{P}^m$. For the models with two regimes, the transition matrix contains two free parameters $\pi_{qt}^m \equiv \Pr[S_t^m = \mathrm{u}|S_{t-1}^m = q, \boldsymbol{z}_{t-1}]$, as they depend on the departure state that can be bullish, $q = \mathrm{u}$, or bearish, $q = \mathrm{d}$. Obviously, in the model with constant transition probabilities, they do not depend on $\boldsymbol{z}_{t-1}$. In the model with time-varying transition probabilities, we use again a logit transformation to link them to predicting variables $\boldsymbol{z}_{t-1}$

$$\pi_{qt}^m = \Lambda(\boldsymbol{\beta}_q^{m\prime}\boldsymbol{z}_{t-1}), \ m = \mathrm{RS2L}. \tag{7}$$

This specification is mathematically similar to the logit models for the rules based approaches in Eq. (3), though it is an integrated part of the regime switching model.

When the Markov switching model has four states, the transition matrix contains parameters

$$\pi_{sqt}^m \equiv \Pr[S_t = s|S_{t-1} = q, \boldsymbol{z}_{t-1}], \ s, q \in \mathcal{S}^m, \ m = \mathrm{RS4C}, \mathrm{RS4L}, \tag{8}$$

where $\mathcal{S}^m = \{\mathrm{u}, \mathrm{d}, \mathrm{b}, \mathrm{c}\}$ denotes the set of states. Of course, the restriction $\sum_{s \in \mathcal{S}^m} \pi_{sqt}^m = 1$ applies. In the model with constant transition probabilities, RS4C, this restriction leaves 12 free parameters to be estimated. If the probabilities are time-varying (model RS4L), we use a multinomial logit model

$$\pi_{sq,t}^m \equiv \Pr[S_t^m = s|S_{t-1}^m = q, \boldsymbol{z}_{t-1}] = \frac{e^{\boldsymbol{\beta}_{sq}^{m\prime}\boldsymbol{z}_{t-1}}}{\sum_{\varsigma \in \mathcal{S}} e^{\boldsymbol{\beta}_{\varsigma q}^{m\prime}\boldsymbol{z}_{t-1}}}, \ s, q \in \mathcal{S}^m, \ m = \mathrm{RS4L} \tag{9}$$

with $\exists s \in \mathcal{S} : \boldsymbol{\beta}_{sq} = \mathbf{0}$ to ensure identification.

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 resulting regime switching model by means of the EM-algorithm of Dempster et al. (1977). To determine the optimal parameters describing the distribution per state, we follow the standard textbook treatments (e.g., Hamilton, 1994, Ch. 24). In appendix A we extend the method of Diebold et al. (1994) to estimate the parameters of the multinomial logit model.

3.3   Variable selection

We consider in total eight variables that can help predicting the future state of the stock market. Not all these variables might be helpful in predicting specific transitions. Therefore, we propose a specific-to-general procedure for variable selection. In both the rules-based and the regime switching approaches we start with a model with only constants included. Next, we calculate for each variable and transition combination the improvement its inclusion would yield in the likelihood function. We select the variable-transition combination with the largest improvement and test whether this is significant with a likelihood ratio test. If the improvement is significant, we add the variable to our specification for that specific transition and repeat the search procedure with the remaining variables-transition combinations. 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 RS4L-model, we would need to estimate a model with $4 \cdot 3 \cdot 13 = 156$ 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   Comparing two filters

An important aim of this paper is to analyze how different the results of the different approaches are. Though the design of the different methods is considerably different, the results can still be quite the same. In this section we propose a theoretical framework to compare them. The comparison can concentrate on the identification that the different algorithms produce or on the predictions that they make. We discuss both.

The different approaches that we apply in this paper can all be seen as filters. Each algorithm m applies a filter $F^m(t, \Omega)$ to an information set Ω to determine the likelihood of each state s at a point in time t. The information set typically contains a prices series, a set of explanatory variables and a set of coefficients. So, we can interpret F^m as a function on Ω that yields a vector of probabilities. If the set Ω contains the available information up to time $t-1$, this likelihood can be interpreted as a forecast probability. If the information set comprises all available information, denoted by Ω_T, the likelihood corresponds with identification, and we call it an inference probability. In case of the rules based approaches, the state at time t is identified as either bullish or bearish, so $\boldsymbol{p}^i_{m,t} = (1,0)'$ or $(0,1)'$ for $m = \text{LT}, \text{PS}$. If regime switching models are used, the identification comes from the smoothed inference probabilities (see Hamilton, 1994, Ch. 22).

Comparing the results of two different filters is equivalent to comparing the two resulting probability vectors. So, we should compare probability vectors $\boldsymbol{p}$ and $\boldsymbol{q}$, both of size n. As a first step, we define a distance measure $d : [0,1]^n \to \mathcal{R}^+$. Specifically, we consider the L1-norm, based on absolute difference

$$d^{\text{L1}}(\boldsymbol{p}, \boldsymbol{q}) = \sum_{s=1}^{n} |p_s - q_s|. \tag{10}$$

Of course, we can only compare two filters, if their states correspond. For instance, we can compare the outcomes of the RS2C-model with the RS2L-model and the LT04-filter, but not with those of the RS4L model. We cannot measure the difference between p_s and q_s 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 rules-based approaches.

The difference between $\boldsymbol{p}$ and $\boldsymbol{q}$ that we observe varies over time. This variation can

come from different realizations of the latent state S_t, and since S_t is latent also from variations in the true probability of each state ϕ_s. Therefore, we want to determine the expected distance between $\boldsymbol{p}$ and $\boldsymbol{q}$. First, we condition on the true probability of each state. Since the distribution of the state follows a categorical distribution, the expected absolute distance equals

$$\mathrm{E}\big[d^{\mathrm{L1}}(\boldsymbol{p}, \boldsymbol{q})|\boldsymbol{\phi}\big] = \sum_{s=1}^{n} \phi_s |p_s - q_s|. \tag{11}$$

Integrating over all possible values for ϕ_s produces the unconditional expected value

$$\mathrm{E}\big[d^{\mathrm{L1}}(\boldsymbol{p}, \boldsymbol{q})\big] = \int_{[0,1]^n} \sum_{s=1}^{n} \phi_s |p_s - q_s| \, \mathrm{d}G_{\boldsymbol{\phi}}, \tag{12}$$

where $G_{\boldsymbol{\phi}}$ is the density function associated with $\boldsymbol{\phi}$. The expected value will equal zero if $\boldsymbol{p} = \boldsymbol{q}$ with almost certainty. Its upper bound equals one. The expression can be interpreted as an integrated absolute difference, similar to the integrated square difference in Pagan and Ullah (1999) and Sarno and Valente (2004). For the binomial case the above expression simplifies to $\mathrm{E}\big[d^{\mathrm{L1}}(\boldsymbol{p}, \boldsymbol{q})\big|\boldsymbol{\phi}] = \phi_1|p_1 - q_1| + \phi_2|p_2 - q_2| = \phi_1|p_1 - q_1| + (1 - \phi_1|1 - p_1 - (1 - q_1)| = |p_1 - q_1|$.

For a given sample of probabilities $\{\boldsymbol{p}_t\}_{t=1}^{T}$ and $\{\boldsymbol{q}_t\}_{t=1}^{T}$ we can estimate the expectation in Eq. (12) by its sample equivalent

$$\widehat{d^{\mathrm{L1}}}(\boldsymbol{p}, \boldsymbol{q}) = \frac{1}{T} \sum_{t=1}^{t} \sum_{s=1}^{n} \phi_{st} |p_{st} - q_{st}|. \tag{13}$$

In the binomial case, the unobserved ϕ_{st} is irrelevant as $\mathrm{E}\big[d^{\mathrm{L1}}(\boldsymbol{p}, \boldsymbol{q})\big|\boldsymbol{\phi}] = |p_1 - q_1|$ and we can simply calculate the average absolute difference over p_{1t} and p_{2t}. In the multinomial case we can circumvent the observations $\phi_{t,s}$ by assuming that either $\boldsymbol{p}$ or $\boldsymbol{q}$ equals the true probability vector $\boldsymbol{\phi}$. $\widehat{d^{\mathrm{L1}}}(\boldsymbol{p}, \boldsymbol{q})$ will depend on the choice for $\boldsymbol{\phi}$, in a similar way as the Kullback-Leibler divergence (see Kullback and Leibler, 1951). As in Sarno and Valente (2004), bootstraps can be used to determine confidence intervals for $\widehat{d^{\mathrm{L1}}}(\boldsymbol{p}, \boldsymbol{q})$.

We can also use this approach to compare the probabilities p_s and q_s for a specific state s. When we focus on a specific state s we actually simplify the set of states to s and all

other states. So, we end up in a binomial case and have

$$
\begin{aligned}
\mathrm{E}\big[d^{\mathrm{L1}}(p_s, q_s)\big] &= \int_{[0,1]} \mathrm{E}\big[d^{\mathrm{L1}}((p_s, 1-p_s)', (q_s, 1-q_s)')|(\phi_s, 1-\phi_s)'\big]\, \mathrm{d}G_{\phi_s} \\
&= \int_{[0,1]} |p_s - q_s|\, \mathrm{d}G_{\phi_s}.
\end{aligned}
\tag{14}
$$

We adjust the distance measure slightly when we compare the inference probabilities from a rules-based algorithm to probabilities from a regime-switching model. In the former case the inference probability p_s is either 0 or 1, while q_s can take all values between zero and one in the latter. However, when p_s is one and $q_s > 1/2$, the inference of the two approaches is as close as possible. In that case we would like to have a zero distance, which means replacing p_s by q_s. By a similar logic, we replace p_s by $1 - q_s$ when p_s is one and $q_s \leq 1/2$. Together, it means we replace the L1-norm by the function

$$
\tilde{d}^{\mathrm{L1}}(p_s, q_s) =
\begin{cases}
0 & \text{if } (p_s = 1 \wedge q_s > 1/2) \text{ or } (p_s = 0 \wedge q_s < 1/2) \\
|1 - 2q_s| & \text{if } (p_s = 1 \wedge q_s \leq 1/2) \text{ or } (p_s = 0 \wedge q_s \geq 1/2)
\end{cases}
\tag{15}
$$

5 Full sample results

In this section, we show the results from the different approaches when we use the full sample of data that we have available. We build up this section as follows logically from the regime switching models. So, we first consider the estimated parameters for the case of constant and of time-varying transition probabilities and after that the identifications and their comparison. For the rules-based approach, identification is actually the first step. After identification, we estimate means, volatilities, and parameters for the transition probabilities. For the purpose of comparison we discuss these implications of rules-based identification first.

5.1 Moments per regime and transition probabilities

Table 1 shows the means and volatilities of the different regimes under the different approaches. In the case of the rules-based approaches, the means during bull and bear regimes are quite distinct. Bull regimes show an average increase of 0.34–0.35% per week, but during bear markets the decrease is on average 0.70–0.78%. Volatilities are lower during bull

markets (around 1.95%) than during bear markets (around 2.80%). When we use a regime switching model with two states and constant transition probabilities, we also see higher means and lower volatilities during bull markets compared to bear markets. However, the difference in means is less pronounced (0.16% bull versus -0.30% bear), while the difference in volatilities has increased (1.56% bull versus 3.41%). Standard errors indicate that all means differ from zero for a confidence level of 90%. The estimates change only slightly when we include time-variation in the transition probabilities.. Because the regime switching models make inferences based on the distribution function, both means and volatilities matter. The rules-based approaches ignore volatility. Therefore, the differences in volatility are higher for the regime switching model with lower standard errors, but smaller for the means.

[Table 1 about here.]

The last two columns of Table 1 show the estimates for the regime switching model with four states, as specified in (6). Besides a regular bull and bear regime based on normal distributions, this model contains a boom and a crash regime, based on log-normal distributions. We put the lower bound for booms at 2%, and the upper bound for crashes at -2%. In Table 1 we report the means and volatilities for these states as implied by their parameter estimates. As can be expected, crashes imply large losses of on average 4.09% per month with a volatility of 2.31%. Booms on the other hand yield large expected gains of 3.88% per month with a volatility of 0.67%. The difference in volatility of booms and crashes makes losses exceeding 4.09% much more likely than gains exceeding 3.88%. This gives rise to the familiar skewness of market returns. Compared to the RS2C-model, the presence of a crash regime leads to a slightly higher mean estimate for the bear regime of -0.15% and a lower volatility estimate of 3.25%. For the bull regime we also observe a smaller volatility, but a slight increase in the mean. Overall, we see that the extra regimes lead to higher standard errors. When we accommodate time-variation in the transition probabilities, we see some small changes compared to the RS4C-case. The volatility of the crash regime increases, but it decreases for all other regimes. The mean decreases for the crash regime and the bull regime, but increases for the bear and boom regime. As a consequence, the difference between the bull and bear regimes becomes smaller but the

differences between all other regimes become larger. Based on the implied moments of the RS4-models, we conclude that the additional crash and boom states are clearly identified. We consider the statistical improvement later in this section.

As well as the average returns and volatilities in different market states, the sequence of the states is relevant for investors. In Table 2, we consider the transition probabilities under the assumption that they are constant over time. We also calculate the unconditional probabilities $\bar{\pi}_{\mathrm{u}}^m = \Pr[S_t^m = \mathrm{u}]$ and $\bar{\pi}_{\mathrm{d}}^m = \Pr[S_t^m = \mathrm{d}]$. These probabilities satisfy $\bar{\boldsymbol{\pi}}^m \boldsymbol{P}^m = \bar{\boldsymbol{\pi}}^m$. When the market can switch between two states, both states are quite persistent. With probabilities of around 0.95 or higher, the current state prevails for another week. We see that bull states tend to be slightly more persistent than bear states. As a consequence, the unconditional probability for a bull state of around 0.70 exceeds that for a bear state. So, the two-state approaches identify approximately 70% of the observations as bullish and 30% as bearish. The differences between the LT and PS approaches or the RS2C-model are small. So, while we find quite some differences in the moments, the sequences of bull and bear markets under the different approaches seem to share common regularities.

When four regimes are possible as in the RS4C-model, the transition probabilities show quite a different picture. Persistence of the bull and bear regimes decreases a bit, but remains high (0.948 for the bull regime, 0.931 for the bear regime). Most remarkable is the clear pattern the probabilities imply. If the state process leaves the bull state, it moves to crash state and big losses occur. When leaving the bear state, again the crash state is most likely, though a switch to the boom state is also possible. From the crash state, a switch to the bull state or bear state is most likely (probabilities 0.432 and 0.378), though another crash or a rebound by the boom state can also occur (probabilities 0.086 and 0.104). After the boom state, the state process switches to the bull state with probability 0.887 or to the crash state with probability 0.113. So a crash is a strong signal of the end of a current bull or bear market. Unfortunately, the market can move in any direction after a crash. The boom state signals a likely switch to the bull state, though the risk of a crash is present. Direct switches from bull to bear markets are not likely at all, so extreme positive or negative returns on financial markets are a clear signal of upcoming changes. The unconditional probabilities indicate that bullish periods remain dominant, and prevail approximately 70% of the time. Some periods that were earlier identified as bearish are

20

now identified as a boom (2.0%) or a crash (4.3%). Compared to the RS2C-model, the unconditional probability of a bear market decreases from 0.308 to 0.236.

[Table 2 about here.]

As an alternative to constant transition probabilities, we link the probabilities to predicting variables, which makes them time-varying. In the rules-based approaches, we first use the LT- or PS-algorithm to label periods as bullish or bearish. We then use these labeled periods as input for the estimation of the Markovian logit model in (7). The two-state regime switching model uses the same logistic transformation to link the transition probabilities to predicting variables. For the four-state regime switching model, the logistic transformation is extended to the multinomial logistic transformation in (9). We determine the variables to include for specific departure-destination combinations by the specific-to-general procedure proposed in Section 3.3. We use a significance level for the likelihood ratio test of 10%.

The (multinomial) logit transformation we apply is non-linear, which complicates a direct interpretation of the economic effect of the predicting variables. Therefore, we calculate the marginal effect for each variable, evaluated at a specified values for $\boldsymbol{z}$. When logit transformations $\pi = \Lambda(\boldsymbol{z})$ are used, the marginal effect of variable 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 A. As a reference point for the marginal effect, we use the average forecast probability

$$\bar{\pi}_{sq} = \frac{\sum_{t=1}^{T} \Pr[S_{t+1} = s|S_t = q, \boldsymbol{z}_{t-1}] \Pr[S_t = q|\Omega_t]}{\sum_{t=1}^{T} \Pr[S_t = q|\Omega_t]}, \tag{16}$$

where Ω_t denotes the information set (predicting and dependent variables) up to time t. In this expression, each forecast probability $\Pr[S_{t+1} = s|S_t = q, \boldsymbol{z}_{t-1}]$ 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.

Table 3 reports the results for time variation in the transition probabilities. The number of non-zero coefficients is small with a maximum of six when we apply the LT-algorithm.

It indicates that the predicting variables offer only limited help to predict switches to bull or bear markets.

[Table 3 about here.]

[Table 3 (continued) about here.]

We can compare the coefficients for the two-state methods in more detail, as these models allow for the same switches. In all cases, the dividend yield is an important driver of changes in transition probabilities. An increase in the dividend yield leads to an increase in the probability of a continuation of a prevailing bull market. For the rules-based approaches it also leads to a higher likelihood of a switch from a bear to a bull market. Because a high dividend yield indicates that stocks are relatively cheap, the appetite for stocks can be expected to grow, resulting in rising prices and thus a bull market. The marginal effects indicate that the dividend yield is economically meaningful. A one-standard deviation decrease more than doubles the probability of a bull-bear switch, from 1-0.9927 = 0.0073 to 0.025 (LT) or 1 - 0.9935 = 0.0065 to 0.016 (PS). A one-standard deviation increase has a similar effect on bear-bull switches, which increase in likelihood from 0.018 to 0.052 and 0.013 to 0.032. Counterintuitively, for the RS2L-model, an increase in the dividend yield leads to an increase in the probability that a bear market continues.

For the other variables, we see more variation. The T-bill rate is selected in the rules-based approaches when the market is bullish, with approximately the same effect. An increase in the T-bill rate is a bad sign, as it increases the probability of a bull-bear switch. In the RS2L-model this variable is not selected. Inflation is also included in the rules-based methods, though with alternating signs and limited impact. In the RS2L-model, a rise in unemployment increases the likelihood of a bull-bear switch. An increase in the yield spread decreases the likelihood of a bear-bull switch, but has no significant effect when the market is bullish. During bear markets such an increase may lead to higher costs of capital when the alternative of equity issuance is unattractive. In the LT-approach, a high credit spread makes continuation of a bull market more likely, but a bear-bull switch less likely. During bull markets a higher credit spread may reflect a higher demand for credit because of favorable economic conditions, while during bear markets it may correspond

with lower supply. The trade-weighted exchange rate and industrial production growth are never included.

We can compare the average forecast probabilities in the last row of Table 3 to the probabilities in Table 2. For the rules-based approaches, persistence generally remains high when transition probabilities can vary over time. The impact of the predicting variables is limited when the market is bullish, but becomes larger when the market is bearish. In the RS2L-model, bull and bear markets are less persistent than in the RS2C-model, in particular bear markets (transition probability of 0.828 vs. 0.948). The marginal effect of the predicting variables is larger than in the rules-based approaches.

The second part of Table 3 shows how the predicting variables affect transitions when four states are possible. The patterns we observed in Table 2 change slightly. A bull market can switch to the crash state or the boom state, but hardly ever directly to a bear state. Its average persistence has gone down slightly. From a bear state, most switches are to the boom state, and with comparably small probability directly to the boom state. In this case, average persistence has increased. From the crash state, switches to all other states remain likely, with only minor changes compared to Table 2. From the boom state, a switch to the bull state remains most likely, with a low probability of a crash and an even lower probability for a boom state to follow.

The effect of the predicting variables is concentrated in the bull and bear states, and does not show up in the crash and boom states. Since the markets state process leaves crash and boom states quickly, these switches are apparently not related to the slow-moving predicting variables. An increase in inflation makes a continuation of a bull market more likely. An rise in credit spreads mainly increases the likelihood of a crash after a bull market. An increase in inflation is a bad sign when the market is bearish, as the probability of continuation rises. An increase in the T-bill rate makes a bear-bull switch more likely. The marginal effects of these changes are moderate ranging from 0.014 for the T-bill rate to -0.057 for the credit spread. It is remarkable that the dividend yield is completely absent in the RS4L-model. This absence might be related to the general low likelihood of direct switches from bull to bear markets or vice-versa in this specification.

To judge the quality of the different models, we calculate and compare log likelihood values in Table 4. Introducing time-variation in the transition probabilities leads to con-

siderable improvements in the likelihood values. By construction, the improvement are all significant. For the rules-based approaches we compute McFadden R^2-values of 26% and 23%, which shows that the predicting variables are valuable. For the regime switching models, improvement in the likelihood values are a bit smaller, also because less variables are selected. We can also compare the likelihood values of the constant regime switching models with each other. The likelihood ratio statistic of RS4C versus RS2C has an impressive value of 56.4. Unfortunately, the statistic does not have a standard χ^2 distribution due to nuisance parameters under the null hypothesis.[5] The rules-based likelihood values cannot be compared with those of the regime switching models. For the rules-based approaches, they are based on the binary series of bull and bear markets, while for the regime-switching models the return distributions under the different regimes are taken into account as well.

[Table 4 about here.]

5.2 Identification

The different methods lead to different characteristics of bull and bear markets, both for the moments of the return distributions and the persistence of the states. To get a fuller understanding of these differences we analyze the identification of the different approaches. For the rules-based approaches the identification is actually the first step, and the characteristics of the previous subsection are derived subsequent to it. For the regime-switching models, identification and estimation of the characteristics which are model parameters are conducted jointly. To visualize the identification we calculate smoothed inference probabilities $\Pr[S_t = s|\Omega_T]$, which are based on the full-sample information set comprising market returns and predicting variables. We also use the identification to calculate the number of bull and bear markets and their average length. To summarize the differences in identification in one number, we determine the integrated absolute differences.

---

[5]The RS2C-model results from the RS4C-model by six restrictions: two on the initial probabilities, $\xi_b = \xi_c = 0$, and four on the transition probabilities, $\pi_{bu} = \pi_{cu} = \pi_{bd} = \pi_{cu} = 0$. Under the null-hypothesis of the RS2C-model, 10 nuisance parameters are present.

Figure 2 shows the identification for the different approaches. The rules-based approaches produce a binary series indicating during which weeks markets are bullish or bearish. Since both approaches are based on peaks and troughs, switches all take place at maxima and minima. The well-known bearish periods of the early '80s, 1989-1990, the deflation of the IT-bubble from 2000 to 2002 and the credit crisis of Fall 2007 till Spring 2009 are all present. In both approaches, also the crash of October 1987 qualifies as a short-lived bear market. The identification of the two rules-based approaches is, however, not identical. Opposite to the LT-approach, the PS-algorithm indicates a bear market at the beginning of our sample period. The LT-approach produces a bull market at the beginning and a switch after 8 weeks. The PS-algorithm cannot identify switches this early in the sample. Also, the PS-approach does not classify the decrease from July to October 1998 as a bear market, because it was too brief. On the other hand, it considers the period from February to August 2008 as a bear market. Since the total decrease of 7.3% does not exceed the limit of 15%, the LT-algorithm does not pick it up. According to both approaches, the bear market of the credit crisis ended with its nadir on March 4, 2009.

[Figure 2 about here.]

Figure 2(c) show the identification that results from the regime switching models with two states and constant transition probabilities. The identification is based on smoothed inference probabilities. We plot the probability for a bull market by a thin red line. To compare the result to the binary rules-based identification, we round the smoothed inference probabilities and show the result again by purple bars. For further analysis or the formulation of an investment strategy, this rounding is of course not necessary. We see that the probability for a bull market is either close to one or close to zero, and rarely equal to values around 0.5. This indicates that the two regimes are quite distinct, and that the approach gives a clear indication which regime prevails.

As we have seen in Table 1, the identification is now based on means and volatilities. Volatile periods with big price drops are classified as bearish, and tranquil periods with prices increases as bullish. For many weeks, identification by rules or by regimes-switching models is the same. However, we also spot periods where identification differs. Highly volatile periods with small price changes are marked as bearish, for example the periods

July 28 1999 – November 10 1999 (weekly average return -0.092% and weekly volatility 2.9%), and December 29 1999 – June 6 2000 (average -0.084%, volatility 2.8%). Investing during this period would have meant negligible returns but considerable risk. Still, the rules based approaches label these periods as bullish. Of course, the reverse also occurs, as low volatile periods with price decreases are classified as bullish. This happens during December 24 1980 – July 29 1981 (average -0.43%, volatility 1.4%) and more recently during the credit crisis, from March 5 2008 – August 27 2008 (average -0.32%, volatility 1.7%).

In Figure 2(d) we plot the smoothed inference probabilities when the transition probabilities can vary over time. While the overall pattern resembles Figure 2(c), the probabilities are more jittery. So during some periods, this regime switching is uncertain about the actually prevailing state. This applies in particular to the period October 13 1982 – June 29 1983. During this period, the relevant predicting variables decrease considerably. The unemployment rate first increases with about 2% per year which slows down to 0.5% by the end of this period. The yield spread decreases from 3.4% to 1.9%. The period May 1986 – January 1987 shows considerable variation as well. We quantify the differences between the model with constant and with time-varying transition probabilities in the next subsection.

In Figure 2(e) we see what happens when we introduce crash and boom regimes next to a bull and a bear regime. To compare the result of this four-regime model to the other approaches, we have summed the smoothed inference probabilities for the boom and bull regime. The resulting probabilities are plotted with a thin red line. The combination of boom and bull regimes can be seen as extended bullish. If the probability of this combination lies below 0.5, an extended bearish regime (crash / bear) prevails. These periods are again indicated with a purple area. Comparing the purple areas with those of the rules-based approaches or the two-state regime switching models shows some differences for the first part of our sample period. The period July 16 1986 – March 4 1987 is no longer qualified as a bear market as by the RS2C model, but as a bull market interrupted by a few crashes. In turquoise, we have plotted the smoothed inference probability for the crash regime. We see many weeks that qualify to some extent as a crash, though probabilities hardly exceed 0.4. These values represent the difficulty to predict crashes, which we also

see in Table 2.

Figure 2(f) finishes the set of graphs with the identification by the four-state regime switching models with time-varying transition probabilities. Compared to the two-state model with time-varying transition probabilities or the four-state model with constant transition probabilities, this model produces a clearer identification. The red line is again either close to one or close to zero. Bear markets seem to last longer in this model. For example, the RS4L qualifies the period June 24 – November 18 1981 as bearish (weekly average -0.78%, weekly volatility 2.35%), while the RS4C- model limits a bear market to August 19 – October 28 1981 (average -1.33%, volatility 2.39%)). As a second example, the RS4L identifies as bear market from August 27 to December 24 1997 (average -0.12%, volatility 2.45%), contrary to a bear market from October 29 to December 3 1997 by the RS2L model (average +0.11%, volatility 3.4%).

The graphs in Figure 2 give a first impression of the differences in identification, but they do not really enable a statement on the extent of similarity. In Table 5 we count the number of spells of a specific regime that the different methods produce. We also consider the duration of these spells. For the RS4-models we base this analysis on the four states, and on a reduced set of two states, extended bull (bull and boom) and extended bear (bear and crash) markets. The number of bull and bear markets that result from the rules-based approaches is substantially lower than from the regime-switching approaches. The rules-based approaches identify about 8 periods of each, with bull markets easily exceeding an average duration of two years, and bear markets lasting on average about 1 year. The standard deviations of the duration are large, in particular compared to the average duration. Together with the deviations of the medians from the means, this indicates that all identification methods produce some very long spells of bull or bear markets and some very short ones.

[Table 5 about here.]

The regime-switching models all identify more spells of the different regimes, that consequently last shorter. The RS2C-model produces 18 spells of bull and bear markets, that last on average 61 and 25 weeks, both about half of what results from the rules-based methods. When we allow for time-varying transition probabilities and/or more states, the

27

number of spells increase substantially. We find 50 to 66 bull market periods, and up to 50 bear market periods. Moreover, the four-state models identify 50 crashes and 25 to 36 booms, both lasting one or two weeks. Because the regime switching models combine identification and estimation, they inherently identify spells that are as much alike as possible. Compared to the rules-based approaches, means and medians differ less, and the standard deviations of duration are smaller.

The results for the extended bull and bear markets (columns "RS4C*" and "RS4L*") shows that crashes are often an interruption of bull markets and not their end. If a crash would trigger a bear market, the number of extended bear markets (bear or crash) should only be slightly higher than the number of pure bear markets. Booms are also mainly an interruption of bull markets, since the number of extended bull markets marginally either exceeds the number of pure bull markets (RS4C) or is lower (RS4L). We also observe that the results for the RS2L model are quite close to those for the extended bull and bear markets. This can indicate that the additional flexibility of time-variation in the transition probabilities serves to capture short-lived interruptions of bull markets.

In Table 6 we focus on the weekly differences between the approaches. We use the integrated absolute difference of Section 4 to quantify these differences. All these differences have the interpretation of an average probability. As expected, we observe the largest differences of 0.20–0.25 between the rules-based approaches on the one hand and the regime-switching models on the other hand. Differences between the LT- and PS-approaches or between the difference regime-switching models are a lot smaller. In the case of the rules-based approaches, the distance corresponds directly with the probability of a different identification, so in 4% of the cases the LT-approach leads to a different identification than the PS-approach. For the regime-switching models, we see that the effect of changing the number of regimes from two to four, or adding predicting variables is about the same.

[Table 6 about here.]

We conclude that the results from the rules-based approaches differ substantially from those from the regime-switching models. While the rules-based approach tend to produce relatively long periods of bull and bear markets, the regime-switching models exhibit periods when bull (bear) markets dominate with short interruptions of bear (bull) markets.

28

A similar result motivates the model put forward by Maheu et al. (2009), which also allows for rallies during bear markets and corrections during bull markets. The differences between the various regime switching models are smaller but still present. The large number of switches between extended bull and bear indicate that bull markets are frequently interrupted by crashes and that bear markets sometimes exhibit a rebound for a week. Of course, profiting from short rallies and avoiding crashes is desirable to any investor. The regime-switching models may be able to do so.

6 Predictions and investments

In the full-sample analysis of the previous section, we concentrated on the differences between the ex-post identification of bull and bear markets and their characteristics that resulted from the different methods. We have spotted substantial differences as well as some similarities, but these results on themselves do not present a clear preference for one method over the others. In this section we investigate which method performs best, when we base an investment strategy on it. The strategies that we consider are dynamic, as the model parameters are regularly updated to incorporate new information. We compare the different methods based on their statistical performance, their investment performance and look at the risk-return trade-off they offer. We use the period from July 7, 1994 to July 1, 2009 to evaluate the different methods out-of-sample.

6.1 Predictions and parameter updates

We consider an investor that chooses one method from the four different methods for identification (LT, PS, a two-state regime switching model or a four-state regime switching model), and decides whether or not to use macro-financial variables to help predicting the future state of the market. She uses past data (starting December 26, 1979) to identify past bullish and bearish periods and to estimate models to make predictions with. Every week, the investor updates her inference on the actual state of the market and makes a one-step-ahead forecast. Every thirteen weeks (so roughly four times per year) she updates the model parameters, using an expanding window. In case of the rules-based approaches,

she determines new parameters for the Markovian logit models. When she uses a regime-switching model, she re-estimates all parameters of these models. In both methods, she follows the specific-to-general procedure for variable selection. Though these steps apply to all models, we discuss below in more detail how we conduct the estimation and construct predictions.

In the LT-approach, the sequence of bull and bear market periods runs until the last observed extreme price, which can be a peak or a trough. The investor does not know how to qualify the period from the last extremum onwards. Therefore, she bases the estimation of the Markovian logit models on the observations up to the last extreme price. She applies the incremental selection procedure of Section 3.3 to determine the variables to include in the Markovian logit model. To make a prediction, the investor uses the recursion in (4), starting at the last extreme price. Every week, the investor checks whether she has observed a new extreme price and whether a switch has taken place. She uses the parameters from the last estimation step to construct a new prediction.

The PS-approach includes a censoring step, in which switches in the last 13 weeks are removed. Therefore, the investor only uses the sequence of bullish and bearish periods up to 13 weeks prior to her current point in time for estimation. Every week, the investor applies the PS-algorithm completely, to determine the state of the market 13 weeks ago. From that point onwards, she use the recursion in (4) to produce a new prediction.

The regime switching models do not need specific rules to treat observations at end of the sample period. The investor simply uses the filter procedure to determine the state of the market at each point in time. Multiplying the last inference probabilities with the (time-varying) transition matrix produces the forecast for the next period. When 13 weeks have passed, all parameters of the regime switching models are newly estimated. In the two-state regime switching model with time-varying transition probabilities, the investor follows the incremental selection procedure for variable selection. When the regime-switching model has four states, the full procedure would be too time-consuming. Therefore, we restrict variable inclusion to switches from and to bull or bear markets. So the parameters β_{sq} can take on non-zero values for $s, q \in \{\mathrm{d}, \mathrm{u}\}$, but are zero when $s \in \{\mathrm{c}, \mathrm{b}\}$ or $q \in \{\mathrm{c}, \mathrm{b}\}$. This means that the transition probabilities from crash and from boom states are constant. This restriction shrinks the number of variable-transition combinations considerably, and

30

speeds up the estimation part. The restriction agrees with the result on the RS4L-model in the previous section.

6.1.1 Parameter evolution

Before we examine the predictions of the different methods and their quality in more detail, we first turn to the evolution of the parameters and model characteristics. This analysis can help our understanding of the predictions. Second, it shows how robust the parameters and characteristics are when more information becomes available. If parameters and characteristics vary strongly over time, this may indicate low quality predictions. Of course, little variation does not necessarily imply better predictions.

In Figure 3 we plot the evolution of the means and volatilities of the different regimes. In the rules-based approaches, we first do the identification, and estimate means and volatilities based on that. In the regime-switching approaches, we estimate means and volatilities directly. Generally, we see that the means and volatility 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 regime-switching models. The regime switching models exhibit more variation in their means and volatilities, in particular for shorter estimation windows. For the RS4C- and RS4L-model the bear regime has a positive mean for estimation windows ending before 2001. Before 1999, the mean of the bear regime actually exceeds the mean for the bull regime. However, the volatility of the bull regime is consistently lower for all estimation windows. Also in the volatility of the boom and crash regimes of the RS4-models, we see quite some variation for the shorter estimation windows. The stability of the parameters for longer estimation windows indicates that a regime-switching model with more states needs more data for reliable parameter estimates.

[Figure 3 about here.]

The evolution of the transition probabilities when assumed constant within an estimation window are in Figure 4. The methods with two states produce transition probabilities

31

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 4 about here.]

For the RS4C-model, different, less stable pictures emerge, with the exception of transitions from the bull state. When the estimation windows are still relatively short (up to July 1997), the bear state can be followed by the bear state again, the bull state or the crash state, but switches to a boom state are not likely at all. After 1997, either continuation of the bear state, or a switch to the boom state become the likely transitions. When the process is in the crash state, the boom state is most likely to follow when the estimation window is short. If the windows becomes longer, the probabilities for a switch to the bull regime become largest, followed by a bear switch. The most abrupt break in the transition probabilities happens for the boom regime as departure state. Up to July 1997, the boom regime is followed with a probability of 0.8 by the bear regime with the bull regime as alternative. After July 1997, the process switches to the bull regime with a probability around 0.9 with the crash regime as alternative. Again, a longer estimation window may be better for larger regime switching models.

Figure 5 shows the dynamics of the models for the time-varying transition probabilities. Here we can not only judge whether the parameters of the models are stable, but also if the same variables are selected for the different estimation windows. At first sight, we conclude that the regime-switching models exhibit more variation in the selected variables and the associated coefficients than the rules-based approaches. However, also the rules-based methods show some differences over time. Consistent with our conclusions in the static analysis, the dividend yield is always an important predictor in each regime in the rules-based approaches. The T-bill rate is consistently relevant when switching from a bull market. The yield spread and industrial production are selected for most of the estimation windows in the LT-approach for switches from the bull state. The unemployment rate and the credit spread are chosen for a few specific estimation windows. For switches from a bear market, only the dividend yield seems relevant. At the very last estimation windows,

other variables show up. While not as stable as the earlier results for the rules-based approaches, these results boost the confidence we have in the predictions of these models.

[Figure 5 about here.]

[Figure 5 (continued) about here.]

The results for the regime-switching models are less reassuring. All variables are selected at least once for the RS2L-model. The yield spread, and to a lesser extent the T-bill rate, is selected most often for switches from the bull regime. The dividend yield is selected only a few times. For switches from the bear regime, sometimes no variables are selected, and no variables is selected with some consistency. The patterns for the transitions in the RS4L model also exhibit variations, though less than in the RS2L-case. Again the yield spread is most consistently selected to predict switches from the bull state, though in less than 50% of the cases. The T-bill rate is more consistently chosen for bear-bear switches. For short estimation windows, some predictability for bear-bull switches is found, but this disappears for longer windows. The oscillating patterns for short estimation windows for both the two- and four state models can again indicate the necessity for longer windows to get stable estimates. When the selection procedure does not select any variable, predictability may be on the boundary of the chosen significance level.

6.2 Building an investment strategy

The investor can use the predictions of the different models and approaches in several ways to build an investment strategy in futures. The most straightforward one is to make a binary decision. If the model predicts a bull market, the investor goes long one futures contract, and if it is a bear market, she goes short one futures contract.

While this strategy is clear-cut, it ignores the strength of the model prediction. A prediction of a bull market with probability 0.95 is a stronger signal than a probability prediction of 0.65. As a second strategy, the investor can take the actual predicted probability for a bull market at time $t+1$ by method m, $\xi_{u,t+1}^m$ into account by taking a position of $2\xi_{u,t+1}^m - 1$. A certain prediction of a bull market ($\xi_{u,t+1}^m = 1$) or of a bear market

$(\xi_{u,t+1}^m = 0)$ leads to a full position in a futures contract. If the investor is less sure on the direction of the market, her investment is a fraction proportional to the probability.

Using only probabilities for an investment strategy goes well with the rules-based approaches, as they produce binary series of bullish and bearish periods. To the contrary, the regime-switching models also use information on the means and volatilities of the regimes to produce inference probabilities. We can combine the predicted probabilities with means and volatilities to construct the predicted density. If we estimate means and volatilities based on the binary series of bull and bear markets of the rules-based approaches, we can do similar computations.

We devise two strategies that take the moment predictions into account. As one strategy, the investors pays attention to the sign of the predicted mean. If it is positive (negative), the investor takes a long (short) position of one futures contract. The prediction for the mean, μ_{t+1}^m is the weighted average of regime-specific means

$$\mu_{t+1}^m \equiv \mathrm{E}[r_{t+1}|\boldsymbol{z}_t] = \sum_{s\in\mathcal{S}} \xi_{s,t+1}^m \mu_s^m. \tag{17}$$

In the last strategy that we consider, we assume that the investor takes a position in futures contracts with a nominal value equal to a percentage w of her wealth W_t. She chooses w to optimize the expectation of a utility function $U(W_{t+1})$ over the next-period wealth $W_{t+1} = W_t(1 + wr_{t+1})$. We approximate this utility function to the second order at her current wealth W_t

$$U(W_{t+1}) \approx U(W_t) + U'(W_t)W_t wr_{t+1} + \frac{1}{2}U''(W_t)W_t^2 w^2 r_{t+1}^2. \tag{18}$$

Since the current utility level does not influence the optimization, we can ignore the first term. Dividing by $U'(W_t)W_t$ produces a standardized utility function

$$\tilde{U}(W_{t+1}) = wr_{t+1} + \frac{1}{2}\frac{U''(W_t)W_t}{U'(W_t)}w^2 r_{t+1}^2 = wr_{t+1} - \frac{1}{2}\gamma_t w^2 r_{t+1}^2 \tag{19}$$

where γ_t is the coefficient of relative risk aversion. The optimal portfolio is given by $w_t^m = \mu_{t+1}^m/(\gamma_t \omega_{t+1}^m)$, where ω_{t+1}^m is the raw second moment

$$\omega_{t+1}^m \equiv \mathrm{Var}[r_{t+1}|\boldsymbol{z}_t] = \sum_{s\in\mathcal{S}} \xi_{s,t+1}^m \mathrm{E}[r_{t+1}^2|S_{t+1} = s, \boldsymbol{z}_{t+1}]. \tag{20}$$

This portfolio reflects both the expected return but also the risk during the coming period. We choose $\gamma_t = 5$.

6.2.1 Predictive Accuracy

For each week, every method produces a one-step-ahead prediction of the regime to occur. To determine the quality of the predictions, we compare them with the identification based on the full sample as shown in Figure 2. In the case of methods that distinguish just two states, a natural way to measure the statistical quality of the predictions is the hit rate, which we report in Table 7. Looking at the overall predictive quality, the LT-approach performs best, with hit rates of approximately 89%. The two-state regime switching models come in second with hit rates of around 86%. The PS-approach performs worst, with hit rates around 72%. Looking at price changes, as the LT-method and the regime-switching models do, seems to yield better results than using duration to determine whether a change could take place, as in the PS-method.

[Table 7 about here.]

Besides the percentage of total correct predictions and its increase relative to the benchmark, we also consider the Kuipers Score. We calculate the Kuipers Score as the percentage of correctly predicted bull markets minus the percentage of incorrectly predicted bear markets. By construction, the same value results from subtracting the percentage of incorrectly predicted bull markets from the percentage of correctly predicted bear markets. The Kuipers Score balances the percentage of hits with the percentage of false alarms under the assumption that the benefits of a hit equal the costs of a false alarm (see Granger and Pesaran, 2000, for a more elaborate discussion). Table 7 shows that the RS2-models yield the highest Kuipers Score, with the LT-approaches close behind. The PS-method follow at some distance, though scores are still well above zero. Overall, the Kuipers Scores are high, because all methods capture the persistence in bull and bear market spells.

The LT-approach outperforms all other two-state methods when predicting bull markets, though the quality of predicting bullish weeks by the RS2-models does not lag far behind. The RS2-models are best at predicting bear markets. The LT-method slightly outperforms the PS-method, but both lag the RS2 models by 6–14%-points. If we compare the models with a benchmark of always predicting a bull market, the regime-switching models show the largest improvement of about 21%-points. The improvement by the LT-based

predictions is slightly lower at 18%. The PS approach shows smaller improvements of less than 5%.

A striking result is that the models without predicting variables (suffix C) perform slightly better than the models with predicting variables. Differences are relatively small, but for all models the hit rates are lower when predicting variables are included. The same applies to the Kuipers Scores. From a statistical point of view, we conclude that the added value of these variables is low. This may be related to the variation we have seen before in the variables that are selected for the different transition models.

For the four-state regime switching models, hit rates are not that informative, since rounding the predicted or inferred probabilities to zero-one variables is typically too crude. Therefore, we use again the integrated absolute differences of Section 4. We calculate the average distances per state, as well as the total distance which weighs the states distances by the likelihood of their occurrence. The results for the two-state approaches in Table 8 confirm our conclusions based on hit rates. The LT-approach performs best, with all other methods at some distance. By definition, the performance for predicting bull markets is the same as for bear markets for the two-state methods. For the LTC approach, we find that the average distance is 0.092. We can compare this to 1 minus the hit rate, which equals 0.107. This number reflects the fraction of observations where prediction and identification are at odds. The IAD being lower than 0.107 indicates that the predictions when wrong are a bit removed from an extreme 0 or 1 forecast. For the PS-method we also see this pattern. For the RS2-models, the numbers for IAD are a bit larger than 1 minus the hit rate, indicating that these models give less extreme predictions than the rules based methods.

[Table 8 about here.]

The last columns of Table 8 report the performance of the four-state regime switching models. The total difference are a bit higher than the for the two-state regime switching models. Of course, making predictions for four states is inherently more difficult than for two. When we look just at bull or bear markets, we see that the RS4-models perform worse in predicting bull markets, but that for bear markets the difference is marginal. Crashes

and booms are predicted relatively well, though they do not happen often, which also drives the average distance down.

6.2.2 Investment performance

Of course, the real test of the different methods should be based on their performance when used in an investment strategy. In the statistical analysis, the implicit loss function of making false predictions may not reflect the actual economic cost. Therefore, we consider the four investment strategies: the binary strategy based on the rounded predicted probabilities, the proportional probability strategy, the sign strategy that uses the sign of the predicted mean and the utility strategy that uses the predicted density. Every week, the investor takes a new position in futures contracts, using the newest predictions. We do not take transaction costs into account, as these are typically low and future contracts are highly liquid. The returns to the strategies are all in excess of the risk-free rate because we use futures contracts. For the RS4-models we only consider the sign and mean-variance strategy. For the early part of our out-of-sample period, the mean of the bearish regime exceeds that of the bullish regime, which makes it difficult to apply the rules for the binary and proportional probability strategy.

In Table 9 we report the yearly mean, yearly volatility and the Sharpe ratio of all combinations of investment strategies and methods. The benchmark for all combinations is a continuously rolled-over long contract on the index. The return to this position would be 2.4% per year (in excess of the risk-free rate) with a yearly volatility of 17.5% and a Sharpe ratio of 0.14. The LT-strategies beat this benchmark quite well with average returns exceeding 6.6% and Sharpe ratios up to 0.6. The strategies that use the PS-algorithm also consistently beat the benchmark, though less convincing than the LT-based strategies. The performance of the regime switching models varies between slightly negative and slightly better than the benchmark. When the average return is negative, the Sharpe ratio is not a proper performance measure. The four-state models outperforms the two-state models. However, all methods fall short of producing a clearly significantly better performance than the benchmark. For the LT-strategies the best results are marginally significant, as the probabilities of obtaining a lower average return than the benchmark are around 0.10. Of

course, this simply shows the well-known fact that outperforming an index for a whole market is difficult.

[Table 9 about here.]

To determine the economic value of the different investment strategies, we follow the procedure of Fleming et al. (2001) and Marquering and Verbeek (2004). If we use the quadratic approximation of the utility function in (18), and assume that the approximation is always taken with respect at the same point W_0, the utility gain V_j of strategy j over the period $t = \tau + 1$ to $t = T$ equals

$$V_j = U'(W_0)W_0 \sum_{t=\tau+1}^{T} \left(w_{jt}r_t - \frac{1}{2}\gamma w_{jt}^2 r_t^2 \right). \tag{21}$$

We calculate the economic value of strategy j by the maximum fee Δ_j as a fraction of wealth that the investor is willing to pay every period to switch to it from the buy-and-hold strategy. If the investor pays this fee, the utility gain of strategy j is equal to the utility gain of the benchmark,

$$U'(W_0)W_0 \sum_{t=\tau+1}^{T} \left((w_{jt}r_t - \Delta_j) - \frac{1}{2}\gamma(w_{jt}r_t - \Delta_j)^2 \right) = U'(W_0)W_0 \sum_{t=\tau+1}^{T} \left(r_t - \frac{1}{2}\gamma r_t^2 \right). \tag{22}$$

We use this equality to solve for Δ_j.

The fees in Table 9 show that the economic value of all strategies based on the LT-algorithm is substantial. An investor would be willing to maximally pay a yearly 4.1% up to 12.3% to use one of these strategies instead of a static buy-and-hold strategy. For the other strategies, the maximum fees are close to zero, or even negative. These results are particularly disappointing for the regime-switching models. These models take both means and volatilities into account for identification and prediction. While this sounds appealing from a utility perspective, the models fail to actually deliver economic value.

The effect of including predicting variables is mixed when we look at the monetary performance, in contrast to the overall statistical deterioration of the predictions we saw earlier. For the rules-based methods, using predicting variables mostly leads to a worse performance, but for all regime-switching models predictive variables increase the performance. So while the statistical quality of the predictions goes down, the actual losses due to the errors in the predictions actually become smaller.

Comparing the different strategies shows that the binary strategy, and the closely related sign strategy produce the best results. The utility strategy sometimes yield higher average returns, but this has to do with the typically larger position that this strategy advises. When we look at the Sharpe ratio, we see that this strategy generally underperforms the binary strategy. This result is consistent with Christoffersen and Diebold (2006) who relate the predictability of the sign of asset returns to predictability of volatility. For assets with a positive expected return, higher volatility increases the probability of a negative return. To a large extent, an increase in the predicted probability of a bear market implies a higher predicted volatility. However, we also find that the proportional strategy does not outperform the binary strategy, nor that the utility strategy outperforms the sign strategy. This indicates that the accuracy of the predictions is limited.

In Figure 6 we track the performance of the different strategies combined with different methods over time. We also show the predictions for bull markets, and the identification of bear markets. We show the performance of the binary, proportional and utility strategies for the two-state methods. We leave out the sign strategy for the two-state methods, because it is almost identical to the binary strategy.

[Figure 6 about here.]

[Figure 6 (continued) about here.]

Figures 6(a) and 6(b) correspond with the LT-approach. The probabilities for a bull market clearly show a saw-pattern, which is caused by the recursive predictions since the last observed extremum. This approach needs some time before it signals a switch in the investment strategy. For instance, the bear market after the burst of the IT-bubble starts on April 5, 2000, but is picked up only after October 18. The method is quicker in picking up the start of a bull market: switch on October 16 2002, picked up on December 4 2002. The use of predicting variables does not make a difference in these two cases. The performance of the different strategies shows the same pattern over time. The proportional strategy shows smaller changes than the other strategies, because it decreases its position when the uncertainty over the prevailing regime increases. The utility strategy shows the largest changes because it can take more extreme positions than the other two.

In Figures 6(c) and 6(d) we see the predictions and the performance of the PS-methods. The predictions show an almost binary pattern, with very abrupt switches. These breaks result from the way the PS-algorithm evaluates peaks and troughs. If the last trough has been passed quite some time ago, a peak will easily qualify as a switch to a bear market, even if prices have not decreased much since the peak. Of course, censoring makes sure that prices have been lower for some period (in our case 13 weeks). The same reasoning applies to troughs marking a switch to bull markets. Often these switches turn out to be false alarms, more often than in the LT-approach. A positive effect is that true switches may be indicated sooner. The PS-algorithm signals a bear market after the burst of the IT-bubble on July 12, 2000. However, it signals the subsequent switch to a bull market at January 22, 2003, later than the LT-algorithm. Moreover, the PS-predictions gave two false alarms in the mean time. The performances of the different strategies are relatively close, with some large deviations at the end of the sample.

The predictions of the two-state regime switching models in Figures 6(e) and 6(f) show a more jittery pattern, but they pick up changes in regimes faster than the rules-based approaches. As we remarked earlier, some periods with price increases but a high volatility are labelled as bear markets. We can now see, that these periods (for example the second half of 1998) lead to a loss for most strategies. For the utility strategy these losses are smaller than for the other strategies, because the increased variance reduces the positions. However, when the market goes really down (for example in 2001), the mean-variance strategy fails to take advantage of this, compared to the binary strategy. As we can see in Table 9, this leads to a considerably lower variance of the strategy, but also an average return below the benchmark, close to zero. Including predicting variables yields a better performance, in particular during the boom and bust of the IT bubble and the credit crisis.

Figures 6(g) and 6(h) show the results for the RS4-models. Here, we plot the predictions for the extended bull state (bull and boom), and for the crash state, as well as the identification of extended bear states. We do not consider the binary and proportional strategies, because the interpretation of the regimes is not clear for the early weeks of the out-of-sample period. The predictions in this model also fluctuate strongly, but a bit less than for the RS2-models. For many weeks, we see a small though relevant predicted probability for the crash state. Its relevance is shown by the flat evolution of the utility strategy,

40

where the probability of a crash state leads to a lower predicted mean and higher predicted variance, and hence a reduced position in the futures contract. Also for these models the resulting volatility of the mean-variance strategy in Table 9 is considerably lower. For the RS4C model, the sign strategy closely follows the benchmark, with the exception of the beginning and the end of the out-of-sample period, where it underperforms the benchmark. In the RS4L case we see an outperformance of the benchmark, in particular during bearish periods. During the second half of 2008 and the first half of 2009, both strategies are able to profit from the market turmoil.

Taking the results of the statistical quality and the investment performance together, we conclude that predicting bull and bear markets is not easy. While the statistical analysis indicates a clear improvement by the LT-approach and the regime switching models over the benchmark of a long position in index futures, resulting investment strategies yield at best a marginally significant outperformance. The LT-approach performs best, outperforming the benchmark and the other methods.

Crucial for the performance of the rules-based approaches is the speed with which they can pick up a true switch in market sentiment. The LT-approach is clearly better than the PS-approach but sometimes still lags half a year. The regime switching are much better at picking up a switch in the market, but a switch in regime may be related more to volatility than to means, which means that risky though profitable opportunities are missed. In the rules-based approaches, the added value of the predicting variables is negative or limited, both from a statistical and investment perspective. For the regime-switching models, they lower the statistical quality of the predictions but increase the investment performance.

Finally, we conclude that binary strategies based on either probabilities or the sign of the predicted means perform best. The strategy that invests proportional to the predicted probability, or uses the predicted means and variances do not lead to superior performance. In case of the regime switching models, we clearly see that the utility strategy curbs the position in the futures contract when riskiness increases, but this not lead to improved Sharpe ratios or higher performance fees.

7 Conclusion

In this article we compare the identification and prediction of bull and bear markets by four different methods. One way to address identification and prediction is by formulating rules to determine bullish and bearish periods, and then as a second step use binary models for prediction. In this category, we consider the approaches of Lunde and Timmermann (2004) and Pagan and Sossounov (2003). Both base identification on peaks and troughs in price data. To find switch points Pagan and Sossounov (2003) impose restrictions on the length of cycles and phases, whereas Lunde and Timmermann (2004) impose restrictions on price changes. As an alternative, an investor can formulate a model that simultaneously handles identification and prediction. We consider a simple regime switching model with a bull and a bear state, and an extended version that also includes boom and crash states.

From the identification we conclude that the rules based approaches produce more or less the same results. A market that has exhibited price decreases since the last peak is bearish; price increases after the last trough qualify as bullish. To the contrary, the regime switching models also pay attention to volatility. Period with low volatility but price decreases are identified as bull markets, while volatile price increases are classified as bearish. For a risk-averse investor, volatile price increases may indeed be less attractive, which justifies identification as bearish. However, low-volatile price decreases are in no way attractive for investors, so identification as bullish is not desirable. We show that a more extensive specification with two additional regimes for booms and crashes improves identification.

To determine which method works best in an investment strategy, we evaluate all methods for a weekly investment in futures contracts on the MSCI index for the US stock market. Crucial for the performance is the speed with which a method can identify a switch to increasing or decreasing prices. The LT-method is best in this respect and yields 6.6% to 15.5%. The regime switching are quicker in picking up regime switches, but they face difficulties warning against low-volatile price decreases. As a consequence, their results are comparable to a continuously rolled-over long position in futures contracts. The PS-method generates losses, because it has difficulties identifying an ongoing bull or bear market.

In line with the somewhat depressing results on return predictability in Welch and

Goyal (2008), we also find that the inclusion of predictive variables is limited and subject to changes. For the rules-based approaches, the effect is clearly detrimental, with performance uniformly worse. For the regime switching models we observe improvements in performance, but the variables that are selected for predictions and their coefficients vary considerably over our sample period.

Harding and Pagan (2003a,b) and Hamilton (2003) have already discussed the difference between rules-based and model-based approaches, applied to dating business cycles. As the resulting identifications were largely similar, the main differences were the larger transparency for the rules-based approaches versus the deeper insight into the data generating process for the regime switching models. For financial time series, differences are larger where the rules-based approaches purely reflect the tendency of the market, while the regime switching models reflect the risk-return trade-off. For an actively managed strategy, volatility is of secondary importance. The strategy should in the first place indicate correctly whether prices will increase. Higher volatility may lead to a less risky position, but success hinges critically on the accuracy of the sign prediction. In this particular situation, an investor should prefer rules for identification.

Table 1: Return Characteristics of Bull and Bear Markets

regime		LT		PS		RS2C		RS2L		RS4C		RS4L	
bull	mean	0.35	(0.05)	0.30	(0.05)	0.16	(0.05)	0.16	(0.05)	0.23	(0.06)	0.18	(0.05)
	vol.	1.92	(0.07)	1.95	(0.08)	1.56	(0.05)	1.53	(0.05)	1.42	(0.05)	1.36	(0.05)
bear	mean	−0.78	(0.13)	−0.62	(0.13)	−0.30	(0.17)	−0.29	(0.17)	−0.15	(0.19)	−0.03	(0.16)
	vol.	2.85	(0.24)	2.81	(0.24)	3.41	(0.16)	3.43	(0.14)	3.25	(0.18)	3.03	(0.15)
crash	mean									−4.09	(0.40)	−4.36	(0.45)
	vol.									2.31	(0.67)	2.83	(0.85)
boom	mean									3.88	(0.20)	3.69	(0.15)
	vol.									0.67	(0.15)	0.58	(0.12)

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 prices 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) as in (5), and with 4 regimes (bull, bear, boom and crash, columns RS4C and RS4L) as in (6). 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 2: Constant Transition Probabilities

(a) Probability Estimates

from	to	LT	PS	RS2C	RS4C
bull	bull	0.991	0.993	0.977	0.948
	bear	0.009	0.007	0.023	< 0.001
	crash				0.052
	boom				< 0.001
bear	bull	0.019	0.017	0.052	< 0.001
	bear	0.981	0.983	0.948	0.931
	crash				0.064
	boom				0.006
boom	bull				0.432
	bear				0.378
	crash				0.086
	boom				0.104
crash	bull				0.887
	bear				< 0.001
	crash				0.113
	boom				< 0.001

(b) Unconditional Regime Probabilities

	LT	PS	RS2C	RS4C
bull	0.686	0.722	0.692	0.701
bear	0.314	0.278	0.308	0.236
crash				0.043
boom				0.020

This table shows the transition probabilities between the different regimes under the different approaches and the resulting unconditional 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 4 regimes (RS4C). The unconditional probabilities $\bar{\boldsymbol{\pi}}$ satisfy $\bar{\boldsymbol{\pi}}^m \boldsymbol{P}^m = \bar{\boldsymbol{\pi}}^m$.

Table 3: Time-varying transition probabilities, (multinomial) logit models

model	LT		PS		RS2L	
from	bull	bear	bull	bear	bull	bear
to	bull	bull	bull	bull	bull	bull
constant	7.53	−6.39	6.30	−6.66	3.06	−1.55
inflation	−0.58	0	0.10	0.36	0	0
	[−0.004]		[0.001]	[0.006]		
unempl.	0	0	0	0	−0.83	0
					[−0.057]	
prod. growth	0	0	0	0	0	0
t-bill rate	−1.61	0	−1.52	0	0	0
	[−0.012]		[−0.010]			
yield spread	0	0	0	0	0	−1.13
						[−0.161]
credit spread	1.78	−0.59	0	0	0	0
	[0.013]	[−0.010]				
trade w. fx	0	0	0	0	0	0
div. yield	2.38	1.96	1.42	1.71	0.82	−2.55
	[0.017]	[0.034]	[0.009]	[0.029]	[0.056]	[−0.363]
$\bar{\pi}_{sq}$	0.993	0.018	0.993	0.017	0.926	0.172

This table shows the estimated coefficients and marginal effects of the predicting variables in Table B.1, when they are linked to the transition probabilities by (multinomial) logit models. The predicting 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 Markovian logit model as in (3), where the coefficients depend on the departure state. In the two-state regime switching model, RS2L, the logistic transformation in (7) is used to link the predicting variables to the transition probabilities. For the four-state regime switching model, RS4L, the multinomial logistic transformation in (9) is used. In that case the coefficients for a switch to the boom 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. The marginal effects in brackets are calculated for the average forecast probability $\bar{\pi}_{sq}$ reported in the last row of the table. The average forecast probability is calculated as in (16). For the two-state approaches, the marginal effect of variable i is calculated as $\bar{\pi}_{sq}(1 - \bar{\pi}_{sq})\beta_{qi}$. For the four-state approaches, the marginal effect is given by (23).

Table 3: Time-varying transition probabilities, (multinomial) logit models – *continued*

RS4L

model from to	bull bull	bull bear	bull crash	bear bull	bear bear	bear crash	crash bull	crash bear	crash crash	boom bull	boom bear	boom crash
constant	4.40	−11.98	1.41	−2.81	3.34	−13.44	1.64	1.29	0.35	3.11	−21.09	0.63
inflation	0.54 [0.033]	0 [0.000]	0 [−0.026]	0 [−0.006]	0.83 [0.040]	0 [0.000]	0	0	0	0	0	0
unempl.	0	0	0	0	0	0	0	0	0	0	0	0
prod. growth	0	0	0	0	0	0	0	0	0	0	0	0
t-bill rate	0	0	0	1.90 [0.014]	0 [−0.014]	0 [0.000]	0	0	0	0	0	0
yield spread	0	0	0	0	0	0	0	0	0	0	0	0
credit spread	−0.94 [−0.057]	0 [0.000]	0	0	0	0	0	0	0	0	0	0
trade w. fx	0	0	0 [0.046]	0	0	0	0	0	0	0	0	0
div. yield	0	0	0	0	0	0	0	0	0	0	0	0
$\bar{\pi}_{sq}$	0.935	0.000	0.052	0.008	0.949	0.000	0.459	0.325	0.126	0.887	0.000	0.074

For table notes, see the first part of the table.

Table 4: Log likelihood values of different model specifications

model	LT	PS	RS2	RS4
constant	-87.4	-82.7	-3310.2	-3282.0
time-varying	-65.0	-63.9	-3300.4	-3269.1
LR	44.8	37.6	19.8	25.8
d.f.	6	5	4	4
p-value	< 0.0001	< 0.0001	0.0006	< 0.0001

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 Markovian logit models as in (3). For the regime switching models with two and four states, we report the likelihood of the complete model. The transition probabilities can be constant (first row, corresponding with Table 2) or time-varying (second row, corresponding with Table 3). 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.

Table 5: Number and duration of market regimes

approach		LT	PS	RS2C	RS2L	RS4C	RS4C*	RS4L	RS4L*
bull	number	9	8	18	50	51	55	66	53
	avg. duration	121.3	133.8	61.0	22.4	21.6	20.5	15.2	19.7
	med. duration	77	83	25	6.5	13	12	6.5	9
	std. dev. duration	131.0	154.4	94.4	50.7	25.0	24.7	26.7	29.6
bear	number	8	8	18	50	25	55	18	54
	avg. duration	56.1	58.9	24.6	8.4	14.4	7.5	24.3	9.2
	med. duration	55.5	55.5	24.5	1	8	1	23	1
	std. dev. duration	44.6	41.5	15.7	14.4	16.1	13.0	17.0	15.0
crash	number					50		50	
	avg. duration					1.1		1.2	
	med. duration					1		1	
	std. dev. duration					0.30		0.49	
boom	number					25		36	
	avg. duration					1.0		1.1	
	med. duration					1		1	
	std. dev. duration					0		0.23	

This table shows for every method the number of spells of the different market regimes, their average and median duration and the standard deviation of the duration. For the two-state regime switching models, a period is qualified as bullish if the smoothed inference probability for the bull regime exceeds 0.5 and bearish otherwise. For the four-state regime switching models, the highest smoothed inference probability determines the prevailing regime. In the columns labeled "RS4C*" and "RS4L*", we reduce the four states to two by joining the bull and boom states, and the bear and crash states.

Table 6: Integrated absolute differences for two-state approaches

	PS	RS2C	RS2L	RS4C	RS4L
LT	0.040	0.212	0.196	0.221	0.238
PS		0.228	0.216	0.236	0.255
RS2C			0.087	0.094	0.098
RS2L				0.121	0.124
RS4C					0.069

This table reports the integrated absolute distance between the identification of the different approaches, based on two states. We calculate the distance as the sample-analogue to (12), if both identification approaches yield probabilities that are strictly between zero and one. If one of the two approaches can yield 0 or 1, we use (15). For the RS4 models we add the smoothed inference probabilities of the bull and boom regimes and of the bear and crash regimes together.

Table 7: Predictive accuracy of two-state methods

	LTC	LTL	PSC	PSL	RS2C	RS2L
bull correct	532	532	411	412	468	465
bull wrong	22	22	129	128	45	48
% bull correct	96.0%	96.0%	76.1%	76.3%	91.2%	90.6%
bear correct	167	159	168	149	218	204
bear wrong	62	70	75	94	52	66
% bear correct	72.9%	69.4%	69.1%	61.3%	80.7%	75.6%
total correct	699	691	579	561	686	669
total wrong	84	92	204	222	97	114
% correct	89.3%	88.3%	73.9%	71.6%	87.6%	85.4%
benchmark	70.8%	70.8%	69.0%	69.0%	65.5%	65.5%
improvement	18.5%	17.5%	5.0%	2.7%	22.1%	19.9%
Kuipers Score	69.0%	65.5%	45.2%	37.6%	72.0%	66.2%

This table shows the predictive accuracy of the different methods that distinguish bull and bear states. Every week t a one-step-ahead forecast for week $t + 1$ is made for the probability of a bull and of a bear regime. The first prediction is made for July 6, 1994 and the last for July 1, 2009, giving a total of 783 predictions. Inferences on the state of the market in week t use the information up to week t by simply applying the identification rules (LT- and PS-approach), taking their limitations into account, or by applying the filter procedure for the regime-switching models based on the last available model parameters. The parameters in the Markovian logit models for predictions in the rules-based approaches and the parameters of the regime switching models are updated every 13 weeks. The predicted probabilities are rounded, and compared with the identification that results from the full sample. For the regime switching models the resulting smoothed inference probabilities are rounded, too. "Bull (bear) correct" gives the number of true bullish (bearish) weeks that were correctly predicted. "Bull (bear) wrong" gives the number of true bullish (bearish) weeks that were wrongly predicted. The percentages are calculated with respect to the number of true bullish (bearish) weeks. The row "benchmark" reports the percentage of total correct predictions, when the models would always predict a bullish week. The row "improvement" shows by how much a method's percentage of correct predictions exceeds the benchmark. The Kuipers Score is calculated as the percentage of correctly predicted bull markets minus the percentage of incorrectly predicted bear markets.

Table 8: Integrated absolute differences between prediction and identification

	LTC	LTL	PSC	PSL	RS2C	RS2L	RS4C	RS4L
bull	0.092	0.098	0.181	0.265	0.161	0.155	0.207	0.198
bear	0.092	0.098	0.181	0.265	0.161	0.155	0.173	0.178
crash							0.069	0.065
boom							0.034	0.039
total	0.092	0.098	0.181	0.265	0.161	0.155	0.217	0.221

This table shows the integrated absolute differences between the predicted probabilities for the different regimes, and the probabilities that result from the identification based on the full sample. The predictions are constructed as in Table 7. The distances per regime for the regime-switching models are based on (14). The total distance uses (13), where we use the smoothed inference probabilities as the true regime probabilities ϕ. For the rules-based approaches we apply (15).

Table 9: Monetary performance of different investment strategies based on different methods

strategy		LTC	LTL	PSC	PSL	RS2C	RS2L	RS4C	RS4L
binary	mean	10.5	6.6	4.1	1.6	-0.2	2.5	-	-
	vol.	17.4	17.4	17.4	17.5	17.5	17.5	-	-
	Sharpe	0.601	0.377	0.236	0.093	-0.009	0.146	-	-
	prob.	0.107	0.257	0.404	0.558	0.650	0.500	-	-
	fee	8.0	4.1	1.7	-0.8	-2.6	0.1	-	-
proportional	mean	7.2	7.0	3.4	1.8	0.1	2.7	-	-
	vol.	15.3	15.5	12.8	15.8	12.2	12.9	-	-
	Sharpe	0.472	0.454	0.263	0.116	0.007	0.212	-	-
	prob.	0.212	0.217	0.433	0.549	0.657	0.486	-	-
	fee	4.8	4.6	1.0	-0.6	-2.3	0.3	-	-
sign	mean	10.5	10.5	4.1	1.1	-3.0	1.4	0.8	3.6
	vol.	17.4	17.4	17.4	17.5	17.5	17.5	17.5	17.5
	Sharpe	0.601	0.601	0.236	0.066	-0.171	0.078	0.049	0.209
	prob.	0.107	0.110	0.404	0.588	0.779	0.565	0.765	0.368
	fee	8.0	8.0	1.7	-1.3	-5.4	-1.1	-1.6	1.2
utility	mean	13.7	15.1	5.7	3.5	-1.3	1.3	-0.6	1.6
	vol.	30.2	31.4	24.9	29.4	9.2	12.1	8.2	11.4
	Sharpe	0.453	0.482	0.228	0.119	-0.142	0.109	-0.076	0.141
	prob.	0.123	0.098	0.351	0.463	0.787	0.595	0.833	0.586
	fee	10.9	12.3	3.1	0.8	-3.6	-1.1	-3.0	-0.8

This table shows the results of different investment strategies based on the different methods to predict bull and bear markets. The predictions are constructed as in Table 7. The binary strategy goes long in a futures contract if the probability of a bull market exceeds 0.5 and short otherwise. For a predicted probability $\xi_{u,t+1}^m$, the proportional strategy takes a position of $2\xi_{u,t+1}^m - 1$ in the futures contract. The sign strategy goes long in the futures contract, if the predicted mean μ_{t+1}^m is positive and short otherwise. The utility strategy uses both the predicted mean μ_{t+1}^m and the predicted second moment ω_{t+1}^m to construct the optimal position $w_{t+1}^m = \mu_{t+1}^m/(\gamma_t \omega_{t+1}^m)$. We take the coefficient of risk aversion γ_t equal to five. For each strategy and each method, we report the average return and the volatility in % on a yearly basis, and the Sharpe ratio. The rows labelled "prob." report the probability that the average return of a strategy is lower than the average return of a continuously rolled-over long position in a futures contract. We use 10,000 bootstraps to calculate this probability. The rows labelled "fee" give the maximum fee (in % per year) an investor is willing to pay each period to switch to a strategy from the benchmark buy-and-hold strategy. The fee is calculated to solve (22).

Figure 1: Performance US Market

This figure show the weekly observations of the US stock market in excess of the risk-free rate over the period December 26, 1979 until July 1, 2009 (12/31/1979 = 100). The excess stock market index is calculates as the ratio P_{it}/I_{it}, where P_{it} is the value of the stock market index of country i and I_{it} is the cumulation of continuously compounded risk free rate, $I_t \equiv \exp \sum_{\tau=0}^{t-1} r_\tau^{\mathrm{f}}$. For the stock market index, we use the MSCI US Index. The risk-free rates is the Financial Times/ICAP 1-Week Euro rate series. All data series are obtained from Thompson DataStream.

Figure 2: Identification of bull and bear markets

(a) LT

(b) PS

(c) RS2C

(d) RS2L

(e) RS4C

(f) RS4L
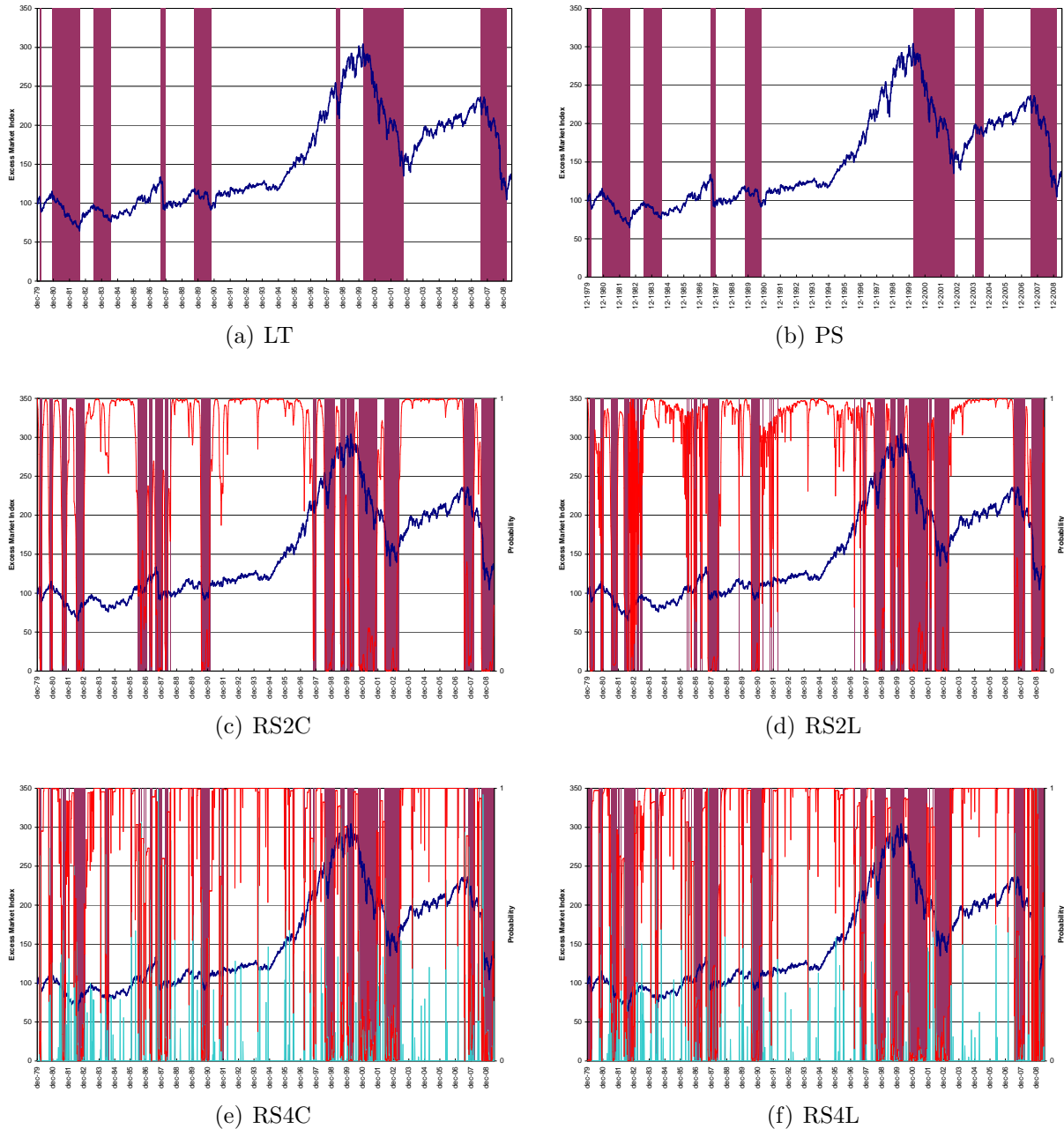
This figure shows the identification of bull and bear periods for the US, based on the different approaches. The thick blue line plots the excess stock market index (left y-axis). Purple areas indicate bear markets, and white areas correspond with bull markets. In panel (a), bull and bear markets are identified by the LT-algorithm, and in panel (b) by the PS-algorithm. Panels (c–d) shows the smoothed inference probability of a bull market (right y-axis) with a thin red line, based on a two-state regime switching model with constant transition probabilities (panel c) and time-varying transition probabilities (panel d). If this probability falls below 0.5, the month is identified as a bear market. In panels (e–f) the thin red line shows the smoothed inference probability for the extended bull regime (bull state or boom state) in a four-state regime switching model with constant transition probabilities (panel e) and time-varying transition probabilities (panel f). If this probability drops below 0.5, the month is marked as (extended) bearish. The thin light blue line shows the smoothed inference probability for the crash regime.

Figure 3: Evolution of means and volatilities per regime
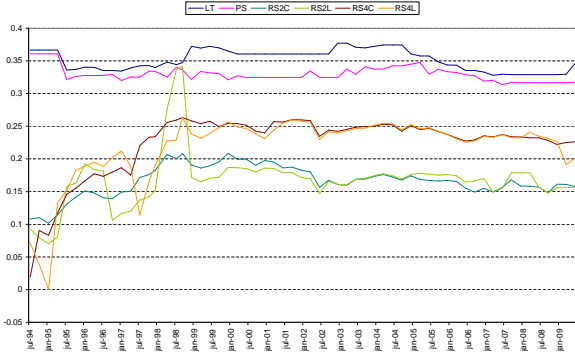
(a) mean, bull regime
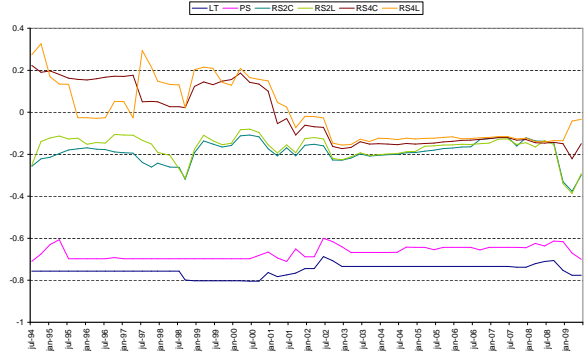
(b) mean, bear regime
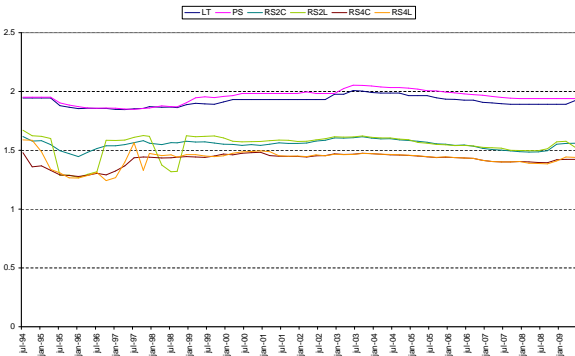
(c) volatility, bull regime
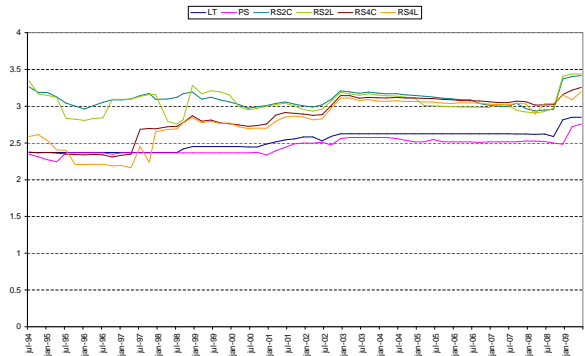
(d) volatility, bear regime
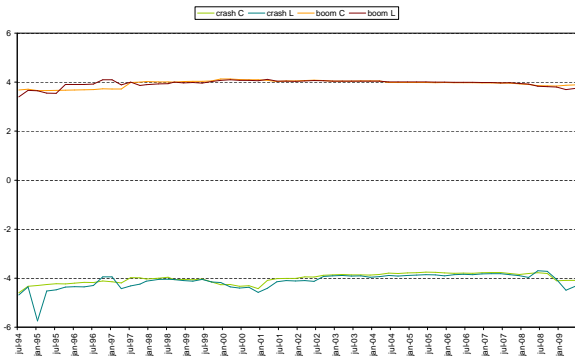
(e) mean, boom and crash regimes
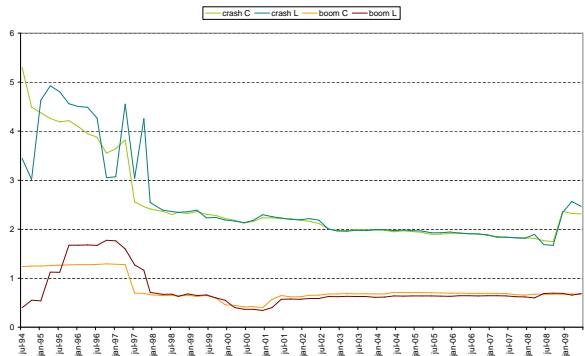
(f) volatility, boom and crash regimes
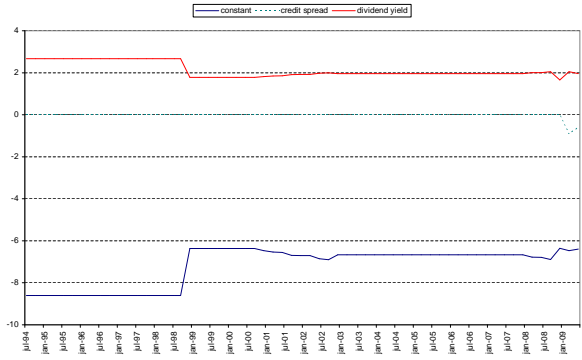
This figure shows the evolution of the means and volatilities when estimated with an expanding window (end date on the x-axis). The first window comprises the period January 2 1980 – June 30 1994 (757 observations), and is continuously expanded with 13 weeks until we reach July 1, 2009. In the approaches 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 2 regimes or 4 regimes, and constant or time-varying transition probabilities. The means and volatilities of the bull and bear market regimes follow directly from the estimation. The means and volatilities of the crash and boom regimes are constructed from their specifications in (6).

Figure 4: Evolution of constant transition probabilities
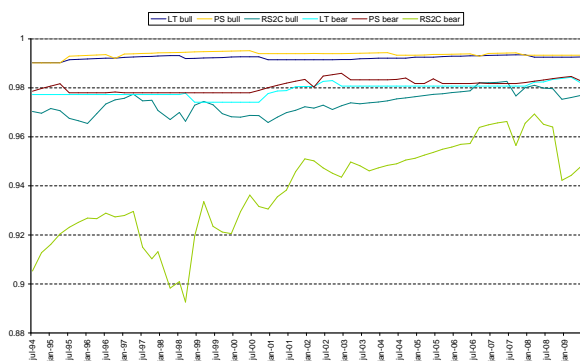
(a) 2-state models
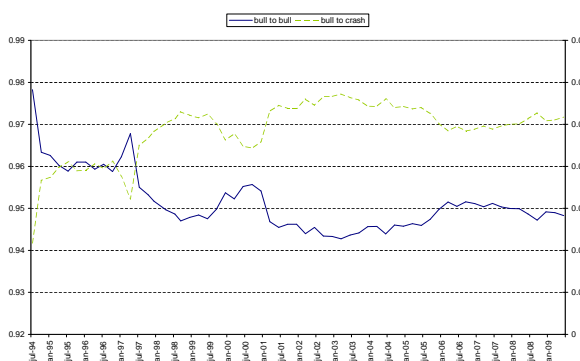
(b) RS4C, from bull regime
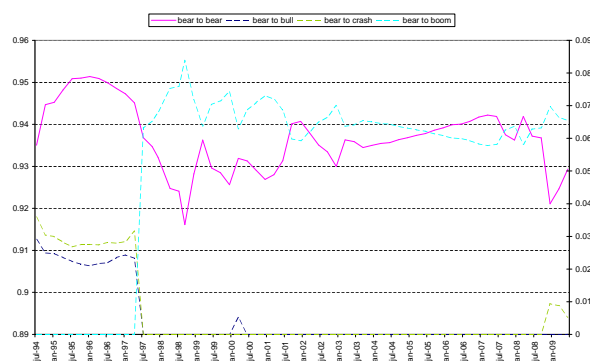
(c) RS4C, from bear regime
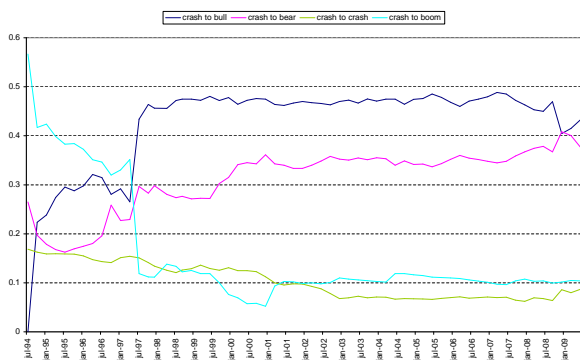
(d) RS4C, from crash regime
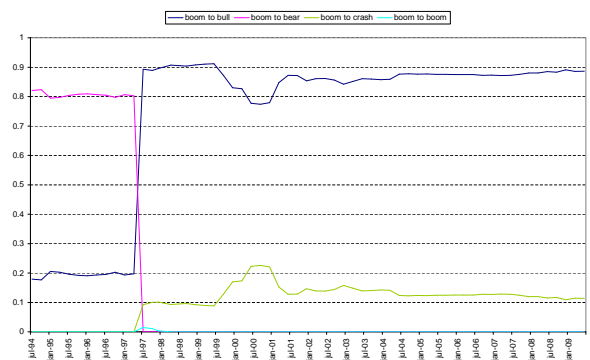
(e) RS4C, from boom regime

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 2 1980 – June 30 1994 (757 observations), and is continuously expanded with 13 weeks until we reach July 1, 2009. 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 2 regimes (RS2C) or 4 regimes (RS4C). For the methods with two states, we plot the probabilities of a bull-bull and a bear-bear switch. For the RS4C we include a subfigure for each departure state. Dashed lines correspond with the secondary y-axis. We do not show transition probabilities that never exceed 0.001.

Figure 5: Evolution of parameters in logit models
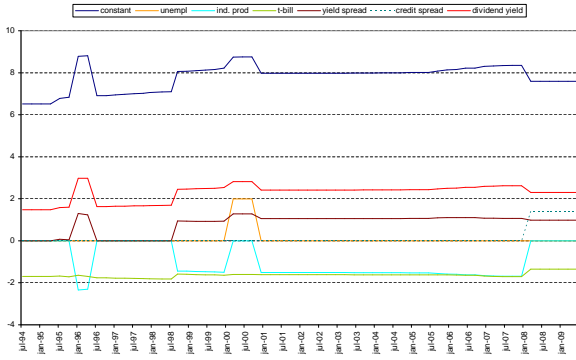
(a) LT, from bull to bull

(b) LT, from bear to bull
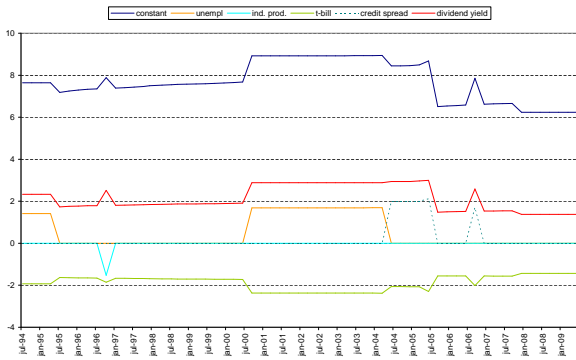
(c) PS, from bull to bull
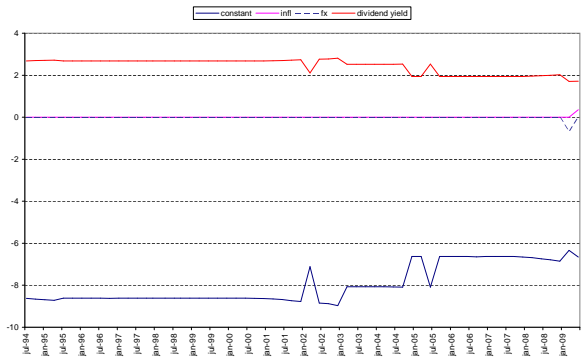
(d) PS, from bear to bull
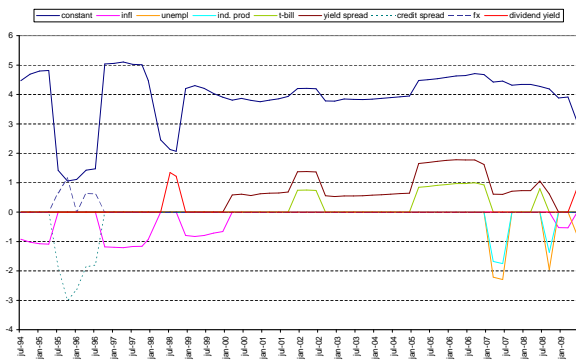
(e) RS2L, from bull to bull
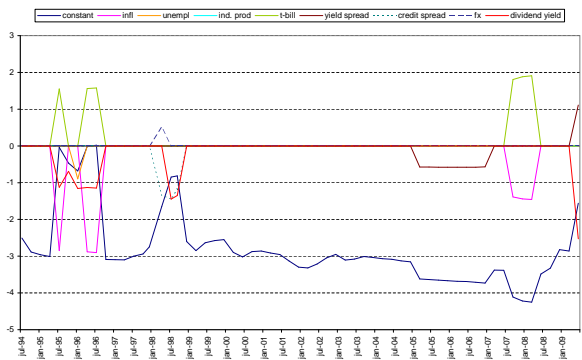
(f) RS2L, from bear to bull
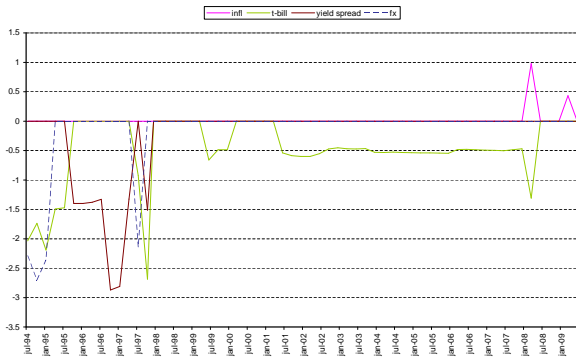
Figure note on next page.

58

Figure 5: Evolution of parameters in logit models – *continued*
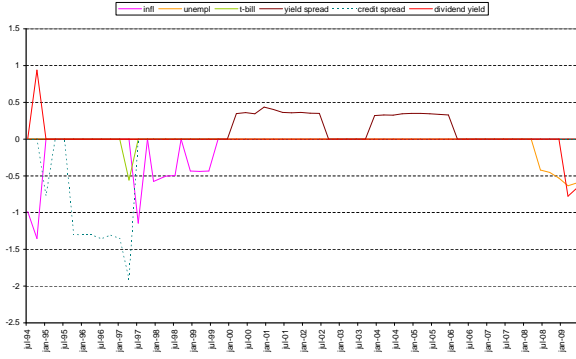

(g) RS4L, from bull to bull
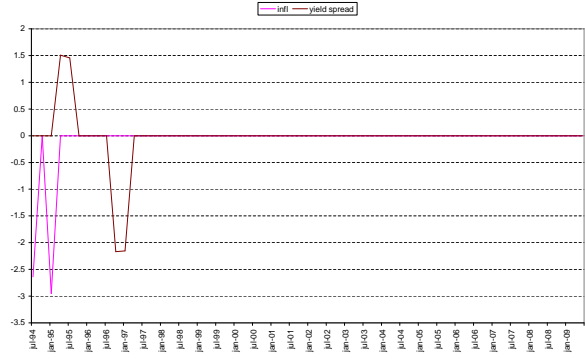

(h) RS4L, from bear to bull

(i) RS4L, from bear to bear

This figure plots the evolution of the coefficients in the (multinomial) logit transitions for the predicting variables in Table B.1, when estimated with an expanding window (end date on the x-axis). The first window comprises the period January 2 1980 – June 30 1994 (757 observations), and is continuously expanded with 13 weeks until we reach July 1, 2009. The predicting 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 Markovian logit model as in (3), where the coefficients depend on the departure state. In the two-state regime switching model, RS2L, the logistic transformation in (7) is used to link the predicting variables to the transition probabilities. For the four-state regime switching model, RS4L, the multinomial logistic transformation in (9) is used. For identification, all coefficients for a switch to the boom regime have been fixed at zero. The inclusion of predicting variables is restricted to transitions from and to bull or bear regimes. 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.
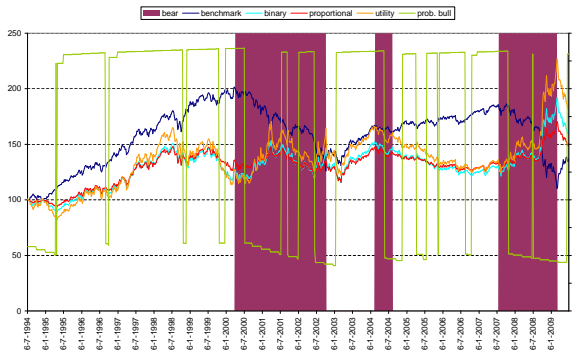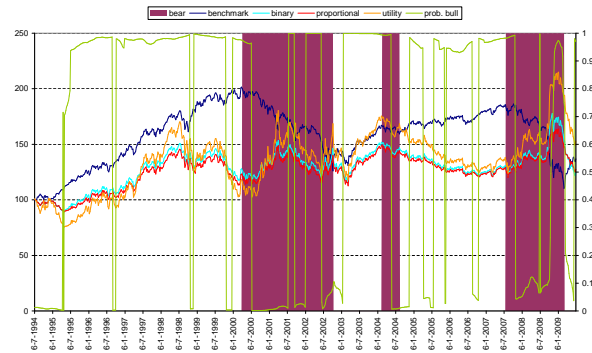
Figure 6: Predictions and performance
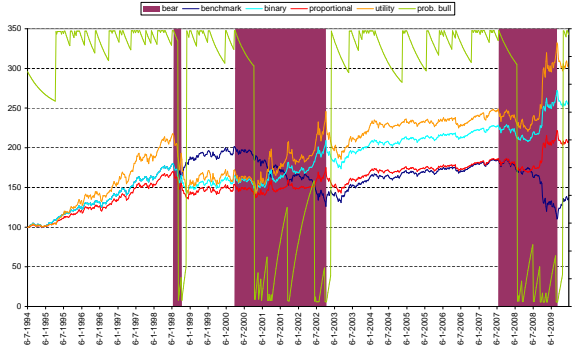
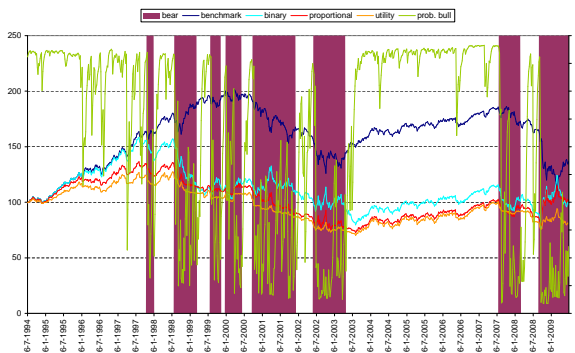(a) LT, constant transition probabilities
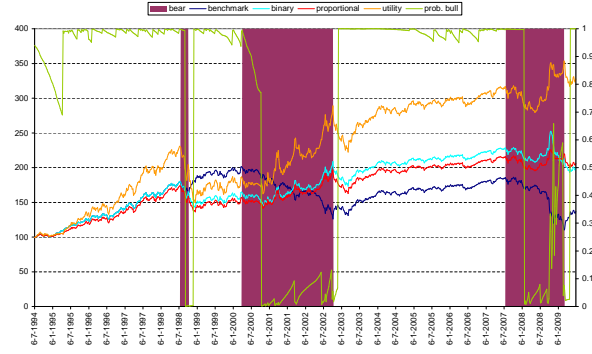
(b) LT, time-varying transition probabilities
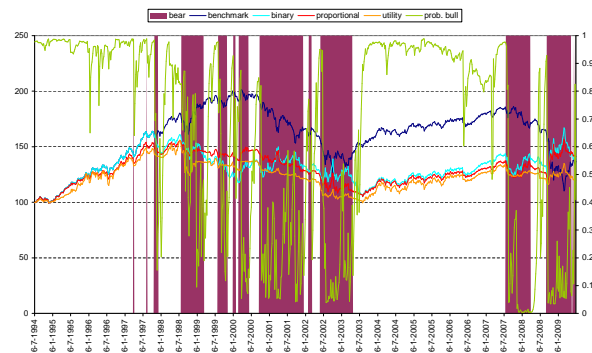
(c) PS, constant transition probabilities
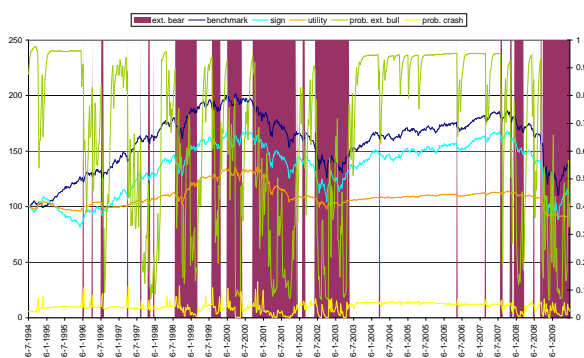
(d) PS, time-varying transition probabilities

(e) RS2C

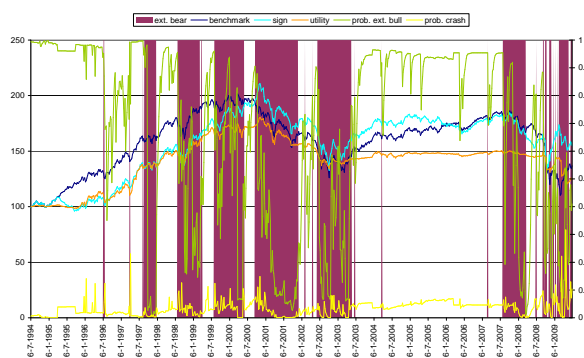(f) RS2L

Figure continues on next page.

60

Figure 6: Predictions and performance – *continued*

(g) RS4C

(h) RS4L

In this figure we show for each method the predictions and the evolution of the performance of the different strategies over time. The predictions are constructed as in Table 7. For the two-state methods, we plot the probability of a bull market (secondary y-axis). For the four-state methods, we plot the sum of the predicted probabilities for a bull period and a boom period. In that case, we also plot the probability of a crash state. The investments are determined as in Table 9. We assume that each week a base endowment of 100 is available. The result of preceding weeks is not reinvested. We also plot the result of the benchmark strategy. This strategy takes a long position in a futures contract on the market index every week. The purple areas indicate the periods of bear markets (extended bear markets for the RS4-models) as identified based on the full sample as in Figure 2.

A Multinomial logit transitions

In Section 3.2 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 predicting variables $\boldsymbol{z}_{t-1}$ by a multinomial logit transformation

$$\pi_{sq,t} \equiv \pi_{sq}(\boldsymbol{z}_{t-1}) \equiv \Pr[S_t = s | S_{t-1} = q, \boldsymbol{z}_{t-1}] = \frac{\mathrm{e}^{\boldsymbol{\beta}_{sq}' \boldsymbol{z}_{t-1}}}{\sum_{\varsigma \in \mathcal{S}} \mathrm{e}^{\boldsymbol{\beta}_{\varsigma q}' \boldsymbol{z}_{t-1}}}, \ \ s, q \in \mathcal{S}, \tag{23}$$

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

A.1 Estimation

To estimate the parameters $\boldsymbol{\beta}_{sq}$, 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. We maintain 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. Therefore, we can focus on the part of the log likelihood function related to the parameters $\boldsymbol{\beta}_{sq}$. The transition part of the expectation of the likelihood function is given by

$$\ell(B) = \sum_{t=1}^{T} \sum_{s \in \mathcal{S}} \sum_{q \in \mathcal{S}} \xi_{sq,t} \log \pi_{sq,t}, \tag{24}$$

where $B = \{\boldsymbol{\beta}_{sq} : s, q \in \mathcal{S}\}$ is the set of all parameters $\boldsymbol{\beta}_{sq}$ and $\xi_{sq,t} \equiv \Pr[S_t = s | S_{t-1} = q, \Omega_T]$ is a smoothed inference probability. These probabilities are based on the complete data set of returns and predictive 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 likelihood function. We derive the first order conditions that apply to $\boldsymbol{\beta}_{sq}$ by differentiating Eq. (24)

$$\frac{\partial \ell(B)}{\partial \boldsymbol{\beta}_{sq}} = \sum_{t=1}^{T} \sum_{\varsigma \in \mathcal{S}} \xi_{\varsigma q,t} \frac{1}{\pi_{\varsigma q}} \frac{\partial \pi_{\varsigma q}}{\partial \boldsymbol{\beta}_{sq}}.$$

Based on Eq. (23) we find

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

Combining these two expressions yields the first order condition

$$\sum_{t=1}^{T} \left(\xi_{sq,t} - \xi_{q,t-1}\pi_{sq,t}\right)\boldsymbol{z}_{t-1} = \boldsymbol{0} \quad \forall q, s \in \mathcal{S} \tag{25}$$

where $\xi_{s,t} = \Pr[S_t = q|\Omega_T]$. 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 $B_q = \{\boldsymbol{\beta}_{sq} : s \in \mathcal{S}\}$. Numerical techniques can be used to find parameters $\boldsymbol{\beta}_{sq}$ that solve this system.

A.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 $\bar{\boldsymbol{z}}$. The marginal effect is given by the first derivative of (23) with respect to z_i:

$$\left.\frac{\partial \pi_{sq}(\boldsymbol{z})}{\partial z_i}\right|_{\boldsymbol{z}=\bar{\boldsymbol{z}}} = \pi_{sq}(\bar{\boldsymbol{z}})\left(\beta_{sqi} - \sum_{\varsigma \in \mathcal{S}}\pi_{\varsigma q}(\bar{\boldsymbol{z}})\beta_{\varsigma qi}\right), \tag{26}$$

where β_{sqi} 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, $\pi_{sq}(\bar{\boldsymbol{z}})(1 - \pi_{sq}(\bar{\boldsymbol{z}}))\beta_{sqi}$.

B   Additional Empirical Results

B.1   Predictive variables

For each predictive variable we test whether it has a unit root. We report the results in Table B.1. For inflation, the yield spread and the credit spread, we reject this hypothesis, and we include these variables untransformed in our analysis. As expected, the unemployment rate is non-stationary. To circumvent seasonal effects, we transform this variable into yearly differences. For industrial production, we consider yearly growth rates. Unexpectedly, we do not reject the hypothesis of a unit root for this variable. This may be due to the autocorrelation in the series that enters by construction. We do not transform this variable any further. For the T-Bill rate and the dividend yield, we also find evidence for a unit root. These two variables show a clearly downward sloping pattern from 1980 onwards. To construct a stationary series we subtract from each observation the prior one-year average. Taking a weekly difference would produce a very erratic series, which will not be useful in our analysis. Because the traded weighted exchange rate is also non-stationary, we apply the same procedure to this variable. Table B.1 also reports the mean and standard deviation of the (transformed) series. Means and variances vary considerably over the series.

[Table B.1 about here.]

References

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

Avramov, D. and Chordia, T. (2006). Predicting stock returns. *Journal of Financial Economics*, 82:387–415.

Avramov, D. and Wermers, R. (2006). Investing in mutual funds when returns are predictable. *Journal of Financial Economics*, 81(2):339–377.

Beltratti, A. and Morana, C. (2006). Breaks and persistency: Macroeconomic causes of stock market volatility. *Journal of Econometrics*, 131(1-2):151–177.

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. 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.

Christoffersen, P. F. and Diebold, F. X. (2006). Financial asset returns, direction-of-change forecasting, and volatility dynamics. *Management Science*, 52(8):1273–1287.

Coakley, J. and Fuertes, A.-M. (2006). Valuation ratios and price deviations from fundamentals. *Journal of Banking & Finance*, 30(8):2325–2346.

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.

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.

Fleming, J., Kirby, C., and Ostdiek, B. (2001). The economic value of volatility timing. *Journal of Finance*, 56(1):329–352.

Gómez Biscarri, J. and Pérez de Gracia, F. (2004). Stock market cycles and stock market development in Spain. *Spanish Economic Review*, 6(2):127–151.

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. 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.

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.

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. (2009). Extracting bull and bear markets from stock returns. Working paper, University of Toronto, CA.

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.

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.

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

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.

Sarno, L. and Valente, G. (2004). Comparing the accuracy of density forecasts from competing models. *Journal of Forecasting*, 23(8):541–557.

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.

Table B.1: Characteristics of Predicting Variables

	Source	Frequency	AR(1)	ADF	p-value	Transformation	Mean	Std. Dev.
Inflation	FRED	monthly	0.588	-9.87	< 0.001		0.295	0.312
Unemployment	IFS	monthly	0.966	-2.17	0.218	yearly change	0.06	1.04
Ind. Prod., yearly growth rate	IFS	monthly	0.967	-2.03	0.275		2.01	4.11
Tbill rate	FRED	weekly	0.999	-1.31	0.627	change to yearly average	-0.18	1.19
Yield spread	FRED	weekly	0.989	-4.44	< 0.001		1.75	1.30
Credit spread	FRED	weekly	0.993	-3.04	0.032		1.12	0.52
Trade weighted FX	FRED	weekly	0.998	-1.18	0.6834	change to yearly average	-0.2666	4.838
Dividend Yield	DataStream	weekly	0.999	-2.00	0.287	change to yearly average	-0.05	0.31

This table shows the set of predicting variables with its source and frequency. For each variable we conduct an adjusted Dickey-Fuller test. We report the first order autocorrelation coefficient, the ADF test-statistic and the p-value for the 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. The series run from January 1980 until July 2009. The FRED data are obtained from the Federal Reserve Bank of St. Louis. The IFS data have been downloaded from DataStream.

68