Stock Volatility Predictability in Bull and Bear Markets

Xingyi Li* and Valeriy Zakamulin†

This revision: April 12, 2019

Abstract

Recent literature on stock return predictability suggests that it varies substantially across economic states being strongest during bad economic times. In line with this evidence, we document that stock volatility predictability is also state dependent. In particular, using a large data set of high-frequency data on individual stocks and a few popular time-series volatility models, in this paper we comprehensively examine how volatility forecastability varies across bull and bear states of the stock market. We find that the volatility forecast horizon is substantially longer when the market is in a bear state than when it is in a bull state. In addition, the volatility forecast accuracy is higher and forecast bias is lower when the market is in a bear state. Our study concludes that the stock volatility predictability is strongest during bad economic times proxied by bear market states.

Key words: stock markets, volatility forecasting, state dependence, high-frequency data, meta-analysis

JEL classification: C22, C53, G17

*School of Business and Law, University of Agder, Service Box 422, 4604 Kristiansand, Norway, Tel.: (+47) 38 14 13 38, E-mail: xingyilis@gmail.com

†School of Business and Law, University of Agder, Service Box 422, 4604 Kristiansand, Norway, Tel.: (+47) 38 14 10 39, E-mail: valeri.zakamouline@uia.no

1 Introduction

Volatility forecasting is crucial for portfolio management, risk management, and pricing of financial derivatives. Specifically, the volatility of a financial asset is a primary input to the optimal portfolio choice problem. Volatility forecasting is a mandatory risk-management exercise for many financial institutions and banks around the world. Volatility is the most vital input variable in the valuation of derivative securities. Specifically, to price an option one needs to know the future volatility of the underlying asset till the option maturity.

There is now an enormous body of research on the properties of volatility, volatility modeling and forecasting (for a good review of this literature, see Poon and Granger (2003)). It is well documented in financial econometric literature that volatility is forecastable over short horizons up to 1 month into the future. However, there is still a controversy in the literature about how far ahead the volatility is forecastable. On the one hand, the results of the studies by Cao and Tsay (1992), Alford and Boatsman (1995), Figlewski (1997), and Green and Figlewski (1999) seem to suggest that volatility is forecastable over long-term horizons that extend to several years. On the other hand, Christoffersen and Diebold (2000), Galbraith and Kisinbay (2005), and Raunig (2006) demonstrate that the volatility is not forecastable beyond horizon of about 30-40 trading days.

Traditionally, volatility forecasting is performed using daily data. Recent availability of intraday data and appearance of different measures of realized daily volatility makes possible to forecast volatility with higher accuracy. In fact, there is ample evidence[1] that forecasting models that use intraday data provide better forecast accuracy as compared with that delivered by the models that use daily data. Exactly how much better is still unknown.

While there is no doubt among researchers that the stock volatility is predictable, there has always been heated debate among researchers on whether the stock returns are predictable. Welch and Goyal (2008) convincingly demonstrate that the stock returns are not predictable in out-of-sample tests. This study seem to put an end to the long-standing debate. However, since that time a number of studies (examples are Schmeling (2009), Rapach, Strauss, and Zhou (2009), Henkel, Martin, and Nardari (2011), Dangl and Halling (2012), Garcia (2013),

[1]See, among others, Andersen, Bollerslev, and Lange (1999), Blair, Poon, and Taylor (2001), Andersen, Bollerslev, Diebold, and Labys (2003), Andersen, Bollerslev, and Meddahi (2005), Andersen, Bollerslev, and Diebold (2007), Martens (2002), Koopman, Jungbacker, and Hol (2005), Andersen et al. (2007), Corsi (2009), Han and Park (2013), Hansen, Lunde, and Voev (2014), and Hansen and Huang (2016).

Golez and Koudijs (2018), Cheema, Nartea, and Man (2018), and Hammerschmid and Lohre (2018)) provide evidence that the stock return predictability depends on the market state and is largely concentrated in "bad economic times". That is, the stock return predictability exists during economic recessions, bear states of the market, and periods with low investor sentiment. Moreover, the concentration of predictability in bad economic times seems to be a common empirical phenomenon in financial time series forecasting. For example, Gargano and Timmermann (2014) report that commodity return predictability varies substantially across economic states, being strongest during economic recessions. Gargano, Pettenuzzo, and Timmermann (2017) find that the degree of predictability of bond returns rises during recessions.

To the best knowledge of the authors, state dependency in volatility predictability is only accounted for in the context of Regime-Switching GARCH (RS-GARCH) models (see, among others, Klaassen (2002), Haas, Mittnik, and Paolella (2004), Marcucci (2005), Pan, Wang, Wu, and Yin (2017), and Haas and Liu (2018)). The main motivation for using RS-GARCH models is the existence of bias in forecasting future volatility using standard GARCH models. Specifically, Klaassen (2002) and then Marcucci (2005) argue that GARCH forecasts are biased and the sign of the bias depends on whether the market volatility is high or low. As a remedy for the GARCH forecast bias, these authors suggest using two-state Markov RS-GARCH models and demonstrate that RS-GARCH models deliver a better forecast accuracy as compared to that of the standard GARCH models. However, in the previous studies the researchers have not conducted a systematic examination of the GARCH volatility forecast bias[2] and the horizon of volatility predictability was limited to 1 month only.

Motivated by the recent literature on state dependency in return predictability, in this paper we comprehensively examine how the stock volatility predictability varies across the states of the market. Our study focuses on four research issues. First, we quantify the forecast accuracy across horizons. Second, we evaluate the horizon of volatility predictability. Third, we measure the forecast bias across horizons. Forth, we assess the gains from using high-frequency data for volatility forecasting. To address these issues, using realized volatility to estimate conditional volatility and a few popular time-series volatility models, we explore the

[2]In contrast, there are a great number of studies that demonstrate that the implied volatility is a biased predictor of the future volatility, see Poon and Granger (2003), Section 6, and references therein. Specifically, most often the researchers find that the implied volatility of stock index options overestimates the future realized volatility.

horizon of volatility predictability and quantify the volatility forecast accuracy and bias across all states of the market and the bull and bear states separately.

In our study we use a relatively large data set of high-frequency data on individual stocks as compared to most of the existing studies on volatility forecasting. In contrast to the previous studies that employ either one or a few financial assets, we do not report the results for each individual stock because each of these results might not be truly representative. Instead, we combine the results for all individual stocks by performing either a "meta-analysis" or running pooled OLS regressions. A key benefit of this approach is that the aggregation of information on individual stocks leads to a higher statistical power and more robust forecast accuracy and bias estimates than it is possible to obtain from the information on any individual stock.

In our study, the market states are classified into bull and bear markets. The primary reason to use bull and bear states of the market instead of, for example, periods of economic recessions and expansions is that bull and bear markets can often be identified in real time or with a minor time lag. In contrast, economic recessions and expansions are always identified with a substantial time delay. Thus, using bull and bear states is advantageous for forecasting purposes. Our paper utilizes two alternative formal turning-point procedures to objectively identify troughs and peaks in stock index series that indicate the starting and finishing points of bull and bear markets.

Our main findings can be summarized as follows. We find that the volatility forecast horizon is substantially longer when the market is in a bear state than when it is in a bull state. For instance, when high-frequency data are used, the horizon of volatility predictability extends beyond 60 weeks when the market is in a bear state. In contrast, the horizon of volatility predictability is limited to only 12 weeks when the market is in a bull state. Regardless of the state of the market, the forecast accuracy diminishes with horizon. However, the forecast accuracy diminishes much more quickly when the market is in a bull state than when it is in a bear state. Generally, the forecast accuracy is higher in bear states of the market.

The forecast accuracy gains provided by high-frequency data are largest for the shortest horizon and diminish quickly with horizon when the market is in a bull state. Conversely, when the market is in a bear state, the forecast accuracy gains persist regardless of the horizon length. Finally, the volatility forecast bias is smaller when the market is in a bear state. Our results suggest that in a bull state of the market each model tends to overestimate the future volatility and the bias increases as horizon increases. In a bear state of the market, on the other hand, each model tends to underestimate the future volatility when forecast horizon is short; when forecast horizon lengthens, most of the models tend to overestimate the future volatility.

Our study concludes that there is strong evidence that the horizon of volatility predictability, forecast accuracy, forecast bias, and the gains from using high-frequency data are state dependent. As in the previous studies on return predictability, we find that the volatility predictability is also strongest in bad economic times proxied by bear market states.

The rest of the paper is organized as follows. Section 2 describes the data and the empirical methodology. The methodology part covers the bull-bear dating algorithms, computation of measures of realized volatility using high-frequency data, volatility forecasting models, how we measure the forecast accuracy and bias, and how we conduct statistical inference. Section 3 presents the empirical results. Finally, Section 4 concludes the paper.

2 Data and Methodology

2.1 Data

We use high-frequency data on the prices of 31 stocks. The data are obtained from `kibot.com` and cover the period from January 2, 1998, to December 31, 2016. The 31 stocks represent either current or previous components of the Dow Jones Industrial Average (DJIA) index that have price data in the whole sample period. Table 1 lists the stocks included in our analysis. The price quotes are given at a 5-minute frequency from 9:30 to 16:00 Eastern Standard Time (EST) in each trading day. Trading days are usually weekdays from Monday to Friday. Additional non-trading days are 1 of January (New Year) and 25 of December (Christmas) if these days fall on weekdays. We also exclude the data for 11 of September 2001 because the stock market was closed shortly after the terrorist attacks.

The states of the stock market are identified using the data on the Standard and Poor's (S&P) 500 index. This index represents a value-weighted stock index based on the market capitalizations of 500 large companies in the US. The index was introduced in 1957 and intended to be a representative sample of leading companies in leading industries within the US economy. Stocks in the index are chosen for market size, liquidity, and industry group representation.

Table 1: **Stocks included in our analysis**

Ticker	Stock name	Ticker	Stock name
AA	Alcoa Inc.	JNJ	Johnson & Johnson
AAPL	Apple Inc.	JPM	JPMorgan Chase & Co.
AXP	American Express Company	KO	Coca-Cola Company
BA	Boeing Co.	MCD	McDonald's Corp.
BAC	Bank of America Corporation	MMM	3M Co.
C	Citigroup Inc.	MO	Altria Group Inc.
CAT	Caterpillar Inc.	MRK	Merck & Co., Inc.
CSCO	Cisco Systems Inc.	MSFT	Microsoft Corporation
DD	E. I. du Pont de Nemours and Company	NKE	Nike Inc.
DIS	Walt Disney Co.	PFE	Pfizer Inc.
GE	General Electric Company	PG	Procter & Gamble Co.
HD	Home Depot Inc.	T	AT&T Inc.
HON	Honeywell International Inc.	UNH	UnitedHealth Group Inc.
HPQ	Hewlett-Packard Company	UTX	United Technologies Corp.
IBM	International Business Machines Corp.	WMT	Wal-Mart Stores Inc.
INTC	Intel Corporation		

This index is probably the most commonly followed equity index and many consider it one of the best representations of the US stock market. The daily data on the S&P 500 index are obtained from Yahoo Finance.[3]

2.2 Dating of Bull and Bear Market States

It is an old tradition to describe the dynamics of the stock market in terms of alternating sequence of bull and bear market states. Unfortunately, there is no generally accepted formal definition of bull and bear markets in finance literature. There is a common consensus among financial analysts that a bull (bear) market denotes a period of generally rising (falling) prices. However, when it comes to the dating of bull and bear markets, financial analysts are broken up into two distinct groups. One group insists that in order to qualify for a bull (bear) market phase, the stock market price should increase (decrease) substantially. For example, the rise (fall) in the stock market price should be greater than 20% from the previous local trough (peak) in order to qualify for being a distinct bull (bear) market. The other group believes that in order to qualify for a bull (bear) name, the stock market price should increase (decrease) over a substantial period of time. For instance, the stock market price should rise (fall) over a period of longer than 5 months in order to qualify for being a distinct bull (bear) market. Since

[3] `https://finance.yahoo.com/`

there is no unique definition of bull and bear markets, there is no single preferred method to identify the state of the stock market. In our study, to detect the turning points between the bull and bear markets, we employ two[4] alternative dating algorithms: the algorithm of Pagan and Sossounov (2003) and the algorithm of Lunde and Timmermann (2004).

The algorithm of Pagan and Sossounov (2003) adopts, with slight modifications, the formal dating method used to identify turning points in the business cycle (Bry and Boschan (1971)). The algorithm is based on a complex set of rules and consists of two main steps: determination of initial turning points in raw data and censoring operations. In order to determine the initial turning points, first of all one uses a window of length $\tau_{\text{window}} = 8$ months on either side of the date and identifies a peak (trough) as a point higher (lower) than other points in the window. Second, one enforces the alternation of turning points by selecting highest of multiple peaks and lowest of multiple troughs. Censoring operations require: eliminating peaks and troughs in the first and last $\tau_{\text{censor}} = 6$ months; eliminating cycles[5] that last less than $\tau_{\text{cycle}} = 16$ months; and eliminating phases that last less than $\tau_{\text{phase}} = 4$ months unless the market move exceeds $\theta = 20\%$.

The algorithm of Lunde and Timmermann (2004) is based on imposing a minimum on the price change since the last peak or trough. This dating rule is implemented in the following manner. Let λ_{Bull} be a scalar defining the threshold of the movement in stock prices that triggers a switch from a bear state to a bull state, and let λ_{Bear} be the threshold for shifts from a bull state to a bear state. Denote by P_t the stock market price at time t and suppose that a trough in P has been detected at time $t_0 < t$. Therefore the algorithm knows that a bull state begins from time $t_0 + 1$. The algorithm first finds the maximum value of P on the time interval $[t_0, t]$

$$P_{t_0,t}^{\max} = \max\{P_{t_0}, P_{t_0+1}, \dots, P_t\}$$

and then computes the (inverse of the) relative change in P where the maximum value serves

[4]Another approach to dating bull and bear markets is based on using a regime switching model, see, for example, Maheu and McCurdy (2000). However, significant drawbacks of this approach are as follows. First, a detected turning point usually does not coincide with a historical peak or trough in stock prices. Second, using only two states in a regime switching model is usually not enough to describe the number of regimes in a stock market, see Maheu, McCurdy, and Song (2012).

[5]A cycle denotes two subsequent phases, either upswing and consequent downswing, or downswing and consequent upswing.

as the reference value

$$\delta_t = \frac{P^{\max}_{t_0,t} - P_t}{P^{\max}_{t_0,t}}.$$

If $\delta_t > \lambda_{Bear}$, then a new peak is detected at time t_{peak} at which P attains maximum on $[t_0, t]$. The period $[t_0 + 1, t_{\text{peak}}]$ is labeled as a bull state. A bear state begins from $t_{\text{peak}} + 1$.

If, on the other hand, a peak in P has been detected at time $t_0 < t$, then the algorithm finds the minimum value of P on the time interval $[t_0, t]$

$$P^{\min}_{t_0,t} = \min\{P_{t_0}, P_{t_0+1}, \ldots, P_t\}$$

and computes the relative change in P from the minimum value

$$\delta_t = \frac{P_t - P^{\min}_{t_0,t}}{P^{\min}_{t_0,t}}.$$

If $\delta_t > \lambda_{Bull}$, then a new trough is detected at time t_{trough} at which P attains minimum on $[t_0, t]$. The period $[t_0 + 1, t_{\text{trough}}]$ is labeled as a bear state. A bull state begins from $t_{\text{trough}} + 1$.

The application of the algorithm of Lunde and Timmermann (2004) requires making an arbitrary choice of two parameters $\{\lambda_{Bull}, \lambda_{Bear}\}$. Since it is unclear how to make an appropriate choice, Lunde and Timmermann (2004) report the empirical results for many alternative sets of parameters. They consider both symmetrical (for example, $\lambda_{Bull} = \lambda_{Bear} = 20\%$) and asymmetrical (for example, $\lambda_{Bull} = 20\% > \lambda_{Bear} = 15\%$) choices.

2.3 Computation of Daily Return and Realized Volatility

Using the high-frequency stock price data, we compute the intraday logarithmic returns for each trading day t. Denote by n the number of 5-minute intervals in each trading day. The construction of the daily return on day t is given by

$$r_t = r_{t,N} + r_{t,D}, \tag{2.1}$$

where $r_{t,N}$ is the overnight return on day t (the close-to-open return) and $r_{t,D}$ is the intraday return on day t (the open-to-close return). Specifically, the overnight return is the return between 16:00 EST on day $t-1$ and 9:30 EST on day t. The intraday return is the sum of n

5-minute interval returns

$$r_{t,D} = \sum_{i=1}^{n} r_{t,i}, \tag{2.2}$$

where $r_{t,i}$ denotes the intraday return on day t for interval i.

It is commonly acknowledged that squared daily returns provide a poor approximation of actual daily variance. If asset price is assumed to follow a continuous time diffusion process, Merton (1980) showed that the daily variance of the process can be approximated using the sum of intraday squared returns. This result was later generalized by Andersen and Bollerslev (1998) and Andersen, Bollerslev, Diebold, and Labys (2001). Consequently, our first measure of daily realized volatility is the square root of the sum of the squared overnight and intraday returns

$$RV_t^{ON} = \sqrt{r_{t,N}^2 + RV_{t,D}^2}, \tag{2.3}$$

where

$$RV_{t,D} = \sqrt{\sum_{i=1}^{n} r_{t,i}^2}. \tag{2.4}$$

This measure of daily realized volatility is commonly used to estimate the daily realized volatility of an exchange rate. This measure is also used to estimate the daily realized volatility of a stock (see, among others, Blair et al. (2001), Martens (2002), Martens and Zein (2004), Koopman et al. (2005), and Todorova and Soucek (2014)).

Our second estimator of the daily realized volatility was proposed by Hansen and Lunde (2005b). The idea behind this estimator is as follows. Observe that RV_t^{ON} represents the square root of a linear combination of $r_{t,N}^2$ and $RV_{t,D}^2$ with weights $(1, 1)$. Hansen and Lunde (2005b) consider the general case of this linear combination in the following form

$$RV_t^{HL} = \sqrt{\omega_1 r_{t,N}^2 + \omega_2 RV_{t,D}^2}, \tag{2.5}$$

where (ω_1, ω_2) are the weights of $r_{t,N}^2$ and $RV_{t,D}^2$ respectively. The approach entertained in Hansen and Lunde (2005b) is to compute the realized volatility using the optimal combination of $r_{t,N}^2$ and $RV_{t,D}^2$. The optimal weights (ω_1, ω_2) are found by minimizing the mean squared error (MSE) criterion. The solution to the optimal weights is given by (for the detailed procedure of finding the solution, see Hansen and Lunde (2005b))

$$\omega_1 = (1-\varphi)\frac{\mu_0}{\mu_1} \qquad \text{and} \qquad \omega_2 = \varphi\frac{\mu_0}{\mu_2}, \tag{2.6}$$

where φ is a relative importance factor defined by

$$\varphi = \frac{\mu_2^2\eta_1^2 - \mu_1\mu_2\eta_{12}}{\mu_2^2\eta_1^2 + \mu_1^2\eta_2^2 - 2\mu_1\mu_2\eta_{12}}, \tag{2.7}$$

where μ_1, μ_2, and μ_0 denote the expectations of the overnight squared return $r_{t,N}^2$, the realized variance during the active part of the day $RV_{t,D}^2$, and their sum $r_{t,N}^2 + RV_{t,D}^2$ respectively, whereas η_1, η_2, and η_{12} denote the variance of $r_{t,N}^2$, the variance of $RV_{t,D}^2$, and the covariance between $r_{t,N}^2$ and $RV_{t,D}^2$ respectively.

A small drawback of the estimator RV_t^{HL} is that in some rare cases the weight ω_1 may take a negative value. As a consequence, the daily realized variance may potentially be negative when $r_{t,N}^2$ is extremely large. Similarly to Todorova and Soucek (2014), in the seldom cases where the realized daily variance takes a negative value, we replace RV_t^{HL} by RV_t^{ON}. This replacement is equivalent to using weights $\omega_1 = 1$ and $\omega_2 = 1$ in the estimator RV_t^{HL}.

2.4 Forecasting Models

In our study we use four models to forecast conditional volatility. The first two models use only daily returns data to forecast conditional volatility. That is, these forecasts do not use information present in the high-frequency data. The other two models use both daily returns and daily measures of realized volatility to forecast conditional volatility. Each model is able to produce a multi-step ahead conditional volatility forecast by performing a series of one-step ahead rolling forecasts. Comparing the results from several models allows us to assess the sensitivity of results with respect to the choice of a model for conditional volatility. In addition, our choice of models allows us to assess the gains in forecast accuracy from using high-frequency data for conditional volatility forecasting.

We assume that the daily logarithmic return process of a stock is given by

$$r_t = \mu + \sigma_t\epsilon_t, \quad \epsilon_t \sim \text{i.i.d.}(0,1), \tag{2.8}$$

where μ is the daily long-run mean of r_t, σ_t is the daily standard deviation, and ϵ_t is an i.i.d. process with zero mean and unit variance.

Our first forecasting model is the well-known Exponentially Weighted Moving Average (EWMA) model popularized by the RiskMetrics™ group (Longerstaey and Spencer (1996)). The one-step ahead variance forecasting equation in this model is given by

$$\hat{\sigma}_{t+1}^2 = (1-\lambda)(r_t - \mu)^2 + \lambda \hat{\sigma}_t^2, \tag{2.9}$$

where λ is the so-called "decay factor". The optimal decay factor is estimated for each individual stock by minimizing the Mean Squared Error (MSE) of daily forecast. The EWMA model assumes that the conditional variance is highly persistent. Specifically, in this model the conditional variance forecast for day $t+i$ equals that for day $t+1$.

We employ the most widely used GARCH(1,1) model, proposed by Bollerslev (1986), as the second conditional variance forecasting model that uses daily data only. In this model the latent daily conditional variance is assumed to evolve according to the following process

$$\sigma_{t+1}^2 = \omega + \alpha(r_t - \mu)^2 + \beta \sigma_t^2, \tag{2.10}$$

where the parameters α, β, and ω are estimated using daily returns by the maximum likelihood method. The one-step ahead variance forecast for day $t+1$ is given by equation (2.10). The conditional variance for day $t+2$ is forecasted using the fact that $E\left[(r_{t+1} - \mu)^2\right] = \sigma_{t+1}^2$. As a result, beginning from day $t+2$ the rolling one-day ahead variance forecast is given by

$$\hat{\sigma}_{t+i+1}^2 = \omega + (\alpha + \beta)\hat{\sigma}_{t+i}^2. \tag{2.11}$$

Our third forecasting model is the Realized GARCH(1,1) (RealGARCH) model of Hansen, Huang, and Shek (2012) which provides a framework for the joint modelling of returns and realized measures of variance. Specifically, the RealGARCH model relates the observed realized variance to the latent variance via a measurement equation, which also includes asymmetric reaction to shocks. Formally, the log-linear specification[6] of the joint model for returns and

[6]Hansen et al. (2012) advocate that the log-linear specification of RealGARCH model should be preferred to the linear specification.

realized variance is given by equation (2.8) combined with the following two equations

$$\log \sigma_{t+1}^2 = \omega + \alpha \log RV_t^2 + \beta \log \sigma_t^2, \tag{2.12}$$

$$\log RV_{t+1}^2 = \xi + \delta \log \sigma_{t+1}^2 + f(\epsilon_{t+1}) + u_{t+1}, \tag{2.13}$$

where ω, α, β, ξ, and δ are the model parameters. Equation (2.13) relates the observed realized variance to the latent variance and is therefore called the "measurement equation". In this equation $u_{t+1} \sim \text{i.i.d.}(0, \sigma_u^2)$ and function $f(\epsilon_{t+1})$ models the asymmetric reaction to shocks. The model is estimated using the method of maximum likelihood.

In the canonical version of the RealGARCH model, $E[f(\epsilon_{t+1})] = 0$ and, therefore, this function disappears from the forecast equations. The one-step ahead conditional variance forecast for day $t+1$ is given by equations (2.12) and (2.13). Starting from day $t+2$, the rolling one-day ahead variance forecast is specified by the following pair of joint formulas

$$\log \hat{\sigma}_{t+i+1}^2 = \omega + \alpha \log \widehat{RV}_{t+i}^2 + \beta \log \hat{\sigma}_{t+i}^2, \tag{2.14}$$

$$\log \widehat{RV}_{t+i+1}^2 = \xi + \delta \log \hat{\sigma}_{t+i+1}^2. \tag{2.15}$$

Finally, our forth forecasting model represents the original HAR-RV model proposed by Corsi (2009) augmented by two additional regressors. This model is a simple autoregressive-type model where the volatility is forecasted using several past volatilities realized over different time horizons

$$\hat{\sigma}_{t+1} = \beta_0 + \beta_1 RV_t^{(1)} + \beta_2 RV_t^{(5)} + \beta_3 RV_t^{(21)} + \beta_4 RV_t^{(63)} + \beta_5 RV_t^{(126)} + \varepsilon_t, \tag{2.16}$$

where $\varepsilon_t \sim \text{i.i.d.}(0, \sigma_\varepsilon^2)$ and $RV_t^{(\tau)}$ denotes the time t average realized volatility over the past τ days

$$RV_t^{(\tau)} = \frac{1}{\tau} \sum_{i=0}^{\tau-1} RV_{t-i}, \qquad \tau \in \{1, 5, 21, 63, 126\}. \tag{2.17}$$

For example, $RV_t^{(5)}$ denotes the time t weekly realized volatility which is the average realized volatility over 5 consecutive days beginning from day $t-4$. In his original model, Corsi uses only daily, weekly, and monthly realized volatilities as regressors (that is, $\tau \in \{1, 5, 21\}$); the volatility is forecasted for maximum 2 weeks ahead. Since in our empirical study we forecast future volatility over much longer horizons, we augment the original HAR-RV model by two additional regressors: past 3-month and 6-month average realized volatilities. The one-step ahead conditional volatility forecast for day $t+1$ is given by equation (2.16). Starting from day $t+2$, the rolling one-day ahead volatility forecast for day $t+i+1$ is obtained by recursively substituting the forecast $\hat{\sigma}_{t+i}$ for the future daily RV_{t+i} into the relevant regressors $RV_{t+i}^{(\tau)}$ that appear on the right-hand-side of equation (2.16).

2.5 Measuring Forecast Accuracy

Let t denote the present time and let d denote the forecast horizon length which is measured in the number of trading days. As a rule, the procedure for measuring the forecast accuracy across various horizons is performed as follows. One starts with predicting the volatility across a set of increasing horizons $\hat{\sigma}_{t,t+d}$ where the subscript $t, t+d$ denotes the time t predicted volatility over horizon of length $d \in [1, 2, \ldots, T]$. That is, $\hat{\sigma}_{t,t+d} = \sqrt{\sum_{i=1}^{d} \hat{\sigma}_{t+i}^2}$. Then, one computes the realized volatility across the same set of horizons, $RV_{t,t+d} = \sqrt{\sum_{i=1}^{d} RV_{t+i}^2}$. Finally, one compares $\hat{\sigma}_{t,t+d}$ with $RV_{t,t+d}$ and measures the forecast accuracy for each horizon.

The procedure described above has several drawbacks. In the context of our empirical study, two of them deserve mentioning. First, the procedure provides useful, but at the same time misleading information about the model's ability to forecast volatility across a specific horizon. In particular, since the multi-step-ahead variance is the sum of single-period variances, and it is known that the volatility forecastability decays quickly with horizon (see Christoffersen and Diebold (2000) and Galbraith and Kisinbay (2005)), the volatility for each subsequent period is forecasted with decreasing accuracy. As a result, the standard procedure measures the average forecast accuracy over all periods that span the forecast horizon. Therefore, the drawback of the standard procedure is that it is not able to provide answers to the following questions: How accurate is the volatility forecast for each period in the future? What is the maximum horizon at which future single-period volatility is forecastable?

Second, even though in the majority of empirical studies on multi-period volatility predictability the volatility is typically forecasted for a number of trading days ahead, the results of these studies are plagued by the presence of the day-of-the-week (a.k.a. intraweek) volatility effect. For example, the variance of returns over the period from Friday close to Monday close is higher than the variance from Monday close to Tuesday close. Fama (1965) and French and Roll

(1986) estimated that the volatility from Friday close to Monday close was about 20% higher than that between two subsequent trading days. More recent estimates provided by Fleming, Ostdiek, and Whaley (1995) and Hansen and Lunde (2005b) suggest that the volatility from Friday close to Monday close is about 10% higher than the average trading day volatility. In addition, it is known that the average daily volatility is not constant across the different days of the week. Whereas Fleming et al. (1995) observed that the average daily volatility is monotonically decreasing from Monday to Friday, Martens, van Dijk, and de Pooter (2009) reported that the average daily volatility displays a rather pronounced U-shaped pattern, with volatility being lowest on Wednesdays.

To avoid the drawbacks of the standard procedure for measuring the forecast accuracy across various horizons, we evaluate the volatility forecast accuracy for each week h in the future. By a "week" we mean a period of 5 trading days. We denote by $\hat{\sigma}_{t,h}$ and $RV_{t,h}$ the time t predicted and realized volatilities, respectively, for week $h \geq 1$ in the future. These quantities are computed as

$$\hat{\sigma}_{t,h} = \sqrt{\sum_{i=1}^{5} \hat{\sigma}^2_{t+5(h-1)+i}} \text{ and } RV_{t,h} = \sqrt{\sum_{i=1}^{5} RV^2_{t+5(h-1)+i}}. \tag{2.18}$$

Note that by aggregating the conditional volatility over 5 consecutive trading days we remove the intraweek volatility effect in the forecasted data. This is because with weekly aggregation we forecast the conditional volatility from Friday close to Friday close, then from Monday close to Monday close, etc. In addition, by comparing $\hat{\sigma}_{t,h}$ with $RV_{t,h}$ we are able to measure the forecast accuracy for each week in the future and answer the question about the maximum horizon at which conditional volatility for week h in the future is forecastable.

Given the model-based h-week-ahead forecast of weekly conditional volatility $\hat{\sigma}_{t,h}$ and the estimated realized volatility for the same week $RV_{t,h}$, it is non-trivial to evaluate the forecast accuracy. There is not a unique criterion for selecting the best forecast accuracy measure. In the rest of this section, we motivate for our choice of the accuracy measure and describe how we conduct statistical inference about estimated forecast accuracies.

2.5.1 The Choice of Accuracy Measure

The volatility model's ability to make accurate predictions of the realized volatility has often been measured in terms of the R^2 from the following regression (see, among others, Engle and Patton (2001), Hansen and Lunde (2005a), and Poon (2005))

$$RV_{t,h} = \alpha + \beta\,\hat{\sigma}_{t,h} + u_{t,h}. \tag{2.19}$$

However, a serious drawback in using regression (2.19) is that a high R^2 can be obtained even in the presence of a large forecast bias (the forecast is unbiased only if $\alpha = 0$ and $\beta = 1$). Therefore, in the vast majority of studies on forecasting volatility the researchers compare forecasting performance of competing models in terms of loss functions. The problem is that it is not possible to identify a unique and natural criterion for the comparison.

The standard procedure for assessing the forecast accuracy usually starts with evaluating the forecast errors

$$e_{t,h} = RV_{t,h} - \hat{\sigma}_{t,h}. \tag{2.20}$$

Using the forecast errors, one computes one or several evaluation measures. Two most popular evaluation measures used in the literature are the Mean Squared Error (MSE) and the Mean Absolute Error (MAE). The disadvantages of these measures are as follows. First, since they measure the absolute magnitude of errors, they can be used for comparing forecasting models on a single dataset only. Second, even though they allow comparing alternative forecasting models, they do not allow measuring predictive accuracy *per se*. Specifically, if the volatility over some forecast horizon is unpredictable, all model forecasts are likely to be worthless. In this case using, for example, the MSE criterion (to select the best model among the poor ones) creates the illusion of predictability when none is present.

In the context of return predictability, one of the most popular forecast accuracy measures is the out-of-sample R^2 statistic popularized by Campbell and Thompson (2007). A similar measure is used by Galbraith and Kisinbay (2005) in the context of volatility predictability. In our notation, the computation of this measure is given by

$$R^2_{OS}(h) = 1 - \frac{\sum_{t=1}^{M}\left(RV_{t,h} - \hat{\sigma}_{t,h}\right)^2}{\sum_{t=1}^{M}\left(RV_{t,h} - \widetilde{RV}_{t,h}\right)^2}, \tag{2.21}$$

where M is the number of forecasts in the out-of-sample period and $\widetilde{RV}_{t,h}$ is the historical average weekly volatility estimated till time t. It is worth noting that the Campbell and Thompson out-of-sample R^2 statistic, R^2_{OS}, is not really an R^2 statistic. In fact, R^2_{OS} measures the proportional reduction in the sum of squared errors obtainable relative to the unconditional volatility forecast.

The R^2_{OS} statistic overcomes both the drawbacks of using a loss function to measure the forecast accuracy: it is a scale-free measure which value can be conveniently reported in percentages. Thus, it can be used for comparing forecasting models on several datasets and for measuring predictive accuracy *per se*. However, since the idea behind the R^2_{OS} is to compare the conditional volatility forecast with the long-run historical mean volatility (that is, the unconditional volatility forecast), in order to obtain meaningful results one needs to insure that the average volatility in the initial in-sample period is not much different from the long-run historical mean volatility. This requirement can be fulfilled when the length of the in-sample period is sufficiently long. In cases where the length of the in-sample period is rather short and the period is characterized by volatility which is far away from the long-run mean, the use of R^2_{OS} tends to overestimate the forecast accuracy.

The potential estimation bias issue in using the R^2_{OS} can be avoided by replacing the historical mean volatility in the in-sample period by the historical mean volatility in the out-of-sample period. Specifically, to measure the forecast accuracy, we are going to employ the proportion of variance explained by the forecasts (this measure was proposed by Blair et al. (2001)):

$$P(h) = 1 - \frac{\sum_{t=1}^{M} \left(RV_{t,h} - \hat{\sigma}_{t,h}\right)^2}{\sum_{t=1}^{M} \left(RV_{t,h} - \overline{RV}_{t,h}\right)^2} = 1 - \frac{SSE(h)}{TSS(h)}, \tag{2.22}$$

where SSE denotes the Sum of Squared Errors, TSS denotes the Total Sum of Squares, and $\overline{RV}_{t,h}$ is the mean value of the weekly realized volatility in the out-of-sample period

$$\overline{RV}_{t,h} = \frac{1}{M} \sum_{t=1}^{M} RV_{t,h}. \tag{2.23}$$

Notice that the computation of P is similar to the computation of the out-of-sample R^2 in the constrained linear regression model (2.19) with zero intercept and unit slope. Thus, the notion of "out-of-sample R^2" suits better to P than to R^2_{OS}. Additionally, notice that the smaller the respective SSE, the closer P to 100%. Given that P is equivalent to an R^2 in the restricted model, it is likely to be smaller than conventional R^2. The value of P can even be negative since the ratio SSE/TSS can be greater than 1. A negative P indicates that the forecast errors have a greater amount of variations than the actual volatility, which means that a forecasting model does not have any predictive power. Finally, it is worth noting that when the out-of-sample period is sufficiently long, then $\overline{RV}_{t,h} \approx \widetilde{RV}_{t,h}$. In words, the mean value of the weekly realized volatility in the out-of-sample period equals the historical mean value of the weekly realized volatility. Hence, the computation of P is very much alike the computation of R^2_{OS}.

2.5.2 Statistical Inference and Meta-Analysis

To the best knowledge of the authors, in all previous studies that examine the horizon of volatility predictability the researchers employ only a few financial assets and report the results for each individual asset. In contrast, we perform a so-called "meta-analysis" that combines the results for all individual assets. A key benefit of this approach is that the aggregation of information on individual assets leads to a higher statistical power and more robust forecast accuracy estimates than it is possible to obtain from the information on any individual asset.

Our meta-analysis starts with the computation of the forecast accuracy for a h-weeks-ahead forecast of the weekly conditional volatility for all stocks in our data set. Denoting by k the number of individual stocks and by $P_j(h)$ the forecast accuracy for stock j, our pooled estimate of the h-weeks-ahead volatility forecast accuracy is the average forecast accuracy across all stocks

$$\overline{P}(h) = \sum_{j=1}^{k} P_j(h). \tag{2.24}$$

For each stock we conduct statistical inference about estimated forecast accuracies. Specifically, we test the following null hypothesis:

$$H_0 : P_j(h) \leq 0, \quad j = 1, 2, \ldots, k. \tag{2.25}$$

In words, the null hypothesis assumes the absence of predictive ability over horizon of length h weeks. We postpone the description of the procedure for computing the p-value of each individual test until the end of this section. After having computed all p-values, we combine the results of individual tests of the null hypothesis to ask whether there is evidence from the collection of individual tests that might reject the null hypothesis. In other words, we combine k p-values for all stocks to test whether collectively they can reject a common null hypothesis of no predictive ability.

When the individual tests of significance of the forecast accuracy $P_j(h)$ are independent, Fisher's method (Fisher (1925)) of combining the probabilities is asymptotically optimal among essentially all methods of combining the results of independent tests (Littell and Folks (1971)). The method is to compute the following test statistic

$$\Psi(h) = \sum_{j=1}^{k} -2\log\, p_j(h), \tag{2.26}$$

where $p_j(h)$ denotes the p-value of the hypothesis test of the absence of predictive ability over horizon of length h for stock j. Fisher demonstrated that for independent tests the statistic $\Psi(h)$ follows a chi-squared distribution with $2k$ degrees of freedom, $\Psi(h) \sim \chi^2_{2k}$.

Brown (1975) extended the Fisher's method to the case where individual tests of significance are dependent. In the dependent case, the statistic $\Psi(h)$ has the following mean and variance

$$E[\Psi(h)] = 2k, \quad Var[\Psi(h)] = 4k + 2\sum\sum_{m<j} Cov\left(-2\log\, p_m(h), -2\log\, p_j(h)\right), \tag{2.27}$$

where $Cov(x, y)$ represents the covariance between x and y. The covariance between $-2\log\, p_m(h)$ and $-2\log\, p_j(h)$ is a function only of the correlation coefficient between $P_m(h)$ and $P_j(h)$ (see Brown (1975)).

Brown's method is based on the assumption that the distribution of $\Psi(h)$ can be approximated by that of $c\chi^2_{2f}$ where c represents a re-scaling constant and χ^2_{2f} is a chi-squared distribution with $2f$ degrees of freedom. Brown calculated c and f by equating the first two moments of $\Psi(h)$ and $c\chi^2_{2f}$ resulting in

$$f = \frac{E[\Psi(h)]^2}{Var[\Psi(h)]}, \quad c = \frac{Var[\Psi(h)]}{2E[\Psi(h)]} = \frac{k}{f}. \tag{2.28}$$

The combined p-value is then given by

$$\overline{p}(h) = 1.0 - \Phi_{2f}\left(\frac{\Psi(h)}{c}\right), \tag{2.29}$$

where Φ_{2f} is the cumulative distribution function of χ^2_{2f}. The covariances in (2.27) can be evaluated using either a numerical integration or by Gaussian quadrature. We follow the original Brown's method and use a Gaussian quadrature that approximates the covariance $Cov\left(-2\log p_m(h), -2\log p_j(h)\right)$ by two quadratic functions of the correlation coefficient $\rho_{mj}(h)$ between $P_m(h)$ and $P_j(h)$.

To estimate the p-values of individual tests, $p_j(h)$, we need to know the distribution of $P_j(h)$ for all $1 \leq j \leq k$. In addition, to estimate the correlation coefficients $\rho_{mj}(h)$, we need to know the joint distribution of $P_m(h)$ and $P_j(h)$ for all $1 \leq m < j \leq k$. In order to estimate the p-values of the individual tests and the correlation coefficients between individual forecast accuracies, we adapt the pairs bootstrap to handle the dependence structure in our data via bootstrapping using blocks of data.

Specifically, we estimate the joint distribution of all forecast accuracies by resampling the original data used to compute all $P_j(h)$. These data are represented by two $M \times k$ matrices **SSE** and **TSS**. The jth column of matrix **SSE** contains the vector of the squared forecast errors for stock j where the typical element is given by $(RV_{t,h} - \hat{\sigma}_{t,h})^2$. The jth column of matrix **TSS** contains the vector of the squared variations of the realized weekly volatility for stock j where the typical element is given by $\left(RV_{t,h} - \overline{RV}_{t,h}\right)^2$.

Our paired block-bootstrap method is conducted by carrying out $N = 1000$ bootstrap trials in total. Each bootstrap trial consists of 3 steps. First, using the stationary block-bootstrap method of Politis and Romano (1994), a vector $\mathbf{I}^*$ of length M is obtained by resampling (with replacement) blocks of consecutive elements of vector $\mathbf{I} = [1, 2, \ldots, M]$. The optimal block length is selected automatically using the method proposed by Politis and White (2004) and subsequently improved by Patton, Politis, and White (2009).[7] Second, the bootstrapped resamples $\mathbf{SSE}^*$ and $\mathbf{TSS}^*$ are constructed using elements of vector $\mathbf{I}^*$ to address the row indices in the original matrices **SSE** and **TSS**. Third, matrices $\mathbf{SSE}^*$ and $\mathbf{TSS}^*$ are used to compute

[7] In particular, we estimate the optimal block length for each column of matrices **SSE** and **TSS**. The stationary block-bootstrap method is implemented using the average block length over $2k$ estimated optimal block lengths.

the elements of vector $\mathbf{P}^* = [P_1^*(h), P_2^*(h), \ldots, P_k^*(h)]$ which represents a re-sampled version of the original vector $\mathbf{P} = [P_1(h), P_2(h), \ldots, P_k(h)]$. In particular, a re-sampled version of $P_j(h)$ is computed using jth columns in matrices $\mathbf{SSE}^*$ and $\mathbf{TSS}^*$. Finally, after having carrying out all bootstrap trials, in order to estimate the p-value of the individual tests of significance of $P_j(h)$, we count how many times the simulated value for $P_j^*(h)$ happens to be below or equal to 0. Denoting this value by q_j, the p-value is computed as $p_j(h) = q_j/N$. The correlation coefficient $\rho_{mj}(h)$ is estimated using the simulated values for $P_m^*(h)$ and $P_j^*(h)$. It is worth emphasizing that our paired block-bootstrap method retains not only the correlations between various columns of matrices **SSE** and **TSS** but also the dependency between the elements of each column in these two matrices.

2.6 Measuring Volatility Forecast Bias

The bias in the volatility forecast is measured using the following pooled OLS regression model without the intercept

$$RV_{t,h}^j = \beta \, \hat{\sigma}_{t,h}^j + u_{t,h}^j, \tag{2.30}$$

where $RV_{t,h}^j$ and $\hat{\sigma}_{t,h}^j$ are the time t realized and predicted volatilities for week $h \geq 1$ in the future for stock j. The forecast is unbiased if $\beta = 1$. Note that we do not include the intercept in the regression because with only the slope estimation it is much easier to find out whether the volatility model tends to underestimate (if $\beta > 1$) or overestimate (if $\beta < 1$) the future realized volatility. Because of the persistence in volatility and potential issue of heteroskedasticity in the error terms $u_{t,h}^j$, we use the Newey-West method (Newey and West (1987)) for computing standard errors for β.

3 Empirical Results

3.1 Dating of Bull and Bear Market States

We remind the reader that, to detect the turning points between the bull and bear markets, we employ two alternative algorithms: the algorithm of Pagan and Sossounov (2003) and the algorithm of Lunde and Timmermann (2004). In the subsequent exposition, in order to shorten the terminology, we denote the algorithms of Pagan and Sossounov (2003) and Lunde and Timmermann (2004) as the *dating algorithm* and the *filtering algorithm* respectively.

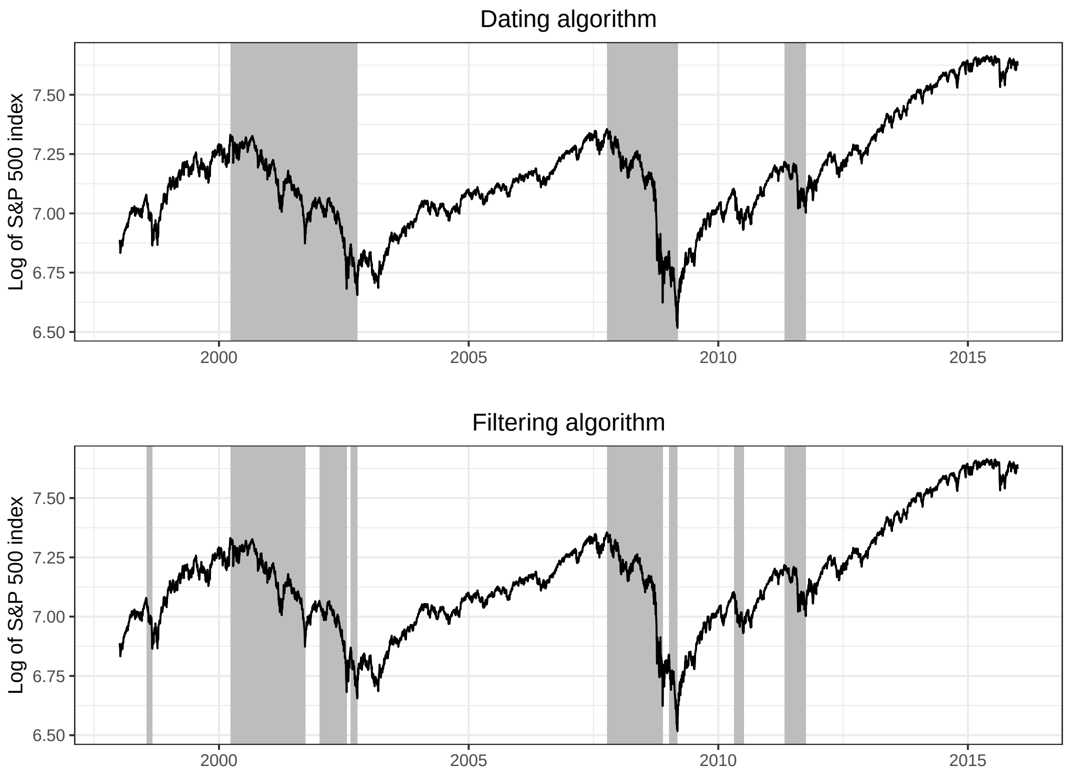

Figure 1: Bull and bear markets over the period from 1998 to 2016. Solid lines plot the log of the S&P 500 index. Shaded areas indicate bear market phases.

Figure 1 illustrates the results of detecting bull and bear market states provided by each algorithm. Each panel in this figure plots the log of the S&P 500 index. Shaded areas indicate bear market phases detected by either dating or filtering algorithm. The following sets of parameters are used in each algorithm: $\{\tau_{\text{window}} = 8, \tau_{\text{censor}} = 6, \tau_{\text{cycle}} = 16, \tau_{\text{phase}} = 4, \theta = 20\%\}$ (each τ_i is measured in months) and $\{\lambda_{Bull} = 20\%, \lambda_{Bear} = 15\%\}$. This figure clearly illustrates the two strongest bear markets in the recent history. The first one is associated with the Dot-com bubble crash of 2000-2002 and the second one with the Global financial crisis of 2007-2008.

Table 2 reports the descriptive statistics of the bull and bear market states identified by the two algorithms. The descriptive statistics of bull and bear states depend on the choice of the algorithm for detecting the turning points between the market states. Despite the differences, the descriptive statistics share of lot of similarities. The first similarity is that bull markets tend to be longer than bear markets. The average bull market duration exceeds the

Statistics	Dating algorithm		Filtering algorithm	
	Bull	**Bear**	**Bull**	**Bear**
Number of states	4	3	9	8
Minimum duration	25.8	5.1	1.1	1.6
Average duration	45.5	11.0	19.9	7.0
Maximum duration	59.9	16.9	59.9	17.8
Mean return	0.93	-1.69	1.26	-2.89
Standard deviation	3.49	6.34	3.60	6.17

Table 2: Descriptive statistics of bull and bear markets. Duration is measured in the number of months. Mean return is the monthly percentage return and standard deviation is the monthly percentage standard deviation.

average bear market duration by a factor of 3-4 depending on the choice of the algorithm. The second similarity is that all bull markets exhibit positive mean return while all bear markets have negative mean return. Finally, the third similarity is that the market volatility during the bear states is almost double as high as the volatility during the bull states. In sum, the descriptive statistics suggest a clear-cut conclusion that the bull market refers to the high-return, low-volatility state, whereas the bear market refers to the low-return, high-volatility state of the stock market.

3.2 Volatility Forecast Accuracy Across Horizons

We remind the reader that our total sample of intraday stock price data covers the period from January 2, 1998 to December 31, 2016. The period from 2 of January 1998 to 31 of December 1999 (2 years) is used as the initial in-sample period. Consequently, the out-of-sample period in our study is from 2 of January 2000 to 31 of December 2016 (17 years) that covers several interchanging bull and bear market states.

All forecasts are obtained using an expanding window scheme. Specifically, given a selected forecasting model, we perform out-of-sample conditional volatility forecasting for every stock in our data set. First, the parameters of a model are estimated using in-sample observations $[1, 2, \ldots, t]$, $t < T$ where T denotes the number of observations in the total sample. Then the conditional volatility is forecasted for 300-days ahead by performing 300 rolling one-day-ahead forecasts. After that, we expand the in-sample period by one day (it becomes $[1, 2, \ldots, t+1]$) and repeat the forecasting procedure. Since the estimation of the parameters of each forecasting model is rather time consuming, to speed up the forecasting process we re-estimate the model's parameters every 50 days only.

In the end of the forecasting process, the forecasted daily conditional volatilities are aggregated to the weekly conditional volatilities. That is, for each time t we obtain 60 weekly volatility forecasts $\hat{\sigma}_{t,h}$, $h \in [1, 2, \ldots, 60]$. Similarly, we aggregate daily realized volatilities to weekly realized volatilities and for each time t we get realized volatilities $RV_{t,h}$ for the future 60 consecutive weeks. Then, given $\hat{\sigma}_{t,h}$ and $RV_{t,h}$, we estimate the unconditional forecast accuracy (that is, the forecast accuracy over all market states) and the forecast accuracy conditional on the bull and bear states. The unconditional forecast accuracy is estimated using equation (2.22). Over the bull and bear states of the market, the conditional forecast accuracy is computed as

$$P_{Bull}(h) = 1 - \frac{\sum_{t=1}^{M} \left(RV_{t,h} - \hat{\sigma}_{t,h}\right)^2 \mathbb{1}_{t\in Bull}}{\sum_{t=1}^{M} \left(RV_{t,h} - \overline{RV}_{t,h}\right)^2 \mathbb{1}_{t\in Bull}}, \tag{3.1}$$

$$P_{Bear}(h) = 1 - \frac{\sum_{t=1}^{M} \left(RV_{t,h} - \hat{\sigma}_{t,h}\right)^2 \mathbb{1}_{t\in Bear}}{\sum_{t=1}^{M} \left(RV_{t,h} - \overline{RV}_{t,h}\right)^2 \mathbb{1}_{t\in Bear}}, \tag{3.2}$$

where $\mathbb{1}_{t\in Bull}$ ($\mathbb{1}_{t\in Bear}$) denotes the indicator function which takes one if the market is in a bull (bear) state on date t and zero otherwise.[8] Note that in the equations for $P_{Bull}(h)$ and $P_{Bear}(h)$ the value of $\overline{RV}_{t,h}$ does not depend on the state of the market. That is, the amount of variations in the weekly realized volatility is always computed with respect to its mean value in the total out-of-sample period (which is largely equivalent to using the historical long-run mean value of the weekly realized volatility if the out-of-sample period is sufficiently long).

Having computed the volatility forecast accuracies for every stock in our data set, we compute the corresponding p-values of the predictive ability test. Finally, for all states of the market and the bull and bear states separately, we compute the average forecast accuracies over all stocks, as well as the p-values of the combined probability tests. Figure 2 plots the average volatility forecast accuracy and the p-value of the combined probability test versus the forecast horizon for the EWMA and GARCH models. These two models use only daily return data to forecast the conditional volatility. Figure 3 plots the average volatility forecast accuracy and the p-value of the combined probability test versus the forecast horizon for the

[8] The motivation for our computation of conditional forecast accuracy is as follows. Most often at time t the state of the market is known. In contrast, at time t the state of the market in week h in the future is unknown. Therefore, the conditional forecast accuracy is computed under the assumption that the forecaster possesses the knowledge of the market state at time t but is completely ignorant about the state of the market in week h in the future.

HAR-RV and RealGARCH models. The latter two models use both daily returns and daily measures of realized volatility to forecast conditional volatility.

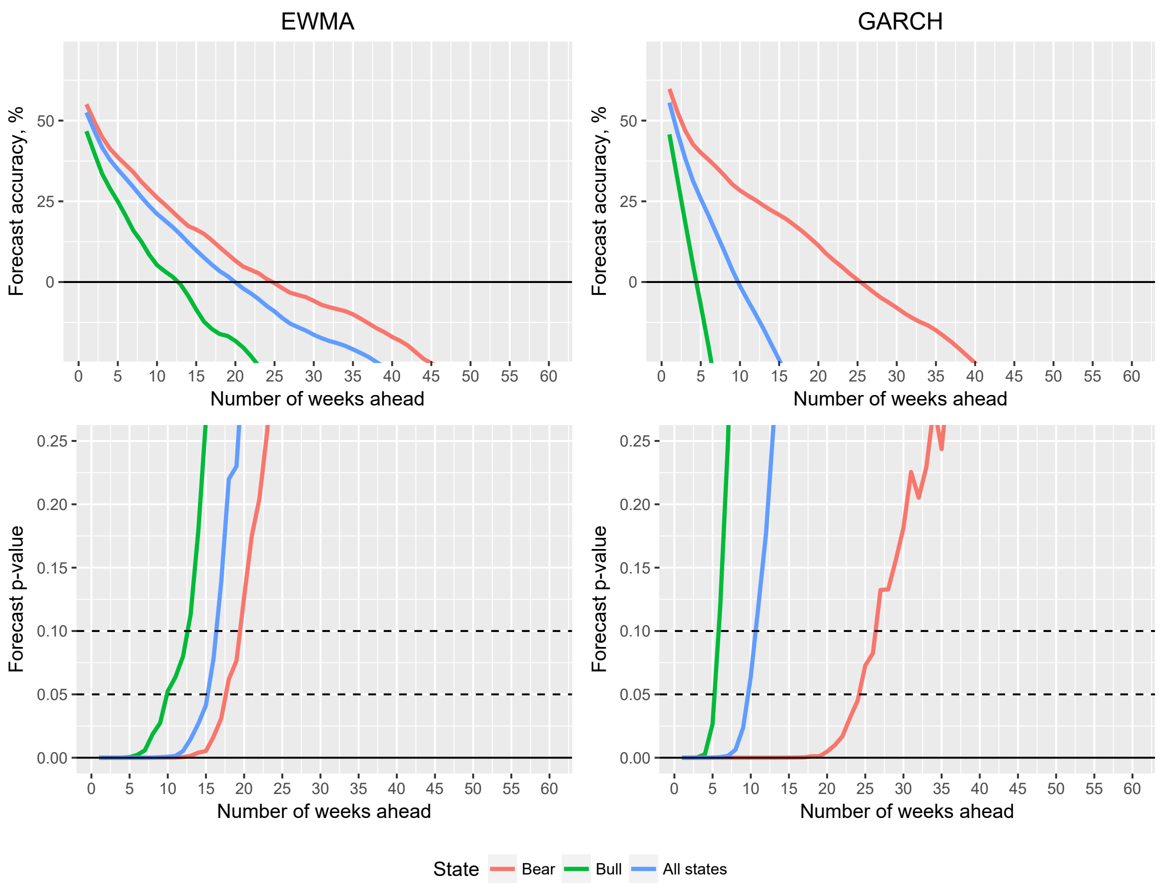

Figure 2: Average volatility forecast accuracy and the p-value of the combined probability test versus the forecast horizon for the EWMA and GARCH models. Bull and bear states of the market are detected using the dating algorithm. All curves are constructed using RV_t^{ON} as the measure of realized volatility.

Specifically, the blue lines in Figures 2 and 3, top panels, show the models' average unconditional forecast accuracy, whereas the green (red) lines show the average forecast accuracy conditional on the bull (bear) state of the market. Bull and bear states of the market are detected using the dating algorithm of Pagan and Sossounov (2003). The lines in the bottom panels in the figures depict the p-value of the combined probability test for the unconditional forecast accuracy and the forecast accuracy conditional on the bull and bear states of the market using the blue, green, and red lines respectively. All curves are constructed using RV_t^{ON} as the measure of realized volatility.

On the basis of the results reported in Figures 2 and 3, the following observations can

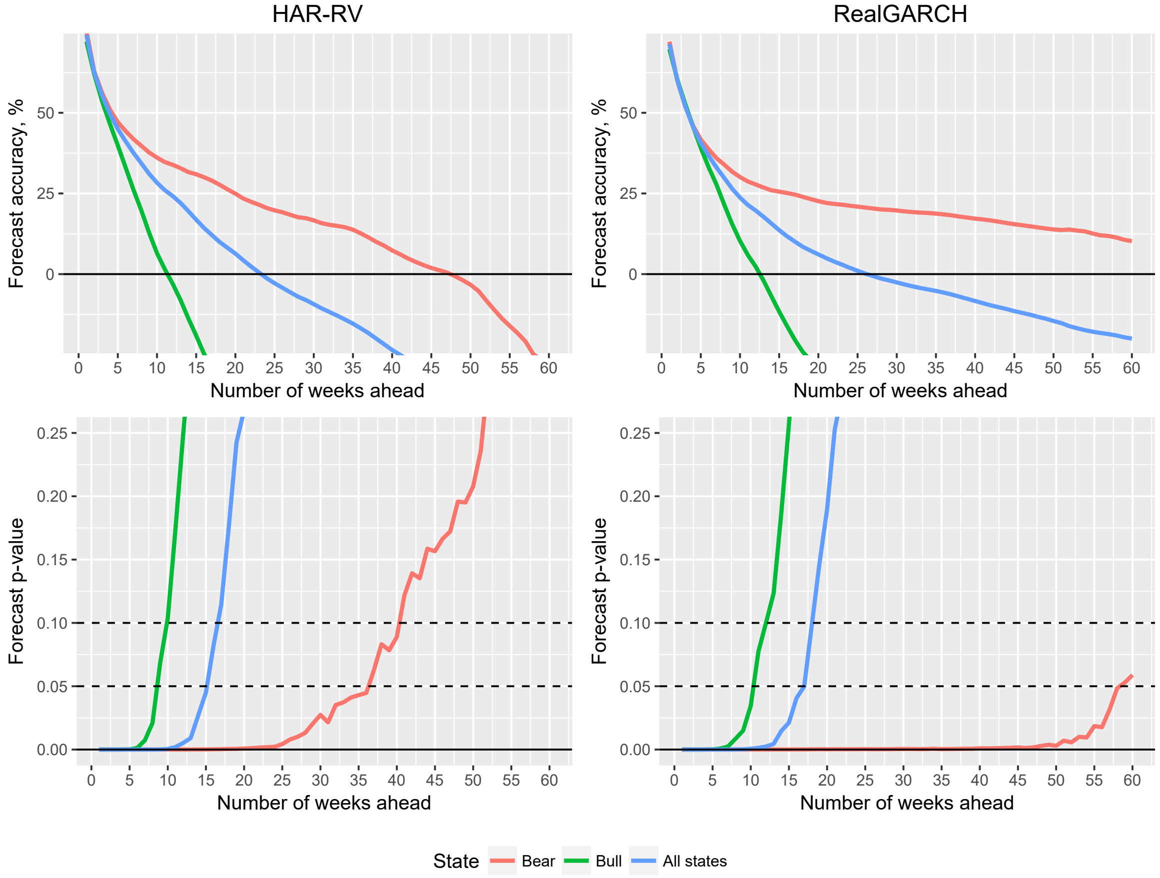

Figure 3: Average volatility forecast accuracy and the p-value of the combined probability test versus the forecast horizon for the HAR-RV and RealGARCH models. Bull and bear states of the market are detected using the dating algorithm. All curves are constructed using RV_t^{ON} as the measure of realized volatility.

be made. First, consider the unconditional forecast accuracy. Notice that, regardless of the choice of volatility model, the forecast accuracy is spread highly unevenly across horizons and the accuracy decreases quickly with horizon. Using a significance level of 10%, the horizon of volatility predictability amounts to 10-11 weeks for the GARCH model and 16-18 weeks for the EWMA, HAR-RV, and RealGARCH model. The forecast accuracy gains provided by intraday data are largest for the shortest horizon and diminish quickly when horizon increases. For instance, as compared to the forecast accuracy provided by the EWMA model (for a visual comparison, see also Figure 4, panel "All states"), the gains in forecasting one-week ahead volatility amount to almost 20%, whereas the gains reduce to about 5% in forecasting volatility beyond a 10-week horizon.

Second, consider the volatility forecast accuracy conditional on the state of the market.

Notice that for the majority of volatility models the horizon of volatility predictability is much longer when the market is in a bear state than when it is in a bull state. Take, as an example, the RealGARCH model. Notice that at a significance level of 10% the horizon of volatility predictability extends beyond 60 weeks when the market is in a bear state. In contrast, the horizon of volatility predictability is limited to only 12 weeks when the market is in a bull state. The EWMA model shows the least difference between the horizons of volatility predictability conditional on the state of the market. For this model, the horizon of volatility predictability amounts to 12 (18) weeks when the market is in a bull (bear) state.

When only daily data are used, regardless of the forecast horizon length the volatility is forecasted with a higher accuracy when the market is in a bear state than when it is in a bull state. When both daily and intraday data are used, over short horizons that are limited to about 4 weeks, the volatility forecast accuracy virtually does not depend on the state of the market. However, beyond a 4-week horizon the volatility forecast accuracy diminishes very quickly with horizon when the market is in a bull state. In contrast, when the market is in a bear state, the volatility forecast accuracy decreases with a slow rate when horizon lengthens.

Figure 4 plots the average volatility forecast accuracies over all states of the market and over bull and bear states separately. In particular, for the sake of visual comparison, each panel in this figure plots the average volatility forecast accuracies provided by each volatility model. The following additional comments can be made regarding the relative performance of the competing models. The EWMA and GARCH models have about the same forecast accuracy when either the forecast horizon is very short or the market is in a bear state. When the market is in a bull state, the forecast horizon length of the GARCH model is much shorter than that of the EWMA model. The HAR-RV and RealGARCH models have about the same forecast accuracy and horizon length when the market is in a bull state. In contrast, when the market is in a bear state, the volatility forecast accuracy is marginally better for the HAR-RV model when the forecast horizon is limited to about 20 weeks. Beyond a 20-week horizon, the RealGARCH model provides a higher forecast accuracy than the HAR-RV model.

For the sake of additional illustration, we employ a simple graphical diagnostic tool[9] that makes it easy to demonstrate how the volatility forecast accuracy depends on the state of the market. In particular, to monitor the predictive power of a volatility forecasting model we

[9]A similar graphical diagnostic tool was suggested in Goyal and Welch (2003).

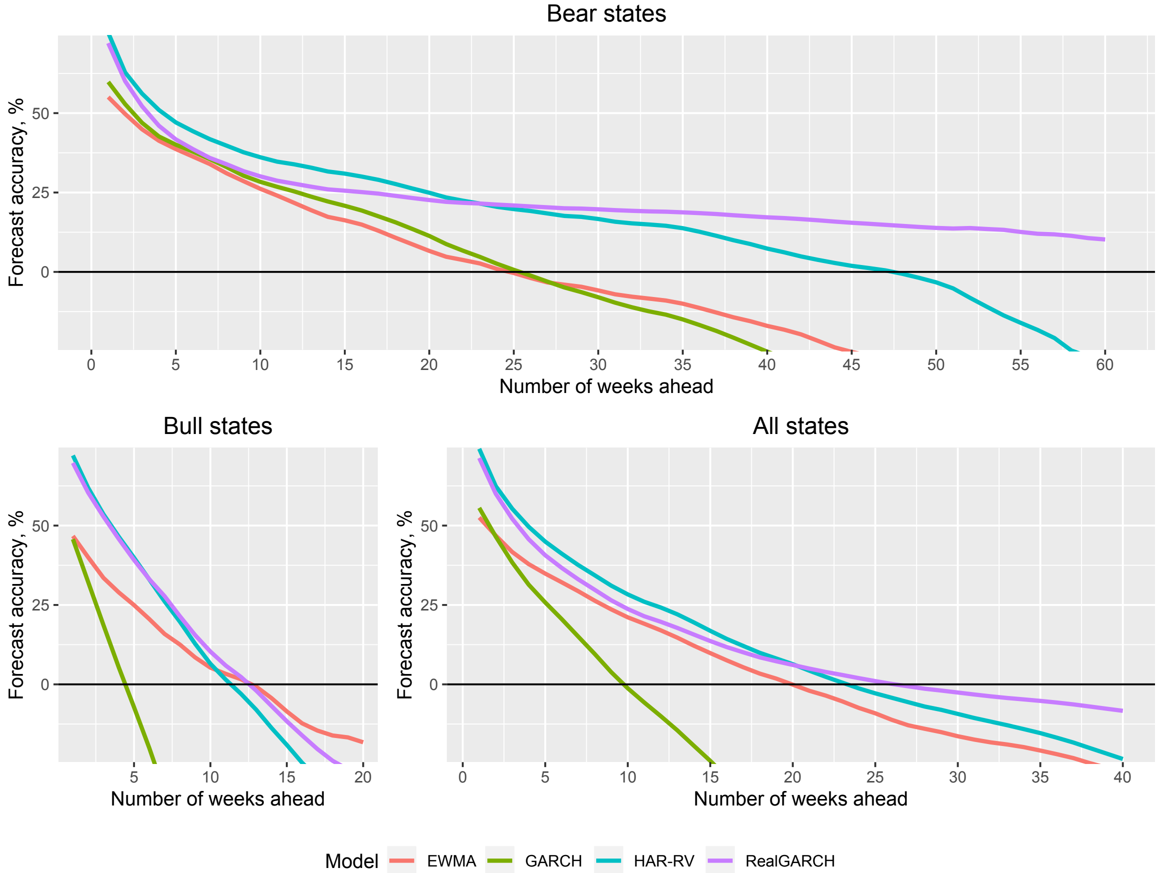

Figure 4: Average volatility forecast accuracies over all states of the market and over bull and bear states separately. Bull and bear states of the market are detected using the dating algorithm. All curves are constructed using RV_t^{ON} as the measure of realized volatility.

use the cumulative difference between the sum of the forecast absolute errors provided by the mean value of the realized volatility (which is similar to computing forecast errors from the historical mean model) and the conditional volatility forecast

$$CUDIF_t = \sum_{s=1}^{t} \left(\left| RV_{s,h} - \overline{RV}_{s,h} \right| - \left| RV_{s,h} - \hat{\sigma}_{s,h} \right| \right), \quad t \leq M.$$

From a visual examination of the graph of $CUDIF_t$, it is easy to understand in which periods the conditional volatility model provides a better forecast than the historical mean model. Specifically, in periods when $CUDIF_t$ increases, the volatility model provides better predictions; in periods when it decreases, the volatility model has worse predictive performance than the historical mean model.

Figure 5 plots $CUDIF_t$ in forecasting 10th week volatility of the Alcoa, Honeywell Inter-

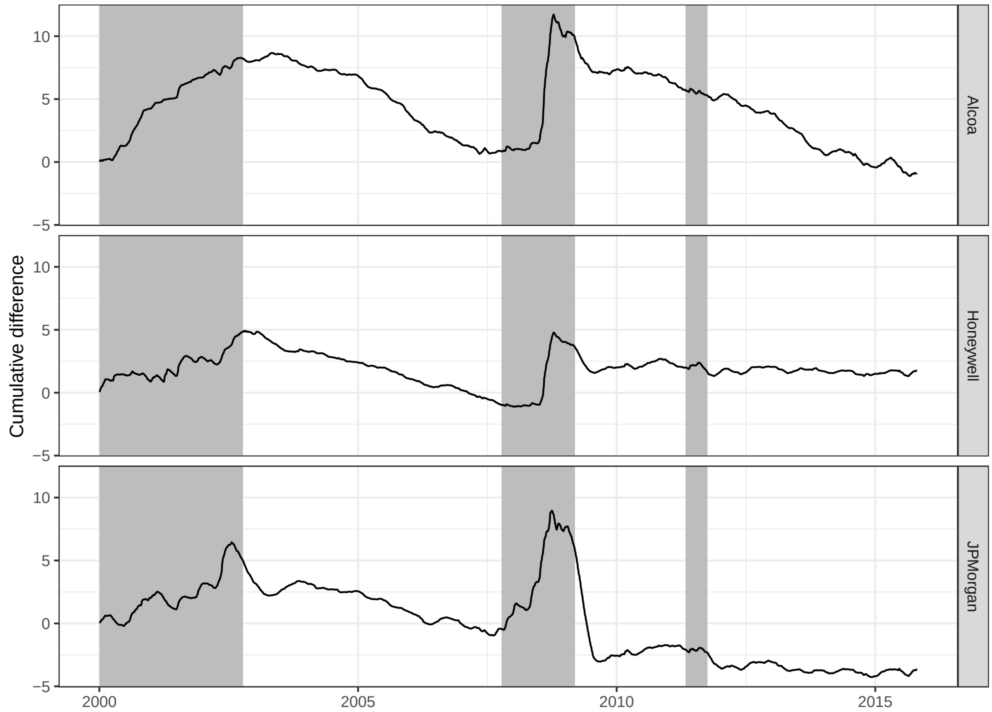

Figure 5: The figure plots $CUDIF_t = \sum_{s=1}^{t} \left(\left| RV_{s,10} - \overline{RV}_{s,10} \right| - |RV_{s,10} - \hat{\sigma}_{s,10}| \right)$ for forecasting 10th week volatility of the Alcoa, Honeywell International, and JPMorgan Chase stocks using the EWMA model. $\left| RV_{s,10} - \overline{RV}_{s,10} \right|$ is the absolute error in forecasting volatility for the 10th week using the historical mean model while $|RV_{s,10} - \hat{\sigma}_{s,10}|$ is the absolute error in forecasting volatility for the 10th week using the EWMA model. Shaded areas indicate bear market states. Bull and bear states of the market are detected using the dating algorithm.

national, and JPMorgan Chase stocks using the EWMA model. Notice that for each stock $CUDIF_t$ tends to increase when the market is in a bear state. In contrast, $CUDIF_t$ tends to decrease or stay at about the same level when the market is in a bull state. That is, the plots in this figure provide a clear-cut illustration of the fact that the volatility forecast accuracy is state dependent.

As a final note to close this section, the results reported in Figures 2 - 4 are obtained using RV_t^{ON} as the measure of realized volatility. Virtually similar results are obtained using RV_t^{HL} as the measure of realized volatility. The results are robust to the choice of the dating algorithm for detecting the turning points between the bull and bear markets. The results are also robust in the sub-samples of data. Specifically, the results are virtually the same for the first (2000-2007) and second (2008-2016) halves of the total out-of-sample period.

3.3 Volatility Forecast Bias Across Horizons

Using the forecasted weekly conditional volatilities $\hat{\sigma}^j_{t,h}$ and weekly realized volatilities $RV^j_{t,h}$, where $j \in [1, 2, \ldots, 31]$ denotes the stock number, we estimate the unconditional forecast bias (that is, the forecast bias over all market states) and the forecast bias conditional on the bull and bear states. The unconditional forecast bias is estimated using the pooled OLS regression model given by equation (2.30). Over the bull and bear states of the market, the conditional forecast bias is estimated using the following pooled OLS regressions

$$RV^j_{t,h}\mathbb{1}_{t\in Bull} = \beta\,\hat{\sigma}^j_{t,h}\mathbb{1}_{t\in Bull} + u^j_{t,h}, \tag{3.3}$$

$$RV^j_{t,h}\mathbb{1}_{t\in Bear} = \beta\,\hat{\sigma}^j_{t,h}\mathbb{1}_{t\in Bear} + u^j_{t,h}. \tag{3.4}$$

We remind the reader that the forecast is unbiased if $\beta = 1$. If $\beta < 1$ ($\beta > 1$), the volatility model tends to overestimate (underestimate) the future realized volatility.

For each model, Figure 6 plots the estimated volatility forecast bias across horizons over all states of the market and over bull and bear states separately. The colored shaded areas show the 99% confidence interval for the estimated bias (β). Consider first the unconditional forecast bias. The patter of this bias is very similar across the various models. Over very short forecast horizons the bias is either absent (for the EWMA and HAR-RV models) or negligible small (for the GARCH and RealGARCH models). However, as the forecast horizon lengthens, all volatility models tend to overestimate the future volatility.

Conditional on the market states, the volatility forecast bias is very different across the bull and bear states of the market. The first apparent conclusion that can be drawn from the visual observation of the conditional forecast bias is that the bias is smaller when the market is in a bear state. Notices that in a bull state of the market each model tends to overestimate the future volatility and the bias increases as horizon increases. In a bear state of the market, on the other hand, each model tends to underestimate the future volatility when forecast horizon is short; when forecast horizon lengthens, all models but the RealGARCH model tend to overestimate the future volatility.

A few final remarks to this section are as follows. The results reported in this section are robust to the choice of the measure of realized volatility and the dating algorithm for detecting

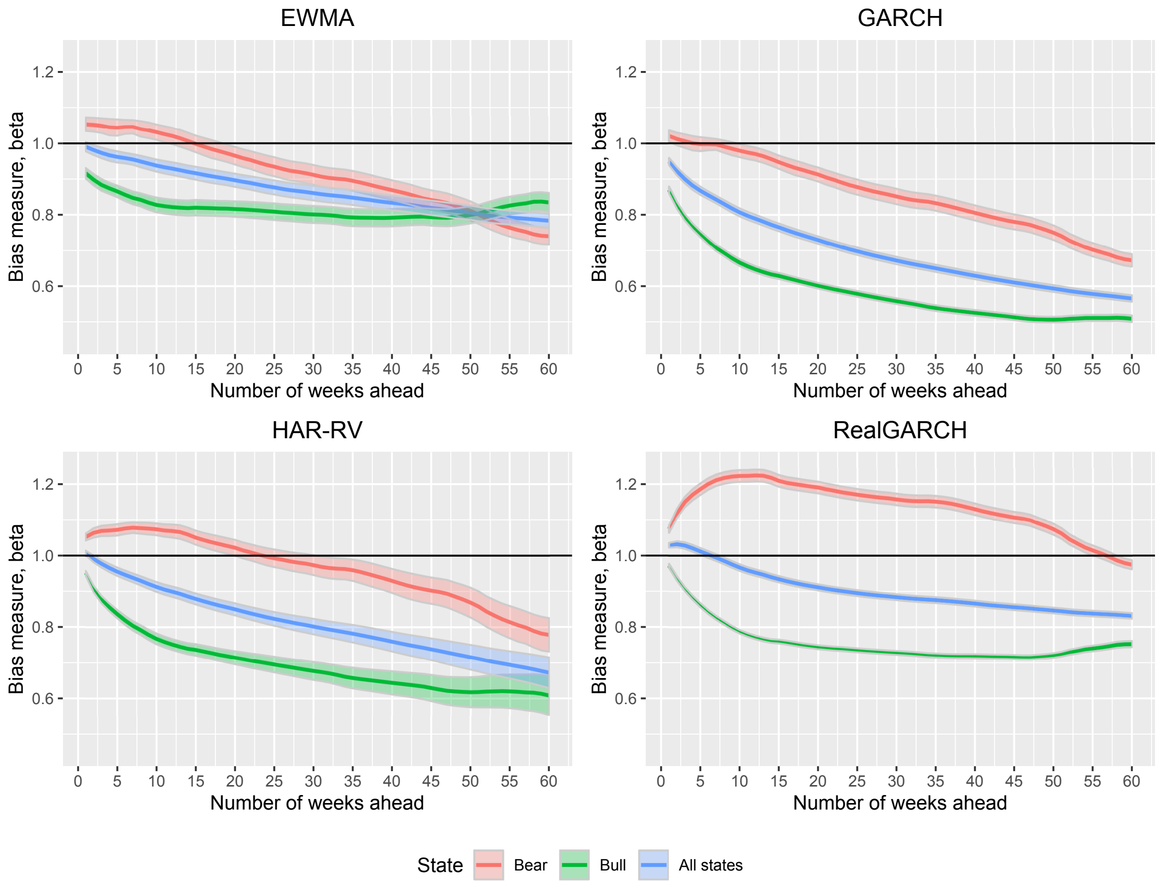

Figure 6: For each model, this figure plots the estimated volatility forecast bias across horizons over all states of the market and over bull and bear states separately. The colored shaded areas show the 99% confidence interval for the estimated bias (β). Bull and bear states of the market are detected using the dating algorithm. All curves are constructed using RV_t^{ON} as the measure of realized volatility.

the turning points between the bull and bear markets. The results are also robust in the sub-samples of data. These results suggest a partial explanation for why volatility forecast accuracy diminishes with horizon: as horizon increases, the volatility forecast bias increases. Since the forecast bias can be estimated, it can be corrected as well. That is, bias correction can potentially increase considerably the horizon of volatility predictability. Importantly, our analysis reveals that the bias correction must be done depending on the market state.

4 Conclusions

Using a large data set of high-frequency data on individual stocks and a few popular time-series volatility models, in this paper we comprehensively examine how volatility forecastability varies across bull and bear states of the stock market. Our results suggest that volatility predictability depends strongly on the state of the market.

Specifically, we find that the volatility forecast horizon is considerably longer when the market is in a bear state than when it is in a bull state. For instance, when the high-frequency data are used, the horizon of volatility predictability extends beyond 60 weeks when the market is in a bear state. In contrast, the horizon of volatility predictability is limited to only 12 weeks when the market is in a bull state. The volatility forecast accuracy is highest for the shortest horizons when it attains 75%. The forecast accuracy diminishes very quickly with horizon when the market is in a bull state. In a bear state, on the other hand, the forecast accuracy diminishes with a slower rate.

The forecast accuracy gains provided by high-frequency data are substantial and amount to almost 20% for the shortest horizon. The gains diminish quickly with horizon when the market is in a bull state. Conversely, in bear states the gains persist regardless of the horizon length. Finally, we find that the volatility forecast bias is also state dependent being larger in bull states. In bull markets each model tends to overestimate the future volatility and the bias increases with horizon. In bear markets the models tend to underestimate (overestimate) the future volatility when forecast horizon is short (long).

In sum, our study concludes that, similarly to the previous studies on return predictability, the volatility predictability is also strongest during bad economic times proxied by bear market states. That is, volatility predictability is best when it is most needed, during periods of high turbulence and uncertainty.

References

Alford, A. W. and Boatsman, J. R. (1995). "Predicting Long-Term Stock Return Volatility: Implications for Accounting and Valuation of Equity Derivatives", *Accounting Review*, *70*(4), 599–618.

Andersen, T. G. and Bollerslev, T. (1998). "Answering the Skeptics: Yes, Standard Volatility Models do Provide Accurate Forecasts", *International Economic Review*, *39*(4), 885–905.

Andersen, T. G., Bollerslev, T., and Diebold, F. X. (2007). "Roughing It Up: Including Jump Components in the Measurement, Modeling, and Forecasting of Return Volatility", *The Review of Economics and Statistics*, *89*(4), 701–720.

Andersen, T. G., Bollerslev, T., Diebold, F. X., and Labys, P. (2001). "The Distribution of Realized Exchange Rate Volatility", *Journal of the American Statistical Association*, *96*(453), 42–55.

Andersen, T. G., Bollerslev, T., Diebold, F. X., and Labys, P. (2003). "Modeling and Forecasting Realized Volatility", *Econometrica*, *71*(2), 579–625.

Andersen, T. G., Bollerslev, T., and Lange, S. (1999). "Forecasting Financial Market Volatility: Sample Frequency Vis-a-Vis Forecast Horizon", *Journal of Empirical Finance*, *6*(5), 457–477.

Andersen, T. G., Bollerslev, T., and Meddahi, N. (2005). "Correcting the Errors: Volatility Forecast Evaluation Using High-Frequency Data and Realized Volatilities", *Econometrica*, *73*(1), 279–296.

Blair, B. J., Poon, S.-H., and Taylor, S. J. (2001). "Forecasting S&P 100 Volatility: The Incremental Information Content of Implied Volatilities and High-Frequency Index Returns", *Journal of Econometrics*, *105*(1), 5 – 26.

Bollerslev, T. (1986). "Generalized Autoregressive Conditional Heteroskedasticity", *Journal of Econometrics*, *31*(3), 307–327.

Brown, M. B. (1975). "400: A Method for Combining Non-Independent, One-Sided Tests of Significance", *Biometrics*, *31*(4), 987–992.

Bry, G. and Boschan, C. (1971). *Cyclical Analysis of Time Series: Selected Procedures and Computer Programs*. NBER.

Campbell, J. Y. and Thompson, S. B. (2007). "Predicting Excess Stock Returns Out of Sample: Can Anything Beat the Historical Average?", *The Review of Financial Studies*, *21*(4), 1509–1531.

Cao, C. Q. and Tsay, R. S. (1992). "Nonlinear Time-Series Analysis of Stock Volatilities", *Journal of Applied Econometrics*, *7*(S1), S165–S185.

Cheema, M. A., Nartea, G. V., and Man, Y. (2018). "Cross-Sectional and Time Series Momentum Returns and Market States", *International Review of Finance*, *18*(4), 705–715.

Christoffersen, P. F. and Diebold, F. X. (2000). "How Relevant is Volatility Forecasting for Financial Risk Management?", *Review of Economics and Statistics*, *82*(1), 12–22.

Corsi, F. (2009). "A Simple Approximate Long-Memory Model of Realized Volatility", *Journal of Financial Econometrics*, *7*(2), 174–196.

Dangl, T. and Halling, M. (2012). "Predictive Regressions with Time-Varying Coefficients", *Journal of Financial Economics*, *106*(1), 157 – 181.

Engle, R. and Patton, A. (2001). "What Good is a Volatility Model?", *Quantitative Finance*, *1*(2), 237–245.

Fama, E. F. (1965). "The Behavior of Stock-Market Prices", *Journal of Business*, *38*(1), 34–105.

Figlewski, S. (1997). "Forecasting Volatility", *Financial Markets, Institutions and Instruments*, *6*(1), 1–88.

Fisher, R. A. (1925). *Statistical Methods for Research Workers*. Oliver and Boyd, Edinburgh.

Fleming, J., Ostdiek, B., and Whaley, R. E. (1995). "Predicting Stock Market Volatility: A New Measure", *Journal of Futures Markets*, *15*(3), 265–302.

French, K. R. and Roll, R. (1986). "Stock Return Variances: The Arrival of Information and the Reaction of Traders", *Journal of Financial Economics*, *17*(1), 5 – 26.

Galbraith, J. W. and Kisinbay, T. (2005). "Content Horizons for Conditional Variance Forecasts", *International Journal of Forecasting*, *21*(2), 249 – 260.

Garcia, D. (2013). "Sentiment During Recessions", *Journal of Finance*, *68*(3), 1267–1300.

Gargano, A., Pettenuzzo, D., and Timmermann, A. (2017). "Bond Return Predictability: Economic Value and Links to the Macroeconomy", *Management Science*, *65*(2), 508–540.

Gargano, A. and Timmermann, A. (2014). "Forecasting Commodity Price Indexes Using Macroeconomic and Financial Predictors", *International Journal of Forecasting*, *30*(3), 825 – 843.

Golez, B. and Koudijs, P. (2018). "Four Centuries of Return Predictability", *Journal of Financial Economics*, *127*(2), 248 – 263.

Goyal, A. and Welch, I. (2003). "Predicting the Equity Premium with Dividend Ratios", *Management Science*, *49*(5), 639–654.

Green, T. C. and Figlewski, S. (1999). "Market Risk and Model Risk for a Financial Institution Writing Options", *Journal of Finance*, *54*(4), 1465–1499.

Haas, M. and Liu, J. (2018). "A Multivariate Regime-Switching GARCH Model with an Application to Global Stock Market and Real Estate Equity Returns", *Studies in Nonlinear Dynamics and Econometrics*, *22*(3), 1–27.

Haas, M., Mittnik, S., and Paolella, M. S. (2004). "A New Approach to Markov-Switching GARCH Models", *Journal of Financial Econometrics*, *2*(4), 493–530.

Hammerschmid, R. and Lohre, H. (2018). "Regime Shifts and Stock Return Predictability", *International Review of Economics and Finance*, *56*, 138 – 160.

Han, H. and Park, M. D. (2013). "Comparison of Realized Measure and Implied Volatility in Forecasting Volatility", *Journal of Forecasting*, *32*(6), 522–533.

Hansen, P. R. and Lunde, A. (2005a). "A Forecast Comparison of Volatility Models: Does Anything Beat a GARCH(1,1)?", *Journal of Applied Econometrics*, *20*(7), 873–889.

Hansen, P. R. and Lunde, A. (2005b). "A Realized Variance for the Whole Day Based on Intermittent High-Frequency Data", *Journal of Financial Econometrics*, *3*(4), 525–554.

Hansen, P. R. and Huang, Z. (2016). "Exponential GARCH Modeling With Realized Measures of Volatility", *Journal of Business and Economic Statistics*, *34*(2), 269–287.

Hansen, P. R., Huang, Z., and Shek, H. H. (2012). "Realized GARCH: A Joint Model for Returns and Realized Measures of Volatility", *Journal of Applied Econometrics*, *27*(6), 877–906.

Hansen, P. R., Lunde, A., and Voev, V. (2014). "Realized Beta GARCH: A Multivariate GARCH Model With Realized Measures of Volatility", *Journal of Applied Econometrics*, *29*(5), 774–799.

Henkel, S. J., Martin, J. S., and Nardari, F. (2011). "Time-Varying Short-Horizon Predictability", *Journal of Financial Economics*, *99*(3), 560 – 580.

Klaassen, F. (2002). "Improving GARCH Volatility Forecasts with Regime-Switching GARCH", In Hamilton, J. D. and Raj, B. (Eds.), *Advances in Markov-Switching Models: Applications in Business Cycle Research and Finance*, pp. 223–254. Physica-Verlag HD, Heidelberg.

Koopman, S. J., Jungbacker, B., and Hol, E. (2005). "Forecasting Daily Variability of the S&P 100 Stock Index Using Historical, Realised and Implied Volatility Measurements", *Journal of Empirical Finance*, *12*(3), 445 – 475.

Littell, R. C. and Folks, J. L. (1971). "Asymptotic Optimality of Fisher's Method of Combining Independent Tests", *Journal of the American Statistical Association*, *66*(336), 802–806.

Longerstaey, J. and Spencer, M. (1996). "RiskMetrics™—Technical Document", Tech. rep., Morgan Guaranty Trust Company of New York: New York.

Lunde, A. and Timmermann, A. (2004). "Duration Dependence in Stock Prices: An Analysis of Bull and Bear Markets", *Journal of Business and Economic Statistics*, *22*(3), 253–273.

Maheu, J. M. and McCurdy, T. H. (2000). "Identifying Bull and Bear Markets in Stock Returns", *Journal of Business and Economic Statistics*, *18*(1), 100–112.

Maheu, J. M., McCurdy, T. H., and Song, Y. (2012). "Components of Bull and Bear Markets: Bull Corrections and Bear Rallies", *Journal of Business and Economic Statistics*, *30*(3), 391–403.

Marcucci, J. (2005). "Forecasting Stock Market Volatility with Regime-Switching GARCH Models", *Studies in Nonlinear Dynamics and Econometrics*, *9*(6), 1–53.

Martens, M. (2002). "Measuring and Forecasting S&P 500 Index-Futures Volatility Using High-Frequency Data", *Journal of Futures Markets*, *22*(6), 497–518.

Martens, M., van Dijk, D., and de Pooter, M. (2009). "Forecasting S&P 500 Volatility: Long Memory, Level Shifts, Leverage Effects, Day-of-the-Week Seasonality, and Macroeconomic Announcements", *International Journal of Forecasting*, *25*(2), 282 – 303.

Martens, M. and Zein, J. (2004). "Predicting Financial Volatility: High-Frequency Time-Series Forecasts Vis-a-Vis Implied Volatility", *Journal of Futures Markets*, *24*(11), 1005–1028.

Merton, R. C. (1980). "On Estimating the Expected Return on the Market: An Exploratory Investigation", *Journal of Financial Economics*, *8*(4), 323 – 361.

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. R. and Sossounov, K. A. (2003). "A Simple Framework for Analysing Bull and Bear Markets", *Journal of Applied Econometrics*, *18*(1), 23–46.

Pan, Z., Wang, Y., Wu, C., and Yin, L. (2017). "Oil Price Volatility and Macroeconomic Fundamentals: A Regime Switching GARCH-MIDAS Model", *Journal of Empirical Finance*, *43*, 130 – 142.

Patton, A., Politis, D. N., and White, H. (2009). "Correction to "Automatic Block-Length Selection for the Dependent Bootstrap" by D. Politis and H. White", *Econometric Reviews*, *28*(4), 372–375.

Politis, D. N. and White, H. (2004). "Automatic Block-Length Selection for the Dependent Bootstrap", *Econometric Reviews*, *23*(1), 53–70.

Politis, D. and Romano, J. (1994). "The Stationary Bootstrap", *Journal of the American Statistical Association*, *89*(428), 1303–1313.

Poon, S.-H. (2005). *A Practical Guide to Forecasting Financial Market Volatility*. John Wiley & Sons.

Poon, S.-H. and Granger, C. W. J. (2003). "Forecasting Volatility in Financial Markets: A Review", *Journal of Economic Literature*, *41*(2), 478–539.

Rapach, D. E., Strauss, J. K., and Zhou, G. (2009). "Out-of-Sample Equity Premium Prediction: Combination Forecasts and Links to the Real Economy", *Review of Financial Studies*, *23*(2), 821–862.

Raunig, B. (2006). "The Longer-Horizon Predictability of German Stock Market Volatility", *International Journal of Forecasting*, *22*(2), 363 – 372.

Schmeling, M. (2009). "Investor Sentiment and Stock Returns: Some International Evidence", *Journal of Empirical Finance*, *16*(3), 394 – 408.

Todorova, N. and Soucek, M. (2014). "Overnight Information Flow and Realized Volatility Forecasting", *Finance Research Letters*, *11*(4), 420 – 428.

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.