

Professor Gheorghe HURDUZEU, PhD

E-mail: gheorghe.hurduzeu@rei.ase.ro

Iulian-Cornel LOLEA, PhD Student

E-mail: loleaiulian@gmail.com

Ana-Maria GIUREA, PhD Student

E-mail: ana.giurea@yahoo.com

Assistant Maria Floriana POPESCU, PhD

E-mail: mariafpopescu@yahoo.com

The Bucharest University of Economic Studies

DOES REAL-TIME MACROECONOMIC DATA ENSURE AN ACCURATE VOLATILITY FORECASTING? A TWO STATES APPROACH FOR THE US EQUITY MARKET

***Abstract:** This article aims to highlight the extent to which stock market volatility can be estimated based on macroeconomic variables. Moreover, it tried and succeeded to emphasize that the accuracy of volatility forecasting models that use real-time data is only marginally different to the performance of those models that are based on revised data. These results were in line with what was previously found by other researchers on this topic. Further, there were considered two volatility regimes, in order to rule out any chance that the results are not significant or are influenced by any statistical discrepancies. The case study was conducted for the US equity market, using monthly data for the period between 1998-2016, which was the longest available period.*

***Keywords:** volatility forecasting, macroeconomic data, U.S. stock market, regression, investors' utility, VIX, PMI.*

JEL Classification: G170

1. Introduction

Volatility is one of the most important variables that characterize the state of financial markets, a very widely used measure of risk. Development of financial markets made possible the emergence of financial instruments through which investors can capture the evolution of volatility, in order to be able to protect or simply take speculative positions. In US do we have the VIX index that showcases the volatility of S & P 500. Moreover, the VIX index represents the underlying

asset for many derivatives (ex. call or put options), which allow investors to take positions in one direction or another regarding the volatility's evolution.

Volatility changes are driven by market events, most frequently by the publishing of new information regarding the economy or profits of listed companies (i.e. following the publication of new data for macroeconomic variables). Thus, it is emphasized that macroeconomic variables are an important determinant of volatility and each market has its peculiarities, being more sensitive to some macroeconomic variables than to others. There were many relevant studies in the literature regarding the correlation between the volatility of the market and business cycles (Hamilton and Lin, 1996; Schwert, 1989), which suggest that there is a clear dependency relationship between volatility in financial markets and macroeconomic variables. This relationship will be tested in this paper and will be highlighted that the use of real-time data would not reduce the value added of an econometric model based on macroeconomic variables.

The literature contains a large number of methodologies and ways that volatility can be estimated for financial markets. Among the most popular types of models are GARCH or GARCH regime-switching models based on Markov chains (Marcucci, 2005). They will however not be the subject of the present research, which will focus on recursive estimation of volatility captured by the VIX index, taking into account a set of representative macroeconomic variables for the US market. It should be recalled that GARCH models, as well as those that allow the regime switches are more efficient in terms of short-term estimations. In this paper, there will be considered monthly observations, for which recursive estimation based on macroeconomic variables may be more compelling.

The present paper will be structured in two parts, which will pursue the objective of highlighting that any use of real-time data is without a decrease in performance of volatility estimation compared to using revised data: i) the first part consists in an estimation of each regression model parameters and the conduct of meaningful tests in order to reveal their statistical significance; ii) the second part includes a comparative analysis based on a statistical criteria (i.e. RMSE, Theil's U), Diebold-Mariano test (Diebold and Mariano 1995) and based on an utility function dependent on the investor's risk aversion (Engle, Kane and Noh, 1996). This approach was also used by Pierdzioch, Dopke and Hartman (2008) and all conclusions have confirmed the initial expectations. Also, the influence of macroeconomic variables on the volatility of the stock market was deeply analysed by Engle, Ghysels, Sohn (2013), which address a technique perspective that could be a directive in which this research might be headed and where it can bring added value.

The novelty brought by this paper consists of two facts. Firstly, it examines the appropriateness of using real-time macroeconomic data for US market in order to estimate volatility measured by the VIX index, for a period not studied yet. In

Does Real-Time Macroeconomic Data Ensure an Accurate Volatility Forecasting? A Two States Approach for the US Equity Market

this paper, there were considered macroeconomic variables extremely important for the US market such as the number of newly created jobs in the non-farming sector (NFP), considering that labour market is a very sensitive issue for the US economy. Unlike the ECB, FED's monetary policy aims at full employment of labour, emphasizing its importance. Along with NFP, there were considered a lot of other variables specific to the US market, applying the proposed methodology to attain the scope of this paper.

Another novelty research path which will be addressed in this research is to identify and highlight if there are any significant differences in the performance estimation between considered models according to two defined regimes: high volatility and low volatility regimes. This approach will try to capture that both models (based on real-time data, as well as those based on revised data) lead similar results for both regimes' volatility. This analysis is based on the concept used in models that estimate volatility for several states (Marcucci, 2005) in order to better capture volatility clusters. There were concerns that in certain periods there may be discrepancies between results using the two types of data, even if considering the whole period when no significant differences were found.

Furthermore, to increase the accuracy of this research, a higher number of statistical indicators will be used in comparison to than other papers in the literature addressed until now. In order to carry out this whole approach, Diebold-Mariano will be used to test to identify whether differences between models are significantly different from zero.

2. Methodology

This paper aims to estimate US stock market's volatility. To put it in other words, it will try to estimate the VIX index, using monthly data. For this purpose, a linear regression will be used that is based on few macroeconomic variables as regressors, assuming a linear relationship between VIX and them.

Recursive modelling of US stock market volatility captured by VIX index

As stated above, macroeconomic variables will be taken into account in order to forecast the index VIX, both for real-time data and revised data. As Pesaran and Timmermann (1995.2000) proceeded in doing, we used the Ordinary Least Squares methodology in order to estimate the linear regression model as it is stated below:

$$y_{t+1} = x_{t,i}\beta_i + \varepsilon_{t+1} \quad (1)$$

y_{t+1} – next period's volatility

$x_{t,i}$ – regressor “i” used estimate the linear regression model. It represents a subset of variables that are used in our approach. This vector of regressors always includes a constant.

β_i – the estimated coefficient for each regressor

ε_{t+1} , - white noise (model's errors)

Based on this equation the regression coefficients will be estimated for each of the variables considered. The results obtained using the two types of data (real-time and revised data) will be compared following two approaches: statistical criteria and a utility based criterion.

Selection Criteria

Statistical Criteria

- i) RMSE (root-mean-squared errors)
- ii) Theil's U (Theil, 1966) which is defined as a ratio of two RMSE (root-mean-squared errors). In the present paper, the U statistic is defined as follows:

$$U = \frac{RMSE^{realtime}}{RMSE^{revised}} \quad (2)$$

The Theil's U is going to be compared to 1, considered as a reference value. If it is above, we can conclude that the estimation based on real-time data conducted to higher errors than the estimation based on revised data.

The next statistical criterion that will be used is iii) Diebold-Mariano Test (Diebold and Mariano, 1995). The Diebold-Mariano test will have the null hypothesis whereby that there is no significant difference between the two estimations and the alternative hypothesis that there is a significant difference between the two estimations. Hence, a significant test result implies that the forecasts obtained using a model based on real-time data are significantly different from the forecasts obtained using a model based on revised data.

Utility-based criteria

The criterion provides an inside into a microeconomic perspective and helps us understand how the estimations using real-time data or revised data behave for different categories of investors (different levels of risk-aversion). This paper will follow the methodology used by West (1993) and further developed by Pierdzioch (2008). A mean-variance approach will be considered in order to maximize investor's utility as in the following equation:

$$U_{t+1} = E_t(W_{t+1}) - \frac{\gamma}{2} Var(W_{t+1}) \quad (3)$$

Where, U_{t+1} represents the utility in the following period, E represents the conditional expectations operator, Var –represents the variance operator and γ shows the investors risk-aversion.

W_{t+1} denotes the investor's wealth in the following period. Taking this into consideration, the investor's wealth dynamic could be written as follows:

$$W_{t+1} = W_t[(1 - f)R_{F,t+1} + fR_{t+1}] \quad (4)$$

Does Real-Time Macroeconomic Data Ensure an Accurate Volatility Forecasting?
A Two States Approach for the US Equity Market

Where “f” denotes the proportion invested in risky assets. $R_{F,t+1}$ represents the gross return of risk-free assets, and R_{t+1} denotes the return on stocks in the considered period. The f proportion which was defined by Pierdzioch, Dopke and Harmann (2008) is going to be used in the same manner in this paper:

$$f = \frac{1 + \delta}{\delta} \left(\frac{\mu_{t+1} - R_{F,t+1}}{h_{t+1}} \right) \quad (5)$$

Where, μ_{t+1} denotes the conditional mean of the risky asset in the considered period. This is assumed to be known. h_{t+1} is considered to be a conditional variance, that could be estimated using a wide range of models, but in this paper we will use simple standard deviation, in order to avoid including too many uncertainties. Also, δ represents the investors’ coefficient of relative risk-aversion. This coefficient is assumed to be constant.

Further, the expected utility for the following period can be computed as follows:

$$E_t(U_{t+1}) = W_t(R_{F,t+1} + d_t u_t), \quad (6)$$

Where $d_t = \left[\frac{(1+\delta)}{\delta} \right] (\mu_{t+1} - R_{F,t+1})^2$ and $u_t = \frac{1}{h_{t+1}} - 0.5h_{t+1}/\widehat{h_{t+1}}^2$

$\widehat{h_{t+1}}$ – estimated volatility through the recursive model

h_{t+1} – observed volatility from the market

Taking all of this into consideration, an average utility could be defined as follows:

$$\bar{U} = \frac{1}{TF} \sum_{t=1}^{TF} W_t(R_{F,t+1} + d_t u_t) \quad (7)$$

where TF denotes the total number of forecasts. This formula leads to a measure that could be used to compare different models based on a utility criteria through one investor’s beliefs regarding risk-aversion.

Hence, we should conclude that: the higher the average utility, the better the model.

The two regimes approach

As outlined previously, we will also apply an approach based on two regimes of volatility, in order to identify how econometric models behave in these periods, using both types of data: real-time data and revised data. Low volatility period is defined here as the period when the VIX index is lower than the median observation period and vice-versa for a high volatility regime. Recall that this approach has not been applied yet and it can be refined later. Thus, the same

methodology will be applied for each of the above two states: high volatility and low volatility state.

3. The data

The main data set used is the one related to the US stock market volatility, measured by the VIX index. The frequency of data is one month. In order to get eloquent results, there was chosen the longest period for which data were available for all considered variables (31/03/1997-31/10/2016). For the VIX index were used observation registered at the end of each month (end of month data).

The VIX index is going to be used to measure market volatility in order to confer this research a touch of applicability. This index is used as an underlying asset for a large range of options and financial instruments, enabling the opportunity to trade based on the results of this research.

Thus, in the considered regression models, the VIX index will be the dependent variable. There will be a total of four independent variables present in all models, such as:

- VIX index's value from the previous period (VIX (-1))
- The difference between S&P 500 and the 50-day moving average and the 100 days: DMA50 and DMA100
- The difference between yields on short-term US government securities and long-term US government securities (T-Bills 3M - 10Y T-Bills). This variable desires to take into account in the model how the yield curve change over time, whether it becomes steeper or flatter.

It should be noted again that all data is on a monthly basis for the same period of time:31/03/1997-31/10/2016

Further, other independent variables will be presented which are going to be added to the model one at a time:

ISM PMI (ISM Purchasing Managers' Index)

The number of newly created jobs in non-farming sectors (NFP)

New industrial orders

Annual inflation rate

It should also be noted that for the last four independent variables, both real-time data and revised data will be used, in order to get a conclusion relative to the scope of this paper: development of volatility forecasts based on real-time data does not lead to results materially different from making a forecast using the revised data.

4. Results

After applying the proposed methodology, we have obtained a set of coefficients for each model. In order to properly implement the estimation algorithm, it is required to test the statistical significance of each estimated

Does Real-Time Macroeconomic Data Ensure an Accurate Volatility Forecasting? A Two States Approach for the US Equity Market

parameter. This verification is performed using the t-Student test which is presented in the below table:

Table 1. T-Statistics for estimated coefficient

t-Stat			t-Stat		
Variables	Real-time data	Revised data	Variables	Real-time data	Revised data
Intercept	2.37	2.12	Intercept	6.12	6.34
3M T-Bills first diff	-2.53	-2.55	3M T-Bills first diff	-2.75	-2.70
VIX(-1)	15.25	15.49	VIX(-1)	14.99	15.09
DMA(50)	-3.28	-3.18	DMA(50)	-3.39	-3.56
DMA(100)	0.99*	0.94*	DMA(100)	0.99*	1.23*
PMI	-2.90	-2.76	NFP	-3.03	-2.50
t-Stat			t-Stat		
Variables	Real-time data	Revised data	Variables	Real-time data	Revised data
Intercept	6.18	6.16	Intercept	5.24	5.24
3M T-Bills first diff	-2.81	-2.78	3M T-Bills first diff	-1.81*	-1.82*
VIX(-1)	16.00	16.01	VIX(-1)	14.95	14.95
DMA(50)	-3.32	-3.30	DMA(50)	-3.30	-3.30
DMA(100)	0.79*	0.77*	DMA(100)	0.58*	0.58*
Industrial Orders	-2.74	-2.64	Inflation	-0.71*	-0.71*
t-Stat			t-Stat		
Variables	Real-time data	Revised data			
Intercept	2.19	1.77*			
3M T-Bills first diff	-2.58	-2.56			
VIX(-1)	14.25	14.49			
DMA(50)	-3.39	-3.43			
DMA(100)	1.15*	1.27*			
PMI	-2.60	-2.58			
NFP	-2.56	-2.28			
Industrial Orders	-2.42	-1.88*			
Inflation	-0.62*	-1.22*			

* indicates coefficients that are not significantly different from zero according to Student test

According to the test's methodology, it can be seen that most coefficients are significantly different from zero, with the exception of those for DMA100 and annual inflation rate. Bearing this in mind, we can conclude that the values obtained can be used to estimate volatility using both, real-time and revised data.

The above-described methodology involves choosing an out-of-sample period. For this period of time, we are going to calculate statistical indicators in order to compare the results using the two types of data.

In-sample: 31/03/1997 – 31/12/2013

Out-of-sample: 31/01/2014-30/10/2016

The coefficients were estimated in-sample and thereafter the forecasts were made for the out-of-sample period. Based on these projections the RMSE (root-mean-squared-errors) and, further, Theil's U were calculated. The ultimate

aim is to highlight whether errors for real-time data are higher than those for revised data.

Table 2. Statistical criteria: RMSE and Theil's U

Model	RMSE		Theil's U
	Real-time data	Revised data	
PMI	24.69	24.59	1.0039
NFP	24.27	24.19	1.0033
Industrial Orders	24.09	23.96	1.0054
Inflation	24.86	24.86	1.0000
All macroeconomic variables	24.29	24.07	1.0092

Findings from the computation of RMSE and Theil's U were as expected and according to the literature. There are no significant differences in terms of estimation errors for any of the considered models. As expected, the largest difference was recorded when all variables were used, while there is a greater source of error that contributed to getting a higher value for Theil's U.

On the other hand, one can see that in the model in which was added only inflation besides the existing four variables, the differences are almost non-existent. This can be explained by the fact that the revised inflation for the United States seldom brought significant changes in the last 20 years. Thus, the two series for inflation: real-time and revised, have many common elements, leading to very close estimates.

There was also computed the Diebold-Mariano test to further highlight that the two types of predictions (based on real-time data or revised data) do not lead to significantly different results. The obtained values for DM test are presented in the following table:

Table 3. Diebold-Mariano statistics

Model	Diebold-Mariano test
PMI	0.73
NFP	0.66
Industrial Orders	0.94
Inflation	0.01
All macroeconomic variables	1.15

After the interpretation of Diebold-Mariano's statistics, it can be concluded that the results of the two types of estimates are not significantly different. Hence,

Does Real-Time Macroeconomic Data Ensure an Accurate Volatility Forecasting?
A Two States Approach for the US Equity Market

forecasts based on the first occurring data (real-time) did not differ greatly from those made on the basis of the revised data when further details are known. In conclusion, real-time data adds extra value, considering how quickly it is published and distributed to all market participants, contributing significantly to the reduction of informational asymmetry.

The approach stated above could conceal weaknesses, which may lead to results less close to reality. It is possible that using real-time data to be successful in certain periods and less in others (e.g. Low volatility or volatility periods), but on average the results would not be significantly different from those obtained by using revised data. Thus, it was deemed appropriate to do the same tests, but considering the separation of observations into two periods (regimes) as was presented in the methodology chapter: one period of high volatility and one period of low volatility.

Table 4. Statistical criteria: RMSE and Theil's U for Low Volatility Regime

Model	RMSE - Low Volatility Regime		Theil's U
	Real-time data	Revised data	
PMI	18.83	18.71	1.0061
NFP	18.18	18.08	1.0056
Industrial Orders	18.29	18.13	1.0085
Inflation	19.33	19.33	1.0000
All macroeconomic variables	18.66	18.33	1.0180

Table 5. Diebold-Mariano Statistics for Low Volatility Regime

Model	Diebold-Mariano test - Low Volatility Regime
PMI	1.01
NFP	0.95
Industrial Orders	1.14
Inflation	0.01
All macroeconomic variables	1.40

Table 6. Statistical criteria: RMSE and Theil's U for High Volatility Regime

Model	RMSE - High Volatility Regime		Theil's U
	Real-time data	Revised data	
PMI	15.97	15.96	1.0008
NFP	16.08	16.07	1.0005
Industrial Orders	15.68		

		15.66	1.0013
Inflation	15.64	15.64	1.0000
All macroeconomic variables	15.61	15.60	1.0008

Table 7. Diebold-Mariano Statistics for High Volatility Regime

Model	Diebold-Mariano test - High Volatility Regime
PMI	0.12
NFP	0.08
Industrial Orders	0.17
Inflation	0.01
All macroeconomic variables	0.11

For each of these previously mentioned regimes there were calculated the same statistics: RMSE, Theil's U and Diebold-Mariano test. The intuition that there are differences depending on the period in which we are (high or low volatility) is correct.

In the case of periods of high volatility, the results revealed very small differences. During periods of low volatility, differences between estimates increase but continue to not be significant, both according to Theil's U and Diebold-Mariano's test. Therefore, it can be said that regardless of the forecasting period, no significant differences between the two types of estimations have been registered. Although higher errors were obtained for periods of low volatility, these were not significant enough for the conclusion of this study to be questioned.

The latter approach will refer to a utility-based criterion which depends on two types of risk-aversion (1 and 10 units):

Table 8. Utility-Based Criterion

Model	Real-time data		Real-time data	
	δ	$\delta = 10$	δ	$\delta = 10$
PMI	40.921	22.324	40.999	22.370
NFP	41.310	22.523	41.302	22.520
Industrial Orders	41.257	22.474	41.247	22.449
Inflation	40.693	22.197	40.692	22.197
All macroeconomic variables	41.171	22.405	41.004	22.313

Does Real-Time Macroeconomic Data Ensure an Accurate Volatility Forecasting? A Two States Approach for the US Equity Market

In the case of this latter approach, it can be seen that the results are in line with those achieved until now: the average utility of an investor will feel does not differ significantly between the two ways of forecasting (using real-time data and revised data). Furthermore, for the out-of-sample period, it can be seen that for each of the estimated models there was obtained a greater utility when it was considered a smaller risk aversion (one unit) (i.e. when it invested much more in risky assets). Getting a higher utility, in this case, was determined by a bull market period as out-of-sample, which favored investments in risky assets.

5. Conclusions

The results obtained through this research are in line with other authors' findings. It was revealed that forecasting volatility using real-time data leads to comparable results to those obtained when using revised data. Also, it was confirmed that the results did not differ significantly between periods of high volatility and low volatility, in line with the results obtained for single period estimations. Moreover, the projections made were based on macroeconomic variables with high potential to generate volatility on the US market, leading to an econometric model which explains in a large manner the variation of the considered dependent variable.

Explanatory variables were chosen based on the practical trading experience and on the economic theory applicable to the economy of the United States. Hence, the investors on the U.S. equity market are very careful about labor market evolution because job creation is a very sensitive issue in the United States, taking into consideration that its population grows at a high pace. Thus, the FED has two important objectives which are hard to be met together: full employment and price stability. Taking into consideration all these facts it is very clear that we had to take into consideration a variable related to the labor market and NFP (non-farm payrolls) is the most relevant one. In addition, the FED's second objective reveals that we should also select inflation rate as an important explanatory variable.

Other variables considered were related to the industrial orders and to the transmission mechanism of the implemented monetary policy (difference between short-term and long-term interest rates). The above mentioned independent variables were used in the proposed model to forecast equity market volatility and the results for real-time data and revised data were compared out-of-sample using 2 statistical indicators (RMSE and Theil's U) and one utility function for a risk-averse and a risk-seeker investor.

Comparisons based on all of these criteria led to the same result that there is no significant difference between forecasting volatility using real-time or revised data, regardless of the type of investor or the volatility regime in which we lie. Therefore, the opportunity cost of waiting for the revised data is not justified and

we should act based on real-time which is a very good proxy for how the real amplitude of the desired explanatory variables.

This work may be continued by extending the research in different fields, but taking into consideration that it should bring a touch of novelty we may propose a new approach. Hence, we have found that there is very little addressed in the relevant literature the idea of comparing the results obtained using regressions based on macroeconomic variables to those that could be obtained through different types of GARCH models with more volatility states based on Markov Chains. It may be interesting to look further into how much closer to reality is each of them and how much they differ from projections made by regression models presented in this paper.

REFERENCES

- [1] Beltratti, A., Morana, C.(2006),*Breaks and Persistency: Macroeconomic Causes of Stock Market Volatility*; *Journal of econometrics*, 131(1-2), 151-177;
- [2] Beetsma, R., Giuliadori, M. (2012),*The Changing Macroeconomic Response to Stock Market Volatility Shocks*; *Journal of Macroeconomics*, 34(2), 281-293;
- [3] Diebold, F.X., Mariano, R.S. (1995), *Comparing Predictive Accuracy*; *Journal of Business and Economic Statistics*, 13, 253-265;
- [4] Engle, R.F., Ghysels, E., Sohn, B. (2013), *Stock Market Volatility and Macroeconomics Fundamentals*; *The Review of Economics and Statistics*, 95,776-797;
- [5] Gonzalez-Rivera, G., Lee, T.H., Mishra, S. (2004), *Forecasting Volatility: A Reality Check Based on Option Pricing, Utility Function, Value-at-Risk, and Predictive Likelihood*; *International Journal of Forecasting*, 20(4), 629-645;
- [6] Hamid, A., Heiden, M.(2015),*Forecasting Volatility with Empirical Similarity and Google Trends*; *Journal of Economic Behavior & Organization*, 117, 62-81;
- [7] Hamilton, J.D., Lin, G.(1996), *Stock Market Volatility and the Business Cycle*; *Journal of Applied Econometrics*, 11,573-593;
- [8] Marcucci, J.(2005), *Forecasting Stock Market Volatility with Regime-Switching GARCH Models*; *Studies in Nonlinear Dynamics & Econometrics*, 9, 1-55;
- [9] Maysami, R.C., Howe, L.C., Rahmat, M.A.(2005),*Relationship between Macroeconomic Variables and Stock Market Indices: Cointegration Evidence from Stock Exchange of Singapore's All-S Sector Indices*; *Journal Pengurusan (UKM Journal of Management)*, 24;
- [10] Nikkinen, J., Omran, M., Sahlström, P., Äijö, J.(2006),*Global Stock Market Reactions to Scheduled US Macroeconomic News Announcements*; *Global Finance Journal*, 17(1), 92-104;

Does Real-Time Macroeconomic Data Ensure an Accurate Volatility Forecasting?
A Two States Approach for the US Equity Market

- [11] Pesaran, M.H., Timmermann A.(2000), *A Recursive Modelling Approach to Predicting UK Stock Returns*; *Economic Journal*, 110,159-191;
- [12] Pierdzioch, C., Dopke, J., Harmann, D.(2008), *Forecasting Stock Market Volatility Macroeconomic Variables in Real Time*; *Journal of Economics and Business*, 60, 256-276;
- [13] Rangel, J.G. (2011),*Macroeconomic News, Announcements, and Stock Market Jump Intensity Dynamics*; *Journal of Banking & Finance*, 35(5), 1263-1276;
- [14] Roh, T.H. (2007), *Forecasting the Volatility of Stock Price Index, Expert Systems with Applications*, 33(4), 916-922;
- [15] Schwert, W.G. (1989), *Why Does Stock Market Volatility Change over Time?*, *Journal of Finance*, 44, 1115-1153;
- [16] West, K.D., EdisonH.J., Cho D.(1993), *A Utility-Based Comparison of Some Models of Exchange Rate Volatility*; *Journal of International Economics*, 35,23- 45.