

stock market volatility

Lohit Adari, Abhinav Thupli, Kai Li



*BIG DATA*

**Contents**

1. **Introduction and Background**
   1. Volatility
   2. GARCH
   3. HMM
2. **Methodology**
3. Workflow
4. Phase A
5. Phase B
6. Phase C
7. Evaluation
8. **Results and Analysis**
9. Data Description
10. DAX Data
11. SMI Data
12. CAC Data
13. FTSE Data
14. Conclusion
15. References

**1 Introduction and Background**

1. **Volatility**

Volatility is defined as the first difference of a series. The value of financial instruments depends on the expected volatility (covariance structure) of returns. Volatility is analogous to risk: financial institutions undertake volatility assessments in their risk analysis exercise. An important research objective in time series analysis is to test hypotheses and estimate the relationships amongst such variables. However, inferences drawn from a static analysis would not be valid if the examined series are realizations of non-stationary processes. Similarly, efficient analysis of a volatile time series can only be accomplished by addressing the issue of time-varying volatility.

1. **GARCH**

If an autoregressive moving average model (ARMA model) is assumed for the error variance, the model is a generalized autoregressive conditional heteroscedasticity (GARCH, Bollerslev (1986)) model. GARCH models preserve the persistence of the process volatility in that small variations tend to follow small variations, and large variations tend to follow large ones. Incorporating GARCH models with hidden Markov chains, where each state (regime) of the chain implies a different GARCH behavior, extends the dynamic formulation of the model and enables a better fit for a process with a more complex time-varying volatility structure. However, a major drawback of such models is that estimating the volatility with switching regimes requires knowledge of the entire history of the process, including the regime path. Incorporating GARCH models with a hidden Markov chain, where each state of the chain (regime) allows a different GARCH behavior and, thus, a different volatility structure, extends the dynamic formulation of the model and potentially enables improved forecasts of the volatility. Unfortunately, the volatility of a GARCH process with switching regimes depends on the entire process history, including the regime path, making the derivation of a volatility estimator impractical. Using a method to calculate variance in fundamental statistics to estimate the current level of volatility, will use a weighting scheme that will assign more weight 1 to recent data.

The variable uses the amount of weight given to the observation *i* days ago. The restrictions on α are the following:

1. ≥ 0

2. Choose when i > j, were less weight is given to older observations.

3. = 1

1. **HMM**

A statistical Markov model called a hidden Markov model (HMM) assumes that the represented system is a Markov process with unobserved (hidden) states. In a hidden Markov model, the form is not immediately apparent, but the output is visual and depends on the state. A probability distribution over all potential output tokens exists for each state. As a result, the order of tokens produced by an HMM provides insight into states' order. The term "hidden" Markov model" refers to the state sequence that the model traverses, not to the parameters of the model; even if the parameters of the model are precisely known, the term "hidden" Markov model" still applies.

**Methodology**

1. **Workflow**

The workflow is depicted in the below figure. The complete methodology of the training process and the forecast estimation has been logically divided into three phases. The following subsections comprehensively describe the functions in each step. The outer loop explains that the training happens again for each forecast by including the forecasted value in the previous iteration. This is required because, with the newly added forecast value, the Markov Regimes might change along with their transition probabilities. This, in turn, results in different GARCH models built on each Regime. The overall approach is that different Markov regimes are identified based on the training data by creating a Hidden Markov Model, and the probability of transition between regimes is calculated. Then separate GARCH models are built on data points belonging to each Regime. Then the current state depicted by the latest point is considered. Then the estimates from each GARCH model built on each Regime are weighted over the respective probability of transition/switching from the current Regime to the Regime on the individual GARCH model created. We are considering a Markov Model with at most two states for the current project.

1. **Phase A**

In this phase, Training data is taken, and Hidden Markov Model is built on it. This provides probabilities of each data point belonging to different regimes. Then we get to know which points belong to which regime. The Transition Probability Matrix (PMAT) contains the transition probabilities from one state to another. Since we have two states, we have 4 possibilities.

1. **Phase B**

In this phase, all the points belonging to each Regime are considered separately because they vary considerably according to the HMM. Hence the hypothesis is that they have a different nature or behavior. Therefore, a different GARCH model is built on each Regime.

The p - the lag of the residual terms (conditional mean terms) and q - the lag of the conditional variances are the two parameters for any GARCH model.

The p and q values should be chosen optimally for best results, as they balance generalization and specification. We can select the optimal p and q values by experimenting with different p and q values (we have tested with p and q values ranging from 1 to 2), then choosing that pair of p and q for which there is the highest Log-likelihood.

Once the optimal p and q values are found, the GARCH model with the optimal p and q values is built on each Regime.

1. **Phase C**

In this phase, we get the forecast from the MRS-GARCH. The estimates should be the current state of the Regime to which the latest data point belongs. The probability of transition from the current Regime to each Regime is taken. The individual GARCH model estimates from each Regime are multiplied by the possibility of transition from the current Regime to that respective Regime on which the GARCH is built. This is a probabilistic weighted average of the individual GARCH estimates of each Regime.

Since the data is highly volatile, which cannot be effectively predicted by a single GARCH model, this MRS-GARCH uses an ensemble approach, where estimates from multiple GARCH models built on different sub-samples of data depending on the nature of the data, then an aggregated forecast is taken which would give a more efficient estimate of the forecasted value.

1. **Evaluation**

* **RMSE:** Root Mean Squared Error is the squared difference between predicted and actual values, where a square root is taken over their mean. It is a measure of error in the forecast.
* **Value Forecast RMSE:** The value forecasts and the data values are plotted, and their RMSE is taken.
* **Residuals Forecast RMSE:** The data residuals and the forecast of residuals are taken. Residual refers to the difference being the current and the previous data value.

**3 Results and Analysis**

1. **Data Description**

The above-explained methodology has been experimented on four different Stock Exchange time series datasets. From each stock exchange, the index values for 500 days after Jan 2009. Each dataset is trained on 500 instances, and the forecast is taken cumulatively for the following five values. The 4 Stock Exchange datasets are:

• **DAX**: Deutscher Aktienindex (German Stock Index) - is a blue-chip stock market index consisting of the 30 major German companies trading on the Frankfurt Stock Exchange.

• **SMI**: Swiss Market Index - Switzerland’s blue-chip stock market index makes it the most important in the country. It comprises 20 largest and most liquid Swiss Performance Index (SPI) large- and mid-cap stocks.

• **CAC**: Cotation Assiste en Continu (French Stock Index) - is a benchmark French stock market index. The index represents a capitalization-weighted measure of the 40 most significant values among the 100 highest market caps on the Euronext Paris (formerly the Paris Bourse).

• **FTSE**: Financial Times Stock Exchange 100 Index - is a share index of the 100 companies listed on the London Stock Exchange with the highest market capitalization. For each dataset, three graphs as presented as results. First, the state probabilities to which each point belongs. They represent which point belongs to which Markov Regime and which possibility; their implied state is the regime of the greater likelihood. Second, the forecasted values and the actual data are plotted to compare the relative distance between actual and projected values, which the RMSE value on the plot can quantitatively present. Third, the data residuals are plotted along with forecasted residuals to compare the relative distance, shown as RMSE value on the plot.

1. **DAX Data**

The figure below presents results with all the three plots for the DAX dataset.

Chart, histogram

Description automatically generated

Probabilities of different points belonging to two different Markov Regimes. Results of MRS-GARCH for DAX Data.

Chart, line chart

Description automatically generated

Plotting the data and the forecasts to compare them along with RMSE value. Results of MRS-GARCH for DAX Data.

Chart, scatter chart

Description automatically generated

Plotting the residuals of the data and the residual forecasts to compare them along with RMSE value. Results of MRS-GARCH for DAX Data.

1. **SMI Data**

The figure below presents results with all the three plots for the SMI dataset.

Chart

Description automatically generated with low confidence

Probabilities of different points belonging to two different Markov Regimes. Results of MRS-GARCH for SMI Data.

Chart, line chart

Description automatically generated

Plotting the data and the forecasts to compare them along with RMSE value. Results of MRS-GARCH for SMI Data.

Chart, scatter chart

Description automatically generated

Plotting the residuals of the data and the residual forecasts to compare them along with RMSE value. Results of MRS-GARCH for SMI Data.

1. **CAC Data**

The figure below presents results with all the three plots for the CAC dataset.

Chart

Description automatically generated

Probabilities of different points belonging to two different Markov Regimes. Results of MRS-GARCH for CAC Data.

Chart, scatter chart

Description automatically generated

Plotting the data and the forecasts to compare them along with RMSE value. Results of MRS-GARCH for CAC Data.

Chart, scatter chart

Description automatically generated

Plotting the residuals of the data and the residual forecasts to compare them along with RMSE value. Results of MRS-GARCH for CAC Data.

1. **FTSE Data**

The figure below presents results with all the three plots for the FTSE dataset.

Chart

Description automatically generated

Probabilities of different points belonging to two different Markov Regimes. Results of MRS-GARCH for FTSE Data.

Chart, scatter chart

Description automatically generated

Plotting the data and the forecasts to compare them along with RMSE value. Results of MRS-GARCH for FTSE Data.

Chart, scatter chart

Description automatically generated

Plotting the residuals of the data and the residual forecasts to compare them along with RMSE value. Results of MRS-GARCH for FTSE Data.

**4 Conclusion**

The MRS-GARCH implementation explained in the report has experimented on 4 different datasets with volatile time series. Forecasts for all the datasets from the MRS-GARCH model show promising results with considerable RMSE values. Hence, this model works effectively for volatile time series datasets.

**5 References**

<https://link.springer.com/chapter/10.1007/978-3-642-51182-0_10>

<https://academic.oup.com/jfec/article/2/4/493/900480>

<https://onlinelibrary.wiley.com/doi/abs/10.1002/for.1119?casa_token=tZ0ozx-iLJQAAAAA:s_lvcftXRw_xNgtGEUzWM1qtNYuxz60_8lQMzbbXIz6ISRPOynsUylX-7uxaEffi_LNUdGNI6rb_xw>

<https://money.cnn.com/data/world_markets/europe/>

<https://rdrr.io/rforge/RHmm/>