A GARCH Tutorial with R

Marcelo S. Perlin, Mauro Mastella, Daniel Vancin, Henrique Ramos

Abstract: In this tutorial paper we will address the topic of volatility modeling in R. We will discuss the underlying logic of GARCH models, their representation and estimation process, along with a descriptive example of a real-world application of volatility modelling: we use a GARCH model to investigate how much time it will take, after the latest crisis, for the Ibovespa index to reach its historical peak once again. The empirical data covers the period between years 2000 and 2020, including the 2009 financial crisis and the current 2020's episode of the COVID-19 pandemia. We find that, according to our GARCH model, Ibovespa is likely to reach its peak once again in 2,5 years. All data and R code used to produce this tutorial are freely available on the internet and all its results can be easily replicated.

Keywords: volatility, GARCH, ibovespa, tutorial

Resumo: Neste artigo tutorial abordaremos o tópico da modelagem de volatilidade na plataforma R. Discutiremos a lógica subjacente dos modelos GARCH, seus processos de representação e estimação, juntamente com um exemplo descritivo de uma aplicação no mundo real: usamos um modelo GARCH para investigar quanto tempo levará, após a última crise, para que o índice Ibovespa volte a atingir seu pico histórico mais uma vez. Os dados empíricos cobrem o período entre os anos 2000 e 2020, incluindo a crise financeira de 2009 e o episódio atual de 2020 da pandemia do COVID-19. Concluímos que, de acordo com nosso modelo GARCH, é provável que o Ibovespa atinja seu pico mais uma vez em 2,5 anos. Todos os dados e códigos R usados para produzir este tutorial estão disponíveis gratuitamente na Internet e todos os seus resultados podem ser facilmente replicados.

Palavras-chave: volatilidade, GARCH, ibovespa, tutorial

Introduction

Modeling uncertainty is a certain element of the financial practice, with important applications in portfolio allocation, risk management and pricing of financial contracts. The simple question of how much uncertainty we can expect for future prices of financial contracts resulted in a large body of literature interested in understanding the statistical properties of price changes and how we can use it to make better predictions (Brockwell & Davis, 2016; Francq & Zakoian, 2019).

The ARCH (Autoregressive Conditional Heterokedasticity) and GARCH (Generalized Autoregressive Conditional Heterokedasticity) family of models constitute a seminal innovation in the field of financial modeling, resulting in a 2003 Nobel prize for its creator, Robert Engle. Its main contribution is to allow uncertainty to be a dynamic process. That is, instead of assuming that the level of future volatility is constant, the ARCH and GARCH family of models acknowledge its time-varying process.

In practice, the implications of dynamic volatility are as follows: financial returns will more often than expected result in large losses --- the "fat tail" effect; uncertain periods tend to cluster, with large price variations within days. As a recent and practical example, on 09/03/2020 the Ibovespa index lost 13,98% of its value in a single day due mostly to the COVID-19 episode. The index recovered 13% of its value just a couple of days later, in 13/03/2020. When considering its historical distribution of price changes, both events are highly unexpected and happened just a couple of days apart.

Ignoring the ARCH effect and underplaying uncertainty can be very costly. Risk control measures in investment portfolios need to assess how likely a large loss can happen. By assuming constant volatility, an analyst underplays the inner risk of financial contracts and, possibly, lead to unexpected losses in the investment portfolio. Likewise, banking regulations, such as Basel III, requires that banks report their level of portfolio risk systematically and periodically. Thus, given that banks act as liquidity hubs, an incorrect calculation of volatility and risk can threaten the stability of a whole financial system.

In the scientific side, there is an extensive body of literature in Brazil relying on GARCH models. A non exhaustive overview of topics recently covered would include the covariance of Brazilian stock exchange (Mastella & Coster, 2014), optimal minimum variance portfolios (Caldeira, Moura, Perlin & Santos, 2017), risk exposure in Brazilian sectoral stock indices (Lobão & Fernandes, 2018; Bernardino, Brito, Ospina & Melo, 2019), correlation between stock returns (Costa, Júnior & Menezes, 2019), among others.

In this paper we will present a practical tutorial for the ARCH/GARCH family of models, presenting its underlying motivations and a reproducible study case in R, with real data and reproducible code. Given the recent 2020's turmoil in financial markets and huge drop in stock prices on the stock exchange, we offer a step by step guide to model financial returns. Our goal is to use a simulated GARCH model to assess when will Ibovespa reach its historical peak once again and, consequently, the time it will take for the market to recover from the current crisis. Thus, our contribution to researchers is to provide a step by step tutorial for beginners to investigate a classical issue in finance and, for experienced researchers, the tutorial allows to easy reformat the code for other applications and models.

The article is organized as follows: we start with a description of the method and brief presentation of the underlying theory, followed by an empirical application of the model in real world data. We end the paper with the usual conclusion section.

A GARCH Model

As noted in the literature, financial time series such as stock prices, inflation rates, and exchange rates present the phenomenon of volatility agglomeration. In other words, they present periods in which the prices/values of these series show significant fluctuation, followed by periods in which there is low variation.

In the modeling side, the most used quantitative tool for researchers in finance is the ordinary least squares (OLS) model. This is a natural choice, because applied econometricians are typically called upon to determine how much one variable will change in response to a change in some other variable (Engle, 2001). Stationarity becomes an issue when modelling financial returns as they tend to have non-constant volatility (heteroskedasticity). In this situations, the use of OLS models could draw misleading conclusions in an empirical study where volatility plays a role.

The typical warning is that, in the existence of heteroskedasticity, the regression coefficients for an ordinary least squares regression are still unbiased, but the standard errors and confidence intervals estimated by traditional procedures will be too narrow, giving a false sense of accuracy. Instead of considering this as a problem to be fixed (through the use of heteroskedasticity-consistent standard errors, for example), ARCH and GARCH models care for heteroskedasticity as a variance to be modelled. Therefore, not only are the shortcomings of least squares corrected, but a forecast is computed for the variance of each error term.

Formally, the assumption of constant variance in a model is called homoskedasticity. The opposite, non-constant variance, is the focus of ARCH/GARCH models. Based on this fact, the estimation and prediction of volatility should be different from those observed in classic time series models, such as the ARMA (Autoregressive Moving Average) by Box & Jenkins (1976). This fact occurs because this model does not accept some of the mentioned stylized facts about the volatility, such as conditional/unconditional non-normality and non-constant conditional variance over time.

To address the issues of heteroskedasticity, Engle (1982), presented the ARCH models in a study of inflation rates. These models seek to estimate time-dependent volatility as a function of the previously observed volatility. The original ARCH model proposed by Engle (1982) modeled the variance of errors in a regression model as a linear function of lagged values of squared regression errors. We can write an ARCH (m) model as:

$$R_t = \sigma_t \epsilon_t \text{ (conditional average)}$$

$$\sigma_t^2 = \alpha_0 + \alpha_1 a_{t-1}^2 + \ldots + \alpha_q a_{t-m}^2 \text{ (conditional variance)}$$

As noted, the ARCH model has a specification for both the conditional average as for the conditional variance. Specifically, a ARCH method models the variance at a time step as a function of the residual errors from a mean process.

Although the ARCH model may be simple, it often demands many parameters/lags to properly explain the volatility process of an asset return (Tsay, 2005). To solve this problem, Bollerslev (1986) extended Engle´s original work by developing a technique that allows the conditional variance to be an ARMA process. In other words, a GARCH model is equivalent to a ARCH model with many, many lags. Formally, we define a GARCH (p, q) model as follows:

$$R_t = \sigma_t \epsilon_t \text{ (conditional average)}$$

$$\sigma_t^2 = \alpha_0 + \Sigma_{i=1}^{q} \alpha_i a_{t-i}^2 + \Sigma_{j=1}^{p} \beta_j \sigma_{t-j}^2 \text{ (conditional variance)}$$

where $\{\epsilon_t\}$ is a sequence of independent and identically distributed random variables with mean zero, variance equal to 1, and $\alpha_0 > 0$ for $i > 0$. The coefficients α_i must satisfy some regularity conditions to ensure that the unconditional variance of a_t is finite. In practice, ϵ_t is often assumed to follow the standard normal or a standardized Student-t or a generalized error distribution. Hence, if all values of β_j equals zero, the GARCH(q, p) model is equivalent to and ARCH(q) model. The benefits of the GARCH model should be clear: a high-order ARCH model may have a more parsimonious GARCH representation that is much easier to identify and estimate. This is particularly true because all coefficients in the equation above must be positive. Likewise, to ensure that the variance is finite, all characteristic roots of the GARCH equation have to lie inside the unit circle. Clearly the more prudent model will result in fewer coefficient restrictions (Enders, 2018).

In practice, the parameters of a GARCH model are estimated from the data. The standard method for the estimation of parameters of a GARCH model is called Maximum Likelihood (ML). The main idea of the method is to find parameters that match, as close as possible, the distribution of predictions from the model against the distribution of the real data. Nowadays, several econometric softwares, including R, have a GARCH toolbox and ML estimation of standard GARCH models takes just a few seconds on a modern computer. From a theoretical viewpoint, ML estimators benefit from being asymptotically optimal under certain conditions. Alternative estimation methods also exists, such as Bayesian methods (Fioruci et all, 2013).

Alternative Volatility Models

As the ARCH and GARCH models set the cornerstone framework for volatility modelling, the reader must be aware that further research improved and extended these models in order to incorporate other stylized facts noted in financial markets. One possible critique to ARCH-GARCH models is to model the conditional variance as being a linear function of the squared past innovations (Francq and Zakoian, 2019).

The Exponential GARCH model (EGARCH) proposed by Nelson (1991) allows asymmetric effects depending on the sign of the random innovation (error term). This follows the idea that volatility may rise in response to 'bad news', and can be reduced after 'good news' (Nelson, 1991). The Threshold-GARCH (TGARCH) model of Zakoian (1994) and the model of Glosten et al. (1993) also deal with the different asymmetric effects of past innovations. Table 01 presents the main formulation of ARCH, GARCH, EGARCH and TGARCH models. Although the scope of this tutorial is not to explore in detail the variety of conditional heteroskedastic models, we believe a novice researcher should be familiar with other types of models.

Table 01 - Conditional Heteroscedastic Models.

Model	Variance Equation
ARCH (q)	$R_t = \sigma_t \epsilon_t,\ \ \sigma_t^2 = \alpha_0 + \alpha_1 R_{t-1}^2 + ... + \alpha_q a_{t-q}^2$
GARCH (p,q)	$R = \sigma_t \epsilon_t,\ \ \sigma_t^2 = \alpha_0 + \Sigma_{i=1}^{q} \alpha_i R_{t-i}^2 + \Sigma_{j=1}^{p} \beta_j \sigma_{t-q}^2$

EGARCH (p,q)	$R_t = \sigma_t \epsilon_t, \quad ln(\sigma_t^2) = \alpha_0 + \Sigma_{i=1}^{q} \frac{\lvert R_{t-i} \rvert + \delta_i R}{\sigma_{t-i}} + \Sigma_{j=1}^{p} \beta_j ln(\sigma_{t-j}^2)$
TGARCH (p,q)	$R_t = \sigma_t \epsilon_t, \; \sigma_t^2 = \alpha_0 + \Sigma_{i=1}^{q} (\alpha_i + \gamma_i N_{t-i}) R_{t-i}^2 + \Sigma_{j=1}^{p} \beta_j \sigma_{t-j}^2$

Regarding the EGARCH model, the δ_i parameter implies the leverage effect of a_{t-i}, in which volatility tends to increase more after negative returns than in positive returns (Black, 1976). In the TGARCH model, N_{t-i} is an indicator variable which assumes value of if $a_{t-i} < 0$ and zero otherwise. Thus, a positive a_{t-i} contributes $\alpha_i a_{t-i}^2$ to volatility (σ_t^2), whereas a negative a_{t-i} contributes $(\alpha_i + \gamma_i) a_{t-i}^2$ to volatility. Considering the parameter $\gamma_i > 0$, the model uses zero as the threshold to separate the impact of past shocks (Tsay, 2005).

Application of a GARCH Model

In this section we will present a practical application of a GARCH model, with a real-world dataset. We will use a GARCH model to answer a simple question: **given the most recent financial crisis of 2020, when will Ibovespa reach its historical peak once again?** The solution to answering the question is to estimate a GARCH model for the index and use it to simulate many different future price paths.

In this application we will not explore alternative GARCH models, but will provide to the reader a hands-on experience in dealing with econometric modelling and simulation in R, a widely used programming platform in academia and financial industry. R is free and can be installed in many different operating systems, facilitating the reproduction of results by the reader.

All the data, code and results presented in this section are available in Github. Simply download the zip file in your computer and extract it to a personal folder. We organized the code as follows:

Table 02 - Description of R Scripts used in the Study

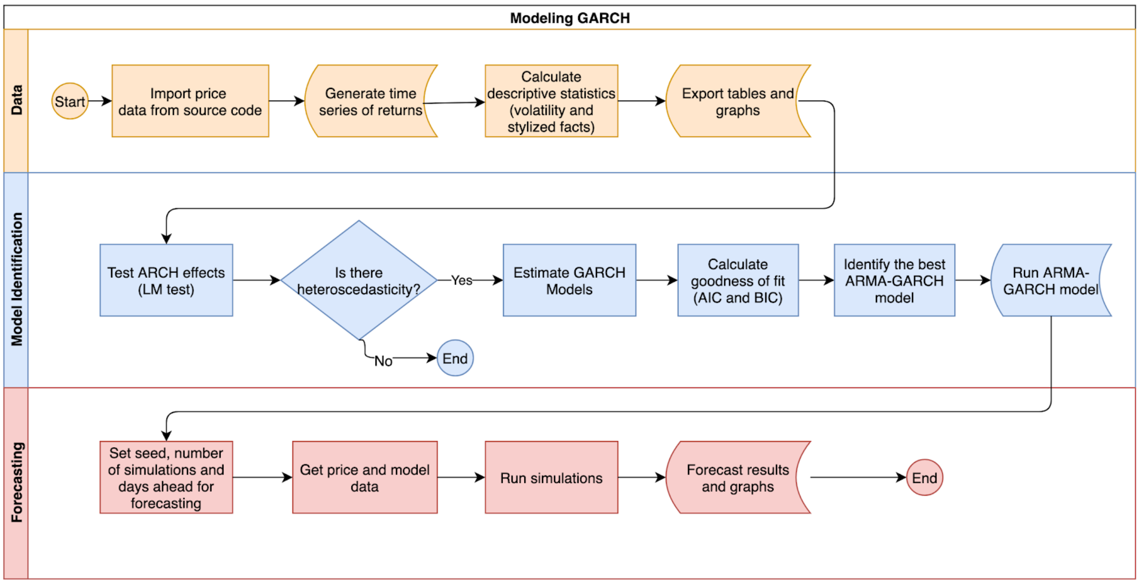
Filename	**Description**

00-Prepare_computer.R	Setup computer by installing all required R packages. **This is a mandatory step**.
01-Get_Index_Data.R	Using the internet and package BatchGetSymbols, imports a dataset of prices of the Ibovespa index.
02-Do_Descriptive_Figures.R	Creates and saves all descriptive figures presented in the paper.
03-Do_ARCH_Test.R	Performs the arch test in the data.
04-Estimate_Simple_Garch_Model.R	Estimate an introductory Garch model and present results.
05-Find_Best_Garch_Model.R	Finds the best ARMA(ar, ma)-GARCH(p,q) model for the dataset.
06-Simulate_Garch_Model.R	Simulates the previous GARCH model and plot simulated paths and probabilities.

The first script, *00-Prepare_computer.R*, will make sure all R dependencies are available in the computer. The second script, *01-Get_Index_Data.R*, will import the index data from the internet and store it as a local file, which will then be used in other scripts. The alphabetical order is not accidental, one should execute each script in the same order. We also took special attention in writing code comments that will guide the user throughout the scripts and help the learning process. It is important to notice that other files and folders available in Github should be downloaded as well as the codes described in Table 02.

We also build a flowchart, Figure 01, that can guide upcoming users to visualize and understand the steps required to reproduce the results in this section.

Figure 01 - Flowchart of R Code



The Data

The chosen data for this example is Ibovespa, a broad market index for the Brazilian equity market. Currently, 20-03-2020, it is composed of approximately 70 stocks and regarded as the main thermometer of the local market, serving as a benchmark for investments and derivatives contracts. Worth pointing out that, without loss, the analysis in this study could be conducted using an individual stock or other international stock index. In fact, the R code was designed so that the user only needs to change the ticker symbol and series name in script *01-Get_Index_Data.R*.

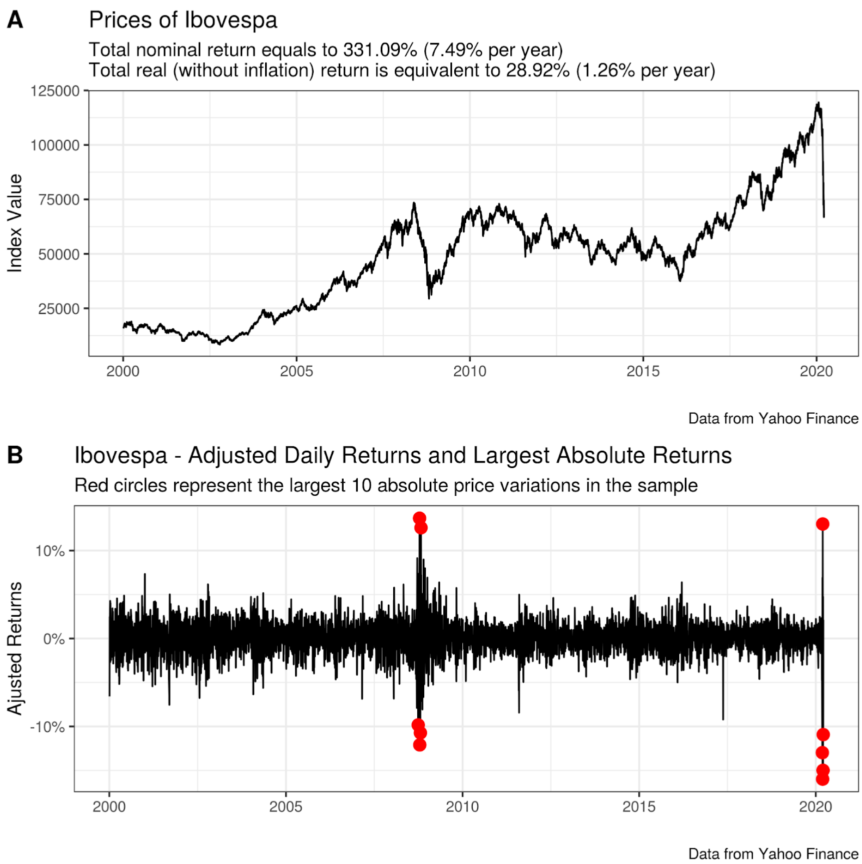
The price data is composed of daily closing values of the index from 01/01/2000 to 20/03/2020, including the 2009 financial crisis and the current 2020's episode of the COVID-19 pandemia. The origin of the data is Yahoo Finance, a vast public repository of financial data. The choice is justified by its open access nature -- anyone can use R or other platform to download daily stock prices from Yahoo Finance.

The first step of the study is to download the dataset of daily values of Ibovespa and manipulate the data. In this example, we are interested in the vector of daily returns, call it R_t. Such vector represents the return of Ibovespa at time *t*, where *t* goes from 1 to the number of observations in the sample. We calculate it based on P_t, a vector of prices, using the following formula:

$$R_t = \log\left(\frac{P_t}{P_{t-1}}\right)$$

The log return is nothing more than the log difference of the value/price from one day to the next. The use of log return is standard in financial research as its properties are convenient and facilitate the calculation of cumulative returns. In the case of GARCH models, such a property is not particularly important, meaning that the arithmetic return (percentage price change) could also be used without any loss. Next, Figure 02, we show the time series value of Ibovespa, panel A, along with its daily log returns, panel B, resulted by the execution of script *02-Do_Descriptive_Figures.R*.

Figure 02 - Prices and Returns of the Ibovespa index.



Historically, an investor of the Brazilian equity market will likely see a positive nominal (without considering inflation) return in their investments (see Panel A of Figure 02). Overall, if someone mirrored the Ibovespa composition at the beginning of 2000, he/she would have seen a total return of 331,09%, equivalent to 7,49% per year. When adjusting for inflation, however, the annual return drops to a meager 1,26% per year. Not surprisingly, the 2020's COVID-19 event of global pandemia has hurt investor's historical returns significantly. While writing this paper, prices have not yet recovered from the big drop in March of 2020. Such an episode resulted in the largest price drop in the sample, with a -14,77% price variation in the date of 12/03/2020.

Looking at panel B of Figure 02, we see that most of the returns are centered around the value of zero. Big price changes -- up or down -- tend to happen within a close period. This is what is commonly called *volatility clustering*. As an example, notice that, out of the ten greatest absolute price changes (red points in the chart), five occurred in the 2009 financial crisis, two of them with positive return and three with negative.

The lesson here is that returns of financial assets present large changes that tend to agglomerate. The ten most extreme price changes have happened mostly in two episodes, 2009's financial crisis and 2020's COVID-19 pandemia. When modelling returns, we should take such an effect into considerations, and not make the mistake of considering a constant volatility of returns. This is exactly the solution that a ARCH/GARCH model offers.

Testing for ARCH Effects

Before we estimate a GARCH model, we need to make sure the effect exists in the dataset. For that, we use the Lagrange Multiplier (LM) test for ARCH effects (Engle, 1982). It works by regressing the squared errors on its lags and testing the hypothesis that all coefficients of the lagged regression are equal to zero.

The test takes as input a time series of returns and a given lag. With both information, it tests the null hypothesis that there are no ARCH effects in the data. As a user, we need to pay attention to the *p-value* resulting from the test, which will indicate the likelihood of no ARCH effects in the data. The lower the *p-value*, the higher the chances of finding the ARCH effect. In order to execute this procedure, the reader should run the code *03-Do_ARCH_Test.R*. In Table 03 we provide the result of the ARCH LM test for several lags.

Tabela 03 - ARCH test for Ibovespa.

Lag	**Statistic**	**P-Value**
1	584.75	0.00%
2	1,262.94	0.00%
3	1,327.44	0.00%
4	1,354.80	0.00%
5	1,442.94	0.00%

Table 03 confirms the expected results, that is, the existence of a strong ARCH effect in the returns of Ibovespa. All reported *p-values* are lower than 5%, which is the usual probability threshold for rejecting the null hypothesis of no ARCH effect. With such result in hand we proceed to estimate our GARCH model.

Estimating a GARCH Model

The first step in estimating a GARCH model is identifying the model, that is, to define the number of used lags in each part. For simplicity, we will estimate a simple ARMA(0,0)-GARCH(1,1) model, that is, we only leave a constant parameter in the mean equation (no ARMA coefficients), and one lag for each term of the GARCH model. In the next section, we will go deeper into this topic and let the data select the best model by using goodness-of-fit indicators.

The estimated model in this section can be represented by the following equations:

$$R_t = \mu + \sigma_t \varepsilon_t$$

$$\varepsilon_t \sim N(0,\ 1)$$

$$\sigma_t^2 = \alpha_0 + \alpha_1 \sigma_{t-1}^2 \varepsilon_{t-1}^2 + \beta\, \sigma_{t-1}^2$$

Next, Table 04, we present the estimated results and statistics generated by *04-Estimate_Simple_Garch_Model.R*.

Table 04 - Estimation Results of GARCH Model.

	GARCH(1,1)
μ	0.001**
α_0	0.000***
α_1	0.079***
β	0.899***
Num. obs.	4927

AIC	-5.425
Log Likelihood	-13.369.111
***p < 0.001, **p < 0.01, *p < 0.05	

The first step after estimating a GARCH model is looking at the significance of its parameters. In Table 04 we see that all coefficients are statistically significant at the 5% level (see asterisks next to the parameter's values). In the mean equation, the value of the intercept, μ, is positive, indicating that, on average, returns tend to be positive. This implies that, as expected, Ibovespa is likely to have a positive return and increase its value in the long run.

In the variance equation, we must pay attention to the value of the α_1 and β. We expect them to be positive and their sum should not be more than one. In our case, their sum equals to 0,98. If this is not the case and the sum of ARCH and GARCH parameter is higher than one, the model has explosive behavior -- will keep increasingly its value infinitely. If this occurs, it is very likely that there was an error in the estimation of the model, probably in the input series.

The last two statistics, *AIC* and *Log Likelihood* are related to the estimation of the model. *AIC* is a measure of goodness of fit, and can be used to select an optimal lag, as we will explain in the next section of the tutorial. The *Log Likelihood* indicates the final value of the log likelihood after being maximized in search of all model's parameters.

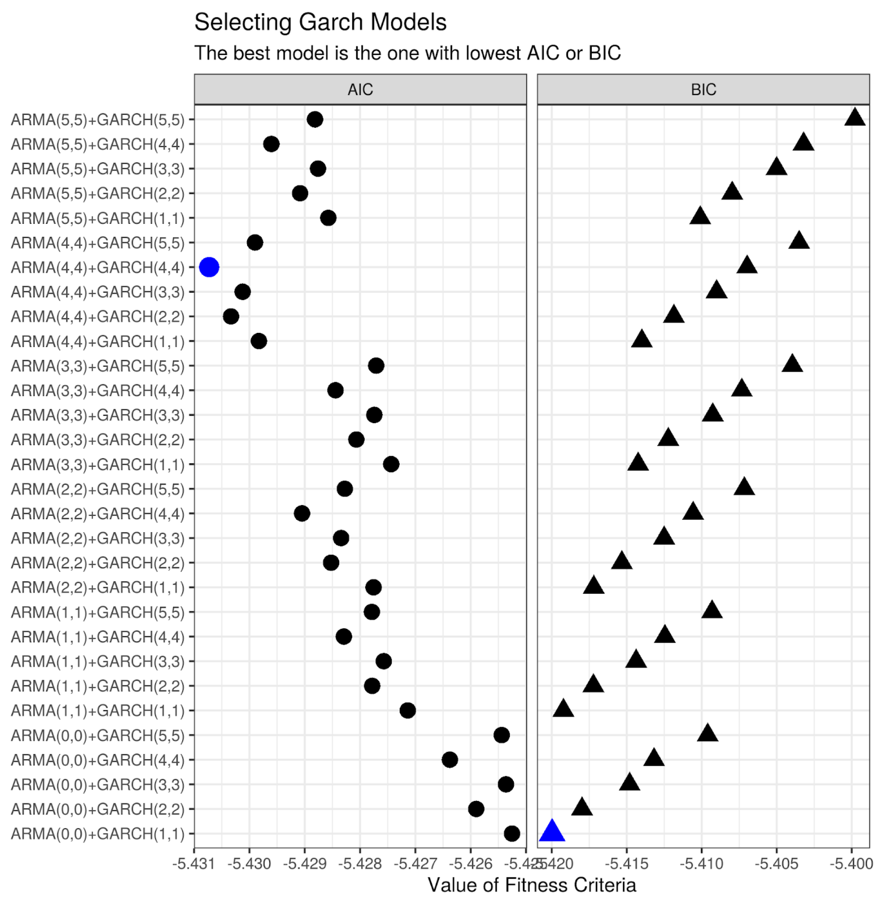
Finding the Best GARCH Specification

In the previous section we've set a GARCH specification to be estimated from the data -- a simple ARMA(0, 0)-GARCH(1, 1). However, for every set of data, we can find a GARCH model with the best fit by comparing measures of goodness of fit. Thus, instead of manually choosing a GARCH model and its lags, we let the data speak for itself. In practice, performing such a parameter search is a good research policy as it removes potential bias from the researcher. As a non-exhaustive list of studies comparing different GARCH models, we suggest Hansen (2005) and Katsiampa (2017).

The most commonly used indicators for selecting models is the AIC (Akaike Information Criteria) and BIC (Bayesian Information Criteria). The rule is: the better the model, the lower the value of AIC or BIC. One must, however, choose which criteria to use. The difference from one to the other is how they penalize the number of coefficients in the model. The AIC is more flexible than BIC and tends to pick models with a large number of parameters. On the other hand, the BIC tends to select parsimonious models, with a relatively small number of coefficients.

Next, Figure 03, we show the result of AIC and BIC for different GARCH models estimated from the data. The specification of the model is presented in the vertical axis. Do notice the use of a maximum lag of five. In the code -- script *05-Find_Best_Garch_Model.R* -- we estimated all possible combinations between lags of the conditional mean, an ARMA(ar, ma) model; and the conditional variance, a GARCH(p, q) model.

Figure 03 - Selecting GARCH Models.



From Figure 03 we see that the best models are ARMA(4, 4)-GARCH(4, 4) when using AIC and ARMA(0, 0)-GARCH(1, 1) when using BIC. Do notice a staircase pattern for the BIC panel, which is explained by the linear penalty on the extra parameters. As the number of lags/parameters increases, the total value of the penalty also increases. In the AIC panel, we see the opposite effect, models with higher number of parameters tend to present better fit and lower value of AIC.

In general, it is best advised to keep GARCH models simple and parsimonious. The benefits come from fast estimations and better volatility forecasts (Hansen & Lunde, 2005). With that in mind, we chose to use the BIC criteria and select the ARMA(0, 0)-GARCH(1, 1) model for the next section of the article.

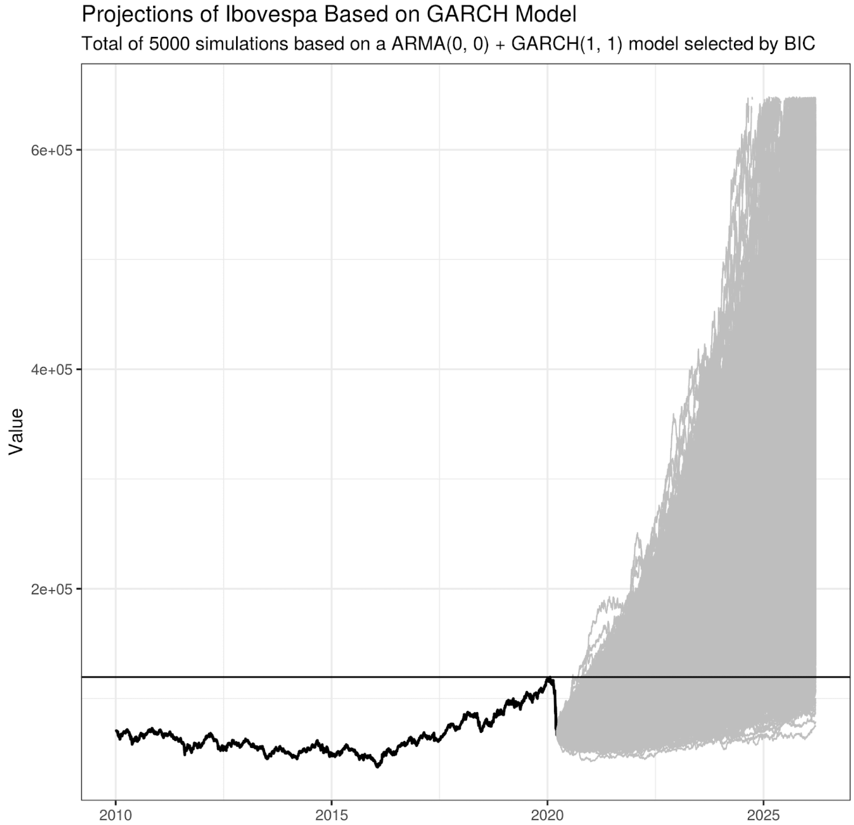
Simulating a Garch Model

Now that we have the best GARCH specification with parameters estimated from the data, it's time to simulate the future time series of returns and possible paths for the Ibovespa index in the upcoming years. This is accomplished by the code in *06-Simulate_Garch_Model.R*.

Simulation with GARCH models works by sequentially inputting the first value of returns in a pre-existing model specification and drawing samples from the distribution of residuals. By doing so, we can build a time series of returns of any length, and the simulated series will have the same properties as the underlying model. This is very convenient as we can project the future for any time frame.

In this application, we are interested in simulating many time series and future paths for Ibovespa. Once that is complete, we will use the simulation paths to better understand the likelihood of the index crossing its historical peak once again. Next, in Figure 04, we show the results for 5000 simulations.

Figure 04 - Simulating the Ibovespa Index

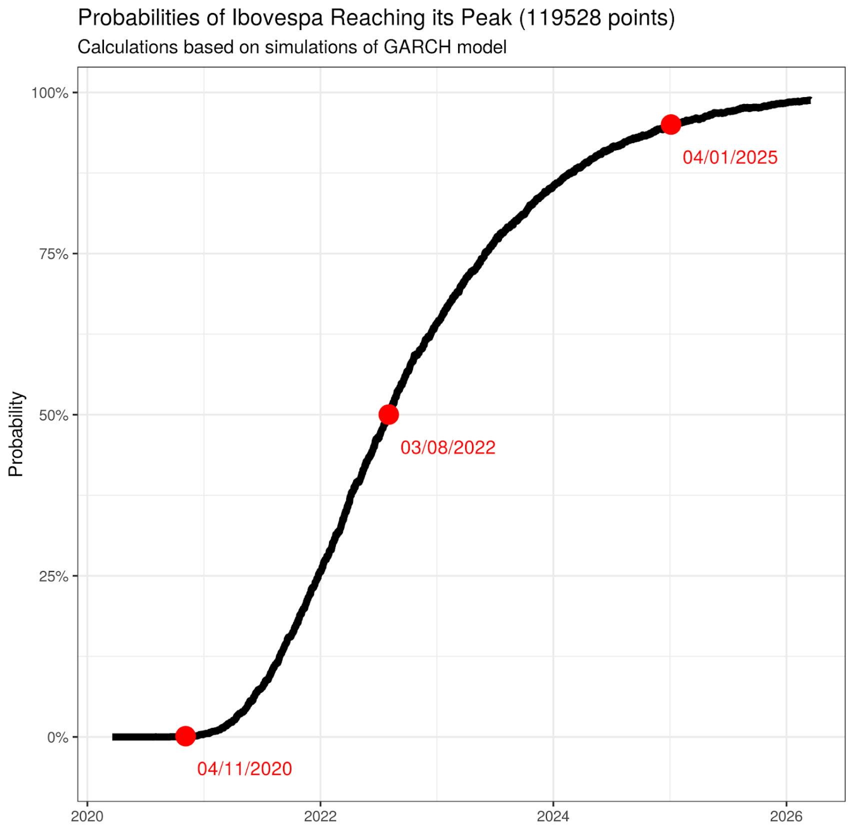


In Figure 04, we restrict the data just after 2010 in order to reduce the scale in the *y-axis*. The historical peak of Ibovespa was reached in the beginning of 2020, with the closing value of 119.528. The last day of the sample of prices is 20/03/2020, which is the first day of the simulation. All gray lines are different and independent simulations of the same ARMA(0, 0)-GARCH(1, 1) model estimated in the previous sections.

The first striking result of the simulation is the upward pattern. In the long run, prices of financial indexes tend to increase in value, and the estimated model was able to capture such an effect. In some unusual cases, the index reached the value of more than 400.000 points. While the simulated index value may drop in the short run, it's chances of passing the historical peak increases with time. From the plot, we see that the first simulated price (grey line) that crosses the peak is at the beginning of 2021.

For a better presentation of the results and accuracy, next, Figure 05, we show the actual probabilities of the future values of Ibovespa reaching the peak. These are calculated by checking, for every simulated point in time, how many simulated cases passed the peak value of 119.528.

Figure 05 - Probabilities of Ibovespa Reaching its Peak



First, as expected, the probabilities increase with time. The first date, 04/11/2020, represents the first probability larger than 0,1%. This means that, according to the model, the chances of Ibovespa passing its peak in the upcoming six months is almost nil. The most likely scenario happens in 03/08/2022, two years and five months from now, where the probabilities are close to 50%. After that, the chances of passing the peak are higher than not passing, reaching a 95% probability in 04/01/2025, approximately five years from now.

One of the messages from this empirical exercise is the destructive nature of a financial crisis. Prices of equity contracts dropped very fast, activating B3's *circuit break*[1] many times within the same week of March 2020. According to our GARCH model, the lost valuation due to the crisis will only be reached with certainty after five years of trading, with a larger chance after two years.

[1] Circuit break is an internal mechanism that halts trading for 30 minutes once the Ibovespa index reaches a 10% negative return within a day.

Conclusions

Volatility models and ARCH/GARCH specifications are one of the main innovations in financial modelling in the last decades, being used extensively in the industry and academic research. In this tutorial paper we presented a brief introduction to the motivation and theory behind the ARCH/GARCH family of models, with an example of empirical application for the Brazilian equity market. Based on a GARCH model and taking into consideration the recent COVID-19 crisis, we investigated how much time it would take for the index to reach its peak value once again. Our GARCH model forecasts that it will take about 2,5 years for Ibovespa to reach its peak again.

But, as a word of caution, we must be honest about the restrictions of our econometric study. The GARCH model is a limited representation of financial returns and no model can perfectly grasp the market participant's state of mind. Reproducing a coined phrase in statistics: "all models are wrong, but some are useful". Despite its shortcomings, the GARCH model has its merits by being able to provide a ballpark answer to our question, meaning that we can **expect** that the Ibovespa market will recover to the COVID-19 crisis in approximately two and half years.

All the code and data used in this study is available in the internet. We motivate the readers to run the scripts and reproduce all results in their own computer. Going further, we also suggest the reader to change the financial data in script *01-Get_Index_Data.R* and reproduce the results for other market index such as SP500 (USA), FTSE (UK) or any other asset available in *Yahoo Finance*.

However, we must point out that only a portion of the topic of volatility modelling was covered here. The idea in this tutorial paper was to introduce the reader to the topic of volatility modelling and provide material for the reproduction of an empirical example in R. We left out many other topics such as: bayesian estimation, multivariate GARCH, among others.

We hope this material will serve as starting point for many students that are learning volatility modelling and financial econometrics. By combining text with actual R code, readers will be able to understand how evey content was produced and, more importantly, use the code as reference for other volatility studies.

References

Bernardino, Wilton, et al. “A GARCH-VaR Investigation on the Brazilian Sectoral Stock Indices.” *Brazilian Review of Finance*, vol. 16, no. 4, 2019, p. 573., doi:10.12660/rbfin.v16n4.2018.74676.

Bollerslev, Tim. “Generalized Autoregressive Conditional Heteroskedasticity.” *Journal of Econometrics*, vol. 31, no. 3, 1986, pp. 307–327., doi:10.1016/0304-4076(86)90063-1.

Brockwell, Peter J., and Richard A. Davis. *Introduction to Time Series and Forecasting*. Springer International Publishing Switzerland, 2016.

Caldeira, João F., et al. “Portfolio Management Using Realized Covariances: Evidence from Brazil.” *EconomiA*, vol. 18, no. 3, 2017, pp. 328–343., doi:10.1016/j.econ.2017.04.002.

Costa, Hudson Chaves, et al. “Um Estudo Empírico Da Dinâmica Da Correlação Do Retorno Das Ações Do Brasil.” *Brazilian Review of Finance*, vol. 16, no. 4, 2019, p. 635., doi:10.12660/rbfin.v16n4.2018.72142.

Enders, Walter. *Applied Econometric Time Series*. Wiley, 2018.

Engle, Robert F. “Autoregressive Conditional Heteroscedasticity with Estimates of the Variance of United Kingdom Inflation.” *Econometrica*, vol. 50, no. 4, 1982, p. 987., doi:10.2307/1912773.

Engle, Robert. “GARCH 101: The Use of ARCH/GARCH Models in Applied Econometrics.” *Journal of Economic Perspectives*, vol. 15, no. 4, 2001, pp. 157–168., doi:10.1257/jep.15.4.157.

Fioruci, José A., et al. “Bayesian Multivariate GARCH Models with Dynamic Correlations and Asymmetric Error Distributions.” *Journal of Applied Statistics*, vol. 41, no. 2, 2013, pp. 320–331., doi:10.1080/02664763.2013.839635.

Francq, Christian, and Jean-Michel Zakoian. *GARCH Models: Structure, Statistical Inference and Financial Applications*. John Wiley & Sons, 2019.

Hansen, Peter R., and Asger Lunde. “A Forecast Comparison of Volatility Models: Does Anything Beat a GARCH(1,1)?” *Journal of Applied Econometrics*, vol. 20, no. 7, 2005, pp. 873–889., doi:10.1002/jae.800.

Jondeau, Eric, and Michael Rockinger. “The Copula-GARCH Model of Conditional Dependencies: An International Stock Market Application.” *Journal of International*

Money and Finance, vol. 25, no. 5, 2006, pp. 827–853., doi:10.1016/j.jimonfin.2006.04.007.

Katsiampa, Paraskevi. “Volatility Estimation for Bitcoin: A Comparison of GARCH Models.” *Economics Letters*, vol. 158, 2017, pp. 3–6., doi:10.1016/j.econlet.2017.06.023.

Katsiampa, Paraskevi. “Volatility Estimation for Bitcoin: A Comparison of GARCH Models.” *Economics Letters*, vol. 158, 2017, pp. 3–6., doi:10.1016/j.econlet.2017.06.023.

Lobão, Júlio, and João Fernandes. “Psychological Barriers in Single Stock Prices: Evidence from Three Emerging Markets.” *Review of Business Management*, vol. 20, no. 2, 2018, pp. 248–272., doi:10.7819/rbgn.v20i2.3049.

Mastella, Mauro, and Rodrigo Coster. “O Impacto Da Crise De 2008 Na Estrutura Temporal De Correlação Condicional Da BM&FBovespa.” *Review of Business Management*, 2014, pp. 110–123., doi:10.7819/rbgn.v16i50.1534.

Nelson, Daniel B. “Conditional Heteroskedasticity in Asset Returns: A New Approach.” *Econometrica*, vol. 59, no. 2, 1991, p. 347., doi:10.2307/2938260.

Zakoian, Jean-Michel. “Threshold Heteroskedastic Models.” *Journal of Economic Dynamics and Control*, vol. 18, no. 5, 1994, pp. 931–955., doi:10.1016/0165-1889(94)90039-6.