

MODELING CONDITIONAL VOLATILITY IN R

Omid Sabbaghi¹

Abstract

Many economic and financial time-series exhibit time-varying volatility. In particular, volatility is an important input for pricing models and portfolio management decisions. This article demonstrates how to estimate volatility using the GARCH (1,1) model through the R analytics software. In addition, this study demonstrates how to employ GARCH-based estimates for the purposes of forecasting volatility.

Key Words: financial econometrics, GARCH, volatility, forecasting

JEL Classifications: C58, G12, C22

Introduction

Time-varying volatility is an important phenomenon in the economics and finance literature. In financial applications, the variance of the return on an asset represents the risk level of those returns, and it is documented that the volatility, or amplitude, of financial returns is time-varying as there are periods of market tranquility as well as periods of market turbulence. In addition to individual stocks and stock market indices, other examples of observed time-varying volatility in our global markets include changes in exchange rates and absolute deviations of the inflation rate from its mean. That is, financial markets data exhibit volatility clustering in that there are periods of high volatility and periods of low volatility.

Precise estimates of volatility are an input for numerous real-world applications in economics and finance. For example, volatility is an important input for forecasting confidence intervals since time-varying volatility translates into time-varying widths for confidence intervals. In addition, precise estimates of volatility enable precise Value-at-Risk (VaR) estimates for portfolios, thereby allowing banks to better estimate the risk of loss for investments. Moreover, volatility is an input variable for the Black-Scholes formula, and thus precise estimates of volatility are important for valuing options. Importantly, the estimation of returns and conditional volatilities is important in identifying investment opportunities that increase returns since expected returns depend on volatility (French et al. 1987). Thus, for these reasons, among others, it is important to obtain robust estimates of volatility through GARCH models.

This paper provides several important lessons in volatility for the undergraduate student studying economics and finance. First, this study provides an introduction to the estimation of time-varying volatility, through the GARCH model of Bollerslev (1986), in the R analytics environment. R is an open source statistical programming language in which it is estimated that nearly 250,000 people work with R on a regular basis, and is considered the second language for individuals emerging from graduate school (Vance 2009). As Li (2019) points out, R is a valuable aspect of an introductory research methods course when taught appropriately. This

¹ Associate Professor of Finance, College of Business Administration, University of Detroit Mercy, 4001 W. McNichols Road, Commerce and Finance 245, Detroit, MI 48221

study demonstrates how to estimate time-varying volatility for a well-known barometer of the U.S. stock market, namely the S&P500 market index. In particular, this article shows how GARCH effects are significant and persistent throughout time.

After estimating the GARCH model in R, this study demonstrates how to extract the time-series of the conditional volatilities, also referred to as the conditional standard deviations, using R analytics. In addition, this paper demonstrates how to construct plots that provide visual confirmation of time-varying volatility. Spikes in the plot convey time periods in which volatility is high.

Finally, this study demonstrates how to produce time-varying confidence intervals for the S&P500 daily returns using the conditional volatilities obtained by the GARCH model. The R environment provides rich graphic plots in the construction of such time-varying intervals. Goodness-of-fit statistics are also discussed when interpreting the statistical output. That is, this study demonstrates how to gauge the adequacy of a fitted GARCH model in R. In addition, this paper concludes with the interpretation of the information criteria as a means of performing model comparison. Figures and tables are in the appendix.

R

The R environment is open-source software, and its capabilities are extended via packages. The base R software, as well as additional packages, may be downloaded freely from the Comprehensive R Archive Network (CRAN) via the internet at <https://cran.r-project.org/>. The fGarch package deals specifically with volatility modeling and is the object of interest for the present study. Upon starting R, entering the following command in the R console enables the fGarch package, and any necessary related packages, to be installed:

```
if(!require(fGarch)) {
  install.packages("fGarch")
}
```

Once the fGarch package is installed, it must be loaded upon initialization of an R working session to ensure that the GARCH model can be estimated in R, either via the Packages submenu, or through entering the following command in the R console:

```
library(fGarch)
```

Sample Data & Empirical Methodology

This study applies the GARCH framework in estimating the conditional volatility for the S&P500 market index. Daily closing levels for the S&P500 index, adjusted for dividends and splits, are extracted from the Yahoo! Finance database. The data ranges from January 1950 through December 2019. The closing price level for the S&P500 at time t is denoted p_t . The daily rate of return for the S&P500 on day t , denoted r_t , is defined as the continuously compounded return in Eqn. (1):

$$r_t = \log(p_t) - \log(p_{t-1}) \quad (1)$$

The GARCH model, first introduced by Bollerslev (1986), describes the conditional return volatility of a financial asset over time. GARCH generalizes the ARCH model which was proposed by Engle (1982) to model the uncertainty of inflation in the United Kingdom.

The present study adopts a GARCH(1,1) model in understanding the time-varying volatility for the S&P 500 index. Research has shown that the GARCH(1,1) model is the most important case for practical applications. For example, Tsay (2010) and Hansen and Lunde (2005) argue that higher-order GARCH models fail to outperform the GARCH(1, 1) model. Engle (2001) states that the GARCH(1,1) model is the most robust of the family of volatility models. Formally the GARCH (1,1) model is given by Eqns. (2) and (3):

$$r_t = \sqrt{h_t} z_t \quad z_t \text{ iid } N(0,1) \quad (2)$$

$$h_t = \omega + \alpha r_{t-1}^2 + \beta h_{t-1} \quad (3)$$

In the GARCH model, the conditional return variance h_t is itself a serially correlated time series process. In addition, tomorrow's return variance h_t is predicted to be a weighted average of today's variance forecast h_{t-1} and the new information in this period as captured by today's squared return r_{t-1}^2 . Conditional return volatility, or also known as conditional standard deviation, is denoted $\sqrt{h_t}$. The errors, z_t , is a sequence of independent and identically distributed (iid) random variables with zero mean and unit variance, and are assumed to be normal in Eqn. (2). The realized errors, also referred to as the standardized returns, are important in diagnostic checking since the standardized returns should no longer show significant volatility clustering.

The economic magnitudes of the α and β coefficient estimates are important. First, in the context of the GARCH(1,1) model, the long-run average variance is given by $\frac{\omega}{1-\alpha-\beta}$, and thus $\alpha + \beta < 1$ as well as $\alpha > 0$, $\beta > 0$, and $\omega > 0$ (Engle 2001). It is required that $\alpha + \beta < 1$ so as to ensure that the volatility is mean-reverting.

The conditional mean represents the average return, and in the pure GARCH(1,1) model of Eqns. (2) and (3), conditional mean return dynamics are negligible. That is, the average daily returns are zero. In the event that daily average returns are statistically distinguishable from zero, Tsay (2010) suggests demeaning the daily returns prior to the GARCH estimation.

As a means of gauging the importance of conditional volatility for the S&P 500 index, this study estimates the GARCH(1,1) model over seven non-overlapping time periods. Since trading volume data for the S&P 500 begins in 1950 based on our data source, this study examines the following non-overlapping time periods: January 1950 – December 1959; January 1960 – December 1969; January 1970 – December 1979; January 1980 – December 1989; January 1990 – December 1999; January 2000 – December 2009; and January 2010 – December 2019. The aforementioned time periods are henceforth referred to as Periods 1, 2, 3, 4, 5, 6, and 7, respectively.

R and the GARCH(1,1) Model

The first step in our analysis consists of inputting the daily returns data. To illustrate the estimation of the GARCH(1,1) model in R, this study considers the S&P 500 returns data for the most recent decade, spanning from January 2010 through December 2019, namely Period 7. In R, we define the Period 7 data with the following commands in the R console:

```
period7data <- read.delim("C:\\data\\period7data.txt",header=T)
period7return <- period7data[,1]
```

In the first command above, note that we are reading in the tab-delimited text data file from its corresponding location via the `read.delim` command. The `header=T` argument is allowing R to read the first row of the data matrix as column names. By applying the `read.delim` command, we are defining the data matrix `period7data` which contains the returns time-series data. The second command above is defining `period7return` as the first column of the `period7data` matrix, namely the returns. Figure 1, Panel A presents a time-series plot of the daily squared returns over the January 2010 – December 2019 time period.² Such a plot is constructed in the R environment via the following command in the R console:

```
ts.plot(period7return^2, main="Panel A: S&P500 Daily Squared
Returns: 2010-2019", ylab="Daily Squared Return")
```

In the R command above, the `ts.plot` command is being applied to the daily squared returns for Period 7. The `main=` subargument is allowing for us to provide a title for the plot, while the `ylab=` argument allows for the labeling of the y-axis. Figure 1 Panel A presents evidence of clustering in return volatility over time. That is, large changes in the daily squared returns are followed by large changes, and small changes in the daily squared returns are followed by small changes. Panel B plots the sample autocorrelation function for the S&P500 daily squared returns over the identical time period, and visually confirms the presence of volatility clustering. The autocorrelation function for the Period 7 squared returns data is constructed in the R environment via the following `acf` command in the R console:

```
acf(period7return^2, lag.max=40, main="Panel B: ACF of S&P500
Daily Squared Returns: 2010-2019")
```

In the above command, the number of lags utilized for the autocorrelation plot is provided by the `lag.max=` subargument. Similarly, the `main=` subargument allows for the title of the plot. Statistically, the return autocorrelations through the 40th lag are statistically distinguishable from zero since the autocorrelations lie outside the dotted lines which represent the significance bounds. That is, the statistically significant autocorrelations suggest that the S&P500 return volatility is serially correlated throughout time.³

The `fGarch` package in R estimates a GARCH(1,1) model for the S&P 500 daily returns. For classroom illustrative purposes, we assume that the conditional mean dynamics for returns

are negligible and that the standardized returns $\frac{r_t}{\sqrt{h_t}}$ are normally distributed, thereby justifying

the estimation of a pure GARCH(1,1) model with normal innovations z_t . The conditional normality assumption may sometimes be violated. However, Bollerslev and Woolridge (1992) show that GARCH parameters are consistently estimated by Gaussian maximum likelihood even when the normality assumption is incorrect.

² Daily squared returns for Periods 1 through 6 are similar in appearance.

³ Autocorrelation functions for the S&P500 daily squared returns in Periods 1 through 6 are similar in appearance.

We estimate the GARCH(1,1) model for the January 2010 through December 2019 time period. Such a model is defined and estimated in R using the following `garchFit` command in the R console:

```
garch11.model <- garchFit(~garch(1,1), data=period7return,
cond.dist = c("norm"), include.mean=FALSE)
```

In the above syntax, the R environment is defining `garch11.model` as an object in its working memory. However, the `garchFit` command requires several inputs. First and foremost, the order of the GARCH model is specified: the `~garch(1,1)` substring conveys the estimation of a GARCH(1,1) model. Higher-order GARCH models may be estimated by inputting different values for this substring. Second, the `garchFit` command requires the vector of daily returns via the `data=` substring. In the above example command, `period7return` is referring to the vector of S&P500 daily returns that reside in Period 7, namely the January 2010 – December 2019 time period. The `cond.dist` substring is referring to the conditional distribution of the standardized returns. In the present study, we assume conditional normality for the standardized returns. In the event that we wish to impose a Student- t conditional distribution, we would alter the substring to `cond.dist = c("std")`. This is especially important for the case of high-frequency intra-day data, in that the conditional density is leptokurtic. In Bollerslev's (1987) conditionally Student's t -GARCH model, the conditional density is assumed to be Student's t , with the degrees of freedom treated as another parameter to be estimated. Finally, the `include.mean` binary substring is set to `FALSE` so that the R environment ignores the estimation of conditional mean dynamics, in particular ARIMA effects.

Following the estimation of the GARCH(1,1) model, a concise summary output of the coefficient estimates and their corresponding significance levels is made possible through the `summary` command. The following command line illustrates the application of the `summary` command to the `garch11.model` object defined within R:

```
summary(garch11.model)
```

Following the application of the `summary` command in the R console, the R environment will produce summary output comprising of coefficient estimates, statistical significance levels, model information criteria, and goodness-of-fit statistics. For the Period 7 data, the `summary` command outputs the following information:

```
Error Analysis:
      Estimate  Std. Error  t value  Pr(>|t|)
omega  3.661e-06   5.491e-07    6.667  2.61e-11 ***
alpha1 1.585e-01   1.704e-02    9.303  < 2e-16 ***
beta1  8.005e-01   1.824e-02   43.886  < 2e-16 ***
---
Signif. codes:  0 '***' 0.001 '**' 0.01 '*' 0.05 '.' 0.1 ' ' 1

Log Likelihood:
8560.509      normalized: 3.403781
```

Standardised Residuals Tests:

			Statistic	p-Value
Jarque-Bera Test	R	Chi ²	616.843	0
Shapiro-Wilk Test	R	W	0.9720951	0
Ljung-Box Test	R	Q(10)	11.2828	0.3359141
Ljung-Box Test	R	Q(15)	18.31018	0.24671
Ljung-Box Test	R	Q(20)	26.78719	0.1413581
Ljung-Box Test	R ²	Q(10)	9.867153	0.4522248
Ljung-Box Test	R ²	Q(15)	16.51333	0.3487793
Ljung-Box Test	R ²	Q(20)	17.58753	0.6145591
LM Arch Test	R	TR ²	11.76193	0.4649839

Information Criterion Statistics:

AIC	BIC	SIC	HQIC
-6.805176	-6.798222	-6.805179	-6.802652

In the R output above, the GARCH(1,1) results are given. The three coefficients in the variance equation are listed as the intercept ω ; the first lag of the squared return α_1 ; and the first lag of the conditional variance β_1 . In particular, the coefficient estimates are listed underneath the Estimate column, and we observe that the coefficient estimates are $\omega = 3.661e-06$, $\alpha = 0.1585$, and $\beta = 0.8005$. Thus, the `garchFit` command has estimated the following GARCH(1,1) model:

$$h_t = 3.66e - 06 + 0.1585r_{t-1}^2 + 0.8005h_{t-1} \quad (4)$$

In Eqn. (4), we discern that the process is mean-reverting since the sum of the coefficients is less than 1. As Engle and Patton (2001) point out, mean reversion in volatility conveys that there is a normal level of volatility to which the volatility process will return.

Having estimated the intercept ω , α , and β , we can employ the coefficient estimates for our GARCH(1,1) model to construct volatility forecasts, since the forecast for tomorrow's volatility $\sqrt{h_{t+1}}$, can be calculated using Eqn. (5):

$$h_{t+1} = \omega + \alpha r_t^2 + \beta h_t \quad (5)$$

To illustrate, consider a trader who thinks that the long-run average daily return standard deviation $\sqrt{\frac{\omega}{1-\alpha-\beta}}$ for the S&P 500 is 1 percent. In addition, suppose today's observed unexpected return (r_t) is 4 percent, and that yesterday, the trader made a forecast of 3 percent for today's volatility ($\sqrt{h_t}$). Thus, based on the estimated GARCH(1,1) model in Eqn. (4), our forecast for tomorrow's volatility is:

$$\sqrt{h_{t+1}} = \sqrt{\omega + \alpha r_t^2 + \beta h_t} = \sqrt{[(1 - \alpha - \beta) * 1] + (\alpha * 4^2) + (\beta * 3^2)} = \sqrt{[(1 - 0.1585 - 0.8005) * 1] + (0.1585 * 4^2) + (0.8005 * 3^2)}$$

$$\sqrt{h_{t+1}} = 3.1275 \text{ percent}$$

Hence, R is useful for estimating the GARCH model and predicting future volatility based on the above output. The R output from above also presents standard errors under the `Std. Error` column, t-values under the `t value` column, and corresponding p-values under the `Pr(>|t|)` column. The t-values are the ratio of the estimated coefficients and standard errors, and convey the degree of statistical significance for the coefficient estimates. Based on the summary output, we find that all of the coefficient estimates are highly significant at the 1 percent level.

The adequacy of a fitted GARCH model warrants attention. For the case of the estimated GARCH(1,1) model above, the output also presents examination of the standardized residuals for serial correlation. If the evidence suggests that serial correlation, or autocorrelation, exists for the standardized residuals, then the GARCH(1,1) model has not captured all of the return volatility dynamics. From the GARCH summary output above, we find that the statistics for the Ljung-Box test for the squared standardized residuals are presented in the rows beginning with `Ljung-Box Test`. Evidently, the Ljung-Box test of interest has a p-value near 0.61 when testing correlation up through the 20th lag. Thus, the estimated GARCH model captures volatility in the returns and there are no more GARCH effects in the residuals since we fail to reject the null hypothesis of no serial correlation for the squared standardized residuals.

The summary output provides several goodness-of-fit metrics. For example, the summary output presents a log-likelihood value of 8560.5 for our estimated GARCH(1,1) model. The log-likelihood is a goodness-of-fit metric since GARCH models that achieve higher log-likelihood values are preferred in practice. In addition, the R summary output also presents the normalized log-likelihood value of 3.40, which is the log-likelihood divided by n . The R summary output also provides the Akaike information criterion (AIC), the Bayesian information criterion (BIC), the Schwarz information criterion (SIC), and the Hannan-Quinn information criterion value (HQIC). The aforementioned information criteria values are helpful for model selection purposes in assessing higher-order GARCH models, such as GARCH(2,2) models. In practice, smaller values of the information criteria reflect preferred models.

As Engle and Patton (2001) point out, persistence in volatility reflects the extent to which today's return exhibits a large impact on future volatility forecasts, and furthermore, present a measure of persistence known as the 'half-life' of volatility: the 'half-life' of volatility is defined as the time taken for the volatility to move halfway back towards its unconditional mean following a deviation from it. Based on data from 2010-2019, this study finds that the half-life of volatility for the S&P500 index is 16.56 days.

A time series plot of the S&P500's in-sample sequence of conditional volatilities may be extracted within the R environment. In particular, the `plot` command provides a rich array of graphical options for the Rmetrics user. For the Period 7 data, the associated GARCH(1,1) conditional volatilities are plotted using the following R command:

```
plot(garch11.model, which=2)
```

In the above example command, the `which` substring is directing the R environment to the relevant plot of interest. For a list of the possible different plots that can be generated by the `fGarch` package, one would enter the following command in the R console:

```
plot(garch11.model)
```

The upper panel in Figure 2 plots the resulting time-series of the S&P500 conditional volatility for the January 2010 through December 2019 time period when adopting a GARCH(1,1) model with normal innovations. The conditional volatilities in Figure 2 are important when constructing dynamic confidence intervals for the S&P500 daily returns. For a given day, a 95% confidence interval for the S&P500 daily return on day t is given by Eq. (5):

$$r_t \pm 2\sqrt{h_t} \quad (5)$$

Since conditional mean dynamics are negligible in the pure GARCH(1,1) model of Eqs. (2) and (3), the daily forecasted return for day t , r_t , is zero. Thus, the 95% confidence interval for the S&P500 daily return on day t is subsequently given by Eq. (6):

$$0 \pm 2\sqrt{h_t} \quad (6)$$

In Eq. (6), it is evident that the conditional volatilities determine the confidence interval over time. In the R environment, the 95% confidence interval for the S&P500 daily returns is plotted via the appropriate `plot` command. Specifically, the `which` substring must be set to 3:

```
plot(garch11.model, which=3)
```

The lower panel in Figure 2 plots 95% confidence intervals for the S&P500 daily returns, superimposed on the daily realized returns for the January 2010 through December 2019 time period. This figure highlights the importance of GARCH volatilities in predicting future returns since the confidence interval captures, to a great extent, the dynamic behavior of the daily S&P500 returns. That is, nearly 95 percent of the realized S&P500 daily returns lie within the upper and lower limits of the confidence bands.

Table 1 presents the results of the estimated GARCH(1,1) model across the different time periods. In Table 1, we observe that all of the GARCH coefficient estimates are highly significant at the 1 percent level across time, further underlining the robustness of the GARCH(1,1) model. In addition, Table 1 reports the long-run average daily return standard deviation as well as the half-life of the S&P500 index over different time periods. In Table 1, we find that the 2000 – 2009 time period exhibits the highest long-run average daily return standard deviation while the 1990 – 1999 time period exhibits the highest half-life of volatility. Specifically, the half-life of volatility for the S&P500 index was 125.68 days for the 1990 – 1999 time period, underlining high persistence during this time period. Thus, in the most recent decade spanning 2010 - 2019, volatility was less persistent relative to earlier decades. In addition, when examining the Ljung-Box test statistics, we find that the GARCH(1,1) model is adequate in capturing the volatility dynamics for all of the remaining time periods.

Concluding Remarks

This study illustrates how to estimate GARCH models for the purpose of modeling volatility using the R analytics platform. In particular, the present study demonstrates the necessary commands and use of the `fGarch` package for R in estimating and forecasting conditional volatility. The R environment is open source software that is flexible, user-friendly, and allows for the estimation of higher order GARCH models with different distributional assumptions.

In an application to the S&P500 returns, this study estimates a GARCH(1,1) model and provides interpretations of the summary output provided by R. In addition, this study demonstrates how to forecast future volatility based on the estimated parameters and, moreover, demonstrates how to extract the time-series of conditional volatility through R's rich graphical capabilities. The present study also underlines the importance of the GARCH(1,1) model in constructing dynamic confidence intervals, and how to assess the adequacy of a fitted GARCH model. Finally, this article presents results on the half-life of volatility for the S&P500 index as a measure of volatility persistence.

Several directions lie ahead for future research in the economics and financial education literature. Firstly, the GARCH model estimates provided by R are useful for conducting dynamic portfolio allocation exercises in university trading rooms and classrooms. For example, a mean-variance investment strategy could be specified for an investor exhibiting quadratic utility preferences. That is, conditional utility would depend on the time-varying return variances. For specific risk aversion levels, the student can then solve for optimal portfolio weights that vary over time on the basis of optimizing the investor's conditional utility. The actual utility obtained by the dynamic portfolio strategy could then be compared to the actual utility of an investor who selects fixed portfolio weights that do not vary over time. Such an exercise allows the student to conduct market timing strategies on the basis of the observed time-varying volatility.

Another avenue for future research surrounds forecasting in the context of the GARCH model. For example, the student may be interested in a 2-day ahead forecast given today's realization of conditional volatility. Thus, a detailed examination of conducting multi-day ahead volatility forecasts is essential.

Third, it is of interest to relate GARCH volatility forecasts to how banks assess future portfolio losses through volatility-based Value-at-Risk (VaR) measures. In particular, out-of-sample forecasts for the conditional volatility, in conjunction with the probability distribution of returns, provide a higher degree of accuracy when estimating potential portfolio losses. In addition, it would be interesting to compare VaR forecasts across different distributional assumptions of the standardized returns.

Finally, the asymmetric class of GARCH models, proposed by Engle and Ng (1993) and Nelson (1991), suggest that volatility tends to be higher on days in which negative returns are realized. That is, asymmetric volatility models allow for good news and bad news to exhibit different predictability for realized volatility. Given the downside risks observed in 2008 through 2010, and in 2018, it would be interesting to explore the estimation of such models in the R environment. These topics are left for future research.

References

- Bollerslev, T. 1986. "Generalized Autoregressive Conditional Heteroskedasticity." *Journal of Econometrics*, 31(3): 307-327.
- Bollerslev, T. 1987. "A Conditional Heteroskedastic Time Series Model for Speculative Prices and Rates of Return." *Review of Economics and Statistics*, 69(3): 542-547.
- Bollerslev, T., and J.M. Woolridge. 1992. "Quasi-Maximum Likelihood Estimation and Inference in Dynamics Models with Time-Varying Covariances." *Econometric Reviews*, 11(2): 143-179.
- Engle, R.F. 1982. "Autoregressive Conditional Heteroskedasticity with Estimates of the Variance of U.K. Inflation." *Econometrica*, 50(4): 987-1008.
- Engle, R.F. 2001. "GARCH 101: The Use of ARCH/GARCH Models in Applied Econometrics."

- Journal of Economic Perspectives*, 15(4): 157 – 168.
- Engle, R.F., and A.J. Patton. 2001. “What good is a volatility model?” *Quantitative Finance*, 1: 237-245.
- Engle, R.F., and V.K. Ng. 1993. “Measuring and Testing the Impact of News on Volatility.” *Journal of Finance*, 48(5): 1749-1778.
- French, K., Schwert, G., and R. Stambaugh. 1987. “Expected stock returns and volatility.” *Journal of Financial Economics*, 19(1): 3-29.
- Hansen, P. R., and A. Lunde. 2005. “A forecast comparison of volatility models: Does anything beat a GARCH(1,1)?” *Journal of Applied Econometrics*, 20(7): 873-889.
- Li, R. 2019. “Teaching Undergraduates R in an Introductory Research Methods Course: A Step-by-Step Approach.” *Journal of Political Science Education*.
- Nelson, D.B. 1991. “Conditional Heteroskedasticity in Asset Returns: A New Approach.” *Econometrica*, 59(2): 347-370.
- Tsay, R.S. 2010. *Analysis of Financial Time Series*. 3rd edition. New York: John Wiley & Sons.
- Vance, A. 2009. “Data Analysts Captivated by R’s Power.” *New York Times*, January 6.

APPENDIX

Figure 1

Panel A plots the time-series of daily squared returns for the S&P500 index over the January 2010 – December 2019 time period. Panel B plots the autocorrelation function (ACF) for the S&P500 daily squared returns over the identical time period.

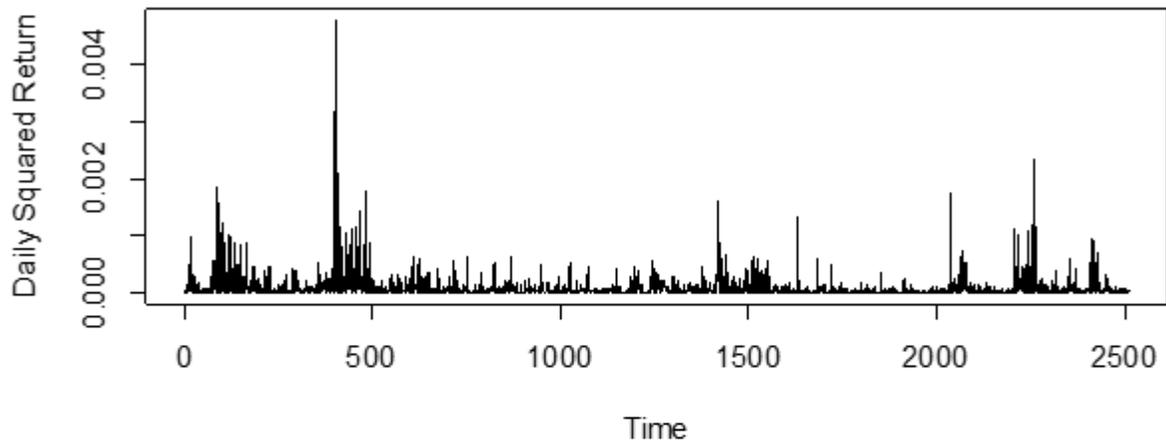
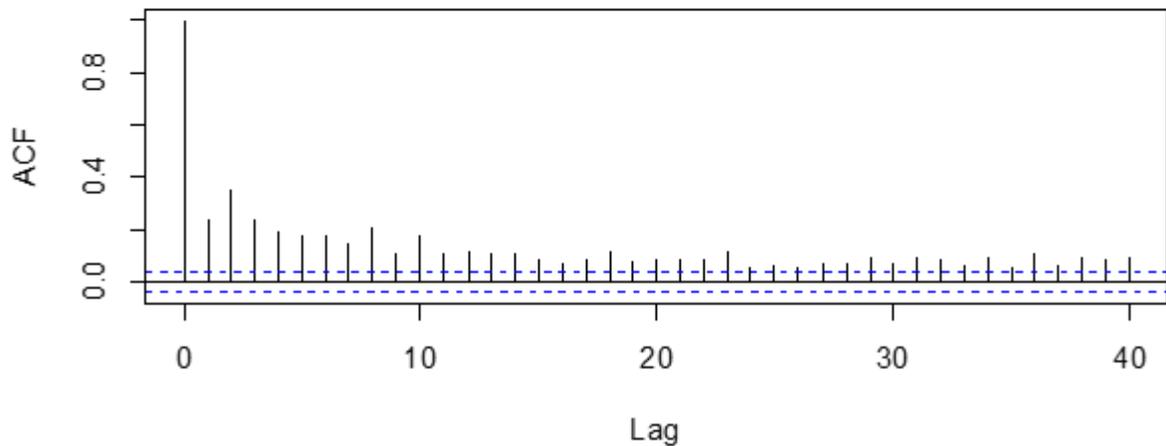
Panel A: S&P500 Daily Squared Returns: 2010-2019**Panel B: ACF of S&P500 Daily Squared Returns: 2010-2019**

Figure 2

The upper panel plots the conditional volatilities for the S&P500 index over the January 2010 through December 2019 time period. The lower panel plots the realized S&P500 daily returns and the associated 95% confidence bands over the identical time period. The conditional volatilities are estimated using a GARCH(1,1) model with Normal innovations.

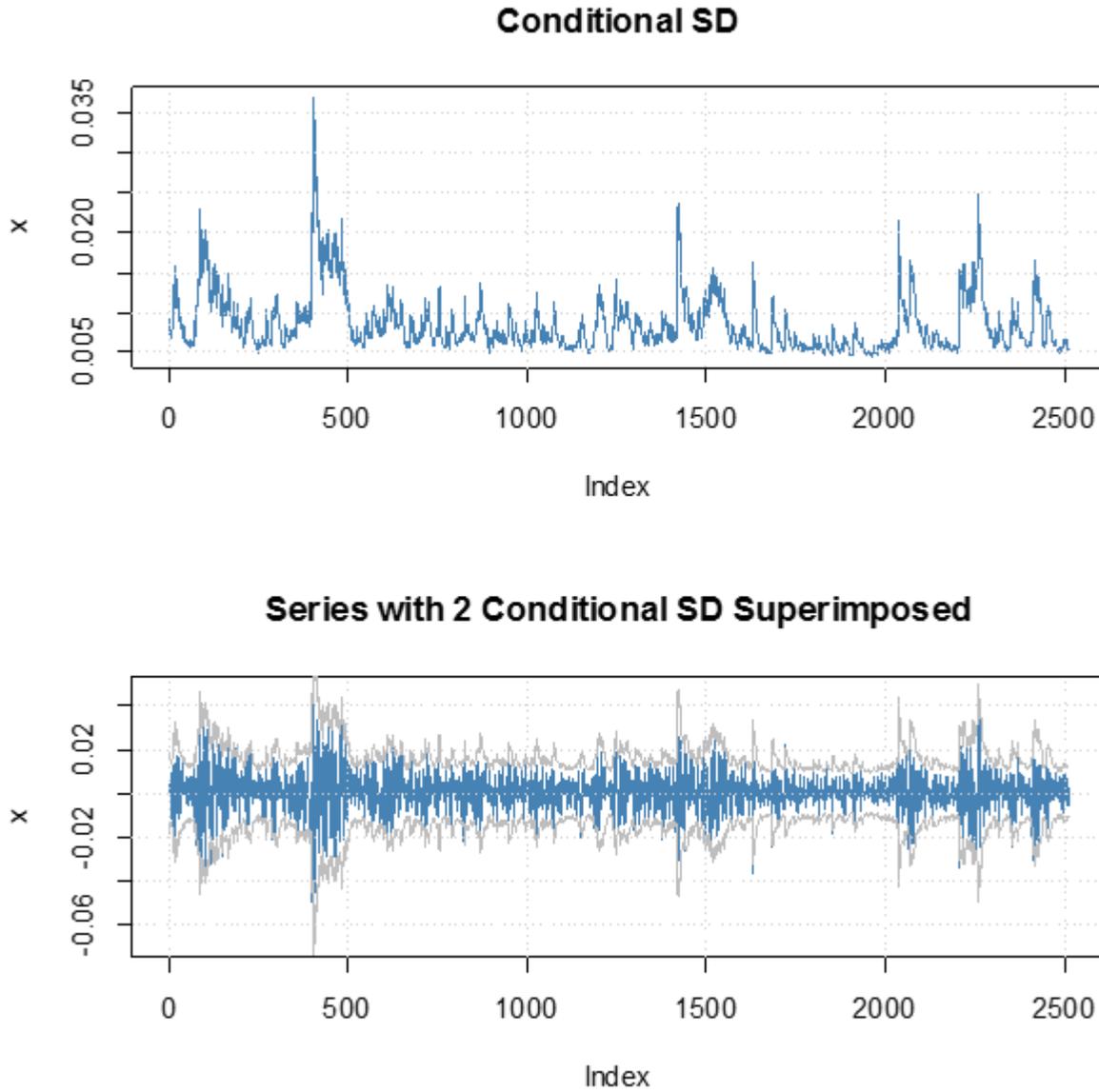


Table 1

This table reports the results of fitting a GARCH(1,1) model with Normal innovations for the S&P500 daily returns using R's `fGarch` package. Quasi-maximum likelihood parameter estimates, the associated t-ratios, log likelihood, Akaike Information Criterion (AIC), and Bayesian Information Criterion (BIC) are reported. This table reports the long-run average daily return standard deviation (SD), persistence, and half-life of volatility for the S&P500 index based on the estimated GARCH(1,1) models across the different time periods. The Ljung-Box test for autocorrelation through the 20th lag for the squared standardized residuals, denoted by Q(20), is presented. The sample period is January 1950 through December 2019. Seven non-overlapping time-periods are examined. Period 1 extends from January 1950 through December 1959. Period 2 extends from January 1960 through December 1969. Period 3 extends from January 1970 through December 1979. Period 4 extends from January 1980 through December 1989. Period 5 extends from January 1990 through December 1999. Period 6 extends from January 2000 through December 2009. Period 7 extends from January 2010 through December 2019. Adjusted closing price data is obtained from

Yahoo! Finance. The GARCH(1,1) model is of the form: $r_t = \sqrt{h_t} z_t$, $h_t = \omega + \alpha r_{t-1}^2 + \beta h_{t-1}$ $z_t \sim N(0,1)$ ^{iid}

Period	1	2	3	4	5	6	7
Coefficient							
ω	1.668e-06 (2.42)**	1.596e-06 (4.06)***	7.515e-07 (3.23)***	4.886e-06 (4.20)***	4.855e-07 (2.75)***	1.065e-06 (3.83)***	3.661e-06 (6.67)***
α	0.0594 (3.91)***	0.1649 (7.90)***	0.0631 (3.23)***	0.0927 (8.20)***	0.0481 (5.85)***	0.0737 (8.18)***	0.1585 (9.30)***
β	0.9103 (35.05)***	0.8012 (30.96)***	0.9260 (94.18)***	0.8632 (44.41)***	0.9464 (103.20)***	0.9196 (99.42)***	0.8005 (43.89)***
Long-run average daily return SD	0.12	0.11	0.13	0.17	0.15	0.20	0.15
Persistence	0.91	0.80	0.93	0.86	0.95	0.92	0.80
Half-life	22.53	20.10	63.24	15.37	125.68	103.11	16.56
Log-Likelihood	8914.1	9382.1	8705.9	8170.4	8597.6	7789.4	8560.5
AIC	-7.10	-7.54	-6.89	-6.46	-6.80	-6.19	-6.81
BIC	-7.09	-7.53	-6.89	-6.46	-6.80	-6.19	-6.80
Q(20)	5.33	21.38	14.99	7.30	10.87	19.18	17.59

***, ** indicate statistical significance at the 0.01 and 0.05 level, respectively