ÁP DỤNG MÔ HÌNH GARCH TRÊN THỊ TRƯỜNG CHỨNG KHOÁN VIỆT NAM

LÊ VĂN TUẤN[1] và PHÙNG DUY QUANG[2]

TÓM TẮT. Bài viết sử dụng mô hình GARCH để mô hình hóa và thực hiện dự báo cho chỉ số VNIndex, chỉ số đại diện cho TTCK Việt Nam. Kết quả thống kê cho thấy mô hình phù hợp nhất để mô hình hóa sự biến động của VNIndex là GARCH(1, 1). Các câu lệnh R được cung cấp đầy đủ tới bạn đọc.

1. Giới thiệu

Mô hình hóa sự bất định là vấn đề căn bản của tài chính định lượng, được ứng dụng trong cả ba mảng chính: phân bổ danh mục đầu tư, quản trị rủi ro và định giá các hợp đồng tài chính. Sự mô hình hóa này đem lại sự hiểu biết về các tính chất thống kê của sự thay đổi giá và cách để dự báo tốt hơn. Các chuỗi dữ liệu theo thời gian được cho là phụ thuộc vào giá trị quá khứ của chính nó (autoregressive), điều kiện của các thông tin trong quá khứ (conditional) và tồn tại phương sai thay đổi (heteroskedastic). Các nghiên cứu cho rằng những biến động của thị trường chứng khoán thay đổi theo thời gian và biến động theo cụm, trong đó một chuỗi thời gian với một số thời kỳ biến động thấp và một số thời kỳ biến động cao được cho là tồn tại biến động theo cụm (volatility clustering)

Mô hình ARCH[3] (Autoregressive Conditional Heterokedasticity) và họ các mô hình tổng quát của nó, GARCH (Generalized Autoregressive Conditional Heterokedasticity), đã mở ra một kỷ nguyên mới trong lĩnh vực mô hình hóa tài chính (trường hợp riêng là cho thị trường chứng khoán). Đóng góp chính của các mô hình này là cho phép mô hình hóa sự bất định là một quá trình động, thay vì giả định sự biến động trong tương lai là hằng số, nó là quá trình biến đổi theo thời gian.

Ở Việt Nam, các nghiên cứu về áp dụng mô hình ARCH/GARCH cho thị trường chứng khoán xuất hiện từ khá sớm. Hoàng (2004) đã tìm kiếm bằng chứng khoa học về hiệu ứng GARCH trên dãy thống kê lợi suất của chỉ số giá thị trường và 10 cổ phiếu đang niêm yết. Kết quả kiểm định đáng khích lệ. Tiên (2017) đã thực hiện các phân tích bằng mô hình GARCH cân xứng và bất cân xứng. Theo tiêu chí AIC và SIC, nghiên cứu chứng minh rằng GARCH (1,1) và EGARCH (1,1) được đánh giá là mô hình thích hợp nhất để đo lường các dao động đối xứng và bất đối xứng của VN-Index. Khoa (2017) đã dự báo những biến động có điều kiện của thị trường chứng khoán Việt Nam. Kết quả cho thấy, mô hình GARCH (1,1) là phù hợp để ước tính sự biến động của thị trường chứng khoán trong nước.

Trong bài viết này, chúng tôi sẽ mô hình hóa sự biến động của thị trường chứng khoán Việt Nam, được đại diện bởi chỉ số VNIndex, qua mô hình ARCH/GARCH. Kết quả thống kê sẽ chỉ ra mô hình

---

[1] Trường Đại học Thương mại - Email: tuanlevan@tmu.edu.vn

[2] Trường Đại học Ngoại thương – Email: quangmathftu@yahoo.com

[3] Tác giả của nó, Robert Engle, đã nhận giải Nobel kinh tế năm 2003 nhờ công trình này.

GARCH nào là phụ hợp để mô tả VNIndex. Bên cạch đó, thông qua giả lập sự biến đổi của VNIndex trong tương lai, chúng tôi cũng đưa ra dự báo về giá trị của chỉ số VNIndex. Các câu lệnh thực hiện trên phần mềm R được trình bày đầy đủ.

2. Cơ sở lý thuyết[4]

2.1. Các khái niệm cơ bản

Ký hiệu chuỗi thời gian là một họ các biến ngẫu nhiên $(X_t)_{t\in Z}$ (còn gọi là quá trình ngẫu nhiên).

Các mô men.

Ta định nghĩa hàm trung bình (mean) và hàm tự hiệp phương sai (autocovariance) của $(X_t)_{t\in Z}$ (nếu tồn tại) là:

$$\mu(t) = E(X_t), \qquad t \in Z$$

$$\gamma(t,s) = E\big((X_t - \mu(t)(X_s - \mu(s))\big), \quad t,s \in Z$$

Tính dừng.

Định nghĩa 1 (dừng chặt). Chuỗi thời gian $(X_t)_{t\in Z}$ được gọi là dừng chặt (strictly stationary) nếu:

$$(X_{t_1},\dots,X_{t_n}) \overset{d}{=} (X_{t_1+k},\dots,X_{t_n+k})$$

với mọi $t_1,\dots,t_n,k \in Z$ và $n \in N$.

Định nghĩa 2 (dừng hiệp phương sai). Chuỗi thời gian $(X_t)_{t\in Z}$ được gọi là dừng hiệp phương sai (covariance stationary – còn gọi là dừng yếu hay dừng bậc 2) nếu các mô men bậc 1 & 2 tồn tại và thỏa mãn

$$\mu(t) = \mu, \qquad t \in Z$$

$$\gamma(t,s) = \gamma(t+k, s+k), \qquad t,s,k \in Z$$

Lưu ý, chuỗi dừng chặt có phương sai hữu hạn là chuỗi dừng hiệp phương sai.

Tự tương quan.

Giả sử $(X_t)_{t\in Z}$ là chuỗi dừng hiệp phương sai, ký hiệu

$$\gamma(h) = \gamma(h,0), \qquad h \in Z$$

Lưu ý, $\gamma(0) = var(X_t),\ \forall t$.

Định nghĩa 3 (hàm tự tương quan). Cho $(X_t)_{t\in Z}$ là chuỗi dừng hiệp phương sai, hàm tự tương quan (autocorrelation) được định nghĩa là

$$\rho(h) = \gamma(h)/\gamma(0), \qquad h \in Z$$

Quá trình nhiễu trắng.

[4] Phần này tham khảo trong McNeil (2005)

Định nghĩa 4 (nhiễu trắng). $(X_t)_{t\in Z}$ được gọi là quá trình nhiễu trắng (white noise) nếu nó là dừng hiệp phương sai và thỏa mãn:

$$\rho(h) = \begin{cases} 1, & h = 0 \\ 0, & h \neq 0 \end{cases}$$

Quá trình nhiễu trắng có trung bình bằng 0, phương sai $\sigma^2 = var(X_t)$ được ký hiệu là $WN(0, \sigma^2)$.

Định nghĩa 5 (nhiễu trắng chặt). $(X_t)_{t\in Z}$ được gọi là quá trình nhiễu trắng chặt (strict white noise) nếu nó là dãy biến ngẫu nhiên độc lập cùng phân phối (idd) có phương sai hữu hạn.

Quá trình nhiễu trắng chặt có trung bình bằng 0, phương sai $\sigma^2 = var(X_t)$ được ký hiệu là $SWN(0, \sigma^2)$.

Lưu ý, quá trình nhiễu trắng chặt là trường hợp riêng của nhiễu trắng.

2.2. Quá trình ARMA

Định nghĩa 6 (quá trình ARMA). Quá trình $(X_t)_{t\in Z}$ được gọi là quá trình ARMA(p, q) trung bình 0 nếu nó là dừng hiệp phương sai và thỏa mãn

$$X_t = \varphi_1 X_{t-1} + \cdots + \varphi_p X_{t-p} + \varepsilon_t + \theta_1 \varepsilon_{t-1} + \cdots + \theta_q \varepsilon_{t-q}, \ \forall t \in Z,$$

với $(\varepsilon_t)_{t\in Z}$ là quá trình $WN(0, \sigma^2)$.

Quá trình $(X_t)_{t\in Z}$ được gọi là quá trình ARMA(p, q) trung bình μ nếu $(X_t - \mu)_{t\in Z}$ là quá trình ARMA(p, q) trung bình 0.

2.3. Quá trình ARCH

Định nghĩa 7. Cho $(Z_t)_{t\in Z}$ là quá trình $SWN(0,1)$. Quá trình $(X_t)_{t\in Z}$ được gọi là quá trình ARCH(p) nếu nó là dừng chặt và thỏa mãn, $\forall t \in Z$:

$$X_t = \sigma_t Z_t$$

$$\sigma_t^2 = \alpha_0 + \sum_{i=1}^{p} \alpha_i X_{t-i}^2$$

với $(\sigma_t)_{t\in Z}$ là quá trình có giá trị dương ngặt; $\alpha_0 > 0$ và $\alpha_i \geq 0, i = 1, \ldots, p$.

2.4. Quá trình GARCH

Định nghĩa 8. Cho $(Z_t)_{t\in Z}$ là quá trình $SWN(0,1)$. Quá trình $(X_t)_{t\in Z}$ được gọi là quá trình GARCH(p, q) nếu nó là dừng chặt và thỏa mãn, $\forall t \in Z$:

$$X_t = \sigma_t Z_t$$

$$\sigma_t^2 = \alpha_0 + \sum_{i=1}^{p} \alpha_i X_{t-i}^2 + \sum_{j=1}^{q} \beta_j \sigma_{t-j}^2$$

với $(\sigma_t)_{t\in Z}$ là quá trình có giá trị dương ngặt; $\alpha_0 > 0$, $\alpha_i \geq 0, i = 1, \ldots, p$ và $\beta_j \geq 0, j = 1, \ldots, q$.

Lưu ý, quá trình GARCH $(X_t)_{t\in Z}$ là dừng hiệp phương sai khi và chỉ khi $\sum_{i=1}^{p} \alpha_i + \sum_{j=1}^{q} \beta_j < 1$. Khi đó, $(X_t)_{t\in Z}$ là nhiễu trắng.

2.5. Mô hình ARMA với sai số GARCH

Ta đã biết rằng quá trình ARMA liên kết với nhiễu trắng $(\varepsilon_t)_{t\in Z}$, và quá trình dừng hiệp phương sai GARCH là một nhiễu trắng. Vì vậy ta có thể kết hợp mô hình ARMA và GARCH bằng cách xem sai số ε_t của mô hình ARMA là $\sigma_t Z_t$.

Định nghĩa 9. Cho $(Z_t)_{t\in Z}$ là quá trình $SWN(0,1)$. Quá trình $(X_t)_{t\in Z}$ được gọi là quá trình ARMA(p_1, q_1) với sai số là GARCH(p_2, q_2) nếu nó là dừng hiệp phương sai và thỏa mãn

$$X_t = \mu_t + \sigma_t Z_t$$

$$\mu_t = \mu + \sum_{i=1}^{p_1} \varphi_i (X_{t-i} - \mu) + \sum_{j=1}^{q_1} \theta_j \left(X_{t-j} - \mu_{t-j}\right)$$

$$\sigma_t^2 = \alpha_0 + \sum_{i=1}^{p_2} \alpha_i (X_{t-i} - \mu_{t-i})^2 + \sum_{j=1}^{q_2} \beta_j \sigma_{t-j}^2$$

với $\alpha_0 > 0, \alpha_i \geq 0, i = 1, \dots, p_2; \beta_j \geq 0, j = 1, \dots, q_2$ và $\sum_{i=1}^{p} \alpha_i + \sum_{j=1}^{q} \beta_j < 1$.

Lưu ý, giả sử $(F_t)_{t\in Z}$ là lọc tự nhiên của $(X_t)_{t\in Z}$, khi đó $\mu_t = E(X_t | F_{t-1})$ và $\sigma_t^2 = var(X_t | F_{t-1})$.

3. Thực hành trên phần mềm R

Câu lệnh R: Cài đặt & gọi các thư viện

```
install.packages("tidyverse");   install.packages("FinTS");    install.packages("fGarch");

library(tidyverse); library(FinTS); library(fGarch)
```

3.1. Dữ liệu

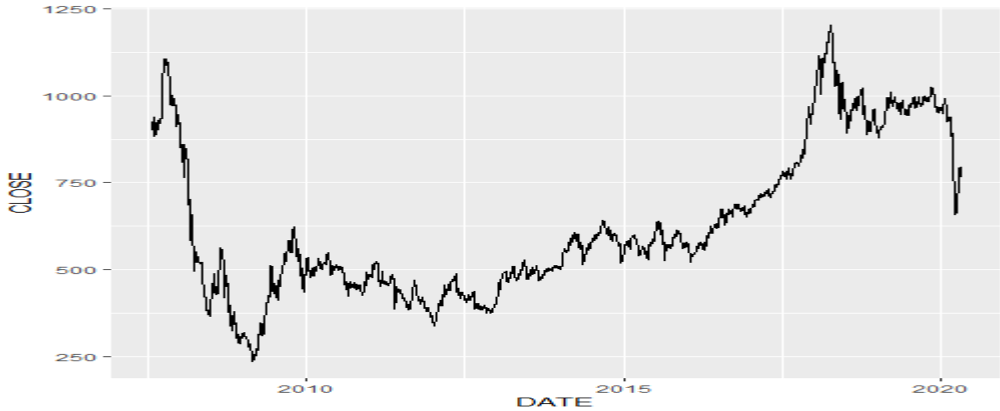
Dữ liệu mô tả chỉ số VNIndex (chỉ số cuối ngày) từ 30/7/2007[5] đến 23/4/2020, dữ liệu bao gồm giai đoạn khủng khoảng tài chính năm 2009 giai đoạn ảnh hưởng của đại dịch Covid-19. Nguồn dữ liệu lấy từ trang chủ của Công ty Cổ phần Chứng khoán VNDIRECT[6].

Câu lệnh R:

```
#Đọc dữ liệu lưu trong file VNIndex.csv trong thư mục D:\data\garch
vnindex<-read.csv("D:\\data\\garch\\VNIndex.csv", header=TRUE)

vnindex<- vnindex[seq(dim(vnindex)[1],1),] #Đảo ngược thứ tự dữ liệu theo chiều tăng thời gian

vnindex$DATE<-as.Date(vnindex$DATE, format = "%d/%m/%Y") #Quy chuẩn thời gian

#Vẽ đồ thị

ggplot(vnindex, aes(x = DATE, y = CLOSE)) + geom_line()
```

[5] Chúng tôi chọn thời điểm xuất phát của dữ liệu từ 30/7/2007 để tránh những năm đầu chưa ổn định của TTCK Việt Nam, và đây là ngày Sở GDCK TP. HCM (HoSE) chính thức áp dụng khớp lệnh liên tục.

[6] https://www.vndirect.com.vn/portal/lich-su-gia/vnindex.shtml

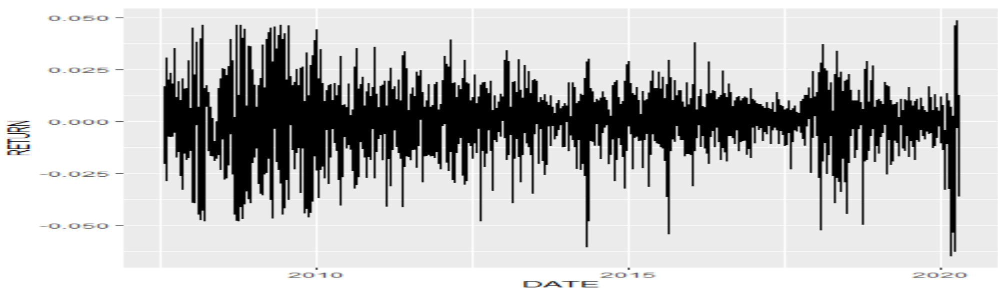


Hình 1. Dữ liệu lịch sử của chỉ số VNIndex

Nhìn chung, hình 1 cho thấy TTCK Việt Nam đã trải qua hai đợt suy giảm mạnh, giai đoạn khủng hoảng tài chính năm 2009 và ảnh hưởng của đại dịch Covid-19. Sau giai đoạn khủng hoảng 2009, thị trường chứng kiến một giai đoạn dài (khoảng 9 năm) tăng trưởng tương đối bền vững.

<u>Câu lệnh R:</u>

```
return <- diff(log(vnindex$CLOSE)) #Tính dãy lợi suất

vnindex$RETURN <-c(0,return) #Bổ sung lợi suất ngày đâu tiên là 0

ggplot(vnindex, aes(x = DATE, y = RETURN)) + geom_line()
```



Hình 2. Dãy lợi suất của VNIndex

Hình 2 cho thấy lợi suất[7] theo ngày của VNIndex dao động xung quanh giá trị 0. Có những biến động mạnh tăng/giảm tập trung gần nhau, gọi là hiện tượng biến động cụm (volatility clustering). Các biến động cụm xuất hiện rõ ràng ở 2 đợt suy giảm. Đây là bằng chứng cho thấy phương sai của lợi suất thay đổi theo thời gian và vai trò của mô hình ARCH/GARCH khi mô hình hóa chuỗi lợi suất.

3.2. Kiểm định hiệu ứng ARCH

Trước khi ước tính mô hình GARCH, cần kiểm tra sự tồn tại hiệu ứng ARCH trong tập dữ liệu. Chúng tôi sẽ sử dụng kiểm định Lagrange Multiplier (LM): hồi quy bình phương sai số theo các độ trễ của nó và kiểm định giả thiết các hệ số của các độ trễ là bằng 0. Phép kiểm định dùng đầu vào là chuỗi lợi suất với một độ trễ cho trước. Giả thiết H_0 là không có hiệu ứng ARCH trong dữ liệu. Giá trị p-value càng nhỏ, càng chứng tỏ tồn tại bằng chứng về hiệu ứng ARCH.

---

[7] Lợi suất (return) được tính theo công thức: $r_t = \log\left(\frac{P_t}{P_{t-1}}\right)$, với P_t là chỉ số cuối ngày.

Câu lệnh R:

```
# Kiểm định cho độ trễ bằng 1, các độ trễ khác làm tương tự

test_out <- ArchTest(vnindex$RETURN, lags = 1)

test_out
```

```
ARCH LM-test; Null hypothesis: no ARCH effects

data:  vnindex$RETURN

Chi-squared = 348.39, df = 1, p-value < 2.2e-16
```

Bảng 1. Kết quả kiểm định hiệu ứng ARCH cho VNIndex

Độ chễ	Thống kê	P-value
1	348.	0
2	506.	0
3	553.	0
4	590.	0
5	643.	0

Bảng 1 cho thấy các giá trị p-value đều nhỏ hơn 5%, đó là bằng chứng cho thấy tồn tại hiệu ứng ARCH trong chuỗi lợi suất của VNIndex.

3.3. Lựa chọn mô hình GARCH

Phần này sẽ trình bày việc lựa chọn mô hình GARCH với độ trễ thích hợp nhất bằng cách so sánh mức độ phù hợp (Goodness Of Fit) của các mô hình. Hai chỉ số quan trọng có thể dùng là AIC (Akaike Information Criteria) hoặc BIC (Bayesian Information Criteria): Chỉ số AIC hoặc BIC nhỏ hơn thì mô hình tốt hơn.

Ta sẽ so sánh mức độ phù hợp của việc mô hình hóa chuỗi lợi suất VNIndex: mô hình ARMA(p_1, q_1) cho trung bình có điều kiện và mô hình GARCH(p_2, q_2) cho phương sai có điều kiện (độ trễ lớn nhất là 5).

Câu lệnh R:

```
# Hàm do_single_garch để kiểm tra cho 1 mô hình ARMA- GARCH

do_single_garch <- function(x, lag_ar, lag_ma, lag_arch, lag_garch) {

my_formula <- formula(paste0('~arma(', lag_ar, ',',lag_ma,')', ' + ','garch(', lag_arch, ',', lag_garch, ')'))

my_garch <- garchFit(formula = my_formula, data = x, trace = FALSE)

est_tab <- tibble(lag_ar, lag_ma, lag_arch, lag_garch,

        AIC = my_garch@fit$ics['AIC'],  BIC = my_garch@fit$ics['BIC'],

         model_name = paste0('ARMA(', lag_ar, ',', lag_ma, ')+','GARCH(', lag_arch, ',', lag_garch, ')') )
return(est_tab)

}
```

```
max_global_lag<-5 #Quy định độ trễ lớn nhất là 5

df_grid <- expand_grid(arma_lag = 0:max_global_lag, garch_lag = 1:max_global_lag)

l_out <- pmap(.l = list(x = rep(list(vnindex$RETURN), nrow(df_grid)),
              lag_ar = df_grid$arma_lag, lag_ma = df_grid$arma_lag,
              lag_arch = df_grid$garch_lag, lag_garch  = df_grid$garch_lag),
              do_single_garch)

tab_out <- bind_rows(l_out)

# Mô hình có AIC nhỏ nhất

idx <- which.min(tab_out$AIC)

best_aic <- tab_out[idx, ]

best_aic
```

```
  lag_ar lag_ma lag_arch lag_garch   AIC   BIC model_name
   <int>  <int>    <int>     <int> <dbl> <dbl> <chr>
1      3      3        1         1 -6.08 -6.06 ARMA(3,3)+GARCH(1,1)
```

```
# Mô hình có BIC nhỏ nhất

idx <- which.min(tab_out$BIC)

best_bic <- tab_out[idx, ]

best_bic
```

```
  lag_ar lag_ma lag_arch lag_garch   AIC   BIC model_name
   <int>  <int>    <int>     <int> <dbl> <dbl> <chr>
1      1      1        1         1 -6.08 -6.07 ARMA(1,1)+GARCH(1,1)
```

Như vậy mô hình tốt nhất là ARMA(3, 3)-GARCH(1, 1) nếu sửa dụng AIC và là ARMA(1, 1)-GARCH(1, 1) nếu sử dụng BIC. Vì mô hình đơn giản hơn sẽ được ưu tiên hơn nên ta sẽ chọn mô hình ARMA(1, 1)-GARCH(1, 1) cho chuỗi lợi suất VNIndex. Các hệ số của mô hình:

<u>Câu lệnh R:</u>

```
my_garch <- garchFit(~arma(1,1)+garch(1,1), data = vnindex$RETURN, trace = FALSE)

my_garch
```

```
Coefficient(s):
        mu        ar1        ma1      omega     alpha1      beta1
```

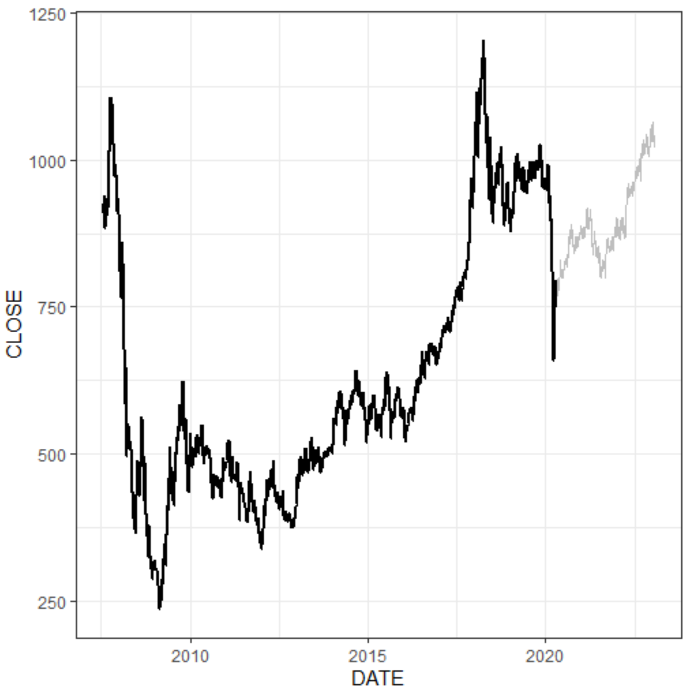
1.9097e-04   3.1102e-01   -1.4948e-01   3.4834e-06   1.3284e-01   8.5160e-01

3.4. Giả lập theo mô hình GARCH

Chúng ta sẽ sử dụng mô hình ARMA(1, 1)-GARCH(1, 1) để giả lập cho các giá trị tương lai của chỉ số VNIndex. Số ngày giả lập trong tương lai là 1000.

Câu lệnh R:

```
df_for <- garchSim(spec = garchSpec(model = coef(my_garch)), n = 1000)

df_sim <- tibble(

        i_t = 0:length(df_for$garch),

        ref_date = last(vnindex$DATE) + i_t,

        sim_log_ret = c(0, df_for$garch),

        sim_arit_ret = exp(sim_log_ret)-1,

        sim_price = last(vnindex$CLOSE)*(cumprod(1+sim_arit_ret)) )

p1 <- ggplot() +

geom_line(data = vnindex,  aes(x = DATE, y = CLOSE), color = 'black', size = 0.75)  +

geom_line(data = df_sim, aes(x =ref_date, y = sim_price, group = 1),color = 'grey', size = 0.35) +

theme_bw()

p1
```



Hình 3. Giả lập cho chỉ số VNIndex trong tương lai

TÀI LIỆU THAM KHẢO

Vương Quân Hoàng. (2004). Hiệu ứng GARCH trên dãy lợi suất thị trường chứng khoán Việt Nam 2000-2003. Tạp chí Ứng dụng toán học tập II, số 1, 2004

Phạm Chí Khoa. (2017). Dự báo biến động giá chứng khoán qua mô hình Arch – Garch. Tạp chí Tài chính, Kỳ 2, 2017, số 6, tr38-39.

Alexander J. McNeil, Rüdiger Frey, Paul Embrechts. (2005). Quantitative Risk Management: Concepts, Techniques, and Tools. Princeton University Press, 26 thg 9, 2005 – 538

Marcelo S. Perlin, Mauro Mastella, Daniel Vancin, Henrique Ramos. (2020). A GARCH Tutorial with R [https://docs.google.com/document/d/1BcIFoM25haxWIVtQfzMMGPnPG--gTt2f3XhGJlSHQus/edit]

Hồ Thủy Tiên, Hồ Thu Hoài, Ngô Văn Toàn. Mô hình hóa biến động thị trường chứng khoán: Thực nghiệm từ Việt Nam. (2017). Tạp chí Khoa học ĐHQGHN: Kinh tế và Kinh doanh, Tập 33, Số 3 (2017) 1-11

Lê Văn Tuấn, Phùng Duy Quang. Áp dụng mô hình GARCH dự báo ảnh hưởng của đại dịch Covid-19 đến thị trường chứng khoán Việt Nam. (2020).