

# Hồi quy phân vị dữ liệu chuỗi thời gian và ứng dụng trong mô hình hóa giá chứng khoán của một số ngân hàng lớn

Lê Thanh Hoa\*, Võ Thị Lệ Uyên, Nguyễn Phát Đạt, Phạm Hoàng Uyên



Use your smartphone to scan this QR code and download this article

## TÓM TẮT

Trong mô hình hồi quy OLS, trung bình của biến phụ thuộc được ước lượng dựa trên trung bình của các biến độc lập. Mỗi quan hệ giữa biến độc lập và biến phụ thuộc cần được xem xét ở nhiều giá trị thay vì chỉ qua giá trị trung bình của biến phụ thuộc, thì mô hình hồi quy phân vị là lựa chọn tối ưu. Trong bài báo này, chúng tôi nghiên cứu về mô hình hồi quy phân vị trong đó rải các mức phân vị của biến phụ thuộc tương ứng từ 10% đến 90%, với khoảng cách 10%, của dữ liệu chuỗi thời gian thông qua mô hình tự hồi quy, nhằm so sánh về độ phù hợp của các mô hình cũng như các sai số của mô hình. Minh chứng các kết quả được chúng tôi ứng dụng vào bộ dữ liệu chuỗi thời gian về giá chứng khoán đóng cửa của bốn mã ngân hàng có mệnh giá lớn nhất (vào ngày 02/08/2021) là VCB, VPB, TCB và BID. Trong đó, các mô hình hồi quy phân vị có độ phù hợp rất cao, tương ứng từ khoảng 80% trở lên. Bên cạnh đó, các tham số ước lượng đều có ý nghĩa thống kê, khác biệt với mô hình hồi quy OLS có một số tham số ước lượng không có ý nghĩa thống kê. Hơn thế nữa, mô hình hồi quy OLS dựa trên trung bình nên dễ bị ảnh hưởng bởi các giá trị đột biến, trong khi đó phân vị không bị ảnh hưởng bởi các giá trị đột biến. Nói cách khác, mô hình hồi quy phân vị khắc phục được yếu điểm này, tức là mô hình hồi quy phân vị không bị ảnh hưởng bởi các giá trị đột biến.

**Từ khóa:** mô hình hồi quy phân vị, mô hình hồi quy phân vị chuỗi thời gian, mô hình giá chứng khoán

## TỔNG QUAN NGHIÊN CỨU

Mô hình hồi quy tuyến tính cổ điển OLS được xây dựng dạng hàm số dựa trên trung bình của biến phụ thuộc, sao cho tổng bình phương các sai lệch giữa giá trị thực tế và giá trị ước lượng bằng mô hình đạt giá trị cực tiểu. Do đó, mô hình ước lượng được theo phương pháp OLS sẽ chỉ phản ánh được giá trị trung bình, không phản ánh được các giá trị khác của biến phụ thuộc, đặc biệt là các trường hợp biến phụ thuộc biến động nhiều. Thật vậy, trong nhiều trường hợp, chúng ta không chỉ quan tâm đến duy nhất giá trị trung bình mà chúng ta cần quan tâm đến các giá trị ở các mức tứ phân vị, thập phân vị... Rõ ràng, khi nghiên cứu nhiều giá trị hơn của các biến phụ thuộc, mô hình hồi quy phân vị là lựa chọn phù hợp trước khi nghiên cứu toàn bộ các điểm của biến phụ thuộc như mô hình hồi quy Bayes. Tắt nhiên, do mô hình hồi quy Bayes có các khó khăn trong quá trình tính toán cũng như các đánh giá đủ sâu về phân phối tiên nghiệm.

Thật vậy, mục đích của mô hình hồi quy là tìm mối quan hệ giữa biến độc lập và biến phụ thuộc (biến cần dự báo) thông qua các thuộc tính chung của các mối quan hệ này bằng giá trị đo lường độ tập trung bao

gồm: trung bình (mean), trung vị (Med) và số yếu vị (Mode). Mô hình hồi quy tuyến tính cổ điển OLS tập trung vào giá trị trung bình, trong đó mỗi quan hệ giữa biến độc lập và biến phụ thuộc được mô tả bởi trung bình thông qua hàm trung bình có điều kiện<sup>1</sup>. Cách tiếp cận mô hình hồi quy thông qua hàm trung bình có điều kiện bao gồm các mô hình như mô hình hồi quy đơn biến, mô hình hồi quy đa biến, mô hình với phương sai sai số thay đổi sử dụng phương pháp bình phương cực tiểu có trọng số hay mô hình hồi quy phi tuyến. Thêm vào đó, một số mô hình cho biến phụ thuộc cũng được nghiên cứu bao gồm mô hình hồi quy cho biến phụ thuộc có hai trạng thái là mô hình logistic – probit, mô hình hồi quy cho biến phụ thuộc dạng số đếm là mô hình Poisson...

Tuy nhiên, đối với mô hình hồi quy cổ điển có một số giới hạn như sau: đầu tiên là mô hình kỳ vọng có điều kiện nghiên cứu cho các trường hợp tuân theo phân phối chuẩn, các trường hợp có đuôi nặng (heavy tail, lower tail, upper tail) chưa được giải quyết trọn vẹn. Thêm nữa là, các giá trị trung tâm mới chỉ ra được một số ít đặc điểm của phân phối tổng thể, trong khi còn rất nhiều các tham số đại diện cho phân phối tổng thể như hình dạng, độ lệch, độ nhọn, các moment bậc cao... Cụ thể là chúng ta cần quan tâm đến các đuôi

Trường Đại học Kinh tế - Luật,  
ĐHQG-HCM, Việt Nam

### Liên hệ

Lê Thanh Hoa, Trường Đại học Kinh tế - Luật,  
ĐHQG-HCM, Việt Nam

Email: hoalt@uel.edu.vn

### Lịch sử

- Ngày nhận: 16-9-2021
- Ngày chấp nhận: 16-8-2022
- Ngày đăng: 30-9-2022

DOI: 10.32508/stdjelm.v6i3.948



### Bản quyền

© ĐHQG Tp.HCM. Đây là bài báo công bố mở được phát hành theo các điều khoản của the Creative Commons Attribution 4.0 International license.



**Trích dẫn bài báo này:** Hoa L T, Uyên V T L, Đạt N P, Uyên P H. Hồi quy phân vị dữ liệu chuỗi thời gian và ứng dụng trong mô hình hóa giá chứng khoán của một số ngân hàng lớn. *Sci. Tech. Dev. J. - Eco. Law Manag.*; 6(3):3015-3032.

dưới, dưới trên của các phân phối xác suất thay vì chỉ nghiên cứu các giá trị trung tâm như trung bình hay trung vị<sup>2</sup>.

Do đó, hồi quy phân vị là lựa chọn phù hợp nhất trong tình huống này, nhằm đảm bảo nghiên cứu nhiều hơn về các giá trị biến động của biến phụ thuộc. Hay nói cách khác, hồi quy phân vị là phương pháp nhằm đạt mục tiêu xác định các phân vị của biến phụ thuộc<sup>3</sup>. Tất nhiên, do hồi quy tuyến tính cổ điển OLS và hồi quy phân vị tập trung vào các đại lượng đặc trưng cho độ tập trung khác nhau, một bên là trung bình, một bên là các phân vị (trong đó bao gồm trung vị) nên các đánh giá sai số cũng phù hợp cho từng tình huống. Nếu như hồi quy tuyến tính cổ điển OLS dựa trên sai số dạng bình phương thì hồi quy phân vị sử dụng sai số dạng trị tuyệt đối<sup>4</sup>. Về bản chất trong ước lượng các tham số của mô hình hồi quy là dựa trên cực tiểu của hàm sai lệch, do đó hiển nhiên việc giải các bài toán này chính là tìm nghiệm của bài toán tối ưu<sup>5,6</sup>. Cũng giống như mô hình hồi quy cổ điển OLS có áp dụng trong trường hợp phi tham số thì mô hình hồi quy phân vị cũng áp dụng cả trường hợp phi tham số<sup>7</sup>. Tức là trong trường hợp này sẽ ước lượng hàm hồi quy dựa vào phân phối thực nghiệm, đặc biệt, các phân vị của phân phối thực nghiệm sẽ có sự thay đổi khi cập nhật thêm dữ liệu. Chính vì vậy, để tốt hơn, chúng ta nên nghiên cứu toàn bộ phân phối xác suất của biến phụ thuộc như trong mô hình hồi quy Bayes. Tuy nhiên, trong mô hình hồi quy Bayes có khá nhiều khó khăn trong tính toán, đặc biệt là việc lựa chọn phân phối tiên nghiệm như thế nào cho phù hợp cũng rất cần các đánh giá thấu đáo. Mặc dù, trong hồi quy Bayes, chúng ta có thể sử dụng tiên nghiệm dạng phi thông tin, nhưng như vậy lại không sử dụng một cách hiệu quả các thông tin hiện có. Để trung hòa tất cả các điều kiện, chúng tôi sử dụng hồi quy phân vị với các mức phân vị đủ nhiều là cách nhau 10%, và khoảng cách đủ rộng là từ 10% đến 90% trong bài nghiên cứu này.

Đối với các dạng dữ liệu của mô hình hồi quy phân vị khá đa dạng, bao gồm cả dữ liệu chéo (ví dụ kết quả của một cuộc khảo sát trong kinh tế, xã hội, tài chính...), dữ liệu dạng chuỗi thời gian (ví dụ như giá gas, giá điện, giá chứng khoán...), và tất nhiên áp dụng được cho dữ liệu mảng<sup>8-10</sup>. Chính vì vậy, mô hình hồi quy phân vị có thể áp dụng được trong nhiều lĩnh vực mà mô hình hồi quy tuyến tính áp dụng được, đặc biệt trong kinh tế - tài chính<sup>11</sup>, thương mại, lao động<sup>12</sup>, sản xuất giữa công nghệ và khí thải carbon<sup>13</sup>, các khuyến nghị về thuốc, phân tích sống còn, các nghiên cứu về kinh tế và tài chính, môi trường<sup>14</sup>... nhằm cùng cấp các dự báo của phân phối có điều kiện của biến phụ thuộc.

Trong bài nghiên cứu này, chúng tôi đề xuất mô hình nghiên cứu hồi quy phân vị ở các mức từ 10% đến 90%, với khoảng cách 10%, so sánh với mô hình hồi quy cổ điển OLS trong mô hình tự hồi quy. Về ứng dụng chúng tôi áp dụng vào bộ dữ liệu giá chứng khoán của bốn ngân hàng lớn Ngân hàng Thương mại Cổ phần Ngoại thương Việt Nam (VCB), Ngân hàng TMCP Việt Nam Thịnh Vượng (VPB), Ngân hàng TMCP Kỹ thương Việt Nam (TCB) và Ngân hàng TMCP Đầu tư và Phát triển Việt Nam (BID), cho thấy các mô hình hồi quy phân vị có tham số tương ứng với các biến độc lập (kể cả các biến hằng số) có ý nghĩa thống kê, mô hình phù hợp với hệ số xác định cao.

## PHƯƠNG PHÁP: HỒI QUY PHÂN VỊ

### Phân vị và hàm phân vị

**Định nghĩa 1.** Phân vị thứ  $p$ , ký hiệu là  $Q^{(p)}$ , của một hàm phân phối xác suất  $F$  là giá trị nhỏ nhất của tập các giá trị  $y$  sao cho:  $F(y) \geq p$ . Hàm số  $Q^{(p)}$  (như là một hàm số của  $p$ ) được gọi là hàm phân vị của  $F$ <sup>1</sup>.

### Mô hình hồi quy tuyến tính

Mô hình hồi quy tuyến tính, trong đó  $y$  là biến ngẫu nhiên liên tục phụ thuộc  $x$ <sup>1</sup>:

$$y_i = \beta_0 + \beta_1 x_{1i} + \dots + \beta_k x_{ki} + \varepsilon_i, \quad (1)$$

Trong đó,  $\varepsilon$  là giống hệt nhau, độc lập và cùng tuân theo phân phối chuẩn với trung bình 0 và phương sai chưa biết  $\sigma^2$ .

Từ giả định trung bình bằng 0, hàm số  $\beta_0 + \beta_1 x_{1i} + \dots + \beta_k x_{ki}$  tương ứng kỳ vọng có điều kiện của  $y$  với  $x$  đã biết, ký hiệu là  $E(y|x)$ . Một điều kiện nữa của mô hình hồi quy tuyến tính là giả định phương sai bằng hằng số  $\sigma^2$ , tức là phương sai có điều kiện  $Var(y|x)$ , trong trường hợp không xảy ra phương sai sai số thay đổi.

Các tham số ước lượng của mô hình hồi quy tuyến tính  $\hat{\beta}_0, \hat{\beta}_1, \dots, \hat{\beta}_k$  là nghiệm của bài toán cực tiểu của bình phương phân dư:

$$\min \sum_i [y_i - (\beta_0 + \beta_1 x_{1i} + \dots + \beta_k x_{ki})]^2. \quad (2)$$

Hay trong trường hợp đơn giản nhất về mô hình hồi quy đơn biến  $y_i = \beta_0 + \beta_1 x_i + \varepsilon_i$  thì  $\hat{\beta}_0$  và  $\hat{\beta}_1$  được ước lượng dựa vào công thức:

$$\begin{cases} \hat{\beta}_1 = \frac{\sum_{i=1}^n (x_i - \bar{x})(y_i - \bar{y})}{\sum_{i=1}^n (x_i - \bar{x})^2}, \\ \hat{\beta}_0 = \bar{y} - \hat{\beta}_1 \bar{x}. \end{cases} \quad (3)$$

Ước lượng phương sai của yếu tố ngẫu nhiên  $\sigma^2$  là  $\hat{\sigma}^2 = \frac{RSS}{n-k-1}$ , trong đó RSS là tổng bình phương các phần dư,  $n$  là tổng số quan sát,  $k$  là tổng số biến độc

lập (trừ biến hằng số  $X_1 = 1$ ). Phân phối xác suất đồng thời của các tham số ước lượng  $\hat{\beta}$  theo phương pháp bình phương cực tiểu là phân phối chuẩn nhiều chiều, với trung bình là giá trị đúng  $\beta$  và ma trận hiệp phương sai  $\sigma^2 (X'X)^{-1}$ . Từ đây, chúng ta sử dụng phân phối xác suất chuẩn nhiều chiều cho các bài toán về ước lượng và kiểm định giả thuyết.

Mô hình hồi quy tuyến tính cổ điển áp dụng cho trường hợp tự hồi quy tổng quát với dữ liệu chuỗi thời gian có dạng<sup>5</sup>:

$$y_t = \alpha_0 + \alpha_1 y_{t-1} + \dots + \alpha_p y_{t-p} + \varepsilon_t \quad (4)$$

Trong đó  $p$  là quá trình tự tương quan bậc  $p$ : AR( $p$ ). Khi đó, phương sai có điều kiện của  $\varepsilon_t$  thông qua mô hình ARCH( $r$ )<sup>5</sup>:

$$E(\varepsilon_t^2 | I_t) = \sigma_t^2 = \gamma_0 + \gamma_1 \varepsilon_{t-1}^2 + \dots + \gamma_r \varepsilon_{t-r}^2, \quad (5)$$

Trong đó  $I_t$  là tập các thông tin tại thời điểm  $t$ . Tức là phương trình ước lượng có dạng:

$$\hat{u}_t^2 = \gamma_0 + \gamma_1 \hat{u}_{t-1}^2 + \dots + \gamma_r \hat{u}_{t-r}^2, \quad (6)$$

Trong đó  $\hat{u}_t$  là phần dư của mô hình AR( $p$ ) và các giá trị độ trễ  $\hat{u}_{t-1}, \dots, \hat{u}_{t-r}$ .

### Mô hình hồi quy phân vị

Mô hình hồi quy phân vị cũng được xác định tương tự mô hình hồi quy tuyến tính<sup>5</sup>:

$$y_i = \beta_0^{(p)} + \beta_1^{(p)} x_{1i} + \dots + \beta_k^{(p)} x_{ki} + \varepsilon_i^{(p)}, \quad (7)$$

Trong đó  $0 < p < 1$  và  $\beta_1^{(p)}$  là phân vị ở mức  $p$  của tham số mô hình hồi quy.

Do đó, phân vị có điều kiện tại mức phân vị  $p$  khi đã biết  $x_i$  là:

$$Q^{(p)}(y_i | x_i) = \beta_0^{(p)} + \beta_1^{(p)} x_{1i} + \dots + \beta_k^{(p)} x_{ki}.$$

Mặt khác, do  $\beta_0^{(p)} + \beta_1^{(p)} x_{1i} + \dots + \beta_k^{(p)} x_{ki}$  là hằng số nên ta có phân vị thứ  $p$  của sai số ngẫu nhiên nhân giá trị 0, tức là:

$$\begin{aligned} Q^{(p)}(y_i | x_i) &= \beta_0^{(p)} + \beta_1^{(p)} x_{1i} + \dots + \beta_k^{(p)} x_{ki} \\ &= \beta_0^{(p)} + \beta_1^{(p)} x_{1i} + \dots + \beta_k^{(p)} x_{ki} + Q^{(p)}(\varepsilon_i). \end{aligned} \quad (8)$$

Các ước lượng và kiểm định của tham số tương tự như trong mô hình hồi quy tuyến tính<sup>1</sup>.

Để đánh giá sự phù hợp của mô hình hồi quy phân vị, chúng ta có các giá trị về độ phù hợp Pseudo R-squared và Adjusted R-squared tương tự như độ phù hợp  $R^2$  và độ phù hợp hiệu chỉnh  $R^2$  của mô hình hồi quy OLS. Các kết quả diễn giải tương ứng được nêu chi tiết trong bài báo<sup>15</sup>. Tất nhiên, cần lưu ý các độ phù hợp của mô hình của mô hình hồi quy phân vị khác độ phù hợp trong mô hình OLS do sử dụng độ

đo khác nhau, một bên dạng trị tuyệt đối và một bên dạng bình phương.

Tương tự trong hồi quy cổ điển OLS có mô hình tự hồi quy, thì mô hình hồi quy phân vị cũng có mô hình tự hồi quy phân vị QARCH( $p$ )<sup>5</sup>:

$$y_t = \alpha_0^{(p)} + \alpha_1^{(p)} y_{t-1} + \dots + \alpha_p^{(p)} y_{t-p} + \varepsilon_t^{(p)}. \quad (9)$$

Tuy nhiên, do độ đo để tính sai số khác nhau nên mô hình đánh giá phương sai trong hồi quy phân vị khác phương sai trong mô hình hồi quy cổ điển OLS. Cụ thể, mô hình QARCH( $r$ ) được biểu diễn dưới dạng sa<sup>5</sup>:

$$\sigma_t = \gamma_0 + \gamma_1 |\varepsilon_{t-1}| + \dots + \gamma_r |\varepsilon_{t-r}|. \quad (10)$$

Hay biểu diễn dưới dạng mô hình ước lượng:

$$|\tilde{u}_t| = \gamma_0 + \gamma_1 |\tilde{u}_{t-1}| + \dots + \gamma_r |\tilde{u}_{t-r}|. \quad (11)$$

## KẾT QUẢ NGHIÊN CỨU: ỨNG DỤNG VÀO BỘ DỮ LIỆU TÀI CHÍNH

### Mô tả dữ liệu

Chúng tôi nghiên cứu giá đóng cửa của bốn mã chứng khoán Ngân hàng Thương mại Cổ phần Ngoại thương Việt Nam (VCB), Ngân hàng Thương mại Cổ phần Việt Nam Thịnh Vượng (VPB), Ngân hàng Thương mại Cổ phần Kỹ thương Việt Nam (TCB) và Ngân hàng Thương mại Cổ phần Đầu tư và Phát triển Việt Nam (BID), theo ngày từ đầu năm 2019, tháng 01/2019, đến hết tháng 7/2021. Bộ dữ liệu giá chứng khoán được thu thập từ Trung tâm Nghiên cứu Kinh tế - Tài chính, Trường Đại học Kinh tế - Luật, Đại học Quốc gia Thành phố Hồ Chí Minh, với phương pháp thu thập dữ liệu thứ cấp từ nguồn có sẵn. Đây là bốn mã chứng khoán ngân hàng có mệnh giá lớn nhất ngày đầu tiên của tháng 08/2021. Định nghĩa về các ngân hàng lớn với mã chứng khoán xét trong bài báo này tương ứng với giá chứng khoán của các mã cổ phiếu của các ngân hàng đó lớn, định nghĩa này không trùng với định nghĩa về ngân hàng lớn thông qua các tiêu chí về vốn điều lệ, tổng giá trị tài sản, mạng lưới hoạt động, hệ thống chi nhánh, phòng giao dịch, số lượng nhân viên, số lượng khách hàng... Tuy nhiên, độ lớn của mệnh giá giá chứng khoán cũng phần nào phản ánh được các tiêu chí trên, minh chứng bốn mã chứng khoán của chúng tôi xem xét trong bài báo này cũng thuộc vào top các ngân hàng lớn nhất Việt Nam trong thời điểm hiện tại.

Theo đồ thị Hình 1, các biến động của các mã chứng khoán này rất khác biệt nhau, nếu như VCB là mã có mệnh giá lớn, tang nhiều và giảm nhiều, so với các giá lịch sử thì giá có xu thế tăng tuyến tính. Trong khi

đó, các mã VPB và TCB trong các giai đoạn trước giá ở mức ổn định, sau đó từ cuối năm 2020 đến nay giá tăng cao đột ngột. Trái lại, với BID, giá chứng khoán đạt đỉnh ở cuối năm 2019, sau đó giảm mạnh ở đầu năm 2020, và hiện nay khá ổn định với xu hướng tăng nhẹ.

Chi tiết các đại lượng thống kê mô tả của dữ liệu về bốn mã chứng khoán này được biểu diễn trong Bảng 1.

So về giá trị trung bình thò mã chứng khoán VCB là cao nhất, sau đó đến BID, còn VPB và TCB xấp xỉ nhau. Xét về độ biến động thì VPB và VCB biến động nhiều nhất, tiếp theo là TCB và cuối cùng là BID. Nếu xem xét về dạng phân phối xác suất của dữ liệu thì cả bốn bộ dữ liệu này đều không tuân theo phân phối chuẩn. Mặc dù các bộ dữ liệu không tuân theo phân phối chuẩn, tuy nhiên do tính phân vị nên với mọi bộ dữ liệu, chúng ta luôn tính toán được các giá trị phân vị thực nghiệm. Hơn thế nữa, các giá trị phân vị thực nghiệm này là các giá trị xuất hiện trong bộ dữ liệu. Như vậy, các giá trị phân vị ưu điểm hơn so với giá trị trung bình, vì trung bình bị ảnh hưởng bởi giá trị đột biến, có thể xảy ra trường hợp giá trị trung bình không phải là giá trị xuất hiện trong bộ dữ liệu.

### Kiểm định nghiệm đơn vị và lựa chọn mô hình tự hồi quy phân vị QAR

Về các bộ dữ liệu mặc dù khác nhau, tuy nhiên chúng cùng là dữ liệu chuỗi thời gian nên chúng ta cần kiểm tra tính dừng của dữ liệu. Việc kiểm tra tính dừng của dữ liệu thông qua kiểm định nghiệm đơn vị. Khi đã xác định được bộ dữ liệu dừng, chúng ta xác định mô hình hồi quy phù hợp thông qua giản đồ tự tương quan trong cả hồi quy cổ điển OLS và hồi quy phân vị.

#### Chuỗi VCB

Kiểm định nghiệm đơn vị thấy dữ liệu gốc VCB không dừng theo kết quả của Bảng 2.

Tuy nhiên, đến sai phân bậc một dữ liệu VCB thì dừng theo kết quả Bảng 3.

Do sai phân bậc một chuỗi dữ liệu VCB đã dừng nên chúng ta sử dụng giản đồ tự tương quan của sai phân bậc một VCB theo Hình 2, làm cơ sở cho việc chọn mô hình hồi quy ARIMA dạng tiếp theo:

Dựa vào kết quả Hình 2, chúng ta có thể thấy sự phù hợp trong lựa chọn mô hình ARIMA(0,1,0) hay chính là AR(1), nên chúng tôi sử dụng mô hình dạng AR(1) cho các phần ước lượng tiếp theo.

#### Chuỗi VPB

Kiểm định nghiệm đơn vị cho thấy dữ liệu gốc VPB không dừng theo kết quả Bảng 4.

Do chuỗi dữ liệu gốc VPB không dừng nên thực hiện tiếp kiểm định nghiệm đơn vị với sai phân bậc một chuỗi dữ liệu VPB, kết quả ở Bảng 5 chỉ ra sai phân bậc một chuỗi VPB dừng.

Dựa vào giản đồ tự tương quan chuỗi sai phân bậc một VPB trong Hình 3, chúng ta có thể chọn mô hình ARIMA(1,1,1), tuy nhiên, để đồng nhất và dễ so sánh với các chuỗi dữ liệu khác, chúng tôi thu nhỏ mô hình về dạng ARIMA(0,1,0), tức là AR(1).

#### Chuỗi TCB

Đối với dữ liệu TCB, chúng tôi cũng kiểm định nghiệm đơn vị với dữ liệu gốc, kết quả thể hiện qua Bảng 6 thấy rằng dữ liệu gốc TCB không dừng.

Tương tự như trên, sai phân bậc một của chuỗi dữ liệu TCB là dừng theo kết quả của Bảng 7.

Chúng tôi cũng xem xét giản đồ tự tương quan của chuỗi sai phân TCB nhằm chọn các tham số trong mô hình ARIMA. Kết quả trong Hình 4 chỉ ra rằng dữ liệu TCB phù hợp mô hình ARIMA(0,1,0) hay chính là AR(1).

#### Chuỗi BID

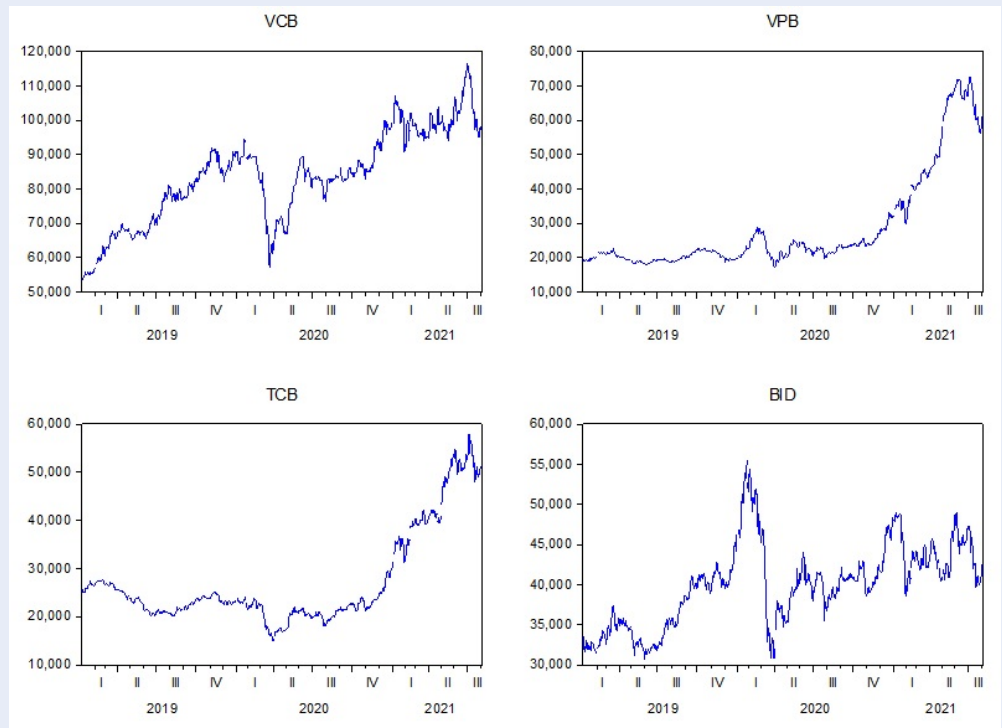
Cuối cùng, chúng tôi xem xét dữ liệu gốc BID, kết quả thể hiện trong Bảng 8 chứng tỏ dữ liệu gốc BID không dừng.

Tương tự như các bộ dữ liệu trên, bộ dữ liệu BID có sai phân bậc một là chuỗi dừng theo kết quả của Bảng 9.

Chúng ta cũng xem xét giản đồ tự tương quan của sai phân bậc một chuỗi dữ liệu BID, kết quả thể hiện qua Hình 5. Kết quả chỉ ra bộ dữ liệu phù hợp với mô hình ARIMA(0,1,0) hay chính là AR(1).

Điểm chung của bốn bộ dữ liệu VCB, VPB, TCB và BID đều có dữ liệu gốc là chuỗi không dừng, sai phân bậc một là chuỗi dừng, giản đồ tự tương quan khá phù hợp với mô hình AR(1). Do đó, trong các phần tiếp theo, chúng tôi sử dụng bộ dữ liệu AR(1) trong ước lượng các mô hình tương ứng. Sau khi có các kiểm định về nghiệm đơn vị đảm bảo đưa dữ liệu về dạng chuỗi dừng thông qua sai phân, chúng ta thấy các bộ dữ liệu này đủ điều kiện cho ước lượng mô hình hồi quy OLS cũng như mô hình hồi quy phân vị. Trong các phân tích tiếp theo, mô hình hồi quy OLS chỉ đưa ra mô hình ước lượng thông qua trung bình của biến phụ thuộc, trong khi đó, mô hình hồi quy phân vị đưa ra các thông tin về mô hình hồi quy ở các mức phân vị của bộ dữ liệu.

### Mô hình hồi quy phân vị và đánh giá độ phù hợp của các mô hình



**Hình 1:** Đồ thị giá đóng cửa của bốn mã chứng khoán theo thời gian từ 01/2019 đến hết 07/2021 (Nguồn: Kết quả nghiên cứu)

**Bảng 1:** Các đại lượng thống kê mô tả của bốn mã chứng khoán theo thời gian

	VCB	VPB	TCB	BID
Trung bình	83408,22	28696,36	27302,09	39996,04
Trung vị	84200,00	22300,00	23300,00	40350,00
Giá trị lớn nhất	116400,0	72700,00	58000,00	55500,00
Giá trị nhỏ nhất	53600,00	16950,00	14900,00	30600,00
Độ lệch chuẩn	13424,31	14500,49	9901,501	5127,963
Hệ số lệch	-0,148017	1,812225	1,498752	0,272998
Hệ số nhọn	2,436162	5,026595	4,089170	2,742400
Jarque-Bera	10,89913	463,4251	273,3541	9,795143
Xác suất	0,004298	0,000000	0,000000	0,007465
Tổng	53798300	18509150	17609850	25797446
Tổng bình phương độ lệch	1,16E+11	1,35E+11	6,31E+10	1,69E+10
Số quan sát	645	645	645	645

Nguồn: Kết quả nghiên cứu

**Bảng 2: Kiểm định nghiệm đơn vị dữ liệu gốc của VCB**

Giả thuyết H0: VCB có 1 nghiệm đơn vị			
Ngoại sinh: Hằng số, Xu thế tuyến tính			
Độ dài trễ: 0 (Tự động – dựa theo SIC, độ trễ tối đa =19)			
		Trị thống kê t	Trị xác suất*
Giá trị kiểm định ADF		-2,804987	0,1960
Các giá trị tới hạn:	Mức ý nghĩa 1%	-3,972368	
	Mức ý nghĩa 5%	-3,416811	
	Mức ý nghĩa 10%	-3,130757	
*MacKinnon (1996) one-sided p-values.			

Nguồn: Kết quả nghiên cứu

**Bảng 3: Kiểm định nghiệm đơn vị dữ liệu sai phân bậc một của VCB**

Giả thuyết H0: D(VCB) có 1 nghiệm đơn vị			
Ngoại sinh: Hằng số, Xu thế tuyến tính			
Độ dài trễ: 0 (Tự động – dựa theo SIC, độ trễ tối đa =19)			
		Trị thống kê t	Trị xác suất*
Giá trị kiểm định ADF		-25,10407	0,0000
Các giá trị tới hạn:	Mức ý nghĩa 1%	-3,972390	
	Mức ý nghĩa 5%	-3,416822	
	Mức ý nghĩa 10%	-3,130764	
*MacKinnon (1996) one-sided p-values.			

Nguồn: Kết quả nghiên cứu

**Bảng 4: Kiểm định nghiệm đơn vị với chuỗi dữ liệu gốc VPB**

Giả thuyết H0: VPB có 1 nghiệm đơn vị			
Ngoại sinh: Hằng số, Xu hướng tuyến tính			
Độ dài trễ: 1 (Tự động – dựa theo SIC, độ trễ tối đa =19)			
		Trị thống kê t	Trị xác suất*
Giá trị kiểm định ADF		-0,991287	0,9431
Các giá trị tới hạn:	Mức ý nghĩa 1%	-3,972390	
	Mức ý nghĩa 5%	-3,416822	
	Mức ý nghĩa 10%	-3,130764	
*MacKinnon (1996) one-sided p-values.			

Nguồn: Kết quả nghiên cứu



**Bảng 5: Kiểm định nghiệm đơn vị với chuỗi sai phân bậc một dữ liệu VPB**

Giả thuyết H0: D(VPB) có 1 nghiệm đơn vị			
Ngoại sinh: Hằng số, Xu hướng tuyến tính			
Độ dài trễ: 0 (Tự động – dựa theo SIC, độ trễ tối đa=19)			
		Trị thống kê t	Trị xác suất*
Giá trị kiểm định ADF		-21,89772	0,0000
Các giá trị tới hạn:	Mức ý nghĩa 1%	-3,972390	
	Mức ý nghĩa 5%	-3,416822	
	Mức ý nghĩa 10%	-3,130764	
*MacKinnon (1996) one-sided p-values.			

Nguồn: Kết quả nghiên cứu

**Bảng 6: Kiểm định nghiệm đơn vị với dữ liệu gốc TCB**

Giả thuyết H0: TCB có 1 nghiệm đơn vị			
Ngoại sinh: Hằng số, Xu hướng tuyến tính			
Độ dài trễ: 0 (Tự động – dựa theo SIC, độ trễ tối đa=19)			
		Trị thống kê t	Trị xác suất*
Giá trị kiểm định ADF		-0,822008	0,9620
Các giá trị tới hạn:	Mức ý nghĩa 1%	-3,972368	
	Mức ý nghĩa 5%	-3,416811	
	Mức ý nghĩa 10%	-3,130757	
*MacKinnon (1996) one-sided p-values.			

Nguồn: Kết quả nghiên cứu

**Bảng 7: Kiểm định nghiệm đơn vị với dữ liệu sai phân bậc một TCB**

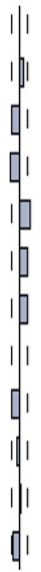
Giả thuyết H0: D(TCB) có 1 nghiệm đơn vị			
Ngoại sinh: Hằng số, Xu hướng tuyến tính			
Độ dài trễ: 0 (Tự động – dựa theo SIC, độ trễ tối đa=19)			
		Trị thống kê t	Trị xác suất*
Giá trị kiểm định ADF		-26,67394	0,0000
Các giá trị tới hạn:	Mức ý nghĩa 1%	-3,972390	
	Mức ý nghĩa 5%	-3,416822	
	Mức ý nghĩa 10%	-3,130764	
*MacKinnon (1996) one-sided p-values.			

Nguồn: Kết quả nghiên cứu

Mẫu: 1/02/2019 8/02/2021

Số quan sát: 644

Hệ số tương quan



Hệ số tương quan riêng phần



Độ trễ	AC	PAC	Thống kê Q	Trị xác suất
1	0,008	0,008	0,043	0,836
2	0,035	0,035	0,8418	0,656
3	-0,073	-0,073	4,263	0,234
4	-0,086	-0,087	9,0823	0,059
5	0,093	0,101	14,691	0,012
6	0,078	0,079	18,656	0,005
7	0,072	0,051	22,002	0,003
8	-0,001	-0,003	22,004	0,005
9	-0,074	-0,053	25,547	0,002
10	-0,022	-0,009	25,863	0,004
11	0,015	0,017	26,011	0,006
12	-0,058	-0,086	28,247	0,005

Hình 2: Giản đồ tự tương quan sai phân bậc một chuỗi VCB (Nguồn: Kết quả nghiên cứu)

Mẫu: 1/02/2019 8/02/2021

Số quan sát: 644

Hệ số tương quan



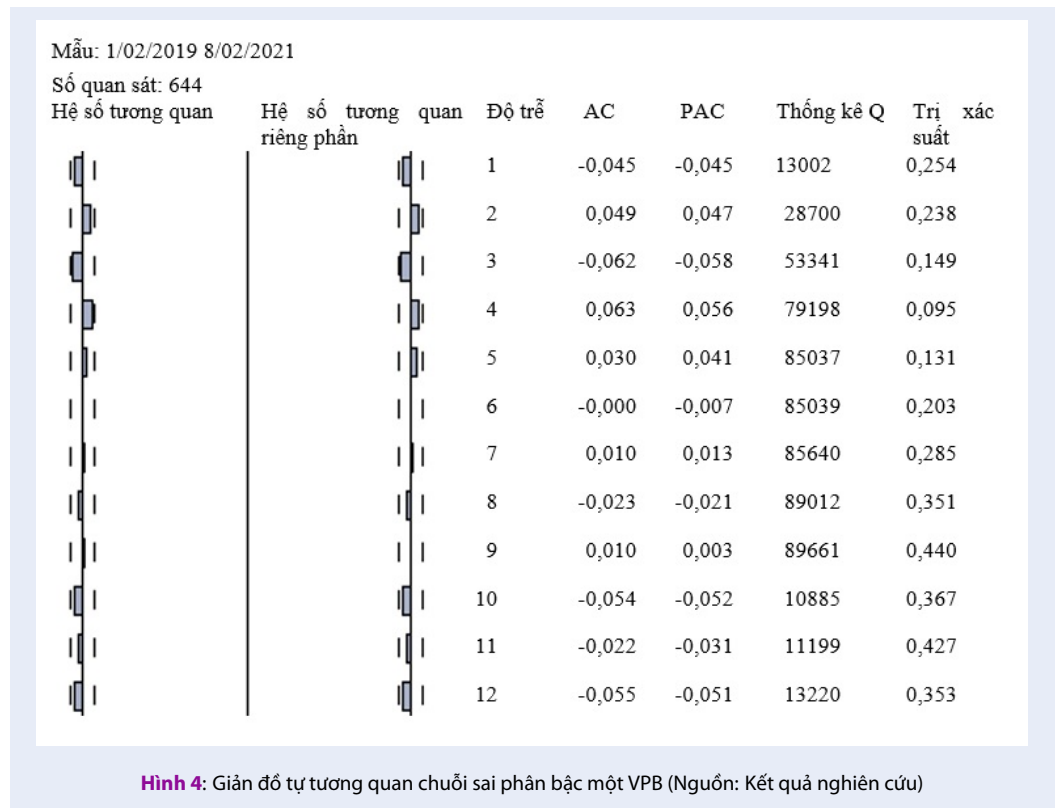
Hệ số tương quan riêng phần



Độ trễ	AC	PAC	Thống kê Q	Trị xác suất
1	0,148	0,148	14,183	0,000
2	0,063	0,042	16,725	0,000
3	0,021	0,006	17,001	0,001
4	-0,046	-0,053	18,355	0,001
5	0,121	0,137	27,897	0,000
6	-0,010	-0,044	27,960	0,000
7	0,043	0,042	29,192	0,000
8	0,040	0,024	30,229	0,000
9	-0,002	0,000	30,232	0,000
10	0,074	0,053	33,835	0,000
11	-0,017	-0,027	34,034	0,000
12	-0,093	-0,104	39,687	0,000

Hình 3: Giản đồ tự tương quan chuỗi sai phân bậc một VPB (Nguồn: Kết quả nghiên cứu)





**Bảng 8:** Kiểm định nghiệm đơn vị với dữ liệu gốc BID

Giả thuyết H0: BID có 1 nghiệm đơn vị		
Ngoại sinh: Hằng số, Xu hướng tuyến tính		
Độ dài trễ: 0 (Tự động – dựa theo SIC, độ trễ tối đa =19)		
		Trị thống kê t
Trị thống kê ADF		-2,783514
Các giá trị tới hạn:	Mức ý nghĩa 1%	-3,972368
	Mức ý nghĩa 5%	-3,416811
	Mức ý nghĩa 10%	-3,130757
*MacKinnon (1996) one-sided p-values.		

Nguồn: Kết quả nghiên cứu

**Chuỗi VCB**

Khi chạy mô hình giá VCB:

$$VCB_t = \beta_0^{(p)} + \beta_1^{(p)} VCB_{t-1} + \epsilon_t^{(p)} \quad (12)$$

Mô hình hồi quy tuyến tính OLS khi ước lượng mô hình trên có các tham số ước lượng với hệ số tự do là 850,98 và hệ số góc là 0,99. Các hệ số hồi quy và tính phù hợp của mô hình hồi quy đều có ý nghĩa thống kê với mức ý nghĩa 5%. Khi đó, độ phù hợp của mô hình khá cao với độ phù hợp là 98,73% thông qua mô

hình:

$$VCB_t = 850,9769 + 0,9906VCB_{t-1} + e_t \quad (13)$$

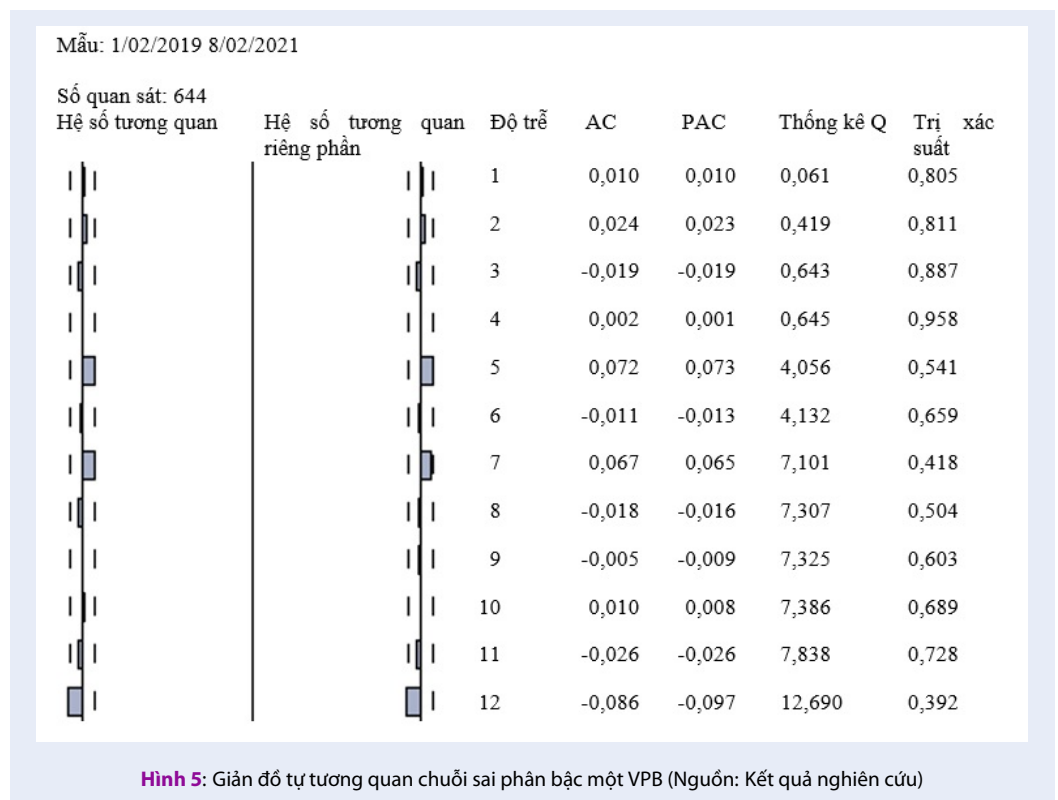
(Se) (374,4965) (0,0044)

Tiếp theo, chúng tôi chạy mô hình tự hồi quy phân vị bậc một dữ liệu VCB thông qua kết quả của Bảng 10, với các mức phân vị từ 10% đến 90%. Các kết quả cho thấy độ phù hợp của mô hình khá cao, từ 87% trở lên chứng minh sự phù hợp của mô hình hồi quy.

**Bảng 9:** Kiểm định nghiệm đơn vị với sai phân bậc một của dữ liệu BID

Giả thuyết H0: D(BID) có 1 nghiệm đơn vị			
Ngoại sinh: Hằng số, Xu hướng tuyến tính			
Độ dài trễ: 0 (Tự động – dựa theo SIC, độ trễ tối đa =19)			
		Trị thống kê t	Trị xác suất*
Giá trị kiểm định ADF		-25,06588	0,0000
Các giá trị tới hạn:	Mức ý nghĩa 1%	-3,972390	
	Mức ý nghĩa 5%	-3,416822	
	Mức ý nghĩa 10%	-3,130764	
*MacKinnon (1996) one-sided p-values.			

Nguồn: Kết quả nghiên cứu



**Chuỗi VPB**

Khi chạy mô hình giá VPB:

$$VPB_t = \beta_0^{(p)} + \beta_1^{(p)}VPB_{t-1} + \epsilon_t^{(p)}. \tag{14}$$

Tương tự, chúng tôi cũng ước lượng các tham số hồi quy dạng OLS như sau:

$$VPB_t = 4,6788 + 1,0021VPB_{t-1} + e_t \tag{15}$$

(Se) (66,0835) (0,0021)

Trong mô hình dạng OLS, các hệ số được ước lượng lần lượt là: hệ số tự do là 4,68 và hệ số góc là 1,00. Mặc

dù hệ số tự do không có ý nghĩa thống kê với mức ý nghĩa 1%, 5% hay 10%, nhưng hệ số góc lại có ý nghĩa thống kê và tính phù hợp của toàn mô hình, thể hiện qua độ phù hợp của mô hình hồi quy là 99,73%.

Tiếp theo, chúng tôi chạy mô hình tự hồi quy phân vị bậc một dữ liệu VPB thông qua kết quả của Bảng 11, với các mức phân vị từ 10% đến 90%. Các kết quả cho thấy độ phù hợp của mô hình khá cao, từ 87% trở lên chứng minh sự phù hợp của mô hình hồi quy. Đồng thời các hệ số trong mô hình tự hồi quy phân vị bậc một chuỗi dữ liệu VPB đều có ý nghĩa thống kê.

**Bảng 10: Mô hình tự hồi quy phân vị bậc một chuỗi dữ liệu VCB**

Phân vị	10%	20%	30%	40%	50%	60%	70%	80%	90%
Tham số									
$\hat{\beta}_0^{(p)}$	1317,81	1000,00	1050,98	604,52	478,26	855,88	600,00	711,19	587,36
$\hat{\beta}_1^{(p)}$	0,97	0,98	0,98	0,99	0,99	0,99	1,00	1,00	1,01
R <sup>2</sup> -Pseudo	0,89	0,90	0,91	0,90	0,90	0,90	0,90	0,89	0,87
R <sup>2</sup> hiệu chỉnh	0,89	0,90	0,91	0,90	0,90	0,90	0,90	0,89	0,87

Nguồn: Kết quả nghiên cứu

**Bảng 11: Mô hình tự hồi quy phân vị bậc một chuỗi dữ liệu VPB**

Phân vị	10%	20%	30%	40%	50%	60%	70%	80%	90%
Tham số									
$\hat{\beta}_0^{(p)}$	82,69	7,13	38,97	-50,00	-118,08	-115,28	-94,59	-98,99	-284,50
$\hat{\beta}_1^{(p)}$	0,98	0,99	0,99	1,00	1,01	1,01	1,01	1,02	1,04
R <sup>2</sup> -Pseudo	0,87	0,91	0,93	0,94	0,95	0,95	0,96	0,97	0,97
R <sup>2</sup> hiệu chỉnh	0,87	0,91	0,93	0,94	0,95	0,95	0,96	0,97	0,97

Nguồn: Kết quả nghiên cứu

**Chuỗi TCB**

Khi chạy mô hình giá TCB:

$$TCB_t = \beta_0^{(p)} + \beta_1^{(p)}TCB_{t-1} + \varepsilon_t^{(p)}. \tag{16}$$

Mô hình hồi quy OLS cho dữ liệu lịch sử TCB như sau:

$$TCB_t = -16,6425 + 1,0020TCB_{t-1} + e_t \tag{17}$$

(Se) (78,4013) (0,0027)

Thông qua kết quả mô hình OLS, các hệ số ước lượng được bao gồm hệ số tự do là -16,64 và hệ số góc là 1,00. Mặc dù hệ số tự do không có ý nghĩa thống kê ở các mức 1%, 5% và 10%, nhưng hệ số góc và tính phù hợp của toàn mô hình đều có ý nghĩa thống kê thể hiện qua độ phù hợp khá cao 99,53%.

Tiếp theo, chúng tôi chạy mô hình tự hồi quy phân vị bậc một dữ liệu TCB thông qua kết quả của Bảng 12, với các mức phân vị từ 10% đến 90%. Các kết quả cho thấy độ phù hợp của mô hình khá cao, từ 88% trở lên chứng minh sự phù hợp của mô hình hồi quy. Đồng thời các hệ số trong mô hình tự hồi quy phân vị bậc một chuỗi dữ liệu TCB đều có ý nghĩa thống kê.

**Chuỗi BID**

Khi chạy mô hình giá BID:

$$BID_t = \beta_0^{(p)} + \beta_1^{(p)}BID_{t-1} + \varepsilon_t^{(p)}. \tag{18}$$

Mô hình hồi quy OLS:

$$BID_t = 655,2726 + 0,9840BID_{t-1} + e_t. \tag{19}$$

(Se) (274,2337) (0,0068)

Dựa vào kết quả mô hình hồi quy OLS với các hệ số góc và hệ số tự do đều có ý nghĩa thống kê với mức ý nghĩa 5%, 10%, trong đó hệ số tự do bằng 655,27 và hệ số góc là 0,98. Đây là mô hình hoàn toàn phù hợp với độ phù hợp cao 97,02%.

Tiếp theo, chúng tôi chạy mô hình tự hồi quy phân vị bậc một dữ liệu BID thông qua kết quả của Bảng 13, với các mức phân vị từ 10% đến 90%. Các kết quả cho thấy độ phù hợp của mô hình khá cao, từ 79% trở lên chứng minh sự phù hợp của mô hình hồi quy.

Khi so sánh về kết quả thực nghiệm giữa mô hình hồi quy OLS và mô hình hồi quy phân vị, chúng ta nhận thấy mô hình hồi quy OLS dựa trên các giá trị trung bình nên dễ bị ảnh hưởng bởi các giá trị đột biến, tức là các giá trị quá lớn hoặc quá nhỏ. Trong khi đó, mô hình hồi quy phân vị dựa trên các mức phân vị nên hầu như không bị ảnh hưởng bởi giá trị đột biến. Thật vậy, các kết quả mô hình OLS khi ước lượng với bộ dữ liệu VPB và TCB đều làm cho các hệ số tự do không có ý nghĩa thống kê, trong khi cũng các bộ dữ liệu này thì mô hình hồi quy phân vị thì các hệ số đều có ý nghĩa thống kê. Mặc dù, các bộ dữ liệu về giá chứng khoán

**Bảng 12: Mô hình tự hồi quy phân vị bậc một chuỗi dữ liệu TCB**

Phân vị	10%	20%	30%	40%	50%	60%	70%	80%	90%
Tham số									
$\hat{\beta}_0^{(p)}$	488,37	161,30	57,47	-50,00	-61,33	-145,18	-188,72	-175,12	-388,58
$\hat{\beta}_1^{(p)}$	0,96	0,98	0,99	1,00	1,00	1,01	1,02	1,02	1,04
R <sup>2</sup> -Pseudo	0,88	0,90	0,92	0,93	0,93	0,94	0,95	0,96	0,96
R <sup>2</sup> hiệu chỉnh	0,88	0,90	0,91	0,93	0,93	0,94	0,95	0,96	0,96

Nguồn: Kết quả nghiên cứu

**Bảng 13: Mô hình tự hồi quy phân vị bậc một chuỗi dữ liệu BID**

Phân vị	10%	20%	30%	40%	50%	60%	70%	80%	90%
Tham số									
$\hat{\beta}_0^{(p)}$	1370,21	796,07	203,70	459,42	$2,49 \times 10^{-1}$	200,00	350,00	566,72	-206,25
$\hat{\beta}_1^{(p)}$	0,94	0,97	0,99	0,98	1,00	1,00	1,00	1,00	1,03
R <sup>2</sup> -Pseudo	0,79	0,83	0,85	0,85	0,84	0,84	0,84	0,84	0,84
R <sup>2</sup> hiệu chỉnh	0,79	0,83	0,85	0,85	0,84	0,84	0,84	0,84	0,84

Nguồn: Kết quả nghiên cứu

của các ngân hàng với mệnh giá lớn thường khá ổn định và ít biến động, cụ thể với bốn mã chứng khoán đang xét, các giá của các ngân hàng trong khoảng thời gian đang xét đều nằm trong mức  $\pm 3\sigma$ .

Trong bài báo này, chúng tôi xem xét bộ dữ liệu giá chứng khoán với mệnh giá lớn, đây là các bộ dữ liệu khá đẹp khi biến động giá thuộc khoảng  $\pm 3\sigma$  nên xem như không có giá trị đột biến, tính hữu ích của mô hình hồi quy phân vị chưa nhiều lắm, ngoài việc hiểu rõ hơn phân phối xác suất của biến phụ thuộc thông qua các mức phân vị. Tuy vậy, kết quả chỉ ra trong mô hình hồi quy phân vị thì các hệ số đều có ý nghĩa thống kê, không như mô hình hồi quy OLS có một số hệ số không có ý nghĩa thống kê. Đặc biệt, mô hình hồi quy phân vị hiệu quả trong trường hợp bộ dữ liệu có nhiều giá trị đột biến, đây cũng là ý tưởng cho các bài báo tiếp theo của chúng tôi.

## THẢO LUẬN: PHÂN PHỐI THỰC NGHIỆM CỦA CÁC THAM SỐ ƯỚC LƯỢNG

### Chuỗi VCB

Đồ thị rải điểm của dữ liệu VCB, tại thời điểm  $t$  so với thời điểm trước đó một ngày là thời điểm  $(t - 1)$  thấy rằng phụ thuộc khá phù hợp dạng tuyến tính. Đây

chính là lý do vì sao mô hình ước lượng dạng tuyến tính có độ phù hợp khá cao. Khi biểu diễn tất cả các đồ thị dạng phân vị của các giá trị, khá khó xem sự khác biệt giữa các mức phân vị thể hiện trong Hình 6:

### Chuỗi VPB

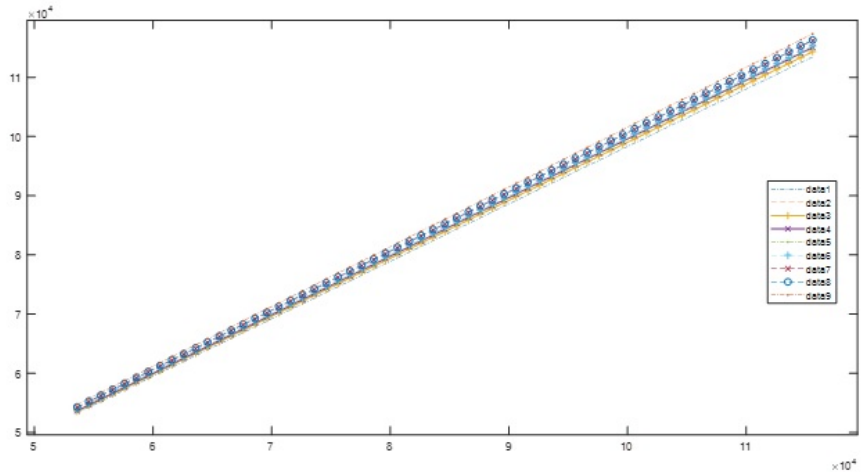
Đồ thị rải điểm VPB tại thời điểm  $t$  so với thời điểm trước đó một ngày là thời điểm  $(t - 1)$  thấy rằng phụ thuộc khá phù hợp dạng tuyến tính. Đây chính là lý do vì sao mô hình ước lượng dạng tuyến tính có độ phù hợp khá cao. Tuy nhiên, hàm hồi quy phân vị theo Hình 7 vẫn khó nhận biết sự khác biệt giữa các mức phân vị.

### Chuỗi TCB

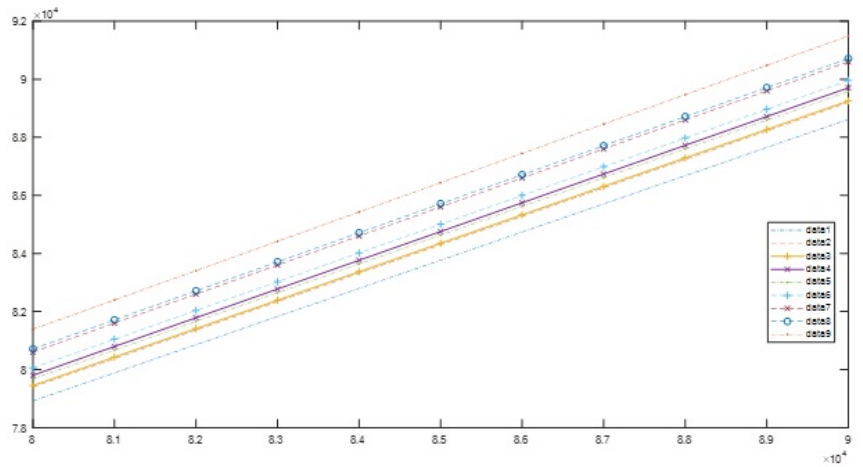
Đồ thị rải điểm TCB tại thời điểm  $t$  so với thời điểm trước đó một ngày là thời điểm  $(t - 1)$  thấy rằng phụ thuộc khá phù hợp dạng tuyến tính. Đây chính là lý do vì sao mô hình ước lượng dạng tuyến tính có độ phù hợp khá cao. Tuy nhiên, hàm hồi quy phân vị theo Hình 8 vẫn khó nhận biết sự khác biệt giữa các mức phân vị.

### Chuỗi BID

Đồ thị rải điểm BID tại thời điểm  $t$  so với thời điểm trước đó một ngày là thời điểm  $(t - 1)$  thấy rằng phụ



a. Toàn bộ dữ liệu



b. Thu hẹp một phần dữ liệu ở giữa từ 80000 đến 90000, để dễ quan sát

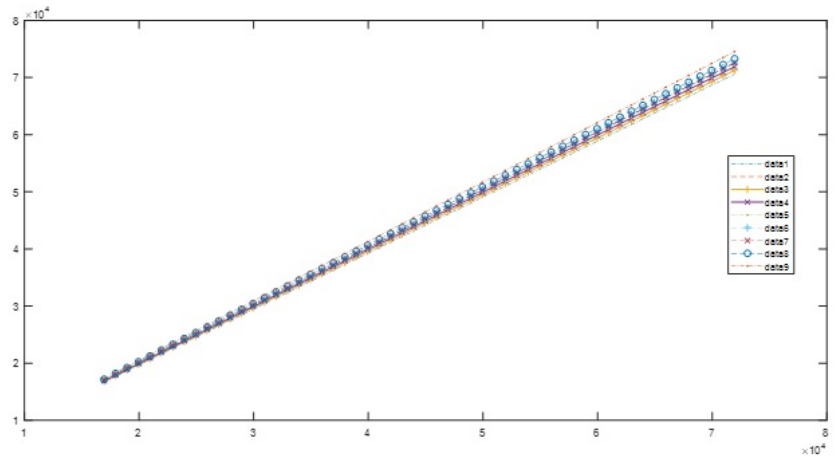
**Hình 6:** Đồ thị dạng phân vị tự hồi quy dữ liệu VCB, trong đó các data từ 1 tới 9, tương ứng 10% tới 90%, tức là data1 là mô hình hồi quy phân vị ở mức phân vị 10%, data2 là mô hình hồi quy phân vị ở mức 20%, đến data9 là mô hình hồi quy phân vị ở mức 90% (Nguồn: Kết quả nghiên cứu)

thuộc khá phù hợp dạng tuyến tính. Đây chính là lý do vì sao mô hình ước lượng dạng tuyến tính có độ phù hợp khá cao. Tuy nhiên, hàm hồi quy phân vị theo Hình 9 vẫn khó nhận biết sự khác biệt giữa các mức phân vị.

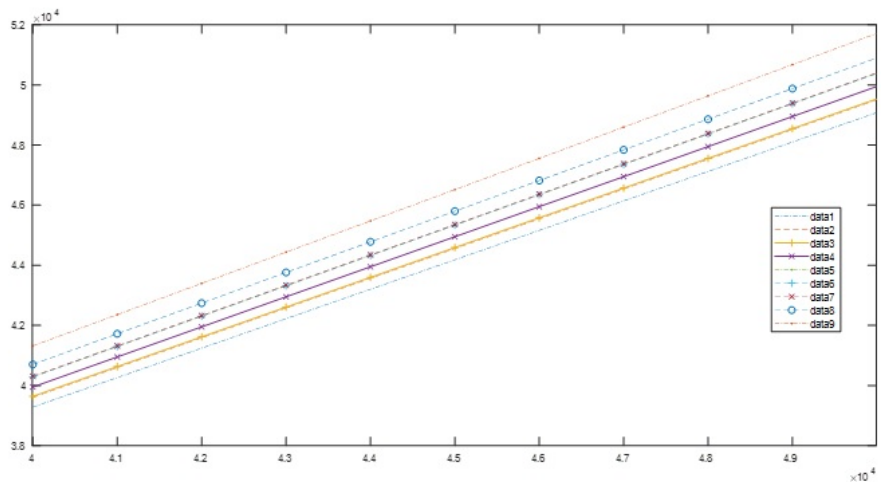
Trong đối sánh giữa mô hình hồi quy phân vị và mô hình hồi quy OLS, chúng ta có thể thấy mô hình hồi quy phân vị tốt hơn mô hình hồi quy OLS do cung cấp nhiều thông tin về biến phụ thuộc hơn mô hình hồi quy OLS. Cụ thể, mô hình hồi quy phân vị cung cấp hàm hồi quy tại các mức phân vị (tứ phân vị, thập

phân vị, hoặc ở bất cứ phân vị nào) của biến phụ thuộc thay vì chỉ duy nhất giá trị trung bình. Thông qua việc cung cấp nhiều thông tin hơn, giúp nhà phân tích có các đánh giá phù hợp, tránh trường hợp chỉ duy nhất giá trị trung bình, trong khi giá trị trung bình lại bị ảnh hưởng bởi các giá trị đột biến, cũng như trong một số trường hợp, các giá trị trung bình không thể xảy ra trong bộ dữ liệu.

Về các sai lệch của hai mô hình khác nhau, nếu như mô hình hồi quy OLS sử dụng sai lệch dạng bình phương thì hồi quy phân vị sử dụng sai lệch dạng trị



a. Toàn bộ dữ liệu



b. Một phần dữ liệu ở giữa, từ 40000 đến 50000 cho dễ quan sát

**Hình 7:** Đồ thị dạng phân vị tự hồi quy dữ liệu VPB, trong đó các data từ 1 tới 9, tương ứng 10% tới 90%, tức là data1 là mô hình hồi quy phân vị ở mức phân vị 10%, data2 là mô hình hồi quy phân vị ở mức 20%, đến data9 là mô hình hồi quy phân vị ở mức 90% (Nguồn: Kết quả nghiên cứu)

tuyệt đối. Do đó, chúng ta không nên so sánh độ phù hợp của hai dạng mô hình, chúng ta chỉ nên đánh giá tính có ý nghĩa thống kê của các hệ số hồi quy, và sự phù hợp với dữ liệu thông qua các độ phù hợp tương ứng nhận giá trị đủ tốt, thông thường là trên 80%.

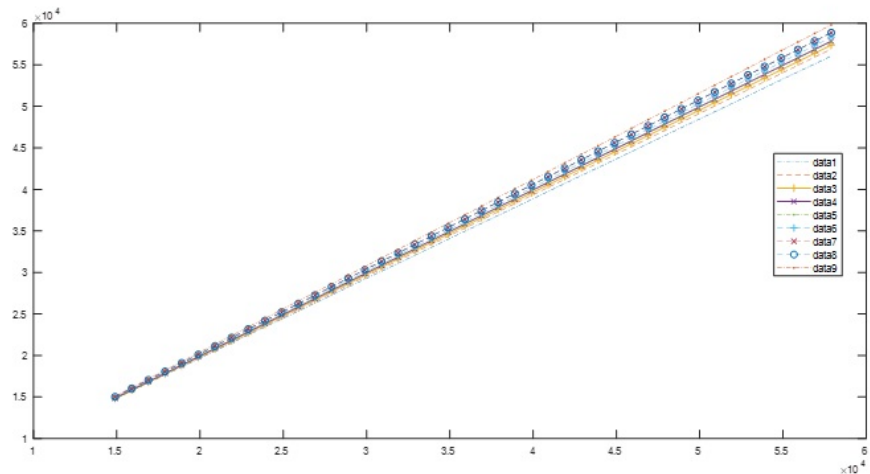
### KẾT LUẬN

Mô hình hồi quy nói chung và hồi quy phân vị nói riêng nhằm mô tả mối quan hệ giữa biến độc lập và biến phụ thuộc, trong đó biến phụ thuộc được giải thích qua các biến độc lập thể hiện ở tính phù hợp của mô hình hồi quy. Nếu như hồi quy tuyến tính cổ

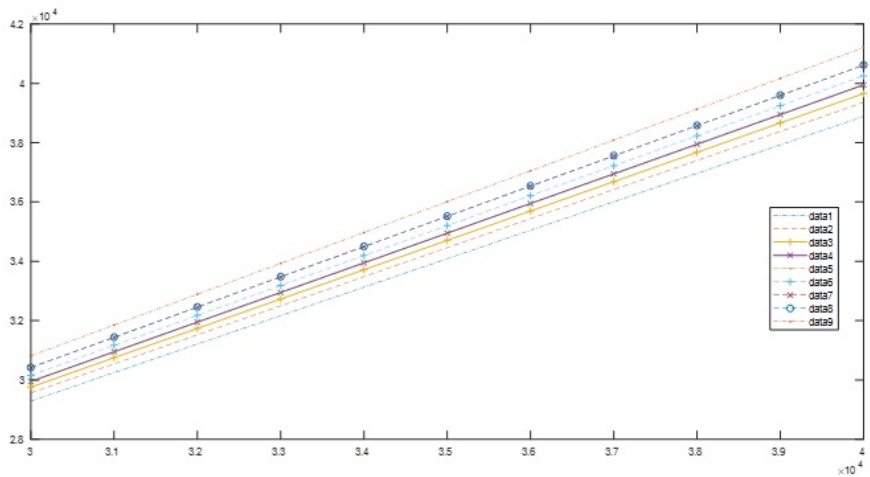
điển chỉ quan tâm đến giá trị trung bình của biến phụ thuộc thì mô hình hồi quy phân vị có ưu điểm hơn thông qua hiểu biết hơn về biến phụ thuộc thông qua các giá trị phân vị. Về tính đa dạng của hồi quy phân vị áp dụng được nhiều dạng dữ liệu như dữ liệu chéo, dữ liệu chuỗi thời gian và dữ liệu mảng.

Kết quả của nghiên cứu khi áp dụng dữ liệu chứng khoán theo thời gian của bốn ngân hàng lớn của Việt Nam là VCB, VPB, TCB và BID phù hợp với mô hình tự hồi quy phân vị. Các kết quả thực tế nghiên cứu chỉ ra mô hình này là phù hợp với dữ liệu về đồ thị rải điểm, tính có ý nghĩa của các tham số hồi quy cũng





a. Toàn bộ dữ liệu



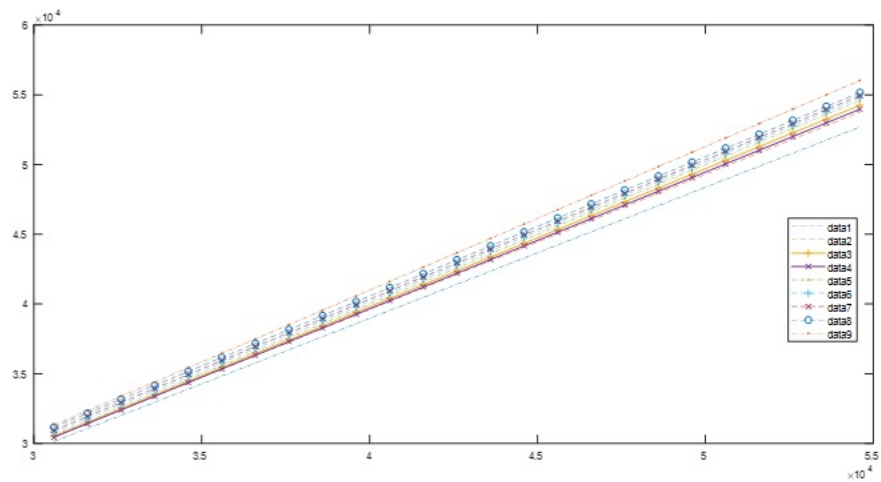
b. Một phần dữ liệu ở giữa, từ 30000 đến 40000 cho dễ quan sát

**Hình 8:** Đồ thị dạng phân vị tự hồi quy dữ liệu TCB, trong đó các data từ 1 tới 9, tương ứng 10% tới 90%, tức là data1 là mô hình hồi quy phân vị ở mức phân vị 10%, data2 là mô hình hồi quy phân vị ở mức 20%, đến data9 là mô hình hồi quy phân vị ở mức 90% (Nguồn: Kết quả nghiên cứu)

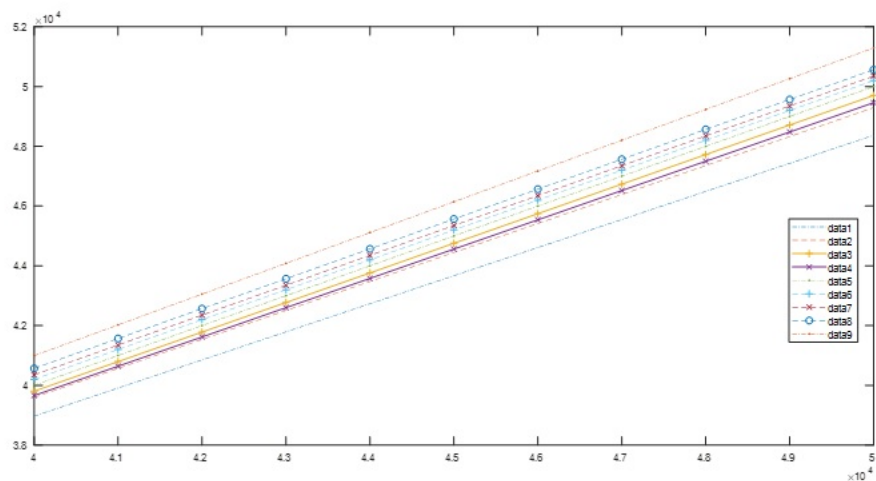
như phù hợp của mô hình. Tuy nhiên, các tham số ước lượng tương ứng với các mức phân vị của biến phụ thuộc không khác biệt nhiều đối với các bộ dữ liệu giá chứng khoán. Mặc dù vậy, các tham số ước lượng được đều có ý nghĩa thống kê so với mô hình hồi quy cổ điển OLS một số tham số không có ý nghĩa thống kê. Các kết quả minh chứng với dữ liệu thực nghiệm chỉ ra mô hình hồi quy phân vị cho chúng ta các kết quả chi tiết hơn về phân phối xác suất của biến phụ thuộc, thể hiện qua các mức phân vị của biến phụ

thuộc được biểu diễn qua các mô hình hồi quy riêng biệt. Hơn thế nữa, mô hình hồi quy phân vị không chịu ảnh hưởng bởi các giá trị đột biến như mô hình hồi quy OLS nên các kết quả hứa hẹn có thể ứng dụng trong mô hình hồi quy phân vị trong nhiều lĩnh vực không chỉ giá chứng khoán mà có thể giá vàng, tỷ giá, các chỉ số kinh tế vĩ mô cũng như các chỉ số kinh tế vi mô.

Trong thời gian tới, nhóm tác giả sẽ mở rộng nghiên cứu đối với các bộ dữ liệu về kinh tế vĩ mô, các kết



a. Toàn bộ dữ liệu



b. Một phần dữ liệu ở giữa, từ 40000 đến 50000 cho dễ quan sát

**Hình 9:** Đồ thị dạng phân vị tự hồi quy dữ liệu BID, trong đó các data từ 1 tới 9, tương ứng 10% tới 90%, tức là data1 là mô hình hồi quy phân vị ở mức phân vị 10%, data2 là mô hình hồi quy phân vị ở mức 20%, đến data9 là mô hình hồi quy phân vị ở mức 90% (Nguồn: Kết quả nghiên cứu)

quả nghiên cứu điều tra thực tiễn... Đồng thời nhóm cũng mở rộng về mô hình hồi quy Bayes so sánh về phân phối xác suất của biến độc lập với các thông tin tiên nghiệm được đánh giá một cách cụ thể thông qua các đánh giá mới về entropy, relative entropy...

### LỜI CẢM ƠN

Nghiên cứu được tài trợ bởi Đại học Quốc gia Thành phố Hồ Chí Minh (ĐHQG-HCM) trong khuôn khổ

Đề tài mã số B2021-34-03.

### XUNG ĐỘT LỢI ÍCH

Nhóm tác giả xin cam đoan rằng không có bất kỳ xung đột lợi ích nào trong công bố bài báo này.

### ĐÓNG GÓP CỦA CÁC TÁC GIẢ

Tác giả Lê Thanh Hoa chịu trách nhiệm về nội dung tổng quan bài báo, các phần kết quả ứng dụng của bài

báo.

Tác giả Võ Thị Lệ Uyên chịu trách nhiệm về các mô hình lý thuyết của bài báo.

Tác giả Nguyễn Phát Đạt chịu trách nhiệm về thu thập dữ liệu và chạy mô hình.

Tác giả Phạm Hoàng Uyên chịu trách nhiệm về ý tưởng bài báo, các phần mục nội dung.

## TÀI LIỆU THAM KHẢO

1. Hao L, Naiman DQ, Naiman DQ. Quantile regression. Vol. 149. SAGE; 2007 Apr 18; Available from: <https://doi.org/10.4135/9781412985550>.
2. Furno M, Vistocco D. Quantile regression: estimation and simulation. Vol. 216. John Wiley & Sons; 2018;.
3. Uribe JM, Guillen M. Quantile regression for cross-sectional and time series data: applications in energy markets using R. Springer. Nature. 2020; Available from: <https://doi.org/10.1007/978-3-030-44504-1>.
4. Davino C, Furno M, Vistocco D. Quantile regression: theory and applications. Vol. 988. John Wiley & Sons; 2013; PMID: 24361876. Available from: <https://doi.org/10.1002/9781118752685>.
5. McMillen DP. Quantile regression for spatial data. Springer Science+Business Media; 2012; Available from: <https://doi.org/10.1007/978-3-642-31815-3>.
6. Koenker R, Hallock KF. Quantile regression. J Econ Perspect. 2001;15(4):143-56. doi: 10.1257/jep.15.4.143; Available from: <https://doi.org/10.1257/jep.15.4.143>.
7. Yu K, Lu Z, Stander J. Quantile regression: applications and current research areas. J Royal Statistical Soc D. 2003;52(3):331-50; Available from: <https://doi.org/10.1111/1467-9884.00363>.
8. Koenker R. Quantile regression for longitudinal data. J Multivariate Anal. 2004;91(1):74-89; Available from: <https://doi.org/10.1016/j.jmva.2004.05.006>.
9. Lamarche C. Robust penalized quantile regression estimation for panel data. J Econ. 2010;157(2):396-408; Available from: <https://doi.org/10.1016/j.jeconom.2010.03.042>.
10. Canay IA. A simple approach to quantile regression for panel data. Econ J. 2011;14(3):368-86; Available from: <https://doi.org/10.1111/j.1368-423X.2011.00349.x>.
11. Gu J, Volgushev S. Panel data quantile regression with grouped fixed effects. J Econ. 2019;213(1):68-91; Available from: <https://doi.org/10.1016/j.jeconom.2019.04.006>.
12. Van HH, Tran TQ. International trade and employment: A quantile regression approach. J Econ Integr. 2017;531-57; Available from: <https://doi.org/10.11130/jei.2017.32.3.531>.
13. Anser MK, Ahmad M, Khan MA, Zaman K, Nassani AA, Askar SE et al. The role of information and communication technologies in mitigating carbon emissions: evidence from panel quantile regression. Environ Sci Pollut Res Int. 2021;28(17):21065-84; PMID: 33405124. Available from: <https://doi.org/10.1007/s11356-020-12114-y>.
14. Anh TTT. Các yếu tố tác động đến chính sách cổ tức của doanh nghiệp Việt Nam: Tiếp cận bằng hồi quy phân vị. Tạp Chí Phát Triển Kinh Tế. 2016; (JED, Vol. 27 (2)):108-27;.
15. Koenker R, Machado JAF. Goodness of fit and related inference processes for quantile regression. J Am Stat Assoc. 1999;94(448):1296-310; Available from: <https://doi.org/10.1080/01621459.1999.10473882>.

# Quantile regression with time-series data and applications in modeling of some big bank stock markets

Le Thanh Hoa\*, Vo Thi Le Uyen, Nguyen Phat Dat, Pham Hoang Uyen



Use your smartphone to scan this QR code and download this article

## ABSTRACT

In the OLS regression model, the mean of the dependent variable is estimated based on the mean of the independent variables. The relationship between the independent variable and the dependent variable needs to be considered in many values instead of just through the mean of the dependent variable, then the quantile regression model is the optimal choice. In this paper, we study a quantile regression model in which the percentiles of the dependent variable are spread from 10% to 90%, with step by 10%, of the time-series data through autoregression model, to compare the fit of the models as well as the errors of the model. To demonstrate the results, we applied it to the time-series data on the closing stock prices of the four largest bank codes (on August 2, 2021) namely VCB, VPB, TCB, and BID. In which, percentile regression models have a very high fit, corresponding to about 80% or more. Besides, the estimated parameters are all statistically significant, different from the OLS regression model where some estimated parameters are not statistically significant. Furthermore, the OLS regression model is based on the mean, so it is susceptible to the outlier values, while the quantile is not affected by the outliers. In other words, the quantile regression model overcomes this weakness, so the quantile regression model is not affected by the outliers.

**Key words:** Quantile regression, Quantile regression with time-series data, stock market model

University of Economics and Law –  
VNU-HCM, Vietnam

## Correspondence

**Le Thanh Hoa**, University of Economics  
and Law – VNU-HCM, Vietnam

Email: hoalt@uel.edu.vn

## History

- Received: 16-9-2021
- Accepted: 16-8-2022
- Published: 30-9-2022

DOI : 10.32508/stdjelm.v6i3.948



## Copyright

© VNUHCM Press. This is an open-access article distributed under the terms of the Creative Commons Attribution 4.0 International license.



**Cite this article :** Hoa L T, Uyen V T L, Dat N P, Uyen P H. **Quantile regression with time-series data and applications in modeling of some big bank stock markets.** *Sci. Tech. Dev. J. - Eco. Law Manag.*; 2022, 6(3):3015-3032.