Tạp chí Khoa học Trường Đại học Cần Thơ
Phần A: Khoa học Tự nhiên, Công nghệ và Môi trường
website: sj.ctu.edu.vn

DOI:10.22144/ctu.jvn.2019.158

SỬ DỤNG MÔ HÌNH MARKOV ẨN ĐỂ PHÂN TÍCH SỰ CHUYỂN ĐỔI TRẠNG THÁI NGẪU NHIÊN CỦA QUÁ TRÌNH GIÁ CỔ PHIẾU

Trần Văn Lý[1*], Đặng Hoàng Tâm[1], Lê Thị Mỹ Xuân[1], Nguyễn Thị Tú Anh[2] và Trần Văn Trọng[3]

[1]*Khoa Khoa học Tự nhiên, Trường Đại học Cần Thơ*

[2]*Lớp Toán ứng dụng - khóa 41, Trường Đại học Cần Thơ*

[3]*Trường Cao Đẳng Kinh Tế - Kỹ Thuật Cần Thơ*

**Người chịu trách nhiệm về bài viết: Trần Văn Lý (email: tvly@ctu.edu.vn)*

Thông tin chung:
Ngày nhận bài: 24/09/2019
Ngày nhận bài sửa: 11/10/2019
Ngày duyệt đăng: 25/12/2019

Title:
Using the hidden Markov model to analyze the random state transition of stock price process

Từ khóa:
Mô hình Markov ẩn, quá trình giá cổ phiếu, thuật toán cực đại hoá kỳ vọng, ước lượng hợp lý tối đa, ước lượng lọc, xích Markov

Keywords:
Expectation maximization algorithm, filter estimation, hidden Markov model, Markov chain, maximum likelihood estimation, stock price process

ABSTRACT

The random state transition of stock price process will be modeled by a hidden Markov model. This model based on a pair (X_h,K_h) of stochastic processes. The process (X_h,) is called the state process, a hidden Markov chain, which represents the random state transition of stock price process. The observation process (K_h) represents the measured sequence of stock price. The parameters of the model were estimated from the real data using maximum likelihood estimation via expectation-maximization algorithm (EM algorithm). Forecast for stock price sequence using simulation data is the proposed application.

TÓM TẮT

Sự chuyển đổi trạng thái ngẫu nhiên của quá trình giá cổ phiếu sẽ được mô hình hóa bởi mô hình Markov ẩn. Đó là một cặp gồm hai quá trình ngẫu nhiên (X_h,K_h). Quá trình (X_h) được gọi là quá trình trạng thái, là một xích Markov ẩn không quan sát được, biểu diễn cho chuỗi thay đổi các trạng thái của giá cổ phiếu. Quá trình (K_h) được gọi là quá trình quan sát, đo cho chuỗi giá cổ phiếu quan sát được. Các tham số của mô hình sẽ được ước lượng từ dữ liệu thực nhờ vào giải thuật cực đại hóa kỳ vọng EM (expectation maximization algorithm). Dự báo cho chuỗi giá cổ phiếu bằng dữ liệu mô phỏng là ứng dụng được đề xuất.

Trích dẫn: Trần Văn Lý, Đặng Hoàng Tâm, Lê Thị Mỹ Xuân, Nguyễn Thị Tú Anh và Trần Văn Trọng, 2019. Sử dụng mô hình Markov ẩn để phân tích sự chuyển đổi trạng thái ngẫu nhiên của quá trình giá cổ phiếu. Tạp chí Khoa học Trường Đại học Cần Thơ. 55(6A): 51-56.

1 GIỚI THIỆU

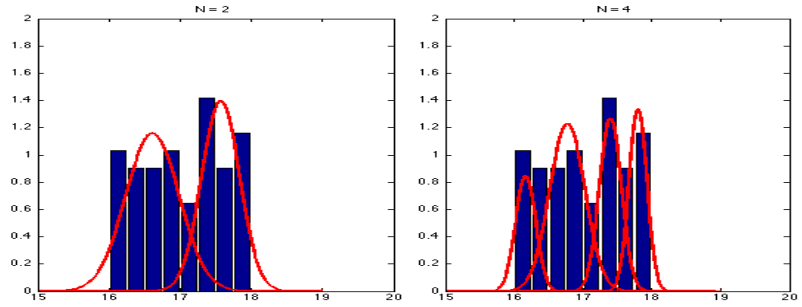
Biến động của thị trường chứng khoán nói chung hay một mã cổ phiếu nào đó luôn là sự quan tâm hàng đầu của các nhà đầu tư cũng như những người nghiên cứu tài chính. Tính không chắc chắn, sự rủi ro, tính biến động lên xuống một cách ngẫu nhiên là thuộc tính cố hữu của thị trường chứng khoán. Qua việc phân tích giá cả trong quá khứ và hiện tại trên thị trường tài chính có thể xác định xu hướng giá với xác suất cao nhất trong tương lai. Các mô hình đã được sử dụng để mô phỏng động thái giá cổ phiếu được biết đến như mô hình chuyển động Brown kinh tế hay Brown hình học (GMB), các mô hình chuỗi thời gian (mô hình trung bình trượt tự hồi quy,…), hồi quy xu thế tuyến tính,… Trong phạm vi bài báo này, mô hình Markov ẩn được sử dụng để thực hiện mô hình hoá

động thái ngẫu nhiên của chuỗi giá cổ phiếu. Đề xuất này được xem xét với giả thuyết là các mức giá cổ phiếu thuộc vào các nhóm phân phối Gauss và sự biến đổi trạng thái giữa các nhóm này có tính chất của một xích Markov. Với đề xuất này, diễn biến quan sát được của chuỗi giá cổ phiếu sẽ được sử dụng để phân lớp các mức giá và xem xét sự biến đổi ngẫu nhiên giữa các lớp này theo dạng một xích Markov.

Dựa vào việc xem xét mối quan hệ giữa dữ liệu đầy đủ và dữ liệu không đầy đủ, giải thuật cực đại hóa kỳ vọng EM (Dempster *et al.*, 1997; Elliott *et al.*, 2010) sẽ được áp dụng để ước lượng các tham số của mô hình.

Số liệu sử dụng để ước lượng minh hoạ số là chuỗi giá đóng cửa của cổ phiếu CTG (Ngân hàng TMCP Công thương Việt Nam) trong quí II năm 2019 (từ 01/4/2019 đến 30/6/2019). Các xử lý dữ liệu thống kê được thực hiện trên phần mềm Matlab.

Các khái niệm liên quan đến thuật ngữ chuỗi giá cổ phiếu được sử dụng theo cách tiếp cận trong lĩnh vực Toán tài chính hiện đại (Trần Hùng Thao, 2013), kỹ thuật và kết quả tính toán lý thuyết ước lượng được thực hiện kỹ thuật đổi độ đo và ước lượng lọc *forward* (Ly, 2016).



Hình 1: Biểu đồ chuỗi giá cổ phiếu quí II năm 2019

2 MÔ HÌNH

Sự phân bố thực nghiệm của một chuỗi giá cổ phiếu trong một thời gian cho thấy nó có thể là một phân phối hỗn hợp Gauss (ví dụ, xem biểu đồ của chuỗi giá cổ phiếu đóng cửa của CTG trong quí II năm 2019 được thể hiện trong Hình 1), mỗi phân phối thành phần Gauss tương ứng với một nhóm mức giá nào đó. Điều này dẫn đến ý tưởng mô hình hóa động thái của chuỗi giá cổ phiếu bằng một mô hình Markov ẩn với thời gian rời rạc, trong đó:

(i) Quá trình trạng thái không quan sát được là một chuỗi Markov thể hiện cho các nhóm giá khác nhau.

(ii) Quá trình quan sát miêu tả cho chuỗi giá cổ phiếu, thuộc vào các nhóm khác nhau, tạo ra bởi các phân phối Gauss với trung bình μ_i và độ lệch chuẩn σ_i ứng với mỗi trạng thái $i,\ i=1,2,...,N.$

2.1 Quá trình trạng thái

Chúng tôi giả định rằng có $N\geq 1$ các trạng thái, trạng thái i được đại diện bởi e_i, vector cột đơn vị của R^n với giá trị 1 tại vị trí $i,\ i=1,2,...,N.$

Các động thái ngẫu nhiên của chuỗi giá cổ phiếu sẽ được mô hình hóa bởi một xích Markov thuần nhất không quan sát được với thời gian rời rạc $(X_h)_{h=0,1,2,...}$, được gọi là quá trình trạng thái, với không gian trạng thái $S=\{e_1,e_2,...,e_n\}$ và ma trận xác suất chuyển $A=(a_{ji})$, trong đó $a_{ji}=P(X_{h+1}=e_j \mid X_h=e_i),\ i,j=1,2,...,N.$ Lưu ý là $a_{ii}=1-\sum_{j\neq i}a_{ji},\ \ i=1,2,...,N.$

Chúng tôi giả định rằng sự phân bố của X_0 là π_0 đã biết.

2.2 Quá trình quan sát và tập tham số

Các giá trị ngẫu nhiên của chuỗi giá cổ phiếu (K_h) được mô hình hóa bởi quá trình quan sát được hình thành như một hàm phụ thuộc vào quá trình trạng thái. Tại mức giá i, khi mà các xích Markov thuộc vào trạng thái e_i $(i=1,2,...,N)$, giá cổ phiếu (K_h) khi đó được xem là thuộc phân phối Gauss $N(\mu_i,\sigma_i^2)$:

$$1_{(X_h=e_i)}K_h=1_{(X_h=e_i)}(\mu_i+\sigma_i w_h),\quad h=1,2,...$$

Trong đó, w_h là các biến ngẫu nhiên độc lập có phân phối chuẩn $N(0,1)$ và μ_i, σ_i là các tham số cần ước lượng. Từ đó, mô hình đề xuất cho chuỗi giá cổ phiếu dưới sự ảnh hưởng ngẫu nhiên của những yếu tố thị trường sẽ là một mô hình Markov ẩn (hidden Markov model) với quá trình trạng thái (X_h) và quá trình quan sát (K_h) được định nghĩa bởi:

$$K_h = \sum_{i=1}^{N} 1_{(X_h = e_i)} K_h = \sum_{i=1}^{N} 1_{(X_h = e_i)} (\mu_i + \sigma_i w_h), \quad h = 1,2,... \quad (3)$$

Ký hiệu "'" biểu thị cho phép chuyển vị, đặt: $\mu = (\mu_1, \mu_2, ..., \mu_N)'$, $\sigma = (\sigma_1, \sigma_2, ..., \sigma_N)'$.

Ta nhận được phương trình quan sát ở dạng $K_h = \langle X_h, m \rangle + \langle X_h, s \rangle w_h$, trong đó $\langle .,. \rangle$ ký hiệu cho tích vô hướng trong.

Tập tham số xác định mô hình là $\theta = \{a_{ji}, 1 \le i \ne j \le N; \mu_1, \mu_2, ..., \mu_N; \sigma_1, \sigma_2, ..., \sigma_N\}$.

2.3 Một số khái niệm

Để ước lượng các tham số của mô hình, một số khái niệm cần thiết được giới thiệu dưới đây:

(i) Số bước nhảy của quá trình trạng thái chuyển từ e_i đến e_j : $J_h^{ij} = \sum_{l=1}^{h} \langle X_{l-1}, e_i \rangle \langle X_l, e_j \rangle$.

(ii) Thời gian lưu trú của quá trình X ở trạng thái $e_i : O_h^i = \sum_{l=1}^{h} \langle X_{l-1}, e_i \rangle$.

(iii) Mức tổng của quá trình quan sát ở trạng thái $e_i : T_h^i(g) = \sum_{l=1}^{h} g(K_l) \langle X_{l-1}, e_i \rangle$.

(iv) Lọc dữ liệu không đầy đủ: $K_h = \sigma\{K_1, K_2, ..., K_h\}$.

(v) Lọc dữ liệu đầy đủ: $G_h = \sigma\{X_0, X_1, X_2, ..., X_h, K_1, K_2, ..., K_h\}$.

(vi) Với $H_h \equiv J_h^{ij}$, O_h^i hoặc $T_h^i(g)$, lọc chuẩn hóa của quá trình H là: $\pi(H_h) = E(H_h \mid K_h)$.

Trong đó, $\sigma\{U\}$ thể hiện σ sinh bởi tập U và $g(K_l) = K_l$ hoặc $g(K_l) = K_l^2$.

3 ƯỚC LƯỢNG THAM SỐ

Trong phần này, chúng tôi trình bày các kết quả dùng để cập nhật ước lượng hợp lý tối đa cho các tham số sử dụng trong mỗi bước lặp của thuật toán EM.

3.1 Thuật toán EM

Ở đây chúng tôi muốn xác định một bộ tham số mới $\hat{\theta}$ làm cực đại hoá *hàm giả hợp lý* dữ liệu đầy đủ thông qua giải thuật EM.

Hàm giả hợp lý được định nghĩa bởi: $Q(\hat{\theta}, \theta) = E_\theta(\log \Lambda_h^{\theta\hat{\theta}} \mid K_h)$, (4)

Trong đó, $\Lambda_h^{\theta\hat{\theta}} = \frac{dP_{\hat{\theta}}}{dP_\theta} \mid G_h$ và P_θ là độ đo xác xuất phụ thuộc vào θ.

Khởi đầu từ một giá trị ban đầu $\hat{\theta}^{(0)}$, các phép lặp củathuật toán EM sẽ tạo ra một chuỗi $\{\hat{\theta}^{(p)}, p \ge 1\}$ ước lượng cho θ.

Mỗi lần lặp của thuật toán EM gồm hai bước sau:

Bước E (Kỳ vọng): Đặt $\theta = \hat{\theta}^{(p)}$ và tính $Q(\hat{\theta}, \hat{\theta}^{(p)}) = E_{\hat{\theta}^{(p)}}(\log \Lambda_h^{\hat{\theta}^{(p)}\hat{\theta}} \mid K_h)$.

Bước M (Tối đa hóa): Tìm $\hat{\theta}^{(p+1)} = \arg\max_{\hat{\theta} \in \Theta} Q(\hat{\theta}, \hat{\theta}^{(p)})$, lặp lại từ **Bước E** với $p = p + 1$ mãi cho tới khi nào kiểm tra thỏa mãn thì dừng.

3.2 Cập nhật tham số

Trong mỗi lần lặp của thuật toán EM, các xác suất chuyển a_{ji} được cập nhật như sau: (Elliott *et al.*, 2010): $\hat{a}_{ji} = \frac{\pi(J_h^{ij})}{\pi(O_h^{ij})}$, $1 \le i \ne j \le N$, (5)

Trong đó, $\pi(J_h^{ij})$ và $\pi(O_h^{ij})$ lần lượt là những quá trình lọc chuẩn hóa của số bước nhảy và thời gian lưu trú.

Bây giờ chúng ta xem xét sự cập nhật cho μ và σ lần lượt bởi $\hat{\mu}$ và $\hat{\sigma}$.

Ta có

$$\Lambda_h^{\theta\hat{\theta}} = \prod_{l=1}^{h} \frac{\langle X_l, \sigma \rangle \varphi\left(\frac{K_l - \langle X_l, \hat{\mu} \rangle}{\langle X_l, \hat{\sigma} \rangle}\right)}{\langle X_l, \hat{\sigma} \rangle \varphi\left(\frac{K_l - \langle X_l, \mu \rangle}{\langle X_l, \sigma \rangle}\right)}, \quad (6)$$

Ký hiệu $\varphi(.)$ là hàm mật độ của phân phối chuẩn $N(0,1)$.

Từ (4) và (6) ta có:

$$Q(\hat{\theta},\theta)=E_\theta\left\{\sum_{l=1}^{h}\left[\log\frac{1}{\langle X_l,\sigma\rangle}-\frac{1}{2}\left(\frac{K_l-\langle X_l,\hat{\mu}\rangle}{\langle X_l,\hat{\sigma}\rangle}\right)^2\right]\mid K_h\right\} \quad (7)$$
$$+R(\theta,K_h),$$

$$Q(\hat{\theta},\hat{\theta}^p)=\sum_{i=1}^{N}E_\theta\left\{\sum_{l=1}^{h}\left[\langle X_l,e_i\rangle\log\frac{1}{\hat{\sigma}_i}-\frac{1}{2}\frac{\langle X_l,e_i\rangle}{\hat{\sigma}_i^2}(K_l-\mu_i)^2\right]\mid K_h\right\}+R(\hat{\theta}^p,K_h)$$
$$=\sum_{i=1}^{N}\left\{\pi(O_h^i)\log\frac{1}{\hat{\sigma}_i}-\frac{1}{2\hat{\sigma}_i^2}\left[\pi(T_h^i(K_h^2))-2\hat{\mu}_i\pi(T_h^i(K_h))+\hat{\mu}_i^2\pi(O_h^i)\right]\right\}+R(\hat{\theta}^p,K_h).$$

Bước M: Ta tìm $\hat{\theta}^{(p+1)}=\arg\max_{\theta\in\Theta}Q(\hat{\theta},\hat{\theta}^p)$:

Với $\hat{\mu}_i, i=1,2,...,N,$ lấy đạo hàm của $Q(\hat{\theta},\hat{\theta}^p)$ ta có

$$\frac{\partial}{\partial\hat{\mu}_i}Q(\hat{\theta},\hat{\theta}^p)=-\frac{1}{2\hat{\sigma}_i^2}\left[-2\pi(T_h^i(K_h))+2\hat{\mu}_i\pi(O_h^i)\right]$$

Với $\frac{\partial}{\partial\hat{\mu}_i}Q(\hat{\theta},\hat{\theta}^p)=0$ ta có $\hat{\mu}_i=\frac{\pi(T_h^i(K_h))}{\pi(O_h^i)}.$ (8)

Tương tự, với $i=1,2,...,N,\frac{\partial}{\partial\hat{\mu}_i}Q(\hat{\theta},\hat{\theta}^p)=0$ ta có

$$\hat{\sigma}_i^2=\frac{1}{\pi(O_h^i)}\left[\pi(T_h^i(K_h^2))-2\hat{\mu}_i\pi(T_h^i(K_h))+2\hat{\mu}_i\pi(O_h^i)\right]. \quad (9)$$

Trong đó, hàm $R(\theta,K_h)$ không phụ thuộc vào $\hat{\theta}$.

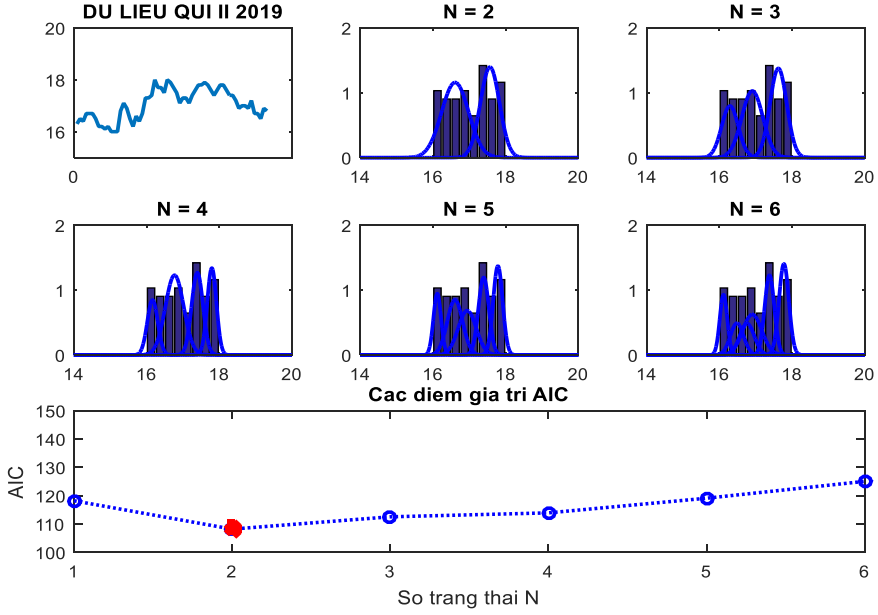
Bước E: Đặt $\theta=\hat{\theta}^{(p)}$ và viết lại (7) như sau

3.3 Thực nghiệm với dữ liệu thực

Từ (5), (8) và (9), các tham số mô hình sẽ được ước lượng từ các dữ liệu thực thông qua giải thuật EM. Số trạng thái N của mô hình được chọn phù hợp với dữ liệu bởi tiêu chuẩn thông tin Akaike (AIC).

3.3.1 Dữ liệu thực nghiệm

Để áp dụng minh họa số, chuỗi giá cổ phiếu CTG (giá đóng cửa, đơn vị tính là nghìn đồng) của Ngân hàng TMCP Công thương Việt Nam trên sàn HOSE trong quí II năm 2019 (đồ thị và tần số của dữ liệu được biểu diễn ở các hàng 1 và 2 của Hình 2) sẽ được sử dụng. Nguồn số liệu được lấy từ *www.cophieu68.vn*.



Hình 2: Các biểu đồ quan sát sự phù hợp của phân bố tần số của dữ liệu với số trạng thái N của mô hình (hàng 1 và 2) và biểu diễn so sánh đánh giá tiêu chuẩn AIC theo N (hàng 3)

Quan sát biểu đồ các giá trị quan sát và đánh giá tiêu chuẩn AIC (lựa chọn mô hình có giá trị AIC nhỏ nhất, hàng 3 của Hình 2) của chuỗi giá cổ phiếu, mô hình có số trạng thái $N=2$ sẽ được chọn để áp dụng minh hoạ.

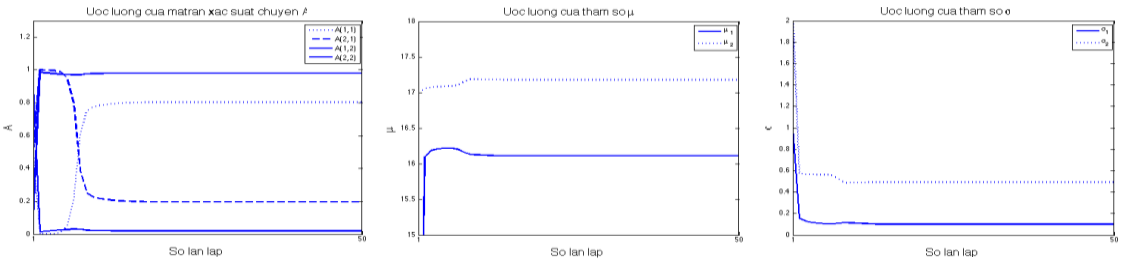
3.3.2 Ước lượng tham số mô hình

Với số trạng thái $N=2$, mô hình được xác định bởi tập tham số $\theta=\{A,\mu,\sigma\}$, trong đó $\mu=(\mu_1,\mu_2)'$, $\sigma=(\sigma_1,\sigma_2)'$ và ma trận xác suất chuyển $A=(a_{11}\quad a_{12};\quad a_{21}\quad a_{22})$.

Các giá trị ban đầu được chọn là: $\mu=(10\quad 17)'$, $\sigma=(1\quad 2)'$, $A=(0.5\quad 0.9;\quad 0.5\quad 0.1)$.

Sau 50 bước lặp của thuật toán EM, chúng tôi thu được kết quả: $\mu=(16.1113\quad 17.1788)'$, $\sigma=(0.1011\quad 0.4925)'$ và $A=(0.8027\quad 0.0204;\quad 0.1973\quad 0.9796)$.

Tiến trình ước lượng của những tham số được biểu diễn ở Hình 3.



Hình 3: Tiến trình ước lượng của ma trận A và các vector μ, σ

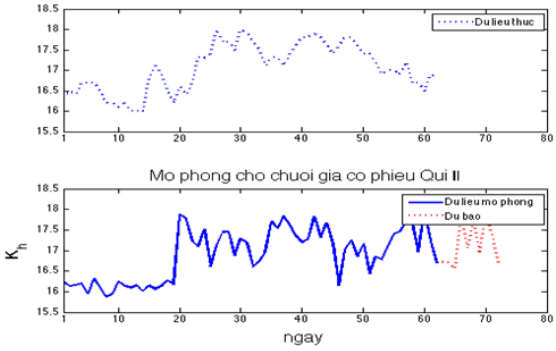
3.3.3 Dự báo

Sử dụng mô hình với các tham số ước lượng được chúng tôi thực hiện các mô phỏng cho dữ liệu đã áp dụng (Hình 4) và tiến hành dự báo giá đóng cửa của CTG cho 10 ngày giao dịch tiếp theo của tháng 7 năm 2019, kết quả được liệt kê ở Bảng 1.

Bảng 1: Dự báo giá đóng cửa của CTG dựa trên mô hình với các tham số ước lượng

STT	Giá dự báo	Giá thực	Sai số tuyệt đối
1	16.7144	17.0	0.29
2	16.7080	17.1	0.39
3	16.5452	17.4	0.85
4	17.8170	17.2	0.62
5	17.0494	17.3	0.25
6	17.6775	17.2	0.48
7	16.9259	17.4	0.47
8	17.8824	17.7	0.18
9	17.3498	18.9	1.55
10	16.7136	18.6	1.89

Các chỉ tiêu đánh giá sai số dự báo như sai số trung bình, sai số phần trăm tuyệt đối trung bình hay sai số bình phương trung bình,… đều khá thấp. Đặc biệt hệ số không ngang bằng Theil nhận được là *U=0.51*, chứng tỏ mô hình đề xuất là khá tốt (lưu ý là trong thực tế nếu $U\le 0.55$, thì các giá trị dự báo được đánh giá là tốt).



Hình 4: Một dãy mô phỏng (dưới) cho chuỗi giá cổ phiếu quí II năm 2019 (trên)

4 KẾT LUẬN

Bài báo giới thiệu một cách tiếp cận mới để mô hình hoá động thái ngẫu nhiên của quá trình giá cổ phiếu, đóng góp thêm một công cụ có ích trong việc nghiên cứu tài chính. Ngoài ứng dụng dự báo, mô hình đề xuất có thể sử dụng để phân lớp các mức giá của các cổ phiếu hoặc có thể sử dụng dữ liệu mô phỏng tạo ra từ mô hình để nghiên cứu phân bố của chuỗi giá cổ phiếu trong tuần, trong tháng, trong quí hay xem xét so sánh phân bố của các biến động đầu tuần, cuối tuần,…. Để phát triển thêm từ hướng nghiên cứu này, mô hình Markov ẩn có thể được kết hợp với mô hình chuyển động Brown kinh tế để mô hình hoá quá trình giá cổ phiếu hay quá trình tài chính theo thời gian liên tục. Về kỹ thuật tính toán ước lượng tham số của mô hình, ước lượng làm trơn (smoothed) có thể sử dụng bằng cách kết hợp ước lượng ước lượng tiến

(forward) hay còn gọi là ước lượng lọc (filtered, đã sử dụng trong bài báo) với ước lượng lùi (backward) để nâng cao chất lượng của mô hình.

TÀI LIỆU THAM KHẢO

Trần Hùng Thao, 2013. Toán tài chính căn bản, Nhà xuất bản Văn hóa thông tin, 386 trang.

Ly, T.V., 2016. Stochastic modeling for daily clearness index sequence in Can Tho city. Can Tho University Journal of Science, 90-99.

Dempster, A.P., Laird, N.M. and Rubin, D.B., 1997. Maximum Likelihood from Incomplete Data via the EM Algorithm, Journal of the Royal Statistical Society, Series B (Methodological), 39(1): 1-38.

Elliott, J.R., Aggoun, L. and Moore, J.B., 2010. Hidden Markov Models: Estimation and control, Springer, 377 pages.