Tạp chí Nghiên cứu Kinh tế và Kinh doanh Châu Á
Năm thứ 29, Số 9 (2018), 53–72

www.jabes.ueh.edu.vn

JABES
Tạp chí Nghiên cứu Kinh tế và Kinh doanh Châu Á

Tạp chí Nghiên cứu Kinh tế và Kinh doanh Châu Á
http://www.emeraldgrouppublishing.com/services/publishing/jabes/index.htm

Ước lượng VaR và CVaR dạng hỗn hợp các phân phối xác suất thông qua mô phỏng Monte Carlo và ứng dụng trong phân tích giá chứng khoán Việt Nam

LÊ THANH HOA [*]

Trường Đại học Kinh tế - Luật, Đại học Quốc gia TP. Hồ Chí Minh

Trường Đại học Khoa học Tự nhiên, Đại học Quốc gia TP. Hồ Chí Minh

THÔNG TIN

Ngày nhận: 07/07/2018
Ngày nhận lại: 04/11/2018
Duyệt đăng: 05/11/2018

Mã phân loại JEL:

C02, C15, C63

Từ khóa:
Ước lượng VaR;
Ước lượng CVaR;
Phương pháp mô phỏng Monte Carlo;
Hỗn hợp các phân phối xác suất;
Giá chứng khoán.

TÓM TẮT

VaR và CVaR là các vấn đề thu hút nhiều quan tâm trong lĩnh vực nghiên cứu đo lường rủi ro đối với dữ liệu tài chính. Tuy nhiên, hầu hết các nghiên cứu chỉ dừng lại ở công thức tính dựa trên xấp xỉ dữ liệu thông qua một phân phối xác suất trong khi phân phối của dữ liệu thực thông thường có dạng hỗn hợp các phân phối xác suất. Chính vì vậy, cần thiết phải có các nghiên cứu tính toán VaR và CVaR trong trường hợp phân phối của dữ liệu có dạng hỗn hợp các phân phối xác suất. Hơn nữa, vì tính toán cũng đòi hỏi thời gian tính toán nhanh nên tác giả ưu tiên sử dụng phương pháp mô phỏng Monte Carlo. Tác giả đề nghị thuật toán tính giá trị VaR và CVaR trong trường hợp phân phối xác suất là dạng hỗn hợp bằng cách hiệu chỉnh phương pháp mô phỏng Monte Carlo cho phù hợp với dạng hỗn hợp các phân phối xác suất. Thêm vào đó, tác giả minh họa kết quả tính toán trong các trường hợp mô phỏng đã biết phân phối xác suất thành phần cũng như trường hợp chưa biết trước các phân phối xác suất thành phần, thông qua bộ dữ liệu thực về giá đóng cửa của chứng khoán Việt Nam bằng hỗn hợp của các phân phối. Các kết quả tính toán dựa trên thuật toán mà tác giả đề nghị được đánh giá thông qua xác suất đuôi cũng như độ lệch chuẩn.

[*] Tác giả liên hệ.
Email: hoalt@uel.edu.vn
Trích dẫn bài viết: Lê Thanh Hoa. (2018). Ước lượng VaR và CVaR dạng hỗn hợp các phân phối xác suất thông qua mô phỏng Monte Carlo và ứng dụng trong phân tích giá chứng khoán Việt Nam. *Tạp chí Nghiên cứu Kinh tế và Kinh doanh Châu Á, 29*(9), 53–72.

Keywords:

Value at Risk (VaR) estimation;
Conditional Value at Risk (CVaR) estimation;
Monte Carlo simulation method;
Mixture of probability distributions;
Stock price.

Abstract

VaR and CVaR are of great concern in risk measurement field applying for financial data. However, studies only stopped at the formula based on an approximation of data through a probability distribution, while the distribution of real data is usually a mixture of probability distributions. Therefore, it is necessary to study the VaR and CVaR estimating in the case of a mixture of probability distributions. Furthermore, the estimations also require short computational time so the author prefers to use the Monte Carlo simulation method. The author proposes the VaR and CVaR calculations in the case of probability distributions that are mixed by calibrating the Monte Carlo simulation method to suit a mixture of probability distributions. In addition, the author illustrates in cases where the simulation is known to distribute the probability of the component as well as the unknown case of the probability distribution of the component via the real data of the closing price of the Vietnam stock by a mixture of probability distributions. The algorithm-based estimation results that the author recommends is evaluated by the tail probability as well as standard deviation.

1. Giới thiệu

Value at Risk, ký hiệu là VaR (hay còn gọi là phân vị – Quantile) (Rockafellar & cộng sự, 2014) là một kỹ thuật được sử dụng rộng rãi trong lĩnh vực quản lý rủi ro, như trong đo lường và kiểm soát rủi ro, quản lý rủi ro hoạt động, quản lý đầu tư, quản lý rủi ro tín dụng, quản lý rủi ro hoạt động, quản lý rủi ro tích hợp… theo các nghiên cứu của Hendricks (1997) và Jorion (1996). Thông qua việc kiểm soát rủi ro nhằm giúp cho các nhà đầu tư, các công ty đưa ra các quyết định như: Quyết định đầu tư, lựa chọn phát triển sản phẩm, phương án kinh doanh, lựa chọn ban điều hành công ty, lựa chọn danh mục đầu tư… Một số phương pháp tính toán VaR thông thường hay được sử dụng như: Phương pháp tính toán dựa vào hàm mật độ thực nghiệm, phương pháp bootstrap, phương pháp mô phỏng Monte Carlo.

- *Thứ nhất*, trong phương pháp tính toán VaR dựa vào hàm mật độ thực nghiệm, Jorion (2001) cùng Martins-Filho và cộng sự (2018) cho rằng đây là cách tiếp cận phi tham số với ưu điểm là không cần biết dạng của hàm mật độ xác suất thực nghiệm $f(.)$. Tuy nhiên, theo Rockafellar và Uryasev (2002), Hendricks (1997), phương pháp này có hạn chế là nếu cỡ mẫu nhỏ thì giá trị VaR sẽ không ổn định khi chỉ cần tăng thêm hay giảm đi một hay một vài quan sát sẽ rất dễ dẫn đến tình trạng giá trị VaR bị tăng lên hoặc giảm xuống nhiều, mà điều này rất hay xảy ra trong thực nghiệm. Còn trong trường hợp cỡ mẫu quá lớn, việc sắp xếp theo độ lớn của dữ liệu với n là cỡ mẫu lớn sẽ tốn rất nhiều thời gian và bộ nhớ, mặc dù Zhao và Luo (2018) đã cải thiện được đáng kể nhưng tính toán vẫn còn khá phức tạp vì bộ dữ liệu cần được sắp xếp từ nhỏ đến lớn, sau đó VaR với xác suất α chính là phần tử ở vị trí thứ $\alpha\% \times n$. Thêm nữa, trong các nghiên cứu của Adams & cộng sự (2014) hay Berkowitz và O'Brien (2002), các tác giả tính toán các giá trị VaR thông qua mô hình ARCH, GARCH, ARMA. Tuy nhiên, đây thực chất là tính VaR của một phân phối thực nghiệm hay còn gọi là phân phối xác

suất của biến phụ thuộc. Phân phối xác suất của biến phụ thuộc là tổ hợp tuyến tính của các phân phối tham số (phân phối chuẩn) với trọng số là giá trị của các biến độc lập (phi ngẫu nhiên) và cộng thêm nhiễu trắng (phân phối chuẩn).

- *Thứ hai,* tính toán VaR bằng phương pháp Bootstrap như nghiên cứu của Hall (1990), phương pháp này dựa trên chọn mẫu ngẫu nhiên có hoàn lại. Điều cần lưu ý là kích thước mẫu thực hiện Bootstrap phải bằng với kích thước mẫu hiện có. Phương pháp này giúp tạo ra một bộ dữ liệu đủ lớn và tin cậy cho các suy diễn thống kê tiếp theo.

- *Thứ ba*, sử dụng phương pháp mô phỏng Monte Carlo. Đây là phương pháp sử dụng cách tiếp cận tham số, trong khi hai phương pháp ở trên đều sử dụng cách tiếp cận phi tham số. Theo Jorion (2001), Engle và Manganelli (2004), đối với phương pháp mô phỏng Monte Carlo, trước tiên cần dựa vào dữ liệu lịch sử nhằm ước lượng bộ dữ liệu tuân theo phân phối xác suất nào rồi tính giá trị các tham số tương ứng các phân phối xác suất đó. Tuy nhiên, một khó khăn của phương pháp mô phỏng Monte Carlo là xem xét ước lượng bộ dữ liệu lịch sử tuân theo phân phối xác suất nào bởi vì đối với dữ liệu thực tế, phân phối của bộ dữ liệu hầu như rất hiếm trường hợp xảy ra hình dạng xấp xỉ phân phối chuẩn (cân đối và đuôi nhỏ) mà thông thường sẽ có hình dạng lệch và đuôi bự (cũng được biểu diễn dưới dạng hỗn hợp các phân phối chuẩn theo Duffie và Pan (1997), Jorion (2001)); hoặc trường hợp phân phối có dạng một đường cong hỗn hợp của nhiều phân phối xác suất. Các phương pháp để kiểm tra xem bộ dữ liệu có tuân theo một phân phối xác suất hay hỗn hợp các phân phối xác suất hay không theo cách trực tiếp như: Phương pháp hợp lý cực đại (Maximum Likelihood), giá trị entropy cực đại (Maximum Entropy), chỉ số AIC (Akaike Information Criterion) (Engle & Manganelli, 2004), chỉ số BIC (Bayesian Information Criterion)… hay theo cách gián tiếp như các bài toán kiểm định giả thuyết thông qua kiểm định Chi bình phương (Chi–square) trong bài toán lựa chọn mô hình phù hợp nhất (The Goodness of Fit), kiểm định Kolmogorov – Smirnov (Duffie & Pan, 1997).

Qua các phân tích ở trên, trong ba phương pháp tính VaR thì phương pháp mô phỏng Monte Carlo vừa đảm bảo được thời gian nhanh chóng vừa kiểm soát được các sai số. Thật vậy, đối với mỗi giá trị VaR đã tính được thì luôn cần một bước kiểm tra lại nhằm đảm bảo tính ổn định của các giá trị VaR. Khi áp dụng vào phương pháp tính VaR dựa vào hàm mật độ thực nghiệm, một trong những bước cần thiết chính là kiểm tra tính ổn định của các giá trị VaR tính được thông qua việc thay đổi dữ liệu. Tuy nhiên, việc thay đổi dữ liệu sẽ dẫn đến hàm mật độ thực nghiệm sẽ thay đổi, đặc biệt, sự thay đổi sẽ đáng kể trong trường hợp dữ liệu ban đầu hạn chế. Chính vì vậy, phương pháp tính VaR dựa vào hàm mật độ thực nghiệm đòi hỏi bộ dữ liệu đủ lớn nhằm đảm bảo các giá trị VaR tính được xấp xỉ nhau, đảm bảo tính ổn định của các giá trị VaR thì mới kết thúc được quá trình tính toán. Còn đối với việc tính VaR dựa vào phương pháp Bootstrap, để thực hiện phương pháp này, trước hết, tác giả sẽ chọn mẫu ngẫu nhiên từ mẫu ban đầu, và vì thế, mẫu ngẫu nhiên được chọn ra sẽ có các xác suất với từng quan sát tương ứng của mẫu, tức là có thể biểu diễn dữ liệu mới tạo ra theo bootstrap có dạng phân phối xác suất rời rạc. Do đó, các giá trị VaR vừa tính được không đảm bảo sự ổn định nên cần phải thực hiện lại các phép tính rồi kiểm tra cho đến khi sai lệch giữa các giá trị VaR đã tính ở mức chấp nhận được. Đến lúc này, giá trị VaR mới có thể sử dụng được. Dựa vào phương pháp thực hiện, tác giả nhận thấy hai phương pháp tính VaR dựa vào hàm mật độ thực nghiệm và bootstrap tốn khá nhiều thời gian và công sức mới sử dụng được kết quả, hơn nữa cũng không đảm bảo được sai lệch giá trị VaR ở lần tính sau thì thấp hơn ở lần tính trước. Tất nhiên, phương pháp tính VaR dựa vào mô phỏng Monte Carlo cũng cần dựa vào hàm mật độ thực nghiệm như phương pháp dựa vào hàm mật độ thực nghiệm. Tuy nhiên, để tính giá trị VaR sau đó thì chỉ cần mô phỏng dữ liệu từ hàm mật độ thực

nghiệm. Do dữ liệu là mô phỏng dựa vào dạng phân phối xác suất nên đảm bảo dữ liệu tuân theo phân phối xác suất liên tục, bởi vậy, các giá trị VaR tính được sẽ không sai lệch nhau nhiều nên có thể kiểm soát được sai số bằng cách tăng số lượng mô phỏng.

Bên cạnh giá trị VaR, một số nghiên cứu về rủi ro như nghiên cứu của Rockafellar & cộng sự (2014) còn sử dụng thêm Conditional Value at Risk, ký hiệu là CVaR hay còn gọi là siêu phân vị (Superquantile), và tính toán khoảng tin cậy với xác suất (1–α) (Confidence Interval) trong đo lường rủi ro. Nếu VaR_α là mức phân vị tại mức α thì $CVaR_\alpha$ thực chất là trung bình của các VaR_β, với $\beta \geq \alpha$, giá trị CVaR có tác dụng hạn chế tính không ổn định của VaR (Rockafellar & Uryasev, 2002). Còn bài toán khoảng tin cậy (1–α) thì khoảng tin cậy chính là tìm hai mức phân vị VaR_η và $VaR_{\eta+(1-\alpha)}$, với η là mức ý nghĩa bất kỳ nhận giá trị trong đoạn $[0; \alpha]$, bài toán này đã được tính toán bằng mô phỏng Monte Carlo như nghiên cứu của Lê Thanh Hoa và cộng sự (2017), Dietz và cộng sự (2016).

Chính vì vậy, bài nghiên cứu này sẽ tập trung theo hướng áp dụng phương pháp Monte Carlo trong tính toán VaR và CVaR. Tuy nhiên, đối với dữ liệu thực tế, hầu như rất hiếm khi xảy ra trường hợp phân phối của bộ dữ liệu tuân theo một hàm mật độ xác suất cụ thể nào đó, thông thường phân phối của bộ dữ liệu thực tế sẽ có dạng một đường cong là hỗn hợp của các phân phối xác suất. Theo Engle và Manganelli (2004), một cách thực hiện khác là thông qua một mô hình hồi quy dạng tuyến tính hoặc dạng phi tuyến. Tuy nhiên, mô hình hồi quy dạng tuyến tính biểu diễn dưới dạng phương trình (1) cũng thực chất là hỗn hợp của các phân phối chuẩn:

$$Y = \beta_0 + \beta_1 X_1 + \beta_2 X_2 + \cdots + \beta_m X_m + \epsilon \quad (1)$$

Trong đó, các tham số ước lượng $\hat{\beta}$ của β xấp xỉ phân phối chuẩn (Engle & Manganelli, 2004) và sai số (nhiễu) ϵ là nhiễu trắng, cũng xấp xỉ phân phối chuẩn trong khi các giá trị $X_i, i = \overline{1,m}$ là các giá trị phi ngẫu nhiên. Hay mô hình hồi quy phi tuyến dạng ước lượng hạt nhân (Fit Kernel), nếu không có các chỉnh sửa gì thêm thì cũng thực chất là hỗn hợp của các phân phối chuẩn, cụ thể là dạng GMM (Gaussian Mixture Model) (Hennig, 2012).

Do đó, về cơ bản, nghiên cứu của Engle và Manganelli (2004) tính VaR_α dựa trên hỗn hợp các phân phối chuẩn và sử dụng phương pháp mô phỏng Monte Carlo dựa trên mô hình GARCH(1,1). Tuy nhiên, mô phỏng Monte Carlo chỉ có tác dụng trong việc ước lượng mô hình nhằm chọn ra mô hình tốt nhất thông qua giá trị hợp lý cực đại, sau đó dựa vào mô hình này để tính ra VaR. Vì vậy, về thực chất phương pháp mô phỏng Monte Carlo trong các nghiên cứu của Engle và Manganelli (2004), Duffie và Pan (1997), Krokhmal và cộng sự (2002) không đánh giá được các giá trị VaR nhận giá trị như thế nào và độ biến động của các giá trị VaR như thế nào. Mô phỏng Monte Carlo với một phân phối như phân phối chuẩn, phân phối gamma, phân phối Weibull, phân phối Pareto… tức là chỉ mô phỏng một dạng phân phối xác suất có đánh giá trung bình (Dietz & cộng sự, 2016) và độ lệch chuẩn (Jorion, 1996).

Trong bài nghiên cứu này, tác giả sẽ hướng đến tính toán VaR và CVaR thông qua mô phỏng Monte Carlo bằng hỗn hợp các phân phối xác suất. Theo đó, các phần tiếp theo của nghiên cứu gồm có: Phần 2 trình bày giá trị VaR và CVaR; Phần 3 trình bày phương pháp lấy mẫu quan trọng (Important Sampling) và phương pháp mô phỏng Monte Carlo trong tính toán VaR và CVaR; Phần 4 sẽ nêu ứng dụng tính toán VaR và CVaR bằng mô phỏng Monte Carlo thông qua hỗn hợp các phân phối xác suất; và phần 5 kết luận.

2. Giá trị VaR và giá trị CVaR

2.1. Định nghĩa 1 – VaR

Trước hết, tác giả giới thiệu định nghĩa VaR dựa theo kết quả nghiên cứu của Rockafellar và Uryasev (2002), Pflug (2000), Huang và cộng sự (2014), Belles-Sampera và cộng sự (2014).

Cho biến ngẫu nhiên X, hàm phân phối xác suất $F_X(.)$. Giá trị VaR_α, $\alpha \in [0,1]$ là giá trị được xác định bởi công thức (2):

$$VaR_\alpha = \inf\{x | F_X(x) \geq \alpha\} = F_X^{-1}(\alpha) \quad (2)$$

Ngoài ra, còn có thêm một đại lượng đo lường rủi ro khi thực hành đó chính là độ đo CVaR, hay còn được gọi là giá trị gặp rủi ro đuôi TVaR (Tail Value-at-Risk) (Belles-Sampera & cộng sự, 2014).

2.2. Định nghĩa 2 – CVaR

Tiếp theo, tác giả giới thiệu định nghĩa CVaR dựa theo kết quả nghiên cứu của Rockafellar và Uryasev (2002), Pflug (2000), Huang và cộng sự (2014).

Cho biến ngẫu nhiên X, hàm phân phối xác suất $F_X(.)$. Giá trị $CVaR_\alpha$ là giá trị được xác định bởi công thức (3) trong trường hợp hàm phân phối xác suất $F_X(.)$ dạng rời rạc:

$$CVaR_\alpha = \frac{\sum_{k=1}^{\hat{n}} VaR_{\beta_k}}{\hat{n}} \qquad (1 \geq \beta_k \geq \alpha) \quad (3)$$

$Trong$ đó, $\hat{n}$ là số lượng mô phỏng.

Trong trường hợp hàm phân phối xác suất $F_X(.)$ dạng liên tục, $CVaR_\alpha$ được xác định bởi công thức (4):

$$CVaR_\alpha = \int_{-\infty}^{+\infty} z \times dF_X^\alpha(z) \quad (4)$$

Trong đó, hàm phân phối xác suất F_X^α của $CVaR_\alpha$ được định nghĩa bởi công thức (5):

$$F_X^\alpha(z) = \begin{cases} 0 & (nếu\ z < VaR_\alpha) \\ \dfrac{F_X(z) - \alpha}{1 - \alpha} & (nếu\ z \geq VaR_\alpha) \end{cases} \quad (5)$$

2.3. Hệ quả 1

Tiếp theo, tác giả giới thiệu mối liên hệ giữa VaR và CVaR thông qua một hệ quả, theo Rockafellar và Uryasev (2002):

Từ định nghĩa VaR và CVaR ở trên, ta có:

$$VaR_\alpha \leq CVaR_\alpha.$$

2.4. Định nghĩa 3 – ES

Bên cạnh các độ đo hay được sử dụng trong đo lường rủi ro như VaR và CVaR thì một độ đo cũng được các nhà nghiên cứu quan tâm đó là thâm hụt dự kiến ES (Expected Shortfall), Acerbi và Tasche (2002).

Giả sử tập sắp thứ tự $x_{(1)} \leq x_{(2)} \leq \cdots \leq x_{(n)}$ và mức ý nghĩa α, giá trị $\omega = [n \times \alpha] = \max\{k | k \leq n \times \alpha, k \in N\}$. Giả sử phân vị mức α là $x_{(\omega)}$. Khi đó, giá trị ES được tính theo công thức (6):

$$ES_\alpha = -\frac{\sum_{i=1}^{\omega} x_{(i)}}{\omega} \tag{6}$$

Thông qua Định nghĩa 2 và 3, tác giả nhận thấy cách tính toán của CVaR và ES gần giống nhau, chỉ có khác nhau là CVaR quan tâm đến các giá trị lớn hơn hoặc bằng VaR, còn ES lại quan tâm đến các giá trị nhỏ hơn hoặc bằng VaR. Do đó, tùy từng yêu cầu của vấn đề đặt ra thì ngoài tính VaR thì chỉ cần tính thêm CVaR hoặc ES.

3. Phương pháp lấy mẫu quan trọng và phương pháp mô phỏng Monte Carlo trong tính toán VaR và CVaR

3.1. Phương pháp lấy mẫu quan trọng

Mô hình hỗn hợp của *m* phân phối xác suất, giả sử hàm mật độ xác suất $f_j(.)$ với xác suất tương ứng là $p_j, j = \overline{1,m}$ (Robert, 2004). Khi đó, với mỗi quan sát X có hàm mật độ xác suất tương ứng là $f(.)$ được biểu diễn dưới dạng công thức (7):

$$f(x) \sim p_1 f_1(x) + p_2 f_2(x) + \cdots + p_m f_m(x) \tag{7}$$

Theo Robert (2004), phương pháp lấy mẫu quan trọng (Important Sampling) là một cách xấp xỉ tích phân trong công thức (8):

$$E_f[h(X)] = \int_X h(x) f(x) dx \tag{8}$$

Dựa trên mẫu tổng quát $X_1, X_2, \ldots, X_n$ từ một phân phối xác suất với hàm mật độ rời rạc *g* xác định trước và xấp xỉ bởi công thức (9):

$$E_f[h(X)] \approx \frac{1}{n} \sum_{i=1}^{n} \frac{f(X_j)}{g(X_j)} h(X_j) \tag{9}$$

Và giá trị ước lượng trong công thức (9) hội tụ về công thức (8). Phương pháp này cũng được dựa trên phương pháp thay thế cho hàm mật độ xác suất *g* trong công thức (8) thông qua công thức (10):

$$E_f[h(X)] \approx \int_\chi h(x) \frac{f(x)}{g(x)} g(x) dx \tag{10}$$

3.2. Phương pháp mô phỏng Monte Carlo trong tính toán VaR và CVaR

Tác giả đề xuất thuật toán tính toán VaR và CVaR bằng mô phỏng Monte Carlo với hàm mật độ xác suất tổng quát xác định theo công thức (11):

$$f(x) = p_1 \times f_1(x) + p_2 \times f_2(x) + \cdots + p_m \times f_m(x) \quad (11)$$

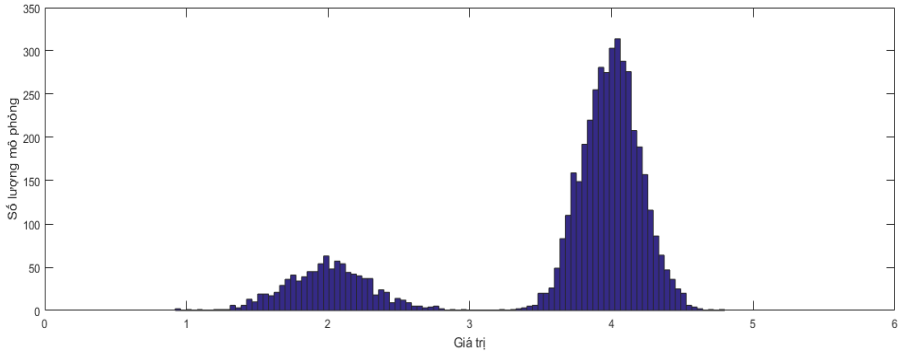
Trong đó, các bước thực hiện được tóm tắt ở Bảng 1.

Bảng 1.

Các bước thực hiện mô phỏng Monte Carlo trong tính toán VaR và CVaR

Đầu vào	Các hàm mật độ xác suất thành phần $f_j(.)$ với xác suất tương ứng $p_j, j = \overline{1,m}$, đồng thời mức ý nghĩa α.
Bước 1	Mô phỏng bộ dữ liệu: Giả sử số lượng cần mô phỏng là n. Bước 1a. Mô phỏng các hàm mật độ thành phần $f_j(.)$ với cỡ mẫu tương ứng là $[n \times p_j]$ dòng và n cột. Bước 1b. Bộ dữ liệu bao gồm m phân phối có tất cả n dòng và n cột. Bước 1c. Sắp xếp bộ dữ liệu theo từng cột với các giá trị từ nhỏ tới lớn.
Bước 2	Xác định các giá trị VaR_α và CVaR_α theo từng cột: Bước 2a. Giá trị VaR_α^j tại mỗi cột là phần tử ở vị trí $[n \times \alpha]$, đối với bộ dữ liệu đã sắp xếp từ nhỏ đến lớn của mỗi cột tương ứng. Bước 2b. Các giá trị của bộ dữ liệu tại mỗi cột là phần tử từ vị trí $[n \times \alpha]$ đến vị trí cuối cùng n, đối với bộ dữ liệu đã sắp xếp từ nhỏ đến lớn của mỗi cột tương ứng. Bước 2c. Giá trị $CVaR_\alpha^j$ là trung bình của các giá trị đã được tính ra từ bước 2b. Bước 2d. Kiểm tra các kết quả vừa tính toán: Tính các xác suất thành phần theo từng cột của hàm mật độ xác suất trên khoảng các giá trị, bao gồm các giá trị nhỏ hơn VaR_α^j, tức là tính $\text{Prob.}\left(\text{X} < \text{VaR}_\alpha^j\right) = \alpha$. Tính xác suất trung bình của các xác suất thành phần $\frac{\sum_j \text{Prob.}\left(\text{X} < \text{VaR}_\alpha^j\right)}{\text{n}}$ Kiểm tra xác suất trung bình có bằng α hay không. Tính độ lệch chuẩn của xác suất thành phần (nhằm kiểm tra tính ổn định).
Đầu ra	Các giá trị ước lượng $VaR_\alpha = \frac{\sum_{j=1}^n VaR_\alpha^j}{n}$ và $CVaR_\alpha = \frac{\sum_{j=1}^n CVaR_\alpha^j}{n}$

Nhận xét: Một lưu ý nhỏ trong các mô phỏng hỗn hợp cần phải được thực hiện với cỡ mẫu dựa trên độ lớn xác suất tương ứng thì mới ra được kết quả chính xác hơn. Minh chứng bằng việc mô phỏng hai phân phối chuẩn $N(4; 0,2^2)$ và $N(4; 0,3^2)$ với xác suất tương ứng là 0,8 và 0,2, có dạng hai đỉnh theo Hình 1 với số lượng mô phỏng n = 5.000.



Hình 1. Hình dạng của phân phối xác suất hỗn hợp $0,8 \times N(4; 0,2^2) + 0,2 \times N(4; 0,3^2)$

Kết quả tính toán $VaR_{0,05}$ và $CVaR_{0,05}$ trong trường hợp số lượng mô phỏng số cột là n = 5.000 và số dòng của từng phân phối bằng n hoặc bằng n x p_j tương ứng được tính toán và trình bày trong Bảng 2.

Bảng 2.

Các kết quả tính toán $VaR_{0,05}$ và $CVaR_{0,05}$ trong trường hợp số lượng mô phỏng số cột n = 5.000 với phân phối xác xuất hỗn hợp là 0,8 và 0,2

Xác suất thành phần khác nhau	Giá trị $VaR_{0,05}$	Độ lệch chuẩn của $VaR_{0,05}$	Giá trị $CVaR_{0,05}$	Độ lệch chuẩn của $CVaR_{0,05}$	Xác suất của hàm f nhận giá trị nhỏ hơn $VaR_{0,05}$	Độ lệch chuẩn của xác suất của hàm f nhận giá trị nhỏ hơn $VaR_{0,05}$
Số dòng bằng n = 5.000	1,6155	0,0073	3,0802	0,0025	0,0200	0,0009
Số dòng bằng $n \times p_j = 5.000 \times p_j$	1,7965	0,0130	3,7038	0,0031	0,0498	0,0027

Rõ ràng, bài nghiên cứu kỳ vọng kết quả phải thể hiện được xác suất của hàm f từ $-\infty$ đến $VaR_{0,05}$ xấp xỉ bằng 0,05. Tuy nhiên, kết quả trong hai cột cuối của Bảng 2 cho thấy chỉ có trường hợp mô phỏng số dòng bằng $n \times p_j$ thì mới đạt kỳ vọng mong muốn, còn trường hợp số dòng bằng n thì không đạt kỳ vọng mong muốn.

Điều này có thể giải thích như sau: Do xác suất của các phân phối khác nhau dẫn đến nếu số lượng mô phỏng bằng nhau thì xác suất mỗi phân phối không còn được đảm bảo, cho nên kết quả sẽ bị sai lệch. Tác giả sẽ thực hiện lại cách tính trên với tỷ lệ xác suất bằng nhau là 0,5 và 0,5.

Bảng 3.

Các kết quả tính toán $VaR_{0,05}$ và $CVaR_{0,05}$ trong trường hợp số lượng mô phỏng số cột là n = 5.000 với phân phối xác xuất hỗn hợp là 0,5 và 0,5

Xác suất thành phần bằng nhau	Giá trị $VaR_{0,05}$	Độ lệch chuẩn của $VaR_{0,05}$	Giá trị $CVaR_{0,05}$	Độ lệch chuẩn của $CVaR_{0,05}$	Xác suất của hàm f nhận giá trị nhỏ hơn $VaR_{0,05}$	Độ lệch chuẩn của xác suất của hàm f nhận giá trị nhỏ hơn $VaR_{0,05}$
Số dòng bằng n = 5.000	1,6154	0,0072	3,0802	0,0025	0,0500	0,0021
Số dòng bằng $n \times p_j = 5.000 \times p_j$	1,6152	0,0102	3,0800	0,0036	0,0500	0,0030

Kết quả tính toán từ Bảng 3 cho thấy rằng trong cả hai trường hợp thì kết quả của $VaR_{0,05}$ và $CVaR_{0,05}$ là xấp xỉ nhau và đều đạt xác suất kỳ vọng từ $-\infty$ đến $VaR_{0,05}$ xấp xỉ bằng 0,05. Tuy nhiên, trong trường hợp mô phỏng với số dòng bằng n thì kết quả có sai lệch nhỏ hơn so với trường hợp mô phỏng với số dòng bằng $n \times p_j$, điều này hợp lý vì trường hợp mô phỏng với số dòng bằng n dẫn đến có số lượng mô phỏng theo mỗi dòng là nhiều hơn (tức là n + n = 2n), còn trường hợp mô phỏng với số dòng bằng $n \times p_j$ dẫn đến có số lượng mô phỏng theo mỗi dòng là ít hơn (tức là $\sum_{j=1}^{m} n \times p_j = 1 \times n = n$).

Tóm lại, trong tất cả các trường hợp với các xác suất thành phần bất kỳ ($0 < p_j < 1, j = \overline{1,m}$), tác giả khuyến khích sử dụng trường hợp mô phỏng số cột bằng n và số dòng bằng $n \times p_j$ sẽ cho kết quả tính tương đối chính xác và cần số lượng mô phỏng không quá nhiều.

4. Ứng dụng tính toán VaR và CVaR bằng mô phỏng Monte Carlo thông qua hỗn hợp các phân phối xác suất

Đối với một bộ dữ liệu, đặc biệt các dữ liệu thực tế như giá chứng khoán, tỷ giá… thì việc bộ dữ liệu chỉ tuân theo một phân phối xác suất là rất hiếm khi xảy ra, kể cả dữ liệu có dạng đối xứng như phân phối chuẩn, phân phối Student hay có dạng lệch như phân phối Gamma, phân phối Weibull, phân phối Beta… Do đó, tác giả cho rằng cần mô hình hóa dữ liệu thông qua hỗn hợp các phân phối xác suất. Và tất nhiên, đối với dữ liệu tuân theo một phân phối xác suất có tính toán các giá trị VaR_α và $CVaR_\alpha$ thì cũng cần tính toán các giá trị này cho hỗn hợp các phân phối xác suất. Trong phần này, tác giả tiến hành tính toán các giá trị VaR_α và $CVaR_\alpha$ trong một số trường hợp đặc biệt như hỗn hợp hai phân phối đối xứng, hỗn hợp hai phân phối lệch, hỗn hợp ba phân phối bao gồm cả phân phối lệch và không lệch. Bên cạnh đó, tác giả cũng tính toán các giá trị VaR và CVaR dựa trên bộ dữ liệu thực về giá chứng khoán Việt Nam.

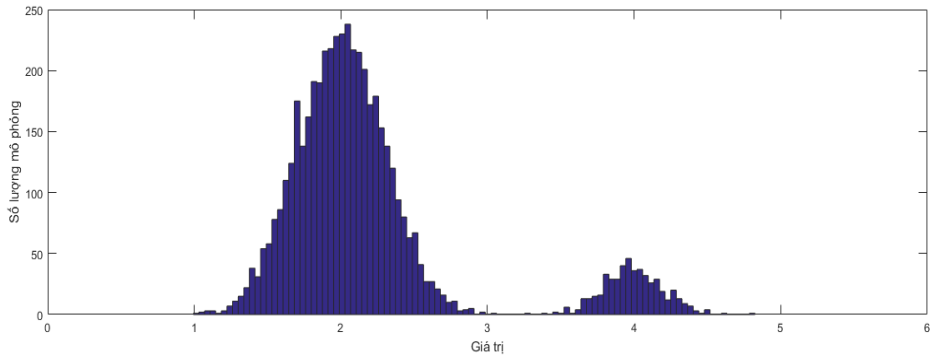
Phương pháp tính toán VaR_α và $CVaR_\alpha$ của tác giả vẫn đúng trong trường hợp bộ dữ liệu chỉ tuân theo duy nhất một phân phối xác suất, tức là:

$$p_{j_0} = 1, p_{j_1} = 0, j_1 \in \{1,2,\dots,j_0-1\} \cup \{j_0+1, j_0+2,\dots,m\}$$

4.1. Hỗn hợp các phân phối xác suất

4.1.1. Hỗn hợp hai phân phối đối xứng

Trước hết, tác giả tiến hành nghiên cứu về hỗn hợp hai phân phối đối xứng là hai phân phối chuẩn là phân phối N(4; $0{,}2^2$) và N(2; $0{,}3^2$) với xác suất tương ứng là 0,1 và 0,9 (tác giả muốn chứng minh trong trường hợp xác suất thành phần chênh lệch nhau quá nhiều thì kết quả vẫn chính xác). Hình ảnh của phân phối hỗn hợp được biểu diễn qua Hình 2.



Hình 2. Hình dạng của phân phối xác suất hỗn hợp $0{,}1 \times N(4; 0{,}2^2) + 0{,}9 \times N(2; 0{,}3^2)$

Kết quả từ Bảng 4 cho thấy xác suất thành phần của các phân phối có nhiều khác biệt thì phương pháp mô phỏng Monte Carlo của tác giả vẫn đảm bảo xác suất đuôi trái xấp xỉ giá trị xác suất mong muốn là 0,1. Hơn nữa, khi số lượng mô phỏng càng lớn thì xác suất đuôi trái càng gần giá trị xác suất mong muốn hơn và càng ổn định hơn thông qua giá trị độ lệch chuẩn càng giảm.

Bảng 4.

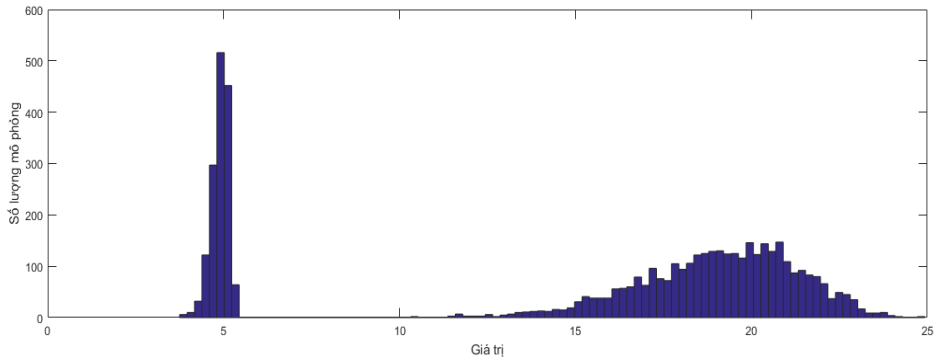
Các kết quả tính toán $VaR_{0,1}$ và $CVaR_{0,1}$ trong trường hợp số lượng mô phỏng số cột lần lượt là n = 100; 500; 1.000; 5.000 với hai phân phối đối xứng

Số lượng mô phỏng n	Giá trị $VaR_{0,1}$	Độ lệch chuẩn của $VaR_{0,1}$	Giá trị $CVaR_{0,1}$	Độ lệch chuẩn của $CVaR_{0,1}$	Xác suất của hàm f nhận giá trị nhỏ hơn $VaR_{0,1}$	Độ lệch chuẩn của xác suất của hàm f nhận giá trị nhỏ hơn $VaR_{0,1}$
100	1,6244	0,0487	2,2675	0,0281	0,0974	0,0261
500	1,6320	0,0239	2,2770	0,0138	0,0996	0,0135
1.000	1,6330	0,0168	2,2784	0,0095	0,0999	0,0095
5.000	1,6335	0,0074	2,2788	0,0042	0,0999	0,0042

4.1.2. Hỗn hợp hai phân phối lệch

Tiếp theo, tác giả sẽ áp dụng phương pháp trên đối với hỗn hợp hai phân phối bất đối xứng hay còn gọi là hai phân phối lệch để kiểm tra tính đúng đắn. Trong trường hợp này, tác giả lấy ví dụ tính toán dựa trên phân phối lệch là phân phối Weibull W(a; b). Trong đó, các tham số a là tham số tỷ lệ (Scale Parameter) và tham số b là tham số hình dạng (Shape Parameter).

Phân phối hỗn hợp của hai phân phối Weibull W(20; 10) và W(5; 25) với xác suất tương ứng là 0,7 và 0,3 với hình dạng thông qua Hình 3.



Hình 3. Hình dạng của phân phối xác suất hỗn hợp 0,7 × W(20; 10) + 0,3 × W(5; 25)

Kết quả tính toán VaR_α và $CVaR_\alpha$ với $\alpha = 0,1$ được biểu diễn trong Bảng 5. Tác giả nhận thấy với hỗn hợp các phân phối lệch, phương pháp mô phỏng Monte Carlo vẫn hiệu quả trong các tính toán VaR_α cũng như $CVaR_\alpha$ và cho kết quả tương đối ổn định giữa các lần tính toán.

Bảng 5.

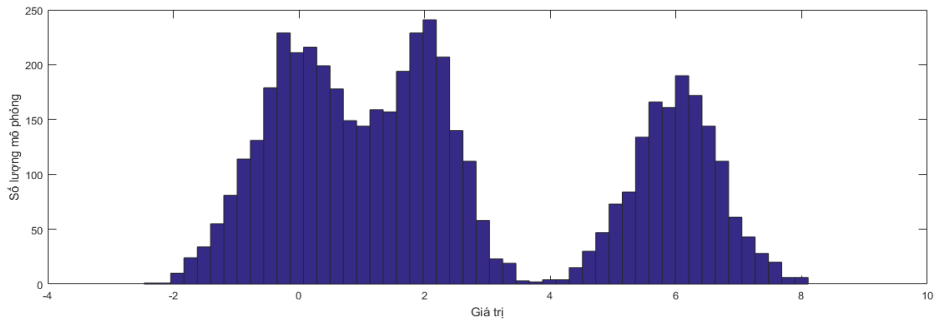
Các kết quả tính toán $VaR_{0,1}$ và $CVaR_{0,1}$ trong trường hợp số lượng mô phỏng số cột lần lượt là n = 100; 500; 1.000; 5.000 với hai phân phối lệch

Số lượng mô phỏng n	Giá trị $VaR_{0,1}$	Độ lệch chuẩn của $VaR_{0,1}$	Giá trị $CVaR_{0,1}$	Độ lệch chuẩn của $CVaR_{0,1}$	Xác suất của hàm f nhận giá trị nhỏ hơn $VaR_{0,1}$	Độ lệch chuẩn của xác suất của hàm f nhận giá trị nhỏ hơn $VaR_{0,1}$
100	4,8133	0,0573	15,7918	0,1989	0,0981	0,0227
500	4,8192	0,0288	15,8895	0,0901	0,0991	0,0119
1.000	4,8206	0,0183	15,9027	0,0650	0,0993	0,0076
5.000	4,8224	0,0087	15,9133	0,0300	0,0999	0,0037

4.1.3. Hỗn hợp ba phân phối đối xứng

Trong trường hợp này, tác giả cũng áp dụng phương pháp tính toán dựa vào mô phỏng Monte Carlo với hỗn hợp các phân phối có nhiều đỉnh hơn, cụ thể là hỗn hợp của ba phân phối chuẩn. Tác giả mô phỏng Monte Carlo trong trường hợp hỗn hợp ba phân phối chuẩn $N(0; 0,75^2)$, $N(2; 0,55^2)$ và

$N(6; 0,7^2)$ với xác suất lần lượt là 0,4, 0,3 và 0,3. Phân phối hỗn hợp có dạng ba đỉnh được biểu diễn thông qua Hình 4.



Hình 4. Hình dạng của phân phối xác suất hỗn hợp
$0,4 \times N(0; 0,75^2) + 0,3 \times N(2; 0,55^2) + 0,3 \times N(6; 0,7^2)$

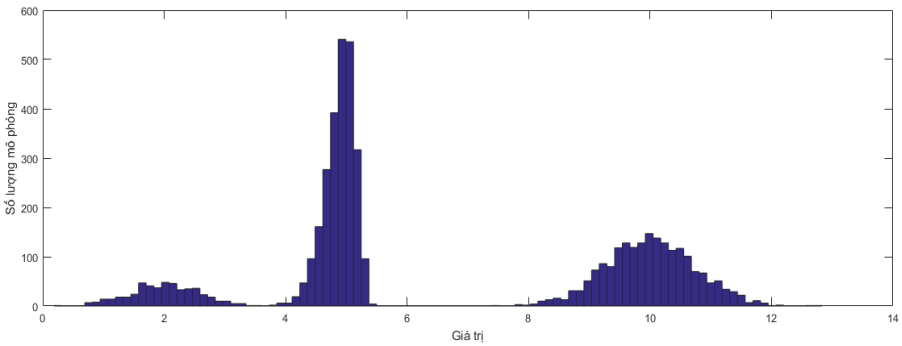
Tác giả cũng tính toán VaR_α và $CVaR_\alpha$ với kết quả được thể hiện trong Bảng 6, và cũng giống như trường hợp hai đỉnh, trong trường hợp ba đỉnh, kết quả tính toán khá chính xác.

Bảng 6.

Các kết quả tính toán $VaR_{0,1}$ và $CVaR_{0,1}$ trong trường hợp số lượng mô phỏng số cột lần lượt là n = 100; 500; 1.000; 5.000 với ba phân phối đối xứng

Số lượng mô phỏng n	Giá trị $VaR_{0,1}$	Độ lệch chuẩn của $VaR_{0,1}$	Giá trị $CVaR_{0,1}$	Độ lệch chuẩn của $CVaR_{0,1}$	Xác suất của hàm f nhận giá trị nhỏ hơn $VaR_{0,1}$	Độ lệch chuẩn của xác suất của hàm f nhận giá trị nhỏ hơn $VaR_{0,1}$
100	–0,5090	0,1456	2,7240	0,0657	0,1010	0,0246
500	–0,5060	0,0720	2,7651	0,0307	0,1004	0,0121
1.000	–0,5086	0,0529	2,7687	0,0218	0,0998	0,0089
5.000	–0,5067	0,0234	2,7717	0,0096	0,0999	0,0040

4.1.4. Hỗn hợp hai phân phối đối xứng và một phân phối lệch



Hình 5. Hình dạng của phân phối xác suất hỗn hợp
$0,4 \times N(10; 0,75^2) + 0,1 \times N(2; 0,55^2) + 0,5 \times W(5; 25)$

Tác giả cũng tính toán VaR_{α} và $CVaR_{\alpha}$ với kết quả được thể hiện trong Bảng 7, và cũng giống như trường hợp ba đỉnh từ các phân phối đối xứng, phương pháp tính toán của tác giả vẫn đưa ra kết quả khá chính xác, tuy nhiên với số lượng mô phỏng càng lớn thì các giá trị ước lượng được càng ít biến động.

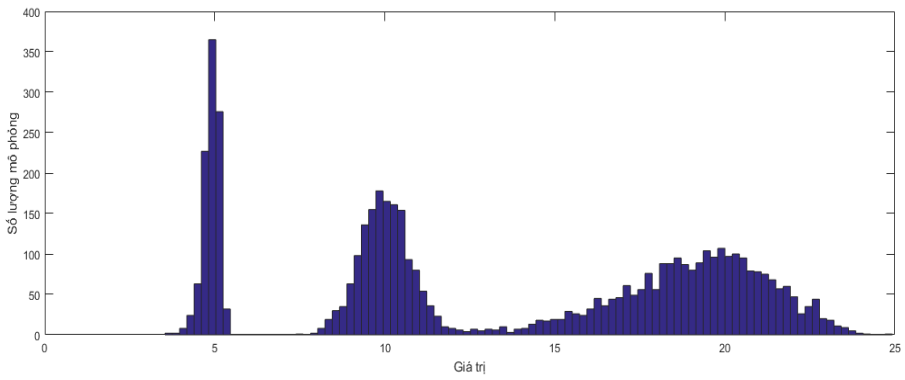
Bảng 7.

Các kết quả tính toán $VaR_{0,1}$ và $CVaR_{0,1}$ trong trường hợp số lượng mô phỏng số cột lần lượt n = 100; 500; 1.000; 5.000 với hai phân phối đối xứng và một phân phối lệch

Số lượng mô phỏng n	Giá trị $VaR_{0,1}$	Độ lệch chuẩn của $VaR_{0,1}$	Giá trị $CVaR_{0,1}$	Độ lệch chuẩn của $CVaR_{0,1}$	Xác suất của hàm f nhận giá trị nhỏ hơn $VaR_{0,1}$	Độ lệch chuẩn của xác suất của hàm f nhận giá trị nhỏ hơn $VaR_{0,1}$
100	2,8575	0,3162	7,0972	0,0595	0,0913	0,0085
500	3,2157	0,2425	7,1535	0,0243	0,0979	0,0022
1.000	3,3601	0,2162	7,1582	0,0179	0,0991	0,0010
5.000	3,5645	0,1230	7,1616	0,0079	0,0999	0,0003

4.1.5. Hỗn hợp một phân phối đối xứng và hai phân phối lệch

Cũng giống như các trường hợp trên, đối với trường hợp hỗn hợp của hai phân phối đối xứng, hai phân phối lệch, ba phân phối đối xứng, hai phân phối đối xứng và một phân phối lệch thì trong trường hợp hỗn hợp của một phân phối đối xứng với hai phân phối lệch được biểu diễn trong Hình 6 và kết quả tính toán trong Bảng 7 cũng cho thấy kết quả của tác giả thông qua mô phỏng Monte Carlo có kết quả tương đối tốt.



Hình 6. Hình dạng của phân phối xác suất hỗn hợp
$0,3 \times N(10; 0,75^2) + 0,5 \times W(20; 10) + 0,2 \times W(5; 25)$

Kết quả tính toán $VaR_{0,1}$ và $CVaR_{0,1}$ với $\alpha = 0{,}1$ được biểu diễn trong Bảng 8.

Bảng 8.

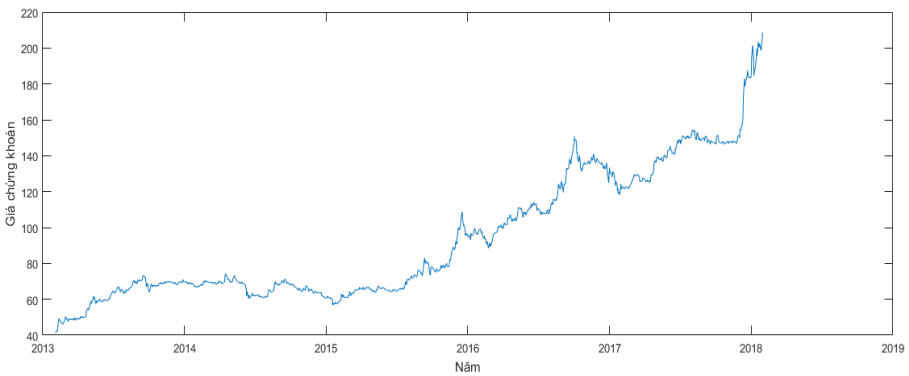
Các kết quả tính toán $VaR_{0,1}$ và $CVaR_{0,1}$ trong trường hợp số lượng mô phỏng số cột lần lượt là n = 100; 500; 1.000; 5.000 với một phân phối đối xứng và hai phân phối lệch

Số lượng mô phỏng n	Giá trị $VaR_{0,1}$	Độ lệch chuẩn của $VaR_{0,1}$	Giá trị $CVaR_{0,1}$	Độ lệch chuẩn của $CVaR_{0,1}$	Xác suất của hàm f nhận giá trị nhỏ hơn $VaR_{0,1}$	Độ lệch chuẩn của xác suất của hàm f nhận giá trị nhỏ hơn $VaR_{0,1}$
100	4,9022	0,0558	14,3531	0,1897	0,0922	0,0186
500	4,9188	0,0303	14,4475	0,0824	0,0973	0,0105
1.000	4,9245	0,0206	14,4587	0,0583	0,0991	0,0072
5.000	4,9265	0,0089	14,4667	0,0264	0,0998	0,0031

4.2. Ứng dụng vào dữ liệu giá chứng khoán Việt Nam

Đối với dữ liệu thực giá chứng khoán, hàm mật độ xác suất không chỉ có dạng một phân phối xác suất, hỗn hợp của hai phân phối xác suất, hỗn hợp của ba phân phối xác suất mà còn có thể là hỗn hợp của nhiều phân phối xác suất hơn nữa.

Trong phần này, tác giả sử dụng giá đóng cửa của mã chứng khoán Công ty Cổ phần Sữa Việt Nam (VNM), lấy từ Trung tâm Nghiên cứu Kinh tế – Tài chính, Trường Đại học Kinh tế – Luật, được cung cấp bởi Thompson Reuters[1] trong khoảng thời gian 5 năm từ 2013–2017. Dữ liệu giá chứng khoán VNM trong khoảng thời gian 5 năm được biểu diễn thông qua Hình 7.



Hình 7. Giá đóng cửa của mã chứng khoán VNM giai đoạn 2013–2017

Một số nghiên cứu thường sử dụng suất sinh lợi *return* khi nghiên cứu giá đóng cửa theo một trong ba công thức từ (12) đến (14), theo Demir & cộng sự (2004). Trong đó, công thức (12) là giá trị tuyệt đối của hai giá đóng cửa ở thời điểm t+1 và thời điểm t. Công thức (13) là giá trị tương đối của

[1] Thompson Reuters (https://www.thomsonreuters.com/en.html)

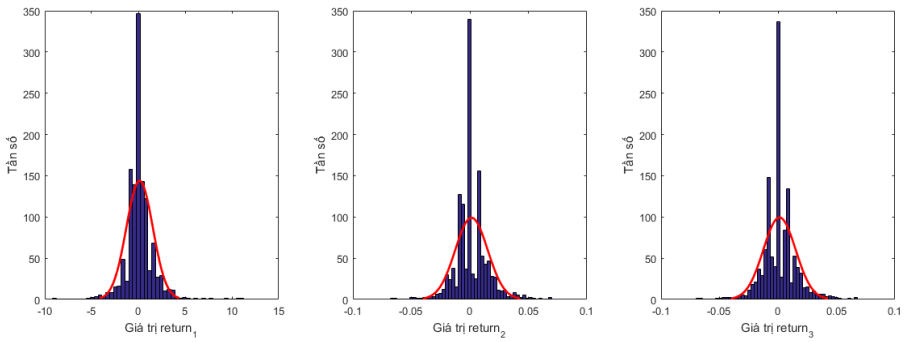
hai giá đóng cửa ở thời điểm t+1 và thời điểm t. Công thức (14) thực chất là hiệu của hai logarit giá đóng cửa ở thời điểm t+1 và thời điểm t.

$$return_1(t) = P(t+1) - P(t) \tag{12}$$

$$return_2(t) = \frac{P(t+1) - P(t)}{P(t)} \tag{13}$$

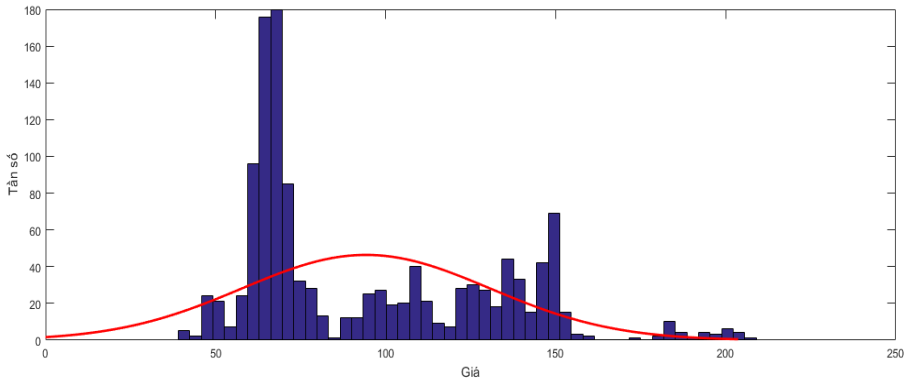
$$return_3(t) = \log\left(\frac{P(t+1)}{P(t)}\right) = \log\big(P(t+1)\big) - \log\big(P(t)\big) \tag{14}$$

Sau đó, tác giả sẽ lấy xấp xỉ suất sinh lợi *return* theo phân phối chuẩn, tuy nhiên, khi tác giả xấp xỉ bộ dữ liệu suất sinh lợi *return* theo phân phối chuẩn thì thấy có sự sai khác tương đối nhiều khi biểu diễn theo Hình 8.



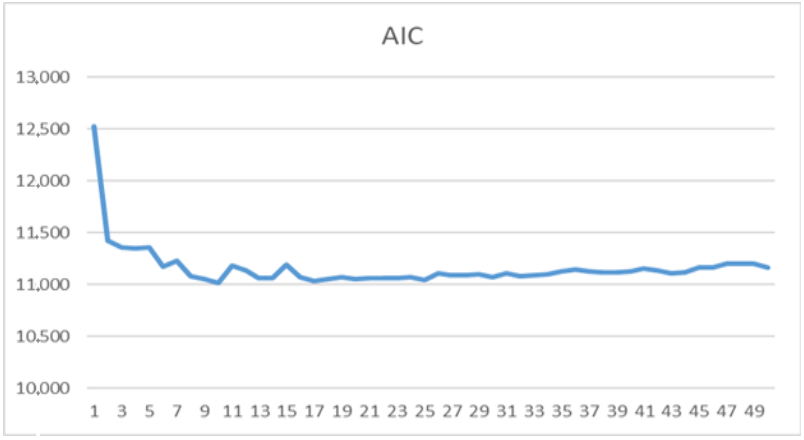
Hình 8. Xấp xỉ suất sinh lợi return của giá đóng cửa mã chứng khoán VNM giai đoạn 2013–2017 theo phân phối chuẩn

Một vấn đề khác, khi tác giả tính VaR_α hoặc $CVaR_\alpha$ cho suất sinh lợi thì chỉ tính được giá trị trung gian ở giữa, tức là cần phải tính thêm một bước nữa mới tính ra được giá trị thực sự của giá đóng cửa. Tuy nhiên, suất sinh lợi lại dựa vào giá đóng cửa ở phiên giao dịch trước đó nữa nên việc tính toán VaR_α và $CVaR_\alpha$ sẽ bị sai lệch nhiều. Chính vì vậy, sự cần thiết phải tính VaR_α và $CVaR_\alpha$ cho giá đóng cửa gốc chứ không phải là suất sinh lợi. Từ đó, tác giả sẽ nghiên cứu hàm mật độ xác suất thực nghiệm của giá đóng cửa VNM thông qua Hình 9. Theo đó, đồ thị cho thấy không thể xấp xỉ giá đóng cửa của mã chứng khoán VNM theo phân phối chuẩn hay các phân phối xác suất một đỉnh khác, mà cần phải xấp xỉ dưới dạng hỗn hợp các phân phối xác suất.



Hình 9. Xấp xỉ giá đóng cửa mã chứng khoán VNM giai đoạn 2013–2017 theo phân phối chuẩn

Tác giả xấp xỉ dữ liệu giá đóng cửa mã chứng khoán VNM dạng hỗn hợp các phân phối chuẩn (Gaussian Mixture Model), sau đó sẽ chọn số lượng phân phối hỗn hợp thông qua giá trị nhỏ nhất AIC. Giá trị AIC được biểu diễn thông qua Hình 10 với số lượng hỗn hợp các phân phối từ 1 đến 50.



Hình 10. Giá trị AIC khi xấp xỉ giá đóng cửa chứng khoán VNM giai đoạn 2013–2017 theo hỗn hợp k phân phối chuẩn.

Do tác giả mong muốn tìm được giá trị nhỏ nhất AIC và Hình 10 đã biểu diễn giá trị AIC có xu hướng giảm xuống sau đó lại tăng lên. Giá trị nhỏ nhất AIC tại k = 10 với giá trị min AIC = 11.023. Khi đó, phân phối thực nghiệm của giá đóng cửa VNM được xấp xỉ thông qua hỗn hợp các phân phối chuẩn với xác suất, trung bình và độ lệch chuẩn tương ứng được biểu diễn trong Bảng 9.

Bảng 9.

Kết quả ước lượng các phân phối chuẩn thành phần

Số lượng các phân phối	Xác suất thành phần	Trung bình phân phối chuẩn thành phần	Độ lệch chuẩn phân phối chuẩn thành phần
1	0,0976	73,5031	23,4555
2	0,1214	101,7594	62,9683
3	0,0475	115,5858	282,8473
4	0,1576	64,7660	11,4777
5	0,0445	48,4524	7,9097
6	0,0751	148,6572	2,3120
7	0,1676	64,4093	12,9181
8	0,0860	69,2342	0,4278
9	0,0288	191,5414	83,3419
10	0,1740	134,5834	101,5937

Kết quả tính toán VaR_α và $CVaR_\alpha$ với $\alpha = 0,1$ với được biểu diễn trong Bảng 10. Tác giả cũng hoàn toàn tính toán được các giá trị VaR_α và $CVaR_\alpha$ và các giá trị này vẫn đảm bảo sai số mong muốn.

Bảng 10.

Các kết quả tính toán $VaR_{0,1}$ và $CVaR_{0,1}$ trong trường hợp số lượng mô phỏng số cột lần lượt là n = 100; 500; 1.000; 5.000 với dữ liệu thực tế từ thị trường chứng khoán Việt Nam

Số lượng mô phỏng n	Giá trị $VaR_{0,1}$	Độ lệch chuẩn của $VaR_{0,1}$	Giá trị $CVaR_{0,1}$	Độ lệch chuẩn của $CVaR_{0,1}$	Xác suất của hàm f nhận giá trị nhỏ hơn $VaR_{0,1}$	Độ lệch chuẩn của xác suất của hàm f nhận giá trị nhỏ hơn $VaR_{0,1}$
100	38,5792	7,8162	105,6066	6,5062	0,0944	0,0255
500	41,3524	2,6976	106,8459	2,8862	0,0992	0,0127
1.000	41,6546	1,7583	107,1012	2,2444	0,0998	0,0088
5.000	41,7913	0,7887	107,1752	0,9495	0,0999	0,0040

Tương tự như vậy, tác giả cũng tính toán các giá trị VaR_α và $CVaR_\alpha$ với α giảm đi một nửa tức là $\alpha = 0,05$. Kết quả được biểu diễn trong Bảng 11. Thông qua kết quả này, tác giả nhận thấy khi số lượng mô phỏng quá ít $n = 100$ trong khi số các phân phối thành phần quá nhiều $m = 10$ thì giá trị $VaR_{0,05}$ nhận giá trị âm, tức là đây không phải là kết quả có thể xảy ra trong thực tế. Tuy nhiên, khi tăng số lượng mô phỏng lên $n = 500$ đến 5.000, các kết quả VaR_α và $CVaR_\alpha$ khá xấp xỉ nhau và đảm bảo xấp xỉ giá trị mong muốn.

Bảng 11.

Các kết quả tính toán $VaR_{0,05}$ và $CVaR_{0,05}$ trong trường hợp số lượng mô phỏng số cột lần lượt là n = 100; 500; 1.000; 5.000 với dữ liệu thực tế từ thị trường chứng khoán Việt Nam

Xác suất thành phần khác nhau	Giá trị $VaR_{0,05}$	Độ lệch chuẩn của $VaR_{0,05}$	Giá trị $CVaR_{0,05}$	Độ lệch chuẩn của $CVaR_{0,05}$	Xác suất của hàm f nhận giá trị nhỏ hơn $VaR_{0,05}$	Độ lệch chuẩn của xác suất của hàm f nhận giá trị nhỏ hơn $VaR_{0,05}$
100	–0,3176	32,2698	100,7468	6,3699	0,0451	0,0189
500	12,8245	11,1787	102,9000	2,9440	0,0488	0,0087
1.000	14,0250	8,0753	103,0629	2,1259	0,0492	0,0065
5.000	15,5072	3,5840	103,2540	0,9409	0,0498	0,0029

5. Kết luận

Giá trị gặp rủi ro VaR hay giá trị gặp rủi ro có điều kiện CVaR không phải là một khái niệm mới trong các tính toán phân tích tài chính, mà còn được sử dụng rất nhiều trong các bài toán đo lường rủi ro. Tuy nhiên, các nghiên cứu chủ yếu dừng lại các tính toán dựa vào một dạng phân phối xác suất (đối xứng hoặc bất đối xứng) theo Rockafellar và Uryasev (2000). Trong khi đối với dữ liệu thực tế, việc dữ liệu chỉ tuân theo một phân phối xác suất hầu như là rất ít xảy ra, thông thường dữ liệu sẽ có dạng tuân theo một hỗn hợp các phân phối xác suất.

Sau khi xác định được dữ liệu tuân theo một hỗn hợp các phân phối xác suất, việc cần thiết tiếp theo là với mỗi mức $\alpha \in (0,1)$ cần xác định VaR_α và $CVaR_\alpha$. Tuy nhiên, một vấn đề cần đặt ra là tính toán các giá trị VaR_α và $CVaR_\alpha$ sao cho vừa đảm bảo nhanh chóng vừa đảm bảo chính xác, tức là các giá trị chênh lệch không nhiều và tương đối ổn định. Một phương pháp được đề nghị là phương pháp mô phỏng Monte Carlo và cần phải hiệu chỉnh phương pháp Monte Carlo trong tính toán này. Trong bài nghiên cứu này, tác giả đã chứng minh sự cần thiết phải hiệu chỉnh phương pháp Monte Carlo nếu muốn đảm bảo sai số xác suất so với giá trị α cho trước là nhỏ nhất (do cần phải kiểm tra lại có thật sự là xác suất từ giá trị nhỏ nhất có thể đến giá trị VaR_α đúng bằng α hay không).

Với phương pháp Monte Carlo đã hiệu chỉnh theo đề xuất của tác giả, các kết quả tính toán khá chính xác được minh họa trong các trường hợp hỗn hợp hai phân phối đối xứng, hai phân phối lệch, ba phân phối đối xứng, hai phân phối đối xứng và một phân phối lệch, một phân phối đối xứng và hai phân phối lệch cũng như hỗn hợp 10 phân phối xác suất trong trường hợp dữ liệu thực tế.

Thêm vào đó, vấn đề khó khăn nhất khi phân tích dữ liệu thực tế đó là cần xét xem dữ liệu thực tế tuân theo hỗn hợp của bao nhiêu phân phối xác suất và các tham số thành phần của phân phối xác suất này bằng bao nhiêu. Trong bài nghiên cứu, tác giả đã sử dụng chỉ số minAIC để lựa chọn phân phối xác suất trong khi thực tế có rất nhiều chỉ số để chọn hỗn hợp các phân phối xác suất phù hợp nhất với bộ dữ liệu. Chính vì vậy, trong các nghiên cứu tiếp theo, tác giả sẽ nghiên cứu các phương pháp tối ưu nhất trong việc lựa chọn hỗn hợp các phân phối xác suất phù hợp với bộ dữ liệu■

Lời cảm ơn:

Bài viết này là kết quả của Đề tài nghiên cứu khoa học cấp C - Đại học Quốc gia, năm 2018 được Đại học Quốc gia Thành phố Hồ Chí Minh tài trợ với tên đề tài: "Mô hình FBayes và ứng dụng trong phân tích Tài chính", mã số: C2018-34-04 (Chủ nhiệm: Phạm Hoàng Uyên).

Tài liệu tham khảo

Acerbi, C., & Tasche, D. (2002). Expected shortfall: A natural coherent alternative to value at risk. *Economic Notes, 31*(2), 379–388.

Adams, Z., Füss, R., & Gropp, R. (2014). Spillover effects among financial institutions: A state–dependent sensitivity value–at–risk approach. *Journal of Financial and Quantitative Analysis, 49*(3), 575–598.

Belles-Sampera, J., Guillén, M., & Santolino, M. (2014). Beyond value-at-risk: GlueVaR distortion risk measures. *Risk Analysis, 34*(1), 121–134.

Berkowitz, J., & O'Brien, J. (2002). How accurate are value-at-risk models at commercial banks?. *The Journal of Finance, 57*(3), 1093–1111.

Demir, I., Muthuswamy, J., & Walter, T. (2004). Momentum returns in Australian equities: The influences of size, risk, liquidity and return computation. *Pacific-Basin Finance Journal, 12*(2), 143–158.

Dietz, S., Bowen, A., Dixon, C., & Gradwell, P. (2016). 'Climate value at risk'of global financial assets. *Nature Climate Change, 6*(7), 676.

Duffie, D., & Pan, J. (1997). An overview of value at risk. *Journal of derivatives, 4*(3), 7–49.

Engle, R. F., & Manganelli, S. (2004). CAViaR: Conditional autoregressive value at risk by regression quantiles. *Journal of Business & Economic Statistics, 22*(4), 367–381.

Hall, P. (1990). Using the bootstrap to estimate mean squared error and select smoothing parameter in nonparametric problems. *Journal of Multivariate Analysis, 32*(2), 177–203.

Hendricks, D. (1997). Evaluation of value–at–risk models using historical data. *Economic Policy Review, 2*(1), 39–69.

Hennig, P., Stern, D., Herbrich, R., & Graepel, T. (2012). Kernel topic models. In *Artificial Intelligence and Statistics*, (pp. 511–519).

Huang, Y., Zheng, Q. P., & Wang, J. (2014). Two–stage stochastic unit commitment model including non–generation resources with conditional value–at–risk constraints. *Electric Power Systems Research, 116*, 427–438.

Jorion, P. (1996). Risk2: Measuring the risk in value at risk. *Financial Analysts Journal, 52*(6), 47–56.

Jorion, P. (2001). *Value at Risk: The New Benchmark for Managing Financial Risk* (3rd ed.). NY: McGraw-Hill Professional.

Krokhmal, P., Palmquist, J., & Uryasev, S. (2002). Portfolio optimization with conditional value–at–risk objective and constraints. *Journal of Risk, 4*, 43–68.

Lê Thanh Hoa, Phạm Hoàng Uyên, & Nguyễn Đình Thiên. (2017). Một phương pháp mới tìm khoảng mật độ hậu nghiệm cao nhất và ứng dụng. *Tạp chí Phát triển Kinh tế, 28*(10), 79–120.

Martins-Filho, C., Yao, F., & Torero, M. (2018). Nonparametric estimation of conditional value–at–risk and expected shortfall based on extreme value theory. *Econometric Theory, 34*(1), 23–67.

Pflug, G. C. (2000). Some remarks on the value-at-risk and the conditional value-at-risk. In *Probabilistic Constrained Optimization* (pp. 272–281). Springer, Boston, MA.

Robert, C. P. (2004). *Monte Carlo Methods*. John Wiley & Sons, Ltd.

Rockafellar, R. T., & Uryasev, S. (2000). Optimization of conditional value-at-risk. *Journal of Risk, 2*(3), 21–42.

Rockafellar, R. T., & Uryasev, S. (2002). Conditional value–at–risk for general loss distributions. *Journal of Banking & Finance, 26*(7), 1443–1471.

Rockafellar, R. T., Royset, J. O., & Miranda, S. I. (2014). Superquantile regression with applications to buffered reliability, uncertainty quantification, and conditional value–at–risk. *European Journal of Operational Research, 234*(1), 140–154.

Zhao, H., & Luo, Y. (2018). *An O (N) Sorting Algorithm: Machine Learning Sorting*. arXiv preprint arXiv:1805.04272.