

---

# HIỆU QUẢ TRONG DỰ BÁO GIÁ DẦU THÔ: MỘT SO SÁNH GIỮA MÔ HÌNH VAR, MÔ HÌNH LASSO VÀ MÔ HÌNH LSTM

**Nguyễn Đức Trung**

*Trường Đại học Ngân hàng Thành phố Hồ Chí Minh*

*Email: trungnd@buh.edu.vn*

**Lê Hoàng Anh**

*Trường Đại học Ngân hàng Thành phố Hồ Chí Minh*

*Email: anhhlh\_ync@buh.edu.vn*

Ngày nhận: 16/3/2021

Ngày nhận bản sửa: 31/5/2021

Ngày duyệt đăng: 05/6/2021

## **Tóm tắt**

*Bài báo này nhằm cung cấp một sự so sánh về hiệu quả dự báo giá dầu thô giữa các phương pháp phổ biến hiện nay. Trong đó, mô hình dự báo giá dầu được nhóm tác giả xây dựng trên các khía cạnh chính của động lực giá dầu là chính sách tiền tệ và tình trạng cạn kiệt dầu thô. Với dữ liệu được thu thập hàng tháng từ tháng 1 năm 1986 đến tháng 5 năm 2020, giá dầu thô được dự báo thông qua 3 mô hình là mô hình vector tự hồi quy (VAR), mô hình hồi quy toán tử co gọn và lựa chọn tối thiểu (LASSO), mô hình mạng bộ nhớ ngắn hạn dài hạn (LSTM). Kết quả cho thấy theo cả ba chỉ số là độ lệch sai số trung bình (RMSE), phần trăm sai số tuyệt đối trung bình (MAPE) và trung bình sai số bình phương (MSE), mô hình LSTM sẽ cho kết quả dự đoán chính xác hơn so với mô hình VAR và LASSO.*

**Từ khóa:** Mô hình VAR, Mô hình LASSO, Mô hình LSTM

**Mã JEL:** C53, C63, E31.

## **Crude oil price prediction efficiency: A comparison between the VAR model, the Lasso Model, and the long short-term memory model**

### **Abstract:**

*This paper aims to compare the efficiency of crude oil price forecasting methods that are currently in use. The oil price forecasting model is based on the main aspects of oil price dynamics, which include monetary policy and crude oil depletion. With data collected monthly from January 1986 to May 2020, Crude oil prices are forecasted through 3 models, including VAR, LASSO, LSTM. The results show that according to all three indicators RMSE, MAPE, and MSE, the LSTM model will provide the most accurate predictive result compared to the VAR and LASSO ones.*

**Keywords:** Var Model, Lasso Model, Long Short-Term Memory Model.

**JEL codes:** C53, C63, E31.

## **1. Giới thiệu nghiên cứu**

Dầu mỏ là một trong những nhiên liệu đầu vào quan trọng nhất của quá trình sản xuất đối với một nền kinh tế. Theo số liệu của công ty BP (2015), 50% lượng dầu mỏ sẽ được dùng để sản xuất năng lượng điện và nhiên liệu cho các phương tiện giao thông để vận chuyển hàng hóa ra thị trường; 50% còn lại được dùng cho hóa dầu để sản xuất các chất dẻo, dung môi, phân bón, nhựa đường, thuốc trừ sâu và nhiều loại sản phẩm khác. Do đó, dầu thô và biến động giá dầu thô có ảnh hưởng lớn đến kinh tế toàn cầu. Nhìn vào thị trường kinh tế thế giới, một điều hiển nhiên là giá dầu đóng góp vào giá trị của bất kỳ sản phẩm và dịch vụ nào đang lưu thông. Từ đó, giá dầu đóng góp lớn vào sự phát triển kinh tế của các quốc gia trên thế giới (Godarzi &

---

cộng sự, 2014).

Biến động của giá dầu thô thường tạo ra những tác động tích cực và tiêu cực đến nền kinh tế của các quốc gia trên thế giới. Khi giá dầu thô tăng cao, thu ngân sách quốc gia có lợi nhưng sẽ làm tăng chi phí đầu vào của nhiều ngành trong nền kinh tế, ảnh hưởng đến tăng trưởng sản lượng, mục tiêu kiềm chế lạm phát và chính sách tài chính - tiền tệ. Ngược lại, khi giá dầu thô giảm, chi phí nhập khẩu xăng dầu các loại giảm, mang lại nhiều lợi ích cho các công ty, vì đây là nhiên liệu đầu vào thiết yếu của nền kinh tế. Hơn nữa, giá dầu giảm tạo điều kiện cho hầu hết người dân tiết kiệm chi phí vận tải và làm tăng tổng cầu của nền kinh tế. Chính vì những lý do này, biến động giá dầu thô và dự báo những biến động đó đang trở thành mối quan tâm của các nước trên thế giới (Nguyen & cộng sự, 2018).

Trong những năm qua, các nhà nghiên cứu đã phát triển nhiều phương pháp dự báo giá dầu thô. Một số phương pháp dự báo biến động giá dầu có thể kể đến như phương pháp vectơ tự hồi quy chuyển mạch Markov (Aloui & Jammazi, 2009), phương pháp phương sai có điều kiện của sai số thay đổi tự hồi quy (ARCH) (Cheong, 2009), phương pháp phương sai có điều kiện của sai số thay đổi tự hồi quy tổng quát (GARCH) (Wei & cộng sự, 2010).

Bên cạnh các phương pháp dự báo truyền thống như đã kể trên, giá dầu cũng thường được ước tính bằng các phương pháp mới, điển hình là các phương pháp co gọn (shrinking). Đặc biệt, trong phân tích chuỗi thời gian, hồi quy toán tử co gọn và lựa chọn tối thiểu (Least Absolute Shrinkage and Selection Operator - LASSO) là kỹ thuật co gọn cung cấp một ước tính hệ số hồi quy gần đúng nhanh chóng và hiệu quả. Hồi quy LASSO thường được sử dụng vì nó có nhiều ưu điểm hơn so với cách tiếp cận bình phương bé nhất (OLS). Cụ thể, mô hình này có kết quả dự đoán chính xác hơn so với hồi quy OLS. Mặc dù, các ước lượng hệ số hồi quy thu được từ OLS là không chệch, nhưng phương sai cao. Trong khi đó, các hệ số hồi quy ước tính bằng phương pháp LASSO có thể sai lệch, nhưng phương sai nhỏ hơn và do đó độ chính xác của dự báo sẽ cao hơn (Zou & Hastie, 2005; Zou, 2006; Li & Tsiakas, 2017). Thêm vào đó, cách tiếp cận LASSO tạo ra các ước lượng không phụ thuộc vào giá trị P-value trong quá trình đánh giá các hệ số hồi quy khác 0. Đồng thời, cách tiếp cận này thường chọn các biến giải thích tốt nhất cho sự biến động của biến phụ thuộc. Trong nghiên cứu này, chúng tôi sử dụng phương pháp LASSO để dự đoán biến động giá dầu dựa trên một loạt các biến vĩ mô và vi mô. Sau đó, kết quả dự báo biến động giá dầu theo phương pháp LASSO được so sánh với phương pháp véc tơ tự hồi quy.

Trong vài năm gần đây, cùng với sự phát triển của khoa học máy tính, các phương pháp dựa vào trí tuệ nhân tạo, học máy bắt đầu được ứng dụng và cho thấy hiệu quả cao hơn các phương pháp trước đây trong dự báo giá dầu thô. Mặc dù các phương pháp vectơ tự hồi quy và LASSO có lợi thế do sự dễ dàng trong tính toán. Tuy nhiên, những phương pháp này thường dựa trên việc xây dựng mô hình từ các lý thuyết kinh tế và các nghiên cứu thực nghiệm. Trên thực tế, các mô hình lý thuyết thường không ổn định do mức độ phức tạp và thay đổi hành vi của các tác nhân trong nền kinh tế. Chính vì vậy, các nhà nghiên cứu đã phát triển các phương pháp dự báo không dựa trên bất kỳ lý thuyết kinh tế nào. Các phương pháp này không có giả định và thường có thể được điều chỉnh cho bất kỳ sự thay đổi hành vi nào. Theo cách này, để dự đoán sự biến động giá dầu, các nhà nghiên cứu đã sử dụng các phương pháp luận trí tuệ nhân tạo như mạng nơ-ron nhân tạo và học máy trong các mô hình dự đoán của họ (Mingming & Jinliang, 2012; Pan & cộng sự, 2009). Nhiều nhà nghiên cứu đã chỉ ra rằng mạng nơ-ron nhân tạo và học máy vượt trội hơn phương pháp tự hồi quy về độ chính xác trong dự báo. Cụ thể, Ramyar & Kianfar (2019) đã thực hiện một nghiên cứu dự báo giá dầu thô với mô hình mạng nơ-ron nhân tạo và mô hình tự hồi quy vector. Kết quả nghiên cứu cho thấy mô hình mạng nơ-ron nhân tạo chiếm ưu thế trong việc dự báo giá dầu thô. Trong nghiên cứu này, chúng tôi sử dụng mô hình bộ nhớ ngắn hạn dài hạn (Long Short-Term Memory model-LSTM), một dạng mạng nơ-ron nhân tạo, để dự báo giá dầu thô. Kết quả dự báo bằng phương pháp LSTM sẽ được so sánh với các kết quả của phương pháp vectơ tự hồi quy và phương pháp LASSO.

Bên cạnh việc ứng dụng các phương pháp truyền thống và hiện đại để dự báo giá dầu, không giống như

các nghiên cứu trước đây thường dự báo giá dầu thô dựa vào các giá trị trong quá khứ, nghiên cứu này xây dựng một mô hình kinh tế để dự báo giá dầu dựa trên sự cạnh tranh dầu thô, dự trữ dầu thô và chính sách tiền tệ. Chúng tôi sẽ thảo luận về tác động của các yếu tố này đến giá dầu thô trong các phần sau.

## 2. Các mô hình dự báo giá dầu

### 2.1. Mô hình VAR (Vector Autoregressive Model)

Mô hình VAR là một trong những mô hình dự báo hiệu quả, có thể mở rộng và dễ sử dụng để phân tích chuỗi thời gian đa biến. Mô hình VAR tỏ ra đặc biệt hữu ích để giải thích và dự báo các chuỗi thời gian kinh tế và tài chính. Mô hình này cũng cung cấp các dự báo vượt trội khi áp dụng với chuỗi thời gian đơn biến và xây dựng các phương trình đồng thời dựa trên lý thuyết.

Mô hình véc tơ tự hồi quy với  $p$  độ trễ (VAR( $p$ )) có dạng như sau:

$$Y_t = c + A_1 Y_{t-1} + A_2 Y_{t-2} + \dots + A_p Y_{t-p} + \varepsilon_t; t = 1, \dots, T \quad (1)$$

Trong đó,  $Y_t = (y_{1t}, y_{2t}, \dots, y_{nt})'$  là một véc tơ ( $n \times 1$ ) của các biến chuỗi thời gian,  $A_i$  là ma trận hệ số hồi quy cấp ( $n \times n$ ) và  $\varepsilon_t$  một véc tơ ( $n \times 1$ ) của các sai số nhiễu trắng có tương quan đồng thời với nhau.

Yêu cầu quan trọng nhất để ước lượng mô hình VAR là các chuỗi thời gian phải có tính dừng. Nói chung, một quá trình ngẫu nhiên được coi là có tính dừng nếu giá trị trung bình và phương sai của nó không thay đổi theo thời gian, và giá trị hiệp phương sai giữa hai giai đoạn chỉ phụ thuộc vào khoảng cách và thời gian trễ giữa hai giai đoạn. Hai khoảng này không phụ thuộc vào phép đo hiệp phương sai thực tế. Trong nghiên cứu này, việc kiểm định tính dừng được thực hiện bằng kiểm định Augmented Dickey-Fuller (ADF). Sau khi kiểm định tính dừng của các chuỗi thời gian, việc tìm ra độ trễ tối ưu cho mô hình VAR được thực hiện dựa trên các tiêu chí như thống kê LR (LR), tiêu chí thông tin Akaike (AIC), tiêu chí thông tin Hannan-Quinn (HQ), tiêu chí thông tin Schwarz (SC), Lỗi dự đoán cuối cùng (FPE).

### 2.2. Mô hình LASSO (Least Absolute Shrinkage and Selection Operator)

Mô hình LASSO có thể được phát triển từ việc xem xét  $N$  quan sát của biến phụ thuộc  $y_i$  và  $p$  biến giải thích  $x_i = (x_{i1}, \dots, x_{ip})$ . Trong đó,  $x_i$  và  $y_i$  tương ứng thuộc  $R^p$  và  $R$ .

Mục đích nghiên cứu của chúng tôi là dự đoán kết quả của biến phụ thuộc dựa trên các biến giải thích. Kết quả sẽ được dự đoán bằng cách sử dụng mô hình hồi quy tuyến tính sau:

$$f(x) = \beta_0 + \sum_{j=1}^p x_{ij} \beta_j \quad (2)$$

Do đó, mô hình hồi quy tuyến tính được tham số hóa với véc tơ trọng số

$B = (\beta_1, \dots, \beta_p)^T \in R^p$  và hệ số chặn  $\beta_0 \in R$ . Ước lượng OLS với cặp  $(\beta_0, \beta)$  dựa trên việc tối thiểu hóa bình phương của sai số như sau:

$$\text{minimize}_{\beta_0, \beta} \left\{ \frac{1}{N} \sum_{i=1}^N \left( y_i - \beta_0 - \sum_{j=1}^p x_{ij} \beta_j \right)^2 \right\} = \frac{1}{N} \|y - \beta_0 \mathbf{1} - X\beta\|^2 \quad (3)$$

Trong đó,  $y = (y_1, \dots, y_N)^T$ ,  $X$  là ma trận cấp ( $N \times p$ ) và  $\mathbf{1} = (1, \dots, 1)^T$ . Lời giải cho (3) như sau:

$$\beta = (X^T X)^{-1} X^T y$$

Hastie & cộng sự (2015) cho rằng có nhiều cách giải thích khác nhau cho sự cần thiết phải sửa đổi mô hình hồi quy OLS. Lời giải thích đầu tiên là ước tính độ chính xác. Hồi quy OLS được ước tính để các ước lượng có độ chệch thấp nhưng phương sai lại lớn. Do đó, độ chính xác của dự báo thường có thể tăng lên bằng cách giảm số lượng hệ số hồi quy hoặc đặt các hệ số nhất định thành 0. Do đó, việc chấp nhận một số sai lệch trong ước tính hệ số hồi quy theo cách làm trên có thể giảm thiểu phương sai của mô hình và do đó tăng độ chính xác của các dự báo. Cách giải thích thứ hai là tập trung vào một số biến giải thích có ảnh hưởng lớn nhất (các biến giải thích có ảnh hưởng thấp sẽ có hệ số hồi quy bằng 0) đến biến phụ thuộc sẽ có

tác dụng dự đoán tốt hơn.

Việc sửa đổi mô hình OLS thành mô hình hồi quy LASSO có thể được thực hiện như sau:

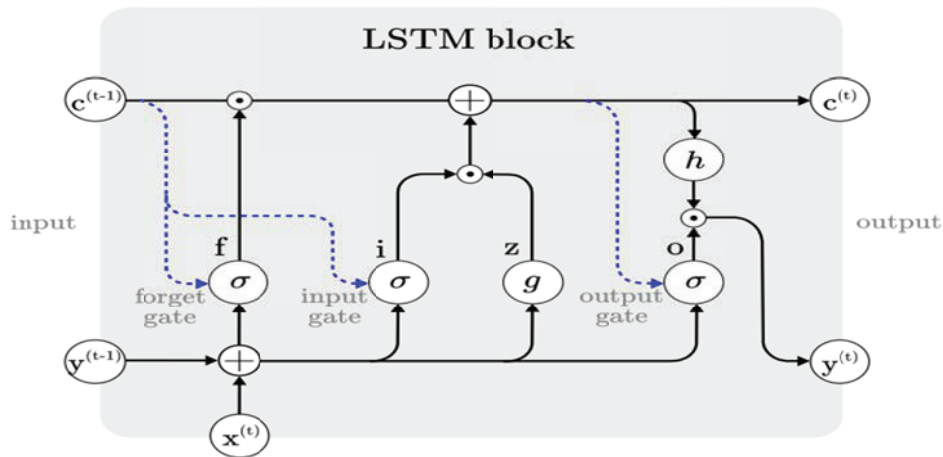
$$\underset{\beta_0, \beta}{\text{minimize}} \left\{ \frac{1}{N} \sum_{i=1}^N \left( y_i - \beta_0 - \sum_{j=1}^p x_{ij} \beta_j \right)^2 + \lambda \sum_{j=1}^p \beta_j \right\} = \frac{1}{N} \|y - \beta_0 \mathbf{1} - X\beta\|^2 + \lambda \|\beta\| \quad (4)$$

Vấn đề chính trong mô hình hồi quy LASSO là chọn giá trị tối ưu của lambda. Việc lựa chọn giá trị lambda tối ưu này thường được thực hiện thông qua các phương pháp như Cross-validation, Theory-driven, và Information Criteria. Trong nghiên cứu này, giá trị tối ưu của lambda được chọn bằng phương pháp Cross-validation. Mục đích của phương pháp Cross-validation là để đánh giá hiệu suất của công cụ ước tính như một công cụ dự đoán ngoài mẫu.

### 2.3. Mô hình LSTM (Long Short-Term Memory)

Trong những năm gần đây, Mô hình LSTM đã được phát triển và điều chỉnh để phù hợp hơn với các dự

Hình 1: Cấu trúc của một tế bào LSTM



Nguồn: Fischer & Krauss (2018).

báo trong kinh tế, tài chính (Fischer & Krauss, 2018; Yan & Ouyang, 2018). Do các đặc điểm động như không tuyến tính, không dừng và tương quan đồng thời, dữ liệu tài chính đặt ra một thách thức đáng kể đối với công tác dự báo. Tuy nhiên, nghiên cứu của Fischer & Krauss (2018) đã chứng minh rằng mô hình LSTM có thể giải quyết các vấn đề khó khăn trên vì phương pháp này không quan tâm đến dạng hàm hoặc các mối quan hệ lý thuyết của các biến số trong mô hình.

Về cơ bản, các LSTM sử dụng khái niệm cổng (gate): một cơ chế nhân thông minh các thành phần xác định hoạt động của từng tế bào (ô nhớ) riêng lẻ. LSTM cập nhật trạng thái tế bào của nó bằng cách kích hoạt các cổng. Đầu vào được cấp cho LSTM được đưa vào các cổng khác nhau chi phối hoạt động của bộ nhớ tế bào: ghi (cổng nhập – input gate), đọc (cổng xuất – output gate), hoặc đặt lại (cổng quên – forget gate).

Việc tính toán giá trị ẩn  $y_t$  của một tế bào LSTM được cập nhật tại mỗi thời điểm  $t$ . Biểu diễn vector (vector biểu thị tất cả các đơn vị trong một lớp) của bản cập nhật của một lớp LSTM được biểu thị là cổng nhập  $i_t$ , cổng quên  $f_t$ , cổng xuất  $o_t$ , ô nhớ  $c_t$  và trạng thái ẩn  $h_t$ .

$$\begin{aligned} f_t &= \sigma(W_f x_t + R_f y_{t-1} + p_f \odot c_{t-1} + b_f) \\ i_t &= \sigma(W_i x_t + R_i y_{t-1} + p_i \odot c_{t-1} + b_i) \\ z_t &= g(W_z x_t + R_z y_{t-1} + b_z) \\ c_t &= z_t \odot i_t + c_{t-1} \odot f_t \\ o_t &= \sigma(W_o x_t + R_o y_{t-1} + p_o \odot c_t + b_o) \end{aligned}$$

trong đó  $W_f, W_p, W_z, W_o, R_f, R_p, R_z, R_o$  là các trọng số tương ứng với  $x_t$  và  $y_{t-1}$ , trong khi  $b_p, b_z, b_o$  là các véc tơ độ chệch.  $\odot$  biểu thị toán tử nhân của 2 véc tơ.  $\oplus$  biểu thị việc thêm thông tin.

Trong các bước trên,  $\sigma, g$  và  $h$  biểu thị cho các hàm kích hoạt. Hàm sigmoid  $\sigma(x) = \frac{1}{1+e^{1-x}}$  được sử dụng như một hàm kích hoạt công, trong khi hàm hyperbolic  $g(x) = h(x) = \tanh(x)$  được sử dụng như một hàm kích hoạt công nhập và công xuất của một tế bào.

Trong nghiên cứu này, các tác giả dự báo giá dầu thô trong những tháng tới theo cả VAR, LASSO và LSTM. Phương pháp dự báo tốt nhất sẽ được chọn dựa trên ba chỉ số là độ lệch sai số trung bình (RMSE), phần trăm sai số tuyệt đối trung bình (MAPE) và trung bình sai số bình phương (MSE):

$$RMSE = \sqrt{\frac{\sum_{i=1}^N (y_i - \hat{y}_i)^2}{N}}$$

$$MAPE = \frac{\sum_{i=1}^N \frac{|y_i - \hat{y}_i|}{y_i}}{N}$$

$$MSE = \frac{\sum_{i=1}^N (y_i - \hat{y}_i)^2}{N}$$

trong đó  $y_i$  là các giá trị thực tế của đầu ra,  $\hat{y}_i$  là các giá trị dự báo thu được từ các mô hình VAR, LASSO, và LSTM.

### 3. Mô hình và dữ liệu nghiên cứu

Nghiên cứu này dự báo giá dầu thô dựa trên các yếu tố mô hình ảnh hưởng đến giá dầu. Như đã nhấn mạnh trong phần giới thiệu, các yếu tố thúc đẩy được tiếp cận trên hai khía cạnh: chính sách tiền tệ và tình trạng cạn kiệt dầu thô.

Chính sách tiền tệ được coi là tác động từ phía cầu đến giá dầu thô. Nó có ảnh hưởng đáng kể đến thị trường hàng hóa và do đó, đến giá dầu thô. Sự co giãn của cầu đối với giá dầu thô, kết hợp với sự khó khăn của nguồn cung, có thể dẫn đến giá dầu thô tăng nhanh. Ảnh hưởng của chính sách tiền tệ đến thị trường dầu thô được thực hiện thông qua hai kênh chính là lãi suất và tỷ giá đô la. Lãi suất có tác động mạnh mẽ đến nhu cầu dầu thô, vì lãi suất thấp dẫn đến hoạt động kinh tế tăng lên (Askari & Krichene, 2010). Ngoài ra, tỷ giá hối đoái đô la Mỹ, là đồng tiền chính được sử dụng để báo giá dầu, sẽ ảnh hưởng trực tiếp đến giá danh nghĩa của dầu (Askari & Krichene, 2010). Trong nghiên cứu này, chúng tôi coi lãi suất của Ngân hàng Dự trữ Liên bang và tỷ giá hối đoái hiệu lực của đồng đô la Mỹ là hai biến số trong mô hình các nhân tố ảnh hưởng đến giá dầu thô. Trong khi đó, do giá dầu thô được quy định tại Sở giao dịch hàng hóa New York, biến động giá dầu thô có thể được coi là một phần của diễn biến thị trường hàng hóa. Do đó, chỉ số giá cả thị trường của hàng hóa cũng được coi là một yếu tố ảnh hưởng đến giá dầu thô.

Hơn nữa, tình trạng cạn kiệt dầu đang trở thành mối quan tâm lớn, đặc biệt là do nguồn dự trữ dầu ở các nước sản xuất dầu lớn đang cạn kiệt. Các địa điểm sản xuất dầu lớn, chẳng hạn như ở Texas và Ả Rập Saudi, đang chứng kiến sự sụt giảm đáng kể về sản lượng. Trên thực tế, một số người đã bắt đầu suy đoán rằng mỏ dầu Ghawar, mỏ dầu thông thường lớn nhất trên thế giới, đang suy thoái. Điều này cho thấy rằng sự suy giảm sản lượng dầu trên toàn cầu có thể xảy ra (Hamilton, 2009). Do đó, chức năng thiết yếu của dầu thô ảnh hưởng đến sản lượng dầu và tồn kho dầu trên mặt đất (Brandt, 2010). Các vấn đề khác, chẳng hạn như căng thẳng địa chính trị và bất ổn chính trị, chẳng hạn như cuộc khủng hoảng ở Iraq và Nigeria, đóng một vai trò quan trọng trong việc sản xuất dầu thô và ảnh hưởng đến sản lượng dầu thô. Do đó, điều quan trọng là phải đưa các yếu tố sản xuất vào mô hình dự báo. Trong nghiên cứu này, sản lượng dầu, nguồn cung dầu và tồn kho dầu được đưa vào mô hình dự báo giá dầu.

Do đó, mô hình đề xuất để dự báo giá dầu thô, như sau:

**Bảng 1: Mô tả các biến số trong mô hình**

Biến số	Mô tả	Nguồn dữ liệu
$P_t$	Giá dầu thô	United States' Energy Information Administration (EIA)
$IR_t$	Lãi suất của Ngân hàng Dự trữ Liên bang	FRED Economic Data
$REER_t$	Tỷ giá hối đoái hiệu lực của đồng đô la Mỹ (REER)	FRED Economic Data
$OP_t$	Sản lượng dầu: được tính bằng khối lượng dầu thô được sản xuất từ các bể chứa dầu. Lượng sản xuất này được đo bằng khối lượng dầu được giao từ các bồn chứa cho thuê (tức là điểm trung chuyển lưu ký) đến đường ống, xe tải hoặc các phương tiện khác để vận chuyển đến các nhà máy lọc dầu.	United States' Energy Information Administration (EIA)
$OS_t$	Nguồn cung dầu: bao gồm sản lượng dầu được sản xuất mới tại các bể chứa, sản lượng được tái tạo, sản lượng dầu đang tồn kho, trừ đi sản lượng dầu đầu vào của các nhà máy lọc dầu.	United States' Energy Information Administration (EIA)
$OC_t$	Sản lượng tiêu thụ: đo lường sản lượng dầu được tiêu thụ bởi cho các hoạt động của nền kinh tế, kể cả đầu vào của các nhà máy lọc dầu.	United States' Energy Information Administration (EIA)
$CI_t$	Chỉ số giá cả thị trường hàng hóa	FRED Economic Data

*Nguồn: đề xuất của tác giả.*

$$P_t = \beta_0 + \beta_1 IR_t + \beta_2 REER_t + \beta_3 OP_t + \beta_4 OC_t + \beta_5 OS_t + \beta_6 CI_t + \varepsilon_t \quad (5)$$

Các biến số trong mô hình được mô tả trong Bảng 1.

Chuỗi thời gian trong Mô hình (5) được thu thập hàng tháng từ tháng 1 năm 1986 đến tháng 5 năm 2020. Như vậy, mỗi chuỗi thời gian sẽ bao gồm 413 quan sát, trong đó 351 quan sát đầu tiên được sử dụng cho các mô hình huấn luyện (VAR, LASSO, LSTM) và 62 quan sát tiếp theo được sử dụng để đánh giá mức độ dự đoán của từng mô hình thông qua các tiêu chí như RMSE, MAPE và MSE.

#### 4. Kết quả nghiên cứu

##### 4.1. Kết quả dự báo bằng mô hình VAR

Trước khi ước lượng mô hình VAR, chúng tôi kiểm định tính dừng của các chuỗi thời gian trong mô hình (5). Kết quả được hiển thị trong Bảng 2.

Kết quả kiểm định tính vững cho thấy tất cả các chuỗi thời gian trong mô hình (5) không dừng ở chuỗi gốc nhưng lại dừng ở sai phân bậc 1. Do đó, sai phân bậc 1 của các chuỗi thời gian sẽ được sử dụng để ước tính mô hình VAR.

**Bảng 2: Kết quả kiểm định tính dừng**

Biến	Thống kê t (p-value)	Biến	Thống kê t (p-value)
CI	-1.033 (0.743)	$\Delta$ CI	-13.019*** (0.000)
IR	-1.447 (0.560)	$\Delta$ IR	-7.109*** (0.000)
OC	-2.061 (0.261)	$\Delta$ OC	-3.071** (0.030)
OP	-2.211 (0.203)	$\Delta$ OP	-2.789* (0.061)
OS	-1.096 (0.719)	$\Delta$ OS	-3.763*** (0.004)
REER	-2.188 (0.211)	$\Delta$ REER	-13.447*** (0.000)
P	-2.078 (0.254)	$\Delta$ P	-12.315*** (0.000)

\*\*\*, \*\*, \* tương ứng với mức ý nghĩa thống kê 1%, 5%, và 10%

Nguồn: kết quả tính toán của tác giả.

**Bảng 3: Kết quả lựa chọn độ trễ tối ưu**

Độ trễ	LogL	LR	FPE	AIC	SC	HQ
0	-849.2070	NA	8.624347	4.992461	5.070782*	5.023658
1	-849.1776	0.057478	8.673317	4.998120	5.087630	5.033774
2	-847.8861	2.515077	8.658623	4.996421	5.097119	5.036532
3	-844.2072	7.143397	8.524458	4.980800	5.092687	5.025368
4	-840.5110	7.155331	8.391536	4.965079	5.088155	5.014103
5	-839.6757	1.612095	8.399655	4.966039	5.100304	5.019521
6	-833.4951	11.89277	8.149778	4.935831	5.081285	4.993770
7	-829.0908	8.448945*	7.989674*	4.915982*	5.072624	4.978377*
8	-829.0100	0.154577	8.032694	4.921341	5.089172	4.988193

\* cho thấy độ trễ được lựa chọn bởi các tiêu chí

LR: tiêu chí LR

FPE: sai số dự báo cuối cùng

AIC: tiêu chí thông tin Akaike

SC: tiêu chí thông tin Schwarz

HQ: tiêu chí thông tin Hannan-Quinn

Nguồn: kết quả tính toán của tác giả.

Để ước lượng mô hình VAR, chúng tôi tiếp tục thực hiện kiểm định độ trễ tối ưu. Kết quả được hiển thị trong Bảng 3.

Bảng 3 cho thấy độ trễ tối ưu cho mô hình VAR được lựa chọn là 7 bởi các tiêu chí lựa chọn là LR, FPE, AIC, HQ.

Do đó, mô hình VAR sẽ được ước lượng với sai phân bậc 1 và độ trễ 7 của các chuỗi thời gian trong mô hình (5). Kết quả ước lượng cụ thể được trình bày tại Bảng 4.

Tiếp theo, tác giả sử dụng mô hình VAR để dự đoán giá dầu thô trong giai đoạn từ tháng 4 năm 2015 đến

**Bảng 4: Kết quả ước lượng mô hình (5) bằng VAR**

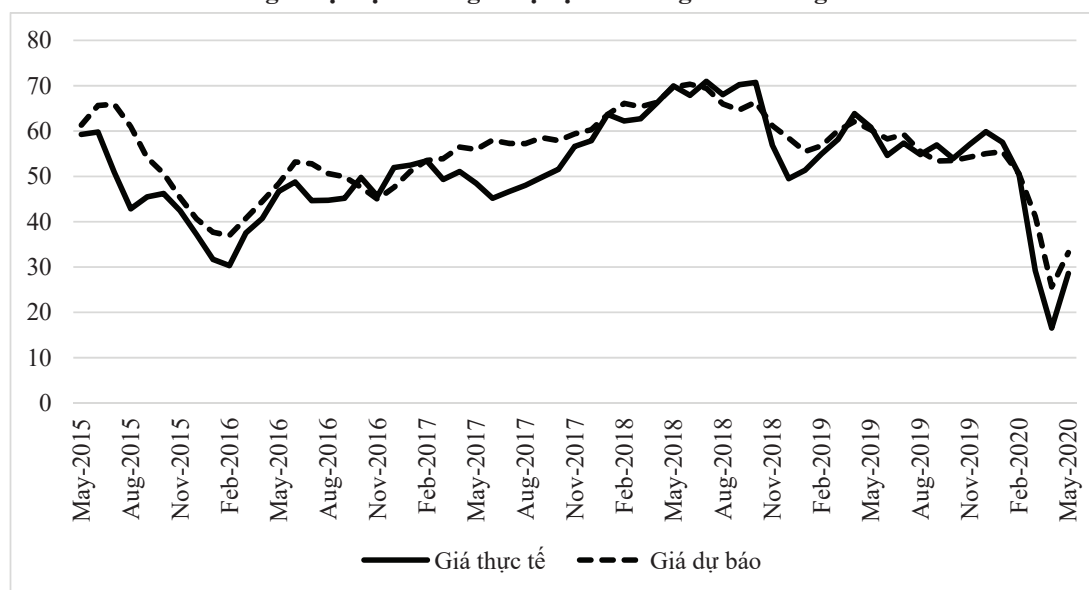
D(P)	Hệ số hồi quy	Sai số chuẩn	Thống kê t
D(P(-1))	-0.011779	0.04385	-0.26864
D(P(-2))	-0.026062	0.04154	-0.62740
D(P(-3))	-0.062937	0.04141	-1.51993
D(P(-4))	-0.082009**	0.04175	-1.96434
D(P(-5))	0.002738	0.04234	0.06466
D(P(-6))	-0.103884**	0.04236	-2.45266
D(P(-7))	-0.119215***	0.04075	-2.92529
C	-0.303259**	0.15319	-1.97965
D(CI)	1.854726***	0.11222	16.5270
D(IR)	0.273187	0.77997	0.35025
D(OC)	7.30E-06	8.7E-06	0.83642
D(OP)	-1.77E-05	2.6E-05	-0.67935
D(OS)	-1.56E-05*	8.3E-06	-1.89466
D(REER)	-0.366765***	0.13295	-2.75870

Ký hiệu  $D(x)$  là sai phân bậc 1 of  $x$ ,  $P(-t)$  là độ trễ thứ  $t$  của biến  $P$

\*\*\*, \*\*, \* tương ứng với mức ý nghĩa thống kê 1%, 5%, và 10%

Nguồn: kết quả tính toán của tác giả.

**Hình 2: giá trị thực tế và giá trị dự báo bằng VAR của giá dầu thô**



tháng 5 năm 2020. Giá trị dự đoán và giá trị thực tế của giá dầu thô cho giai đoạn từ tháng 4 năm 2015 đến tháng 5 năm 2020 được thể hiện trong Hình 2.

#### 4.2. Kết quả dự báo bằng mô hình LASSO

Để dễ dàng so sánh độ chính xác của dự báo với mô hình VAR, mô hình 5 được ước lượng bằng phương pháp LASSO, cũng bao gồm tất cả các biến được ước lượng bằng phương pháp VAR. Cụ thể, mô hình (5) bao gồm độ trễ 7 của các biến. Tất cả các biến này đều ở dạng sai phân bậc 1, như trong mô hình VAR.

Để ước lượng mô hình 5 bằng phương pháp LASSO, cần xác định giá trị tối ưu của tham số lambda. Trong nghiên cứu này, giá trị lambda tối ưu được chọn bằng phương pháp Cross-Validation. Kết quả lựa



**Bảng 5: Kết quả lựa chọn tham số lambda**

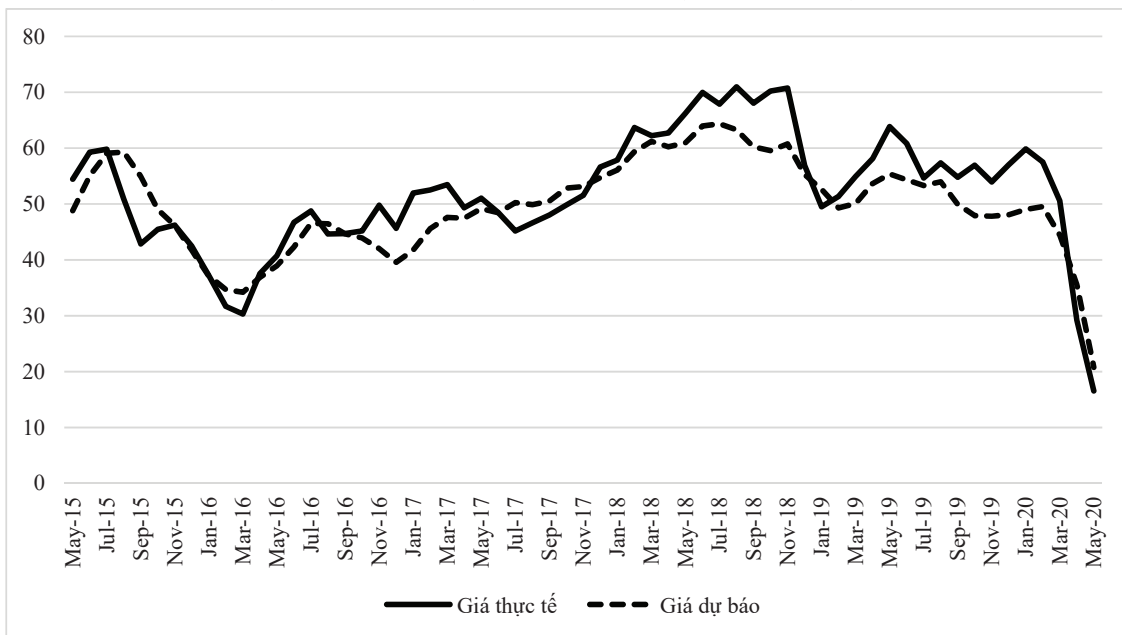
ID	Mô tả	lambda	Số lượng hệ số hồi quy khác 0	R <sup>2</sup> ngoài mẫu	Lỗi dự đoán trung bình CV
1	first lambda	3.1778	0	0.0019	18.5397
40	lambda before	0.0844	9	0.5626	8.0941
*41	selected lambda	0.0769	10	0.5627	8.0924
42	lambda after	0.0701	10	0.5626	8.0950
46	last lambda	0.0483	10	0.5611	8.1217

Nguồn: Kết quả tính toán của tác giả.

chọn giá trị Lambda được hiển thị trong Bảng 5.

Bảng 5 cho thấy giá trị lambda được chọn là 0,0769 với sai số dự đoán trung bình CV là 8,0924. Ngoài ra, tại giá trị lambda này, có 10 biến có hệ số hồi quy có giá trị khác 0. Các biến tương ứng là D(P(-2)), D(P(-3)), D(P(-4)), D(P(-6)), D(P(-7)), D(OC), D(OS), D(REER), D(IR), D(CI).

Tiếp theo, tác giả sử dụng mô hình LASSO để dự đoán giá dầu thô trong giai đoạn từ tháng 4 năm 2015

**Hình 3: giá trị thực tế và giá trị dự báo bằng LASSO của giá dầu thô**

đến tháng 5 năm 2020. Giá trị dự đoán và giá trị thực tế của giá dầu thô cho giai đoạn từ tháng 4 năm 2015 đến tháng 5 năm 2020 được thể hiện trong Hình 3.

#### 4.3. Kết quả dự báo bằng mô hình LSTM

Dữ liệu đầu vào của mô hình LSTM thường được biến đổi để nâng cao hiệu quả dự báo (Basheer và Hajmeer 2000). Dữ liệu được sử dụng trong bài báo này được biến đổi để nằm trong đoạn  $[-1, 1]$  bằng cách sử dụng công thức sau:

$$x_{scale} = 2 \times \frac{x - x_{min}}{x_{max} - x_{min}} - 1$$

Giá dầu thô được thu thập hàng tháng từ tháng 01 năm 1986 đến tháng 5 năm 2020. Như vậy, chuỗi thời gian của Giá dầu thô sẽ bao gồm 413 quan sát, trong đó 351 quan sát đầu tiên được sử dụng để huấn luyện

mô hình LSTM và 62 quan sát tiếp theo được sử dụng để đánh giá mức độ dự báo của mô hình thông qua các tiêu chí như RMSE, MAPE, MSE.

Bây giờ, chúng ta thêm lớp LSTM và lớp Dense chỉ định đầu ra của 1 đơn vị. Tiếp theo, mô hình phù hợp để chạy trên 100 epoch (Một Epoch được tính là khi chúng ta đưa tất cả dữ liệu vào mạng LSTM 1 lần)

**Bảng 6: Tóm tắt mô hình LSTM**

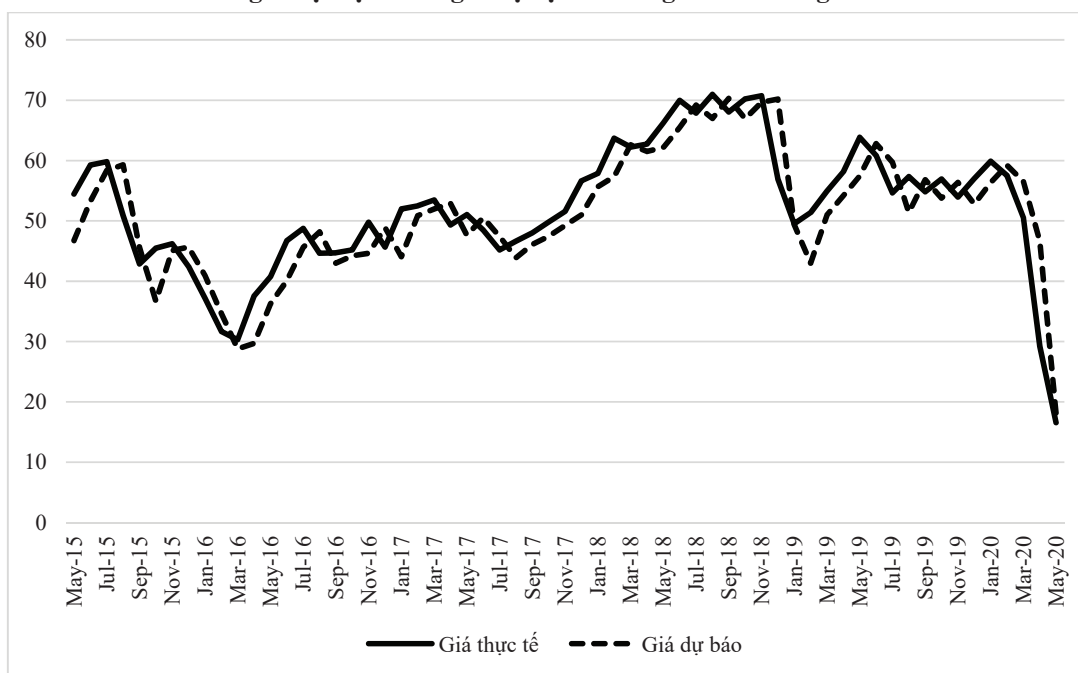
Model: "sequential\_2"

Layer (type)	Output shape	Param #
lstm_2 (LSTM)	(1, 1)	12
dense_1 (Dense)	(1, 1)	2

Total params: 14  
Trainable params: 14  
Non-trainable params: 0

Nguồn: Kết quả tính toán của tác giả.

**Hình 4: giá trị thực tế và giá trị dự báo bằng LSTM của giá dầu thô**



với kích thước batch là 1 (chúng ta không thể đưa hết toàn bộ dữ liệu vào huấn luyện trong 1 epoch, vì vậy chúng ta cần phải chia tập dữ liệu thành các phần (number of batch), mỗi phần có kích thước batch riêng).

Tiếp theo, nhóm tác giả sử dụng mô hình LSTM để dự đoán giá dầu thô trong giai đoạn từ tháng 4 năm 2015 đến tháng 5 năm 2020. Giá trị dự đoán và giá trị thực tế của giá dầu thô cho giai đoạn từ tháng 4 năm 2015 đến tháng 5 năm 2020 được thể hiện trong Hình 4.

Cuối cùng, chúng tôi đo lường các chỉ số RMSE, MAPE và MSE để so sánh độ chính xác dự báo của ba mô hình (VAR, LASSO, LSTM). Kết quả của phép tính được thể hiện trong Bảng 7.

Bảng 7 cho thấy theo cả 3 chỉ số RMSE, MAPE và MSE, mô hình LSTM sẽ cho kết quả dự đoán chính xác hơn so với mô hình VAR và LASSO.

**Bảng 7: Kết quả tính toán RMSE, MAPE và MSE của các mô hình VAR, LASSO và LSTM**

	VAR	LASSO	LSTM
RMSE	5.876137	5.407809	4.971465
MAPE	0.103235	0.086833	0.082162
MSE	34.52898	29.24439	24.71546

Nguồn: kết quả tính toán của tác giả.

## 5. Kết luận

Đóng góp đầu tiên của bài báo này là cung cấp mô phỏng chi tiết về các yếu tố tác động đến giá dầu xuất phát từ khía cạnh chính sách tiền tệ và tình trạng cạn kiệt dầu thô. Cụ thể, kết quả ước lượng mô hình VAR cho thấy giá dầu thô chịu sự tác động có ý nghĩa thống kê của giá dầu trong quá khứ với độ trễ tác động có ý nghĩa thống kê là 4 tháng. Bên cạnh đó, giá dầu thô còn bị tác động bởi chính sách tiền tệ thể hiện thông qua tỷ giá hối đoái hiệu lực của đồng đô la Mỹ (REER) và biến động giá của các hàng hóa khác thể hiện thông qua chỉ số giá cả thị trường (CI). Kết quả về tác động của chính sách tiền tệ và tình trạng cạn kiệt dầu thô đến giá dầu thô một lần nữa được khẳng định thông qua kết quả của mô hình LASSO. Cụ thể, kết quả của mô hình LASSO cho thấy các biến số có ảnh hưởng đáng kể đến giá dầu thô bao gồm giá dầu thô trong quá khứ, sản lượng tiêu thụ, nguồn cung cầu, tỷ giá hối đoái hiệu lực của đồng đô la Mỹ, lãi suất của Ngân hàng Dự trữ Liên bang và chỉ số giá cả thị trường hàng hóa. Như vậy, kết quả ước lượng của mô hình VAR và mô hình LASSO có sự nhất quán với nhau về tác động của chính sách tiền tệ và tình trạng cạn kiệt dầu thô đến giá dầu thô. Các kết quả này cũng phù hợp với nghiên cứu của Cheong (2009), Askari & Krichene (2010).

Đóng góp thứ hai của bài báo này là phân tích hiệu suất dự báo giá dầu thô của các mô hình kinh tế lượng và mô hình trí tuệ nhân tạo. Với tính chất biến động và phi tuyến của giá dầu thô, trong bài báo này, mô hình LSTM như một công cụ hiệu quả để mô tả các quan hệ ảnh xạ phi tuyến được phát triển để dự đoán giá dầu thô. Bên cạnh đó, các mô hình dự báo phổ biến khác cũng được chúng tôi ước lượng như mô hình véc tơ tự động hồi quy (VAR) và mô hình toán tử lựa chọn và thu hẹp tuyệt đối bé nhất (LASSO). Các kết quả dự báo của các mô hình sau đó được so sánh với nhau để tìm ra mô hình dự báo tốt nhất. Kết quả cho thấy theo cả 3 chỉ số RMSE, MAPE và MSE, mô hình LSTM sẽ cho kết quả dự đoán chính xác hơn so với mô hình VAR và LASSO. Kết quả nghiên cứu này cũng phù hợp với nghiên cứu của Fischer và Krauss (2018) khi ứng dụng mô hình LSTM để dự báo các chuỗi dữ liệu tài chính.

## Tài liệu tham khảo

- Aloui, C. & Jammazi, R. (2009), 'The effects of crude oil shocks on stock market shifts behaviour: A regime switching approach', *Energy Economics*, 31(5), 789-799, DOI: <https://doi.org/10.1016/j.eneco.2009.03.009>.
- Askari, H. & Krichene, N. (2010), 'The impact of monetary policy on oil process parameters and market expectations', *Global Finance Journal*, 21(2), 186-200, DOI: <https://doi.org/10.1016/j.gfj.2010.06.005>.
- BP (2015), *BP Statistical Review of World Energy June 2015*, retrieved on May 29<sup>th</sup> 2021, from: <<https://www.ourenergypolicy.org/resources/bp-statistical-review-of-world-energy-june-2015/>>.
- Brandt, R.A. (2010), 'Review of mathematical models of future oil supply: Historical overview and synthesizing critique', *Energy*, 35(9), 3958-3974, DOI: <https://doi.org/10.1016/j.energy.2010.04.045>.
- Cheong, C.W. (2009), 'Modeling and forecasting crude oil markets using ARCH-type models', *Energy Policy*, 37(6), 2346-2355, DOI: <https://doi.org/10.1016/j.enpol.2009.02.026>.

- 
- Fischer, T. & Krauss, C. (2018), 'Deep learning with long short-term memory networks for financial market predictions', *European Journal of Operational Research*, 270(2), 654-669, DOI: <https://doi.org/10.1016/j.ejor.2017.11.054>.
- Godarzi, A.A., Amiri, R.M., Talaei, A. & Jamasb, T. (2014), 'Predicting oil price movements: A dynamic artificial neural network approach', *Energy Policy*, 68, 371-382., DOI: <https://doi.org/10.1016/j.enpol.2013.12.049>.
- Hamilton, J.D. (2009), 'Understanding crude oil prices', *The Energy Journal*, 30(2), 179-206, DOI: <https://doi.org/10.1073/pnas.0703993104>.
- Hastie, T., Tibshirani, R. & Wainwright, M. (2015), *Statistical Learning with Sparsity* (1st Edition), New York: Chapman and Hall/CRC.
- Li, J. & Tsiakas, I. (2017), 'Equity premium prediction: The role of economic and statistical constraints', *Journal of Finance Markets*, 36, 56-75.
- Mingming, T. & Jinliang, Z. (2012), 'A multiple adaptive wavelet recurrent neural network model to analyze crude oil prices', *Journal of Economics and Business*, 64(4), 275-286. DOI: <https://doi.org/10.1016/j.jeconbus.2012.03.002>.
- Pan, H., Haidar, I. & Kulkarni, S. (2009), 'Daily prediction of short-term trends of crude oil prices using neural networks exploiting multimarket dynamics', *Frontiers of Computer Science in China*, 3(2), 177-191, DOI: <https://doi.org/10.1007/s11704-009-0025-3>.
- Wei, Y., Wang, Y. & Huang, D. (2010), 'Forecasting crude oil market volatility: Further evidence using GARCH-class models', *Energy Economics*, 32(6), 1477-1484, DOI: <https://doi.org/10.1016/j.eneco.2010.07.009>.
- Yan, H. & Ouyang, H. (2018), 'Financial Time Series Prediction Based on Deep Learning', *Wireless Personal Communications*, 102, 683-700, DOI: <https://doi.org/10.1007/s11277-017-5086-2>.
- Zou, H. (2006), 'The adaptive lasso and its oracle properties', *Journal of the American Statistical Association*, 101, 1418-1429.
- Zou, H. & Hastie, T. (2005), 'Regularization and variable selection via the elastic net', *Journal of the Royal Statistical Society, Series B (Statistical Methodology)*, 67(2), 301-320.
- Ramyar, S. & Kianfar, F. (2019), 'Forecasting Crude Oil Prices: A Comparison Between Artificial Neural Networks and Vector Autoregressive Models', *Computational Economics*, 53, 743-761, DOI: <https://doi.org/10.1007/s10614-017-9764-7>.