

Ứng dụng các mô hình dự báo tiên tiến trong phân tích chính sách môi trường: Kinh nghiệm thực tiễn từ Vương quốc Anh và hàm ý cho Việt Nam

Đỗ Huyền Vy¹

Ngày nhận bài: 10/11/2025 | Ngày gửi phản biện: 12/11/2025 | Ngày duyệt đăng: 21/11/2025

Tóm tắt: Trước cam kết Net-Zero của Việt Nam, việc xác định đúng các yếu tố tác động môi trường là cấp thiết. Bài viết này sử dụng nghiên cứu tình huống định lượng tại Anh để so sánh hiệu quả của các mô hình dự báo (ARIMA, Hồi quy Lasso) trên các chuỗi dữ liệu vĩ mô về nhiệt độ, khí methane, vận tải quốc tế và tiêu thụ năng lượng. Kết quả chỉ ra Hồi quy Lasso là công cụ vượt trội, có khả năng tự động lựa chọn biến, khi xác định Methane và Vận tải là yếu tố dự báo chính cho nhiệt độ, đồng thời loại bỏ biến Tiêu thụ Năng lượng. Nghiên cứu đề xuất hàm ý phương pháp luận (áp dụng Lasso) và kiến nghị Việt Nam cần đa dạng hóa ưu tiên chính sách, kiểm soát phát thải từ nông nghiệp và giao thông, bên cạnh năng lượng.

Từ khóa: Hồi quy Lasso, Chính sách môi trường, Vương quốc Anh, Việt Nam.

1. Đặt vấn đề

Tại Hội nghị COP26, Việt Nam đã đưa ra cam kết mạnh mẽ về việc đạt được mức phát thải ròng bằng "0" (Net-Zero) vào năm 2050 (Thủ tướng Chính phủ, 2021). Cam kết này đòi hỏi sự tái cơ cấu sâu sắc nền kinh tế, được cụ thể hóa trong Chiến lược quốc gia về biến đổi khí hậu giai đoạn đến năm 2050 (Chính phủ, 2022). Để các chính sách này hiệu quả, việc hoạch định không thể chỉ dựa trên định tính mà phải dựa trên các bằng chứng định lượng và mô hình dự báo đáng tin cậy. Thách thức lớn nhất đối với các nhà hoạch định chính sách tại Việt Nam là xác định các lĩnh vực ưu tiên – năng lượng, giao thông, hay nông nghiệp – để phân bổ nguồn lực một cách hiệu quả nhất nhằm đạt được mục tiêu giảm phát thải.

Việc trả lời câu hỏi trên đòi hỏi phải xác định chính xác các yếu tố kinh tế - xã hội then chốt (key drivers) tác động mạnh mẽ nhất đến các chỉ số môi trường vĩ mô. Nếu các chính sách được ban hành dàn trải, hoặc dựa trên các mô hình phân tích truyền thống vốn có nhiều hạn chế (như vấn đề đa cộng tuyến giữa các biến), Việt Nam có nguy cơ lãng phí nguồn lực mà không đạt được mục tiêu Net-Zero đúng hạn (Trần Văn Tùng, 2023). Điều này tạo ra nhu cầu cấp thiết về việc áp dụng các phương pháp luận dự báo và phân tích tác động tiên tiến, có khả năng "sàng lọc" và xác định các yếu tố quan trọng nhất. Để đạt được mục tiêu này, thay vì phân tích trực tiếp bộ dữ liệu của Việt Nam (vốn có thể còn hạn chế về tính đồng bộ), bài viết sử dụng một nghiên cứu tình huống (case study) điển hình tại Vương quốc Anh (UK). Vương quốc Anh được chọn vì có cam kết khí hậu tương tự Việt Nam và sở hữu bộ dữ liệu môi trường - kinh tế vĩ mô chất lượng cao, công khai.

¹ ThS., Học viện công nghệ Bưu chính Viễn thông; Email: huyenvy279@gmail.com

2. Tổng quan nghiên cứu

2.1. Các mô hình dự báo chuỗi thời gian truyền thống

Trong lĩnh vực dự báo, các mô hình làm mượt hàm mũ (Exponential Smoothing - ES), đặc biệt là phương pháp Holt-Winters (HW), từ lâu đã được chứng minh là hiệu quả cho các chuỗi dữ liệu có cả yếu tố xu thế và tính mùa vụ (Hyndman & Athanasopoulos, 2018). Bên cạnh đó, nhóm mô hình Tự hồi quy Tích hợp Trung bình trượt (ARIMA), được chuẩn hóa bởi Box & Jenkins (1970), đại diện cho một phương pháp luận tinh vi hơn. Nhiều nghiên cứu tại Việt Nam và khu vực Châu Á - Thái Bình Dương đã ứng dụng thành công ARIMA để dự báo nhu cầu tiêu thụ điện (Nguyễn Văn An, 2022) hoặc dự báo lượng khách du lịch (Trần T. Minh Hải, 2025). Tuy nhiên, cả ES và ARIMA đều có một hạn chế cố hữu: chúng là các mô hình đơn biến (univariate), tức là chỉ dự báo dựa trên các giá trị trong quá khứ của chính nó, mà không giải thích được *nguyên nhân* hay *yếu tố tác động* (drivers) đằng sau sự thay đổi đó.

2.2. Các mô hình hồi quy và thách thức trong lựa chọn biến

Để khắc phục hạn chế trên, các nhà nghiên cứu thường sử dụng mô hình hồi quy tuyến tính đa biến, dựa trên phương pháp Bình phương nhỏ nhất (OLS), để lượng hóa tác động của các biến độc lập (ví dụ: GDP, tiêu thụ năng lượng) lên một biến phụ thuộc (ví dụ: phát thải CO₂) (Wooldridge, 2010). Nhiều nghiên cứu về chính sách môi trường tại Việt Nam đã đi theo hướng này.

Tuy nhiên, khi phân tích các biến vĩ mô, mô hình OLS truyền thống thường gặp phải hai vấn đề lớn: (1) Đa cộng tuyến (Multicollinearity), xảy ra khi các biến độc lập (như GDP, năng lượng, vận tải) cùng tăng trưởng và có tương quan mạnh với nhau, làm cho hệ số hồi quy bị chệch; và (2) Quá khớp (Overfitting), xảy ra khi mô hình có quá nhiều biến dự báo, khiến nó mất khả năng dự báo trên dữ liệu mới.

Để giải quyết vấn đề này, các phương pháp hồi quy tiên tiến thuộc lĩnh vực học máy đã được phát triển. Nổi bật trong số đó là Hồi quy Lasso (Least Absolute Shrinkage and Selection Operator), được giới thiệu bởi Tibshirani (1996). Mô hình Lasso giải quyết cả hai vấn đề trên bằng cách thêm một "thành phần phạt" (L1 regularization) vào quá trình hồi quy, buộc các hệ số của những biến không quan trọng hoặc gây nhiễu, phải co về chính xác bằng 0. Nói cách khác, Lasso có khả năng tự động thực hiện việc lựa chọn các biến quan trọng nhất từ một tập hợp lớn các yếu tố tiềm năng.

2.3. Khoảng trống nghiên cứu và đóng góp của bài viết

Tổng quan các nghiên cứu tại Việt Nam cho thấy, mặc dù các mô hình ARIMA và OLS được sử dụng tương đối phổ biến (Nguyễn Văn An, 2022), việc ứng dụng các phương pháp hồi quy tiên tiến như Lasso để phục vụ phân tích và lựa chọn ưu tiên chính sách trong lĩnh vực môi trường vẫn còn rất hạn chế. Như Trần Văn Tùng (2023) đã chỉ ra, tồn tại một khoảng cách giữa khả năng của các mô hình phân tích dữ liệu hiện đại và việc ứng dụng chúng trong thực tiễn hoạch định chính sách tại Việt Nam.

Bài viết này giải quyết khoảng trống đó. Thông qua một nghiên cứu tình huống chi tiết tại Vương quốc Anh, bài viết cung cấp một bản kiểm nghiệm về phương pháp luận, so sánh hiệu quả của các mô hình và làm nổi bật khả năng lựa chọn biến của Hồi quy Lasso. Nghiên cứu này đóng vai trò như một tài liệu tham khảo và khuyến nghị phương pháp luận trực tiếp cho các nhà phân tích tại Việt Nam và các quốc gia khác trong khu vực.

3. Phương pháp nghiên cứu

3.1. Các mô hình dự báo đơn biến

Nghiên cứu tiến hành so sánh hai nhóm mô hình dự báo đơn biến. Nhóm mô hình thứ nhất là Làm mượt hàm mũ (Exponential Smoothing - ES), được sử dụng làm mô hình cơ sở. Các biến thể khác nhau được áp dụng dựa trên đặc tính của chuỗi, bao gồm mô hình Holt's Linear (HL) cho chuỗi MSTA (có

xu thế, không mùa vụ) và mô hình Holt-Winters (HW) Additive & Multiplicative cho các chuỗi có tính mùa vụ (CH4, GMAF, ET12). Một điều chỉnh quan trọng được thực hiện với chuỗi GMAF là giai đoạn COVID-19 (2020-2021), vốn có dữ liệu sụt giảm đột ngột, đã được xử lý bằng phương pháp nội suy (interpolating) dựa trên xu hướng trước đó (2010-2019) để mô hình hội tụ và tránh dự báo sai lệch. Kết quả từ Bảng 1 cho thấy, mô hình HW Multiplicative sau khi điều chỉnh COVID-19 cho chuỗi GMAF đã giảm đáng kể sai số (ví dụ: MAPE từ 15,43% xuống 4,94%).

Bảng 1: So sánh các chỉ số sai số của mô hình Làm mượt hàm mũ (ES)

Chỉ số	MSTA (Holt's Linear)	MSTA (Holt- Winters Add.)	CH4 (Holt- Winters Add.)	GMAF (HW-Mult. Thô)	GMAF (HW-Mult. Đã sửa COVID)	ET12 (HW- Mult.)
MSE	0,0110	0,0110	1,3155	308001,45	79562,85	0,5728
RMSE	0,1051	0,1051	1,1469	555,69	282,06	0,7568
MAE	0,0820	0,0819	0,8858	341,23	196,11	0,5518
MAPE(%)	146,76	146,756	0,0501	15,43	4,94	3,1873
AIC	-5669,15	-5676,73	167,74	6706,15	5990,11	-168,04
BIC	-5648,59	-5591,51	235,01	6774,46	6058,41	-105,91

(Nguồn: Phân tích của tác giả)

Nhóm mô hình thứ hai là ARIMA, được áp dụng cho chuỗi MSTA để so sánh hiệu quả với mô hình ES. Quá trình xây dựng mô hình tuân theo quy trình kinh điển của Box & Jenkins (1970) gồm các bước Nhận dạng, Ước lượng và Kiểm định. Nhiều mô hình đã được thử nghiệm, bao gồm ARIMA(1,1,1), ARIMA(2,1,2) và SARIMA(1,1,1)(1,1,1,12). Như được trình bày trong Bảng 2, mô hình ARIMA(2,1,2) cho thấy hiệu quả tốt nhất khi có chỉ số AIC và RMSE thấp nhất so với các mô hình ARIMA và Holt's Linear khác. Hình 2 cũng minh họa sự phù hợp của mô hình ARIMA(2,1,2) trong việc bám sát các biến động ngắn hạn của chuỗi MSTA.

Bảng 2: Các chỉ số sai số của mô hình ARIMA và Holt's Linear cho chuỗi MSTA

Mô hình	MSE	RMSE	MAE	MAPE (%)	AIC	BIC
Holt's Linear	0,0110	0,1051	0,0820	146,76	-5669,1	5648,5
ARIMA (1,1,1)	0,0110	0,1049	0,0819	146,21	-2098,1	-2082,7
ARIMA (2,1,2)	0,0107	0,1036	0,0806	143,35	-2127,9	-2102,2
SARIMA(1,1,1)(1,1,1,12)	0,0113	0,1066	0,0831	151,58	-2072,1	-2146,5

(Nguồn: Phân tích của tác giả)

3.2. Mô hình hồi quy đa biến và lựa chọn biến

Nhóm mô hình thứ ba, và là trọng tâm của hàm ý chính sách, là Hồi quy đa biến. Nhóm này phân tích các yếu tố tác động đến biến MSTA (Nhiệt độ), với MSTA là biến phụ thuộc và CH4, GMAF, ET12 là các biến độc lập. Quá trình phân tích bắt đầu với mô hình OLS (Ordinary Least Squares). Kết quả kiểm định Breusch-Pagan ($p < 0,001$) và Durbin-Watson (0,640), được tổng hợp trong Bảng 3, cho thấy mô hình vi phạm nghiêm trọng giả định về phương sai sai số không đổi (heteroscedasticity) và tự tương quan (autocorrelation). Các đồ thị chẩn đoán phần dư (residual diagnostics) cũng trực quan hóa vấn đề này, cho thấy phần dư chưa được phân phối chuẩn.

Bảng 3: Hiệu quả và các kiểm định của các mô hình hồi quy

Model	R^2	AIC	BIC	DW	Kiểm định BP
OLS	0,713	-418,4	-402,87	0,640	<0,001
Log-trans	0,627	-401,3	-388,7	0,768	<0,001
GLS	0,713	-418,4	-402,7	0,640	<0,001

(Nguồn: Phân tích của tác giả)

Sau khi thử nghiệm với WLS (Weighted Least Squares), nghiên cứu áp dụng Hồi quy Lasso (L1 Regularization). Đây là phương pháp luận then chốt, như đã đề cập trong Tổng quan nghiên cứu (Tibshirani, 1996), giúp giải quyết vấn đề đa cộng tuyến và tự động lựa chọn biến bằng cách "co rút" (shrinking) hệ số của các biến không quan trọng về chính xác bằng 0. Mô hình này đặc biệt hữu ích khi phân tích các yếu tố chính sách, giúp xác định đâu là các biến có tác động thực sự trong bối cảnh có nhiều biến số phức tạp.

4. Kết quả và Thảo luận

4.1. Kết quả phân tích từ nghiên cứu tình huống (UK Case Study)

Như đã phân tích tại Phần 3, các mô hình hồi quy OLS cơ bản bộc lộ nhiều yếu kém về mặt giả định thống kê, đòi hỏi phải sử dụng các mô hình hồi quy nâng cao (WLS, Lasso) và so sánh với các mô hình dự báo chuỗi thời gian (ARIMA). Bảng 4 trình bày kết quả so sánh các chỉ số sai số giữa mô hình WLS và Hồi quy Lasso.

Bảng 4: So sánh các chỉ số sai số của mô hình Hồi quy WLS và Lasso

Chỉ số	Hồi quy Lasso	Hồi quy WLS
MSE	0,019012	0,018611
RMSE	0,137884	0,136421
MAE	0,106691	0,104523
MAPE (%)	20,867116	20,219544

(Nguồn: Phân tích của tác giả)

Kết quả cho thấy cả hai mô hình đều có hiệu suất dự báo khá tương đồng, mặc dù WLS có chỉ số MAPE (20,22%) thấp hơn một chút so với Lasso (20,87%). Tuy nhiên, giá trị của mô hình Lasso không chỉ nằm ở sai số dự báo, mà còn ở khả năng lựa chọn biến, như sẽ được phân tích dưới đây.

So sánh quan trọng nhất được thể hiện trong Bảng 5, nơi mô hình Lasso được đối chiếu với các mô hình dự báo đơn biến (ARIMA và Holt's Linear). Mặc dù mô hình ARIMA(2,1,2) đạt được RMSE (0,1036) và MAE (0,0806) thấp nhất, cho thấy khả năng dự báo các biến động ngắn hạn rất tốt, nhưng mô hình Hồi quy Lasso lại đạt được chỉ số MAPE (20,87%) thấp hơn một cách vượt trội so với ARIMA (143,35%) và Holt's Linear (146,76%).

Điều này là một phát hiện quan trọng: về mặt tỷ lệ sai số trung bình, các mô hình đơn biến (ARIMA, ES) gần như thất bại trong việc mô tả bản chất của chuỗi MSTA (với sai số hơn 140%). Ngược lại, mô hình hồi quy đa biến (Lasso) đã giải thích được cấu trúc của dữ liệu tốt hơn nhiều, cho thấy rằng các yếu tố ngoại sinh (CH₄, GMAF) là cần thiết để dự báo MSTA.

Phát hiện then chốt nhất của nghiên cứu đến từ chính cơ chế hoạt động của Hồi quy Lasso. Khi được đưa vào mô hình với 3 biến độc lập (CH4, GMAF, ET12) để giải thích cho MSTA (Nhiệt độ), cơ chế chính quy hóa L1 (L1 regularization) của Lasso đã tự động co rút hệ số của biến ET12 (Tiêu thụ năng lượng) về chính xác bằng 0. Đồng thời, mô hình xác định CH4 (Khí methane) và GMAF (Vận tải/Du lịch quốc tế) là hai yếu tố dự báo có ý nghĩa thống kê và quan trọng nhất cho sự biến động của nhiệt độ.

Hình 3 trực quan hóa sự khác biệt trong kết quả dự báo của các mô hình. Có thể thấy, trong khi mô hình Holt's Linear (đường chấm) duy trì một xu hướng tuyến tính cứng nhắc và mô hình ARIMA(2,1,2) (đường gạch) bám sát các biến động ngắn hạn, thì dự báo của mô hình Lasso (đường gạch-chấm) thể hiện một quỹ đạo trung dung, được "neo" bởi các yếu tố dự báo ngoại sinh (CH4 và GMAF), mang lại một cái nhìn giàu tính giải thích hơn.

Bảng 5: So sánh hiệu quả dự báo của Hồi quy Lasso, ARIMA và Holt's Linear

Mô hình	Hồi quy Lasso	ARIMA(2,1,2)	Holt's linear
MSE	0,019012	0,010730	0,011046
RMSE	0,137884	0,103585	0,105101
MAE	0,106691	0,080638	0,081997
MAPE (%)	20,87	143,35	146,76

(Nguồn: Phân tích của tác giả)

4.2. Thảo luận và hàm ý cho Việt Nam

Kết quả nghiên cứu tại Vương quốc Anh cung cấp hai nhóm hàm ý quan trọng cho Việt Nam trong hoạch định chính sách Net-Zero.

Thứ nhất – Về phương pháp luận. Nghiên cứu khẳng định ưu thế của các mô hình hồi quy đa biến có chọn lọc biến (như Lasso) so với OLS hay các mô hình đơn biến (ARIMA, ES). OLS dễ bị sai lệch do đa cộng tuyến giữa các biến vĩ mô, còn ARIMA dù dự báo ngắn hạn tốt nhưng là “hộp đen”, không cho biết yếu tố nào tác động. Lasso khắc phục các hạn chế này, cho phép đưa vào nhiều yếu tố kinh tế - xã hội (GDP, dân số, sản lượng thép, diện tích lúa...) và tự động chọn ra những biến thực sự ảnh hưởng. Vì vậy, Việt Nam nên ưu tiên áp dụng Lasso trong nghiên cứu và hoạch định chính sách để xây dựng các quyết sách dựa trên bằng chứng khoa học.

Thứ hai – Về hàm ý chính sách. Phát hiện đáng chú ý là mô hình Lasso loại bỏ biến Tiêu thụ năng lượng (ET12) và giữ lại Methane (CH4) cùng Vận tải/Du lịch quốc tế (GMAF). Điều này không phủ nhận vai trò của năng lượng, mà cho thấy khi xét về tác động biên, hai yếu tố còn lại có sức giải thích mạnh hơn với biến động nhiệt độ.

Bài học cho Việt Nam rất rõ: nếu chính sách Net-Zero chỉ tập trung vào năng lượng mà bỏ qua các nguồn phát thải khác, mục tiêu 2050 sẽ khó đạt. Bên cạnh chuyển dịch năng lượng (như Quy hoạch điện VIII), Việt Nam cần chiến lược kiểm soát hai lĩnh vực trọng yếu:

- Khí Methane (CH4): Nguồn phát thải lớn từ nông nghiệp, chăn nuôi và rác thải.
- Vận tải và Du lịch (GMAF): Phát thải tăng mạnh cùng sự phát triển của logistics và hàng không.

Do đó, cần có chính sách đồng bộ nhằm kiểm soát phát thải trong hai ngành này, song song với năng lượng, để phân bổ nguồn lực hiệu quả và đảm bảo lộ trình Net-Zero.

5. Kết luận và Kiến nghị

Kết quả khẳng định Hồi quy Lasso (L1 regularization) là công cụ phù hợp trong nghiên cứu biến động khí hậu. Mô hình này khắc phục nhược điểm của OLS, đồng thời vượt trội hơn ARIMA trong khả năng giải thích (MAPE thấp hơn nhiều). Lasso xác định CH₄ và GMAF là các yếu tố tác động chính đến nhiệt độ, trong khi loại bỏ ET12. Từ đó, hai nhóm kiến nghị chính được đề xuất.

Thứ nhất, về phương pháp luận: Các cơ quan hoạch định chính sách như Bộ KH&ĐT, Bộ TN&MT nên sớm ứng dụng Lasso và các kỹ thuật phân tích hiện đại khác để sàng lọc và xác định các “động lực” quan trọng của phát thải, giúp chính sách dựa trên bằng chứng định lượng.

Thứ hai, về chính sách nội dung: Bên cạnh chuyển dịch năng lượng, cần tập trung vào kiểm soát Methane và phát thải vận tải. Đây là hai nguồn gây biến động khí hậu mạnh nhưng chưa được chú trọng tương xứng.

Nghiên cứu thừa nhận giới hạn khi chỉ dựa trên dữ liệu của Anh, nên các kết quả mang tính gợi mở. Trong tương lai, khi Việt Nam có bộ dữ liệu vĩ mô đầy đủ, cần áp dụng trực tiếp phương pháp này để định lượng chính xác tác động của từng ngành, hỗ trợ thực thi hiệu quả cam kết Net-Zero 2050.

Tài liệu tham khảo

1. Chính phủ (2022). *Chiến lược quốc gia về biến đổi khí hậu giai đoạn đến năm 2050*. Quyết định số 896/QĐ-TTg, Hà Nội.
2. Nguyễn Văn An (2022). Ứng dụng mô hình SARIMA dự báo nhu cầu tiêu thụ điện tại Việt Nam. *Tạp chí Khoa học Thống kê*, (số 4), tr 50-58.
3. Thủ tướng Chính phủ (2021). *Phát biểu của Thủ tướng Phạm Minh Chính tại Hội nghị COP26*.
4. Trần, T. Minh Hải (2025). Vận dụng mô hình ARIMA dự báo lượt khách quốc tế đến Việt Nam. *Tạp chí Du lịch*, (số 3), tr 45-50.
5. Trần Văn Tùng (2023). Thách thức trong hoạch định chính sách dựa trên dữ liệu tại Việt Nam. *Tạp chí Kinh tế và Phát triển*, 26(11), tr 80-89.
6. Box, G. E. P., & Jenkins, G. M. (1970). *Time series analysis: Forecasting and control*. Holden-Day.
7. Department for Energy Security and Net Zero (2024). *Inland energy consumption: primary fuel input basis (ET 1.2 monthly)*. Truy cập ngày 03/11/2025, từ <https://www.gov.uk/government/statistics/total-energy-section-1-energy-trends>
8. Hyndman, R. J., & Athanasopoulos, G. (2018). *Forecasting: principles and practice* (2nd ed.). OTexts.
9. Met Office (2024a). *Global Mean Surface Temperature Anomaly (MSTA)*. Truy cập ngày 03/11/2025, từ <https://www.metoffice.gov.uk/hadobs/hadcrut5/data/HadCRUT.5.0.2.0/download.html>
10. Met Office (2024b). *Global Monthly Atmospheric Carbon Dioxide Levels (CH₄)*. Truy cập ngày 03/11/2025, từ https://climate.metoffice.cloud/greenhouse_gases.html
11. Office for National Statistics (ONS) (2024). *International Passenger Survey, UK visits abroad (GMAF)*. Truy cập ngày 03/11/2025, từ <https://www.ons.gov.uk/peoplepopulationandcommunity/leisureandtourism/datasets/internationalpassengersurveytimeseriesspreadsheet>
12. Tibshirani, R. (1996). Regression shrinkage and selection via the lasso. *Journal of the Royal Statistical Society: Series B (Methodological)*, 58(1), 267-288.
13. Wooldridge, J. M. (2010). *Econometric analysis of cross section and panel data*. MIT press.