



Original Article

Forecasting Economic Growth at a Provincial Level in Vietnam: A Systematic Dynamics Model Approach

Dong Manh Cuong*

VNU University of Economics and Business, No. 144 Xuan Thuy, Cau Giay District, Hanoi, Vietnam

Received: February 21, 2022

Revised: July 10, 2022; Accepted: October 25, 2022

Abstract: Forecasting annual economic growth is an important task to help local governments set goals and policies for socio-economic development. This study proposes and develops a forecasting method for economic growth at the provincial level in Vietnam. This method is named Dynamic Systems Modeling. Through the dataset collected from the General Statistics Office of Vietnam from 2016 to 2020, the Dynamic Systems Modeling method is adopted to forecast two important economic growth indicators at the provincial level, which are GRDP and GRDP per capita. The analysis and forecasting evaluation results show that Dynamic Systems Modeling is an effective tool for forecasting economic growth at the provincial level in Vietnam.

Keywords: Forecast, economic growth, PCI, Dynamical System Modeling, Bayesian.

* Corresponding author.

E-mail address: dmcuong@vnu.edu.vn

<https://doi.org/10.25073/2588-1108/vnueab.4756>

Dự báo tốc độ tăng trưởng kinh tế cấp tỉnh tại Việt Nam: Cách tiếp cận mô hình động lực học hệ thống

Đông Mạnh Cường*

Trường Đại học Kinh tế - Đại học Quốc gia Hà Nội, 144 Xuân Thủy, Cầu Giấy, Hà Nội, Việt Nam

Nhận ngày 21 tháng 2 năm 2022

Chỉnh sửa ngày 10 tháng 7 năm 2022; Chấp nhận đăng ngày 25 tháng 10 năm 2022

Tóm tắt: Dự báo tăng trưởng kinh tế hàng năm là một nhiệm vụ quan trọng giúp chính quyền các địa phương xác định mục tiêu và chính sách phát triển kinh tế - xã hội. Nghiên cứu đề xuất và phát triển việc dự báo tốc độ tăng trưởng kinh tế cấp tỉnh tại Việt Nam thông qua phương pháp mô hình động lực học hệ thống. Với bộ dữ liệu điều tra của Tổng cục Thống kê từ năm 2016-2020, nghiên cứu áp dụng phương pháp này để dự báo hai chỉ tiêu tăng trưởng kinh tế quan trọng cấp tỉnh là GRDP và GRDP trên đầu người. Kết quả cho thấy mô hình động lực học hệ thống là một công cụ hữu hiệu cho dự báo tăng trưởng kinh tế cấp tỉnh tại Việt Nam.

Từ khóa: Dự báo, tăng trưởng kinh tế, PCI, động lực học hệ thống, Bayesian.

1. Giới thiệu

Công tác dự báo kinh tế là một trong những công cụ thiết yếu trong điều hành kinh tế vĩ mô, nhằm đưa nền kinh tế và các hoạt động xã hội phát triển theo hướng bền vững, tránh những xáo động lớn có thể làm đảo lộn các mục tiêu ban đầu. Dự báo phát triển kinh tế còn được coi là “hạt nhân” trong các kế hoạch hành động của chính quyền, giúp các cơ quan quản lý chủ động sớm đưa ra quyết sách, giải pháp thích hợp để phát huy những tình huống thuận, đồng thời ngăn ngừa những tình huống nghịch mà các dự báo kinh tế đã đưa ra. Bên cạnh đó, để có thể đối phó và ngăn ngừa những nguy cơ, hậu quả tiêu cực có thể xảy ra thì công tác phân tích và dự báo kinh tế cần phải được đẩy mạnh như một kênh hữu hiệu trong việc giám sát nền kinh tế, hỗ trợ cho công tác điều hành của chính quyền một cách nhanh nhạy, định hướng hoạt động sản xuất

kinh doanh của doanh nghiệp và hành vi của cộng đồng.

Trong thời gian qua, công tác phân tích, dự báo kinh tế đã được Chính phủ cũng như nhiều cơ quan, đơn vị, tổ chức quan tâm và thực hiện. Đối với cấp tỉnh, thành phố trực thuộc Trung ương, trong nhiều năm trở lại đây, sự phát triển mạnh mẽ của các bộ chỉ số đo lường chất lượng hành chính cấp tỉnh như: chỉ số năng lực cạnh tranh cấp tỉnh (PCI), chỉ số hiệu quả quản trị và hành chính công (PAPI), chỉ số cải cách hành chính (PAR), chỉ số sẵn sàng cho chính phủ điện tử (ICT)... cũng như các bộ số liệu từ các cơ quan thống kê đã mang đến nguồn dữ liệu hữu ích giúp phát triển và cải thiện công tác dự báo kinh tế.

Khác với các nghiên cứu đánh giá tác động thông thường, việc lập mô hình dự báo trong nghiên cứu không nhất thiết phải bao gồm đầy đủ các biến tác động hoặc kiểm soát vấn đề nội

* Tác giả liên hệ.

Địa chỉ email: dmcuong@vnu.edu.vn

<https://doi.org/10.25073/2588-1108/vnueab.4756>

sinh, mà mục tiêu quan trọng là tìm được các biến có khả năng dự báo cao nhất, giúp mô hình dự báo tối đa hóa sự chính xác. Theo Ligneul (2021), nghiên cứu đánh giá tác động tập trung vào việc biết điều gì xảy ra với biến phụ thuộc khi thay đổi biến giải thích, trong khi nghiên cứu dự đoán tập trung vào việc tìm ra biến phụ thuộc tiếp theo khi có các biến giải thích tại hiện tại. Thêm nữa, nghiên cứu tác động thường muốn tìm ra các ước lượng không chệch của tham số mô hình, trong khi nghiên cứu dự báo thường chấp nhận một chút sai lệch của tham số nhưng tập trung vào giảm tối đa phương sai của dự đoán.

Hiện nay, phần lớn các nghiên cứu dự báo kinh tế ở Việt Nam, đặc biệt là dự báo tăng trưởng kinh tế, được tập trung thực hiện ở cấp quốc gia. Trong khi đó, việc dự báo kinh tế cấp tỉnh, thành phố còn gặp nhiều hạn chế do thiếu hụt thông tin cần thiết và phương pháp phù hợp. Bên cạnh đó, việc thực hiện dự báo tăng trưởng kinh tế cấp tỉnh cũng được thực hiện rời rạc ở từng tỉnh mà không có sự đánh giá tổng thể toàn bộ hệ thống. Điều này cũng làm khả năng dự báo giảm xuống trong bối cảnh sự kết nối và liên kết kinh tế giữa các tỉnh, địa phương ngày càng lớn. Để giải quyết vấn đề này, nghiên cứu sử dụng cách tiếp cận kết hợp giữa mô hình thống kê truyền thống và phương pháp học máy có tên là mô hình động lực học hệ thống (Dynamical Systems Modelling). Phương pháp này giúp đưa ra các dự báo tăng trưởng kinh tế cấp tỉnh mà không cần quá nhiều chỉ số hay thông tin khó thu thập. Bên cạnh đó, việc phương pháp này được sử dụng để dự báo dữ liệu dạng bảng cũng giúp nâng cao khả năng dự báo do có thể kiểm soát cả dữ liệu chuỗi thời gian lẫn dữ liệu chéo.

Nghiên cứu có những đóng góp quan trọng sau: (1) Đây là nghiên cứu thực nghiệm tiên phong về cải tiến mô hình dự báo phát triển kinh tế cho các tỉnh, thành phố trực thuộc Trung ương; (2) Sử dụng cách tiếp cận dự báo mới kết hợp giữa thống kê truyền thống và học máy, lựa chọn giữa nhiều mô hình khác nhau để tìm ra mô hình dự báo phù hợp thay vì chỉ dùng một mô hình cố định; (3) Kết quả nghiên cứu cung cấp phương pháp và góc nhìn mới giúp nâng cao chất lượng dự báo kinh tế cấp tỉnh.

2. Tổng quan nghiên cứu

Có rất nhiều cách tiếp cận được đề xuất trong việc dự báo tăng trưởng kinh tế. Một số nghiên cứu kinh tế vĩ mô về chủ đề này sử dụng phương pháp tiếp cận dự báo chuỗi thời gian, đặc biệt là sử dụng các kỹ thuật khác nhau của mô hình VAR (Ang và cộng sự, 2006; Brave và cộng sự, 2019; Koop và cộng sự 2020). Đối với các phương pháp này, kết quả dự báo có thể được cải thiện thông qua quy trình Bayesian (Bańbura và cộng sự, 2010). Chauvet và Potter (2013) đưa ra so sánh giữa các mô hình chuyển đổi dạng rút gọn, mô hình tự hồi quy, mô hình VAR, mô hình chuyển đổi Markov và nhận thấy rằng mô hình tự hồi quy theo chuỗi thời gian bậc hai AR(2) hoạt động tốt hơn các mô hình khác trong dự báo tăng trưởng kinh tế của Hoa Kỳ. Baffigi và cộng sự (2002) sử dụng mô hình chuỗi thời gian ARIMA để dự báo tăng trưởng kinh tế cho Vương quốc Anh. Lunde và Torkar (2020) khai thác hơn 120 yếu tố dự báo, sau đó sử dụng phương pháp phân tích thành phần chính (PCA) để rút gọn các yếu tố dự báo.

Về các chỉ số kinh tế tiềm năng được sử dụng làm yếu tố dự báo GDP, rất nhiều yếu tố vĩ mô đã được sử dụng để dự báo tăng trưởng kinh tế như chỉ tiêu trong nước, tỷ lệ thất nghiệp, tiền lương, giá cả, khoáng trữ đầu ra, ngoại thương và cán cân thanh toán... (Estrella và Mishkin, 1996; Koop, 2013; Diebold và cộng sự, 2006). Bên cạnh đó, nhiều tác giả cũng sử dụng đường cong lợi suất để dự báo thông tin về hoạt động kinh tế trong tương lai (Giannone và cộng sự, 2008; Yiu và Chow, 2010). Đặc biệt, Estrella và Hardouvelis (1991) chỉ ra độ dốc của đường cong lợi suất có thể dự đoán các biến động tích lũy trong GDP thực tế tới 4 năm. Một nghiên cứu tương tự được thực hiện bởi Bernard và Gerlach (1998) ở 8 quốc gia cho thấy, mặc dù có sự khác biệt đáng kể giữa các quốc gia, nhưng độ dốc của đường cong lợi suất cung cấp thông tin về khả năng suy thoái trong tương lai. Trong khi đó, Ang và cộng sự (2006) nhận thấy lãi suất ngắn hạn danh nghĩa tốt hơn độ dốc của đường cong lợi suất trong việc dự báo tăng trưởng GDP.

Đối với Việt Nam, phần lớn các nghiên cứu tìm hiểu tác động của các yếu tố vĩ mô đến tăng

trường kinh tế và bỏ qua vấn đề dự báo. Một số ít nghiên cứu có tham gia giải quyết vấn đề dự báo như Lê Tuấn Anh (2018) dự báo tăng trưởng kinh tế Việt Nam bằng việc sử dụng các nhân tố động và bộ dữ liệu có tần số cao thông qua mô hình hồi quy MIDAS. Kết quả nghiên cứu cho thấy các dự báo bằng hồi quy MIDAS có độ chính xác cao hơn các dự báo sử dụng mô hình truyền thống và việc sử dụng các nhân tố có tần số cao sẽ khai thác toàn bộ và hiệu quả các thông tin tài chính để dự báo chính xác tăng trưởng kinh tế. Nguyễn Thị Hồng Trang (2021) sử dụng mô hình hiệu chỉnh sai số VECM và dự báo Winter để mô tả được sự ảnh hưởng, mức độ phụ thuộc của GDP khi kim ngạch xuất, nhập khẩu thay đổi, từ đó đưa ra kết luận về tác động của các biến đến tăng trưởng kinh tế Việt Nam giai đoạn 2009-2020 và dự báo tăng trưởng GDP giai đoạn 2020-2025.

Có thể thấy một số vấn đề tồn tại trong các phương pháp dự báo tăng trưởng kinh tế của các nghiên cứu cả trong và ngoài nước trước đây: (1) Chủ yếu sử dụng mô hình chuỗi thời gian để dự báo cho một đối tượng duy nhất mà bỏ qua mối quan hệ hay liên kết với các đối tượng khác; (2) Cần sử dụng nhiều biến dự báo với tần số dữ liệu cao khiến cho việc thu thập dữ liệu khó khăn, đặc biệt với cấp địa phương; (3) Việc dự báo được thực hiện trên một mô hình duy nhất khiến kết quả dự báo không đảm bảo tối đa hóa sự chính xác; (4) Kết quả dự báo không được kiểm chứng thông qua các công cụ kiểm định khiến kết quả dự báo không đáng tin cậy.

Để giải quyết các vấn đề này, nghiên cứu sử dụng cách tiếp cận mô hình động lực học hệ thống. Đây là phương pháp kết hợp giữa thống kê truyền thống và học máy giúp giải quyết các vấn đề nêu trên. Cụ thể: (1) Phương pháp này được sử dụng để dự báo dữ liệu dạng bảng, giúp kiểm soát cả vấn đề thời gian và không gian của các đối tượng dự báo; (2) Phương pháp này cần sử dụng ít biến dự báo với giả thiết các biến dự báo đều có mối liên hệ với nhau nên không cần quá nhiều biến dư thừa cho việc dự đoán; (3) Việc đưa ra dự báo dựa trên sự so sánh và chọn lọc nhiều mô hình khác nhau; (4) Kết quả dự báo được kiểm định, đảm bảo kết quả dự báo đáng tin cậy.

3. Phương pháp nghiên cứu

Mô hình động lực học hệ thống được sử dụng trong nhiều lĩnh vực khác nhau của cả khoa học tự nhiên và khoa học xã hội để phân tích và dự báo các tương tác theo thời gian giữa nhiều biến trong một hệ thống. Cụ thể, cách tiếp cận cơ bản của mô hình động lực học hệ thống là mô hình hóa những thay đổi của một biến trong khoảng thời gian và dưới dạng một hàm của tất cả các biến mô hình được đưa vào tại thời điểm để hiểu sự tương tác giữa các biến theo thời gian. Phương pháp mô hình động lực học hệ thống áp dụng trong lĩnh vực khoa học xã hội nói chung và trong nghiên cứu kinh tế nói riêng bởi đề xuất của Ranganathan và cộng sự (2014). Cách tiếp cận này được thực hiện dựa trên các thuật toán của phương pháp học máy, trong đó sử dụng dữ liệu có sẵn để lựa chọn mô hình dự báo tốt nhất từ một nhóm các mô hình khả thi (thông qua các tiêu chuẩn so sánh), thay vì chỉ áp dụng một mô hình duy nhất.

Giả sử có một hệ thống dữ liệu bảng với N biến X_i ($i = 1, \dots, N$) của M đối tượng với khoảng thời gian T . Mỗi điểm dữ liệu sẽ được ký hiệu bởi $x_i(j, t)$, trong đó $i = 1, \dots, N$ chỉ số lượng biến, $j = 1, \dots, M$ chỉ số đối tượng và $t = 1, \dots, T$ chỉ thời gian. Sự thay đổi của các biến giữa hai khoảng thời gian liên nhau được biểu thị dưới dạng:

$$dx_i(j, t) = x_i(j, t+1) - x_i(j, t)$$

Dưới đây, tác giả giải thích mô hình động lực học hệ thống của Ranganathan và cộng sự (2014) thông qua ví dụ đơn giản khi mô hình chỉ có 2 biến và sau đó mở rộng ra mô hình chứa nhiều hơn 2 biến.

3.1. Mô hình động lực học hệ thống với 2 biến

Mô hình động lực học hệ thống với 2 biến X_1 và X_2 được thể hiện:

$$\begin{cases} \frac{dX_1}{dt} = f_1(X_1, X_2) \\ \frac{dX_2}{dt} = f_2(X_1, X_2) \end{cases}, \quad (1)$$

với $f_{i(i=1,2)}(\cdot)$ là hàm số chứa 2 biến; X_i thể hiện ma trận $M \times (T - 1)$ với các phần tử $x_i(j, t)$; và dX_i thể hiện ma trận $M \times (T - 1)$ với các phần tử $dx_i(j, t)$. Dạng cụ thể của hàm số này được thể hiện ở phương trình (2) dưới đây:

$$\begin{aligned} f_i(X_1, X_2) = & a_0 + \frac{a_1}{X_1} + \frac{a_2}{X_2} + a_3 X_1 \\ & + a_4 X_2 + \frac{a_5}{X_1 X_2} + \frac{a_6 X_2}{X_1} + \frac{a_7 X_1}{X_2} \\ & + a_8 X_1 X_2 + \frac{a_9}{X_1^2} + \frac{a_{10}}{X_2^2} + a_{11} X_1^2 + a_{12} X_2^2. \end{aligned} \quad (2)$$

Dạng phương trình này sẽ giúp kiểm soát được các mối quan hệ phi tuyến tính giữa biến dự đoán và biến được dự đoán, giúp việc dự báo trở nên chính xác hơn. Bên cạnh đó, cần nhắc lại rằng $dx_i(j, t) = x_i(j, t + 1) - x_i(j, t)$. Nghĩa là, phương pháp này đang tính toán và dự báo tốc độ tăng trưởng tương lai hoàn toàn thông qua các thông tin của biến dự đoán tại thời điểm hiện tại $x_i(j, t)$. Đây là điều khác biệt cơ bản giữa mô hình dự báo và mô hình đánh giá tác động.

Trong phương trình (2) ở trên, hàm số $f_i(\cdot)$ có 13 tham số $(a_0, a_1, \dots, a_{12})$. Đặt m là số lượng các tham số khác 0 trong phương trình (2), như vậy sẽ có $\binom{13}{m}$ phương trình tương ứng với mỗi giá trị của m ($m = 0, 1, \dots, 12$). Vậy sẽ có tổng cộng $2^{13} = 8.192$ lựa chọn cho phương trình (2). Mục tiêu của thuật toán trong mô hình động lực học hệ thống là chọn ra mô hình phù hợp nhất trong số 8.192 lựa chọn này (mỗi mô hình trong số 8.192 lựa chọn này được gọi là một mô hình ứng viên) để trở thành mô hình dự báo. Việc ước lượng các hệ số trong 8.192 mô hình ứng viên này được thực hiện như ước lượng các mô hình hồi quy đa biến dữ liệu bảng thông thường.

Sau khi các mô hình ứng viên được ước lượng, việc lựa chọn mô hình phù hợp nhất để dự báo trải qua 2 bước:

- Bước 1: Tính giá trị log-likelihood cho mỗi mô hình ứng viên thông qua công thức:

$$L = \log P(dX_1 | X_1, X_2, \phi), \quad (3)$$

Trong đó là ϕ tập hợp các tham số xác định của từng mô hình ứng viên cụ thể. Như vậy, với mỗi giá trị của m nêu trên có thể tìm ra giá trị log-likelihood cực đại:

$$L(m) = \log P(dX_1 | X_1, X_2, \phi(m)), \quad (4)$$

Với $\phi(m)$ là tập hợp tham số của mô hình ứng viên có giá trị log-likelihood cực đại. Nói cách khác, trong bước 1 sẽ tìm ra các mô hình có giá trị log-likelihood cao nhất tương ứng với mỗi giá trị của m . Nhờ vậy, từ 8.192 mô hình ứng viên ban đầu, chúng ta rút lại chỉ còn 13 mô hình ứng viên tương ứng với mỗi $m = 0, 1, \dots, 12$.

- Bước 2: Ở bước này, rõ ràng không thể so sánh 13 mô hình ứng viên còn lại thông qua giá trị log-likelihood vì $L(m)$ luôn lớn hơn $L(m - 1)$ do mỗi tham số tăng lên sẽ làm tăng bậc tự do. Hơn nữa, đối với một tập dữ liệu hữu hạn, sự tăng thêm bậc tự do bằng cách thêm quá nhiều tham số và chỉ dựa vào $L(m)$ để so sánh có thể dẫn đến việc mô hình quá khớp với dữ liệu (overfitting), từ đó chấp nhận một mô hình phù hợp với dữ liệu hiện có nhưng mô hình đó lại kém tổng quát cho những dữ liệu chưa thu thập được, khiến mô hình được chọn ít có khả năng dự báo (MacKay, 2003). Vì vậy, ở bước 2, phương pháp thống kê Bayesian được áp dụng để giải quyết vấn đề chênh lệch tham số và bậc tự do.

Cụ thể, bước này sẽ tính khả năng cận biên Bayesian cho mỗi mô hình ứng viên còn lại thông qua công thức:

$$\begin{aligned} B(m) = & \int_{\phi(m)} P(dX_1 | X_1, X_2, \phi(m)) \\ & \times \pi(\phi(m)) \times d\phi(m). \end{aligned} \quad (5)$$

Không giống như phương pháp thống kê thông thường, các tham số thuộc $\phi(m)$ trong suy luận Bayes được coi như các biến ngẫu nhiên được mô tả bởi phân phối $\pi(\phi(m))$. $\pi(\phi(m))$ còn được gọi là phân phối tiên nghiệm do nó thể hiện niềm tin hay sự hiểu biết đối với $\phi(m)$. Nếu không có bất kỳ thông tin hay hiểu biết trước nào về $\phi(m)$, phân phối tiên nghiệm $\pi(\phi(m))$ thường được chọn dưới dạng phân phối đều liên tục. Lợi thế của việc sử dụng khả năng cận biên Bayesian là nó không bị chi phối bởi số lượng tham số và

bậc tự do. Khả năng cận biên Bayesian của mỗi mô hình ứng viên sau đó sẽ được sử dụng để so sánh thông qua nhân tố Bayes (quy trình lựa chọn mô hình thông qua nhân tố Bayes có thể tham khảo trong nghiên cứu của Berger và Pericchi (1996)). Cuối cùng, mô hình có khả năng dự báo cao nhất sẽ được lựa chọn.

3.2. Mô hình động lực học hệ thống với nhiều hơn 2 biến

Tương tự như ở mục 3.1, mô hình động lực học hệ thống với N biến X_i được thể hiện:

$$\begin{cases} \frac{dX_1}{dt} = f_1(X_1, X_2, \dots, X_N) \\ \vdots \\ \frac{dX_N}{dt} = f_N(X_1, X_2, \dots, X_N) \end{cases}, \quad (6)$$

với $f_i(\cdot)$ là các hàm số chứa các biến X_i . Giả sử với 3 biến được đưa vào trong hệ thống, hàm số $f_i(\cdot)$ có dạng:

$$\begin{aligned} f_i(X_1, X_2, X_3) = & a_0 + \frac{a_1}{X_1} + \frac{a_2}{X_2} + \frac{a_3}{X_2} + a_4 X_1 \\ & + a_5 X_2 + a_6 X_3 + \frac{a_7}{X_1 X_2} + \frac{a_8}{X_2 X_3} + \frac{a_9}{X_1 X_3} \\ & + a_{10} X_1 X_2 + a_{11} X_1 X_3 + a_{12} X_2 X_3 + \frac{a_{13} X_1}{X_2} \\ & + \frac{a_{14} X_2}{X_1} + \frac{a_{15} X_1}{X_3} + \frac{a_{16} X_3}{X_1} + \frac{a_{17} X_2}{X_3} + \frac{a_{18} X_3}{X_2} \\ & + \frac{a_{19} X_1}{X_2 X_3} + \frac{a_{20} X_2}{X_1 X_3} + \frac{a_{21} X_3}{X_1 X_2} + \frac{a_{22} X_1 X_2}{X_3} \\ & + \frac{a_{23} X_1 X_3}{X_2} + \frac{a_{24} X_2 X_3}{X_1} + a_{25} X_1 X_2 X_3 + \frac{a_{26}}{X_1 X_2 X_3} \\ & + \frac{a_{27}}{X_1^2} + \frac{a_{28}}{X_2^2} + \frac{a_{29}}{X_3^2} + a_{30} X_1^2 + a_{31} X_2^2 + a_{32} X_3^2, \end{aligned}$$

với 33 tham số (a_0, a_1, \dots, a_{32}). Tương tự như ở mục 3.1, nếu đặt m là số lượng các tham số khác 0 trong phương trình (7), sẽ có tổng cộng $\binom{33}{m}$ phương trình tương ứng với mỗi giá trị của m ($m = 0, 1, \dots, 32$). Với số lượng biến càng nhiều

thì hàm số $f_i(\cdot)$ càng trở nên phức tạp và số lượng mô hình ứng viên càng lớn. Các bước lựa chọn mô hình dự báo giống với mục 3.1.

3.3. Phương pháp dự báo

Mô hình ứng viên có khả năng dự báo cao nhất sẽ được sử dụng để đưa ra các dự báo cho $dX_i(t+1)$. Nghiên cứu này sử dụng phương pháp dự báo cửa sổ trượt (rolling-window forecast) để đánh giá khả năng dự báo của phương pháp này. Cụ thể, dữ liệu của 2 năm đầu tiên sẽ được sử dụng để xây dựng, lựa chọn mô hình và đưa ra dự báo cho năm thứ 3. Số liệu dự báo này sẽ được so sánh với dữ liệu thực của năm thứ 3 để đánh giá độ chính xác. Tiếp theo, dữ liệu của 3 năm đầu lại được sử dụng để xây dựng và lựa chọn mô hình dự báo cho năm thứ 4. Quá trình này tiếp tục cho đến khi kết thúc dữ liệu. Các bước của phương pháp này không chỉ đưa ra kết quả dự báo mà còn so sánh và đánh giá khả năng dự báo của phương pháp so với dữ liệu thực.

Việc đánh giá khả năng dự báo của mô hình được đánh giá thông qua 2 phương pháp:

(1) Hướng chính xác trung bình (Mean direction accuracy - MDA): Thước đo đánh giá sự chính xác về hướng của các dự báo. Cụ thể, khi đưa ra các dự báo tăng trưởng kinh tế cho năm tiếp theo là tăng lên hay giảm đi, phương pháp MDA sẽ đánh giá trung bình số lượng dự đoán hướng đúng. Nói cách khác, MDA so sánh hướng dự báo (lên hoặc xuống) với hướng thực tế thu được. Công thức tính MDA:

$$MDA = \frac{1}{M} \sum_{j=1}^M a_j, \quad (8)$$

Với $a_j = 1$ nếu hướng dự báo chính xác và $a_j = 0$ nếu hướng dự báo không chính xác. Giá trị của a_j nằm trong khoảng $[0, 1]$ với giá trị càng cao thể hiện mô hình càng chính xác trong dự đoán hướng tăng trưởng.

(2) Lỗi trung bình bình phương gốc (Root mean squared error - RMSE): Thước đo mức độ hiệu quả dự báo thông dụng để so sánh kết quả dự báo với giá trị thực thu thập được. Công thức để tính giá trị RMSE:

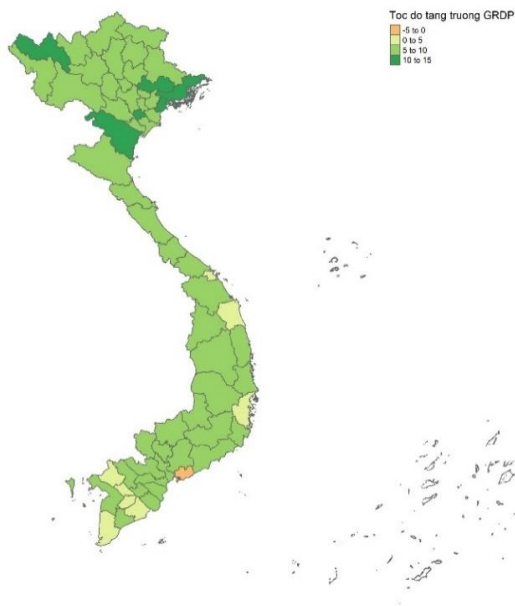
$$RMSE = \sqrt{\frac{\sum_{j=1}^M (\hat{y}_j - y_j)^2}{M}}, \quad (9)$$

Với \hat{y}_j và y_j lần lượt là giá trị dự đoán và giá trị thực của đối tượng được dự đoán. Giá trị của RMSE nằm trong khoảng $[0, +\infty]$ với giá trị càng nhỏ thể hiện mô hình dự đoán càng chính xác.

4. Dữ liệu nghiên cứu

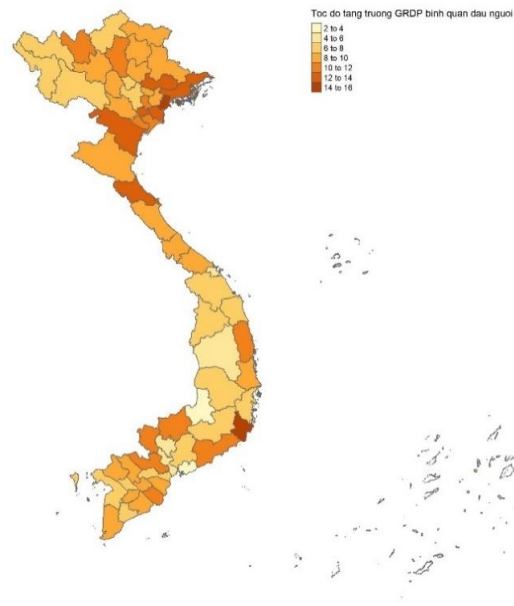
Trong nghiên cứu này, tác giả sử dụng 2 chỉ tiêu để thể hiện tăng trưởng kinh tế các địa phương là tổng sản phẩm trên địa bàn (GRDP) và tổng sản phẩm trên địa bàn bình quân đầu người (GRDP trên đầu người). Theo cách tính

của Tổng cục Thống kê, GRDP là chỉ tiêu kinh tế tổng hợp phản ánh giá trị mới tăng thêm của hàng hóa và dịch vụ cuối cùng được tạo ra trên địa bàn tỉnh, thành phố trong một thời gian nhất định (trong nghiên cứu này là 1 năm). Cụm từ “hàng hóa và dịch vụ cuối cùng” được hiểu theo nghĩa không tính giá trị sản phẩm vật chất và dịch vụ sử dụng ở các khâu trung gian trong quá trình sản xuất sản phẩm. Tương tự, GRDP bình quân đầu người được tính bằng cách chia tổng sản phẩm trên địa bàn trong năm cho dân số trung bình trong năm tương ứng. Dữ liệu của 2 chỉ tiêu này được thu thập từ số liệu của Tổng cục Thống kê giai đoạn 2016-2021. Hình 1a và 1b thể hiện tốc độ tăng trưởng GRDP và GRDP bình quân đầu người trung bình tại các tỉnh, thành phố ở Việt Nam giai đoạn này.



Hình 1a: Tốc độ tăng trưởng GRDP trung bình tại các tỉnh, thành phố trực thuộc Trung ương giai đoạn 2016-2021 (Đơn vị: %).
Nguồn: Tổng cục Thống kê, 2021.

Để đưa ra mô hình dự báo cho tốc độ tăng trưởng GRDP và GRDP trên đầu người của các tỉnh, thành phố trực thuộc Trung ương, tác giả thử nghiệm đưa vào mô hình một số biến dự đoán dễ thu thập như số doanh nghiệp hoạt động trên địa bàn tỉnh (thu thập từ số liệu của Tổng



Hình 1b: Tốc độ tăng trưởng trung bình GRDP trên đầu người tại các tỉnh, thành phố trực thuộc Trung ương giai đoạn 2016-2021 (Đơn vị: %).
Nguồn: Tổng cục Thống kê, 2021.

cục Thống kê) và chỉ số PCI cấp tỉnh (thu thập từ báo cáo PCI của Liên đoàn Thương mại và Công nghiệp Việt Nam). Tất cả các biến này đều được thu thập trong giai đoạn 2016-2021. Bảng 1 thể hiện các thống kê mô tả cho các biến trong dữ liệu.

Bảng 1: Thống kê mô tả của các biến dự báo

Năm	Số doanh nghiệp (ĐV: 10 nghìn)				Chỉ số PCI			
	Trung bình	Độ lệch chuẩn	Min	Max	Trung bình	Độ lệch chuẩn	Min	Max
2016	0,80	2,50	0,05	17,13	62,38	3,80	55,21	72,19
2017	1,04	3,01	0,01	20,27	62,98	3,60	53,27	72,51
2018	1,13	3,32	0,07	22,83	62,23	3,90	52,60	69,55
2019	1,20	3,52	0,06	23,96	66,84	2,66	59,83	74,37
2020	1,06	3,20	0,05	21,86	58,61	4,27	48,14	67,68
2021	1,15	3,37	0,04	20,73	59,22	3,78	48,83	69,81

Nguồn: Kết quả phân tích.

5. Kết quả và thảo luận

Trong phần này, tác giả áp dụng phương pháp động lực học hệ thống được trình bày ở trên để đưa ra dự báo với biến được dự báo là chỉ tiêu tăng trưởng kinh tế (GRDP hoặc GRDP trên đầu người) và các biến dự báo là số doanh nghiệp hoạt động tại mỗi địa phương và chỉ số PCI.

Cụ thể, mỗi biến chỉ tiêu tăng trưởng kinh tế sẽ được đưa vào các hàm đa thức với 2 biến ở phương trình (2) và 3 biến ở phương trình (7). Sau đó, phương pháp động lực học hệ thống sẽ lựa chọn ra mô hình có khả năng dự báo tốt nhất với mỗi trường hợp thông qua giá trị log-likelihood và Bayesian marginal-likelihood. Mô hình được lựa chọn tiếp tục được sử dụng để đưa ra dự báo cho các chỉ tiêu tăng trưởng kinh tế. Kết quả dự báo được so sánh với kết quả thực thông qua MDA và RMSE để xem xét tính chính xác của các dự báo.

Bảng 2 thể hiện kết quả dự báo của mô hình động lực học hệ thống chứa 2 biến đối với chỉ tiêu GRDP. Kết quả dự báo được so sánh với dữ liệu thực và đánh giá độ chính xác thông qua MDA và RMSE. Có thể thấy, với mỗi năm được dự báo, hệ số m không cố định mà luôn thay đổi để tìm ra mô hình dự báo tốt nhất. Điều này thể hiện sự linh động của phương pháp động lực học hệ thống trong việc lựa chọn mô hình và đưa ra dự báo. Bên cạnh đó, việc hệ số m thường khá lớn thể hiện rằng các mô hình thống kê thông thường khó có khả năng dự báo chính xác chỉ tiêu GRDP.

Kết quả kiểm định MDA cho thấy phương pháp mô hình động lực học hệ thống có khả năng dự đoán hướng tăng trưởng khá tốt khi tất cả các dự đoán đều có độ chính xác cao (từ 0,89 đến 1). Tuy nhiên, có thể thấy độ chính xác của kết quả dự báo hướng giảm mạnh vào năm 2020. Điều này là do sự ảnh hưởng đột ngột của dịch COVID-19 khiến nhiều tỉnh, thành phố tăng trưởng GRDP âm trong năm 2020. Đây là vấn đề mà các phương pháp dự báo định tính thường gặp phải khi dự báo kinh tế dài hạn. Kết quả kiểm định RMSE cũng cho thấy kết quả tốt khi các giá trị kiểm định ở trong khoảng 0,89 đến 2,56.

Bảng 3 thể hiện kết quả dự báo chỉ tiêu GRDP trên đầu người cho các tỉnh, thành phố trực thuộc Trung ương sử dụng mô hình động lực học hệ thống chứa một biến dự báo. Có thể thấy kết quả ở đây khá tương đồng với Bảng 2 khi hệ số m thường lớn, kết quả dự báo hướng (MDA) có kết quả chính xác cao trong khi kết quả RMSE cũng khá chính xác. Điều này cho thấy việc thay đổi biến dự đoán không làm sai lệch nhiều kết quả dự đoán của phương pháp mô hình động lực học hệ thống.

Bảng 4 đưa ra kết quả dự báo với 3 biến trong mô hình: 1 biến được dự báo là chỉ tiêu tăng trưởng kinh tế (GRDP hoặc GRDP trên đầu người), 2 biến dự báo là số doanh nghiệp hoạt động trên địa bàn tỉnh và chỉ số PCI. Kết quả cho thấy không có nhiều sự khác biệt trong độ chính xác của dự báo khi cho thêm biến dự báo vào mô hình.

Tóm lại, có thể thấy việc sử dụng cách tiếp cận động lực học hệ thống trong dự báo tăng trưởng kinh tế cấp tỉnh giúp tạo ra sự linh hoạt trong việc lựa chọn mô hình dự báo phù hợp. Bên cạnh đó, các biến dự báo được sử dụng thử nghiệm trong nghiên cứu này (số doanh nghiệp hoạt động và chỉ số PCI cấp tỉnh) đều là các dữ

liệu thứ cấp dễ thu thập nhưng lại mang đến kết quả dự báo khá tốt. Dĩ nhiên kết quả dự báo hoàn toàn có thể được cải thiện thông qua việc điều chỉnh hàm số ở phương trình (2) hoặc (7) phù hợp với từng dữ liệu hay đưa thêm các biến khác cho các mô hình ứng viên.

Bảng 2: Kết quả các mô hình chứa một biến dự báo được lựa chọn để dự báo chỉ tiêu GRDP

Năm được dự báo	Khoảng dữ liệu được sử dụng để đưa ra dự báo	Biến dự báo là số doanh nghiệp			Biến dự báo là chỉ số PCI		
		m	MDA	RMSE	m	MDA	RMSE
2018	2016-2017	8	1	0,89	8	1	1,03
2019	2016-2018	10	0,98	1,12	6	0,98	1,15
2020	2016-2019	7	0,89	2,21	4	0,90	2,56
2021	2016-2020	7	0,95	1,42	6	0,94	1,83

Bảng 3: Kết quả các mô hình chứa một biến dự báo được lựa chọn để dự báo chỉ tiêu GRDP trên đầu người

Năm được dự báo	Khoảng dữ liệu được sử dụng để đưa ra dự báo	Biến dự báo là số doanh nghiệp			Biến dự báo là chỉ số PCI		
		m	MDA	RMSE	m	MDA	RMSE
2018	2016-2017	11	1	1,45	9	1	1,64
2019	2016-2018	4	0,95	1,98	10	0,98	1,54
2020	2016-2019	8	0,90	3,02	5	0,90	3,45
2021	2016-2020	7	0,95	1,75	8	0,87	2,35

Bảng 4: Kết quả các mô hình chứa cả hai biến dự báo được lựa chọn

Năm được dự báo	Khoảng dữ liệu được sử dụng để đưa ra dự báo	Biến được dự báo là GRDP			Biến được dự báo là GRDP trên đầu người		
		m	MDA	RMSE	m	MDA	RMSE
2018	2016-2017	7	1	1,44	6	1	1,83
2019	2016-2018	20	0,95	1,87	15	0,98	2,47
2020	2016-2019	14	0,89	2,99	24	0,90	3,47
2021	2016-2020	10	0,87	2,01	20	0,97	2,91

6. Kết luận

Nghiên cứu đã đưa ra một hướng tiếp cận dự báo mới cho tăng trưởng kinh tế tỉnh, thành phố trực thuộc Trung ương thông qua phương pháp mô hình động lực học hệ thống. Khác với các cách tiếp cận thông thường, phương pháp này kết hợp giữa thông kê truyền thống và học máy để phân tích các tương tác theo thời gian giữa tăng trưởng kinh tế và các yếu tố liên quan giúp đưa ra các dự báo. Kết quả cho thấy phương pháp mô

hình động lực học hệ thống có khả năng dự báo tốt kể cả khi sử dụng các biến dự báo thông dụng, dễ tìm kiếm. Từ đó, nghiên cứu đưa ra hướng dự báo mới giúp nâng cao khả năng dự báo tăng trưởng kinh tế cấp tỉnh, tạo điều kiện thuận lợi cho chính quyền trong điều hành kinh tế vĩ mô.

Tuy nhiên, nghiên cứu còn một số vấn đề có thể cải thiện và mở rộng trong các nghiên cứu tiếp theo. Cụ thể: (1) Dạng mô hình ở phương trình 2 có thể được xem xét cải thiện với các dạng mô hình phức tạp hơn giúp cải thiện khả năng dự

báo; (2) Có thể xem xét thử nghiệm thêm nhiều biến dự báo khác nhau để lựa chọn mô hình dự báo tốt hơn.

Tài liệu tham khảo

- Ang, A. et al. (2006). What Does the Yield Curve Tell Us about GDP Growth? *Journal of Econometrics*, 131(1-2), 359-403.
- Bañbura, M. et al. (2010). Large Bayesian Vector Auto Regressions. *Journal of applied Econometrics*, 25(1), 71-92.
- Berger, J. O., & Pericchi, L. R. (1996). The Intrinsic Bayes Factor for Model Selection and Prediction. *Journal of the American Statistical Association*, 91(433), 109-122.
- Bernard, H., & Gerlach, S. (1998). Does the Term Structure Predict Recessions? The International Evidence. *International Journal of FINANCE & Economics*, 3(3), 195-215.
- Brave, S. A. et al. (2019). Forecasting Economic Activity with Mixed Frequency BVARs. *International Journal of Forecasting*, 35(4), 1692-1707.
- Diebold, F. X. et al. (2006). The macroeconomy and the Yield Curve: A Dynamic Latent Factor Approach. *Journal of Econometrics*, 131(1-2), 309-338.
- Estrella, A., & Hardouvelis, G. A. (1991). The Term Structure as a Predictor of Real Economic Activity. *The Journal of Finance*, 46(2), 555-576.
- Estrella, A., & Mishkin, F. S. (1996). The Yield Curve as a Predictor of Recessions in the United States and Europe. The Determination of Long-Term Interest Rates and Exchange Rates and the Role of Expectations. Basel: Bank for International Settlements.
- Giannone, D. et al. (2008). Nowcasting: The Real-time Informational Content of Macroeconomic Data. *Journal of Monetary Economics*, 55(4), 665-676.
- Koop, G. M. (2013). Forecasting with Medium and Large Bayesian VARs. *Journal of Applied Econometrics*, 28(2), 177-203.
- Koop, G. et al. (2022). Reconciled Estimates of Monthly GDP in the US. Warwick, United Kingdom: Economic Statistics Centre of Excellence.
- Le Tuan Anh (2018). Financial System and Economic Growth: Using a Combination of Financial Variables to Forecast Economic Growth in Vietnam. Master Thesis, University of Economics Ho Chi Minh City.
- Ligneul, R. (2021). Prediction or Causation? Towards a Redefinition of Task Controllability. *Trends in Cognitive Sciences*, 25(6), 431-433.
- MacKay, D.J.C. (2003). *Information Theory, Inference and Learning Algorithms*. Cambridge: Cambridge University Press.
- Nguyen Thi Hong Trang (2021). Impact of Import and Export Activities on Vietnam's Economic Growth in the Period of 2009-2020. Master Thesis, University of Economics Ho Chi Minh City.
- Ranganathan, S. et al. (2014). Bayesian Dynamical Systems Modeling in the Social Sciences. *Plos One*, 9(1), e86468.
- Yiu, M. S., & Chow, K. K. (2010). Nowcasting Chinese GDP: Information Content of Economic and Financial Data. *China Economic Journal*, 3(3), 223-240.