

# ỨNG DỤNG PHƯƠNG PHÁP LỰA CHỌN THUỘC TÍNH VÀ CHỈ SỐ DẪN BÁO TRONG XÂY DỰNG MÔ HÌNH DỰ BÁO CÁC CHỈ TIÊU KINH TẾ VĨ MÔ

TS. Cù Thu Thủy - Ths. Đồng Thị Ngọc Lan\*

**Bài viết đưa ra các bước ứng dụng kỹ thuật lựa chọn thuộc tính trong khi xây dựng mô hình dự báo kinh tế - xã hội theo cách tiếp cận của phương pháp chỉ số dẫn báo kết hợp với mô hình kinh tế lượng. Đây là một hướng nghiên cứu và ứng dụng mới trong lĩnh vực phân tích và dự báo dữ liệu nói chung, xây dựng mô hình dự báo các chỉ tiêu kinh tế vĩ mô nói riêng.**

• Từ khóa: dự báo kinh tế vĩ mô; lựa chọn thuộc tính; chỉ số dẫn báo.

*This article applies the technique of feature selection to build a socio-economic forecasting model based on leading indicators combined with econometric model. This is a new researching approach in the field of data analysis and forecasting in general, and in forecasting macroeconomic targets in particular.*

• Keywords: macroeconomic forecasts; select properties; indicator.

Ngày nhận bài: 2/5/2019

Ngày chuyển phân biên: 10/5/2019

Ngày nhận phân biên: 15/5/2019

Ngày chấp nhận đăng: 20/5/2019

## 1. Giới thiệu vấn đề

Dự báo các chỉ tiêu kinh tế vĩ mô luôn được chính phủ và các tổ chức kinh tế đặc biệt quan tâm. Trong những thập kỷ gần đây đã có rất nhiều nghiên cứu ứng dụng các phương pháp định lượng nhằm xây dựng các mô hình dự báo kinh tế vĩ mô. Hiện tại ở Việt Nam cũng như các nước trên thế giới người ta đã hình thành và ứng dụng nhiều phương pháp khác nhau dự báo các chỉ tiêu kinh tế vĩ mô chủ yếu theo tháng, quý và năm. Trong đó, các phương pháp thường được sử dụng là: phương pháp chuyên gia, phương pháp ngoại suy, phương pháp điều tra, phương pháp chuỗi thời gian, phương pháp chỉ số dẫn báo và mô hình kinh tế lượng (hay hệ phương trình đồng thời).

Với sự phát triển hết sức mạnh mẽ của khoa học công nghệ các thông tin dữ liệu được hình thành trong quá trình hoạt động cũng như trong công tác chỉ đạo

điều hành và quản lý nền kinh tế ngày càng có cơ hội được thu thập đầy đủ và toàn diện. Khi đó những yếu tố kinh tế - xã hội có ảnh hưởng đến sự thay đổi của các chỉ tiêu kinh tế vĩ mô sẽ được nhận diện ngày càng nhiều hơn và mô hình dự báo các chỉ tiêu kinh tế vĩ mô cần phải chứa đầy đủ nhất có thể các yếu tố ấy. Khi đó vấn đề xây dựng các mô hình dự báo các chỉ tiêu kinh tế vĩ mô cần phải được thực hiện trên tập thông tin dữ liệu rất lớn và các phương pháp xây dựng mô hình dự báo kinh tế truyền thống ở trên không thể đáp ứng được yêu cầu này.

Trong những năm gần đây người ta đã tập trung nghiên cứu các phương pháp làm giảm chiều dữ liệu, đó là các phương pháp lựa chọn, biến đổi tập dữ liệu đầu vào ban đầu thành một tập dữ liệu khác nhỏ hơn rất nhiều nhưng cơ bản phản ánh khá đầy đủ các thông tin có trong tập dữ liệu gốc. Xây dựng mô hình dự báo trên tập dữ liệu có số chiều cao (big data) được ứng dụng để mô hình hóa nền kinh tế quốc dân ở những nước phát triển như Mỹ (Lawrence R. Klein, 2009), Nhật Bản (Yoshihisa Inada, 2009), Liên bang Đức (Andrei Roudoi, 2009). Phương pháp luận chung của cả 3 mô hình trên là: thực hiện một số kỹ thuật giảm chiều dữ liệu để tạo ra một số biến mới ít hơn nhiều so với số các biến ban đầu trong khi vẫn nắm bắt được những thông tin trong tập ban đầu nhiều nhất có thể, và sau đó xây dựng mô hình dự báo bằng phương pháp hồi quy nhiều biến trên tập các biến mới.

Việc nghiên cứu xây dựng mô hình dự báo trên tập dữ liệu có số chiều cao (hay big data) đã được một số nhà nghiên cứu Việt Nam quan tâm. Các nghiên cứu của Đỗ Văn Thành và cộng sự đã sử dụng kết hợp một số kỹ thuật giảm chiều dữ liệu, mô hình phương sai thay đổi tự hồi quy và phương pháp hồi quy nhiều

\* Học viện Tài chính

biến để xây dựng mô hình dự báo các chỉ số chứng khoán, mô phỏng và đánh giá rủi ro dự báo giá cổ phiếu của các công ty niêm yết trên thị trường chứng khoán. Kết quả thực nghiệm xây dựng mô hình dự báo chỉ số thị trường chứng khoán dựa vào gần 300 biến kinh tế khác cho thấy độ chính xác dự báo bằng mô hình được xây dựng theo phương pháp này là rất cao.

Bài viết này nhằm nghiên cứu đưa ra các bước ứng dụng kỹ thuật lựa chọn thuộc tính trong khi xây dựng mô hình dự báo các chỉ tiêu kinh tế vĩ mô theo cách tiếp cận của phương pháp chỉ số dẫn báo kết hợp với hệ phương trình kinh tế lượng. Đây là một hướng nghiên cứu và ứng dụng rất mới, thời sự, đang được cộng đồng quốc tế quan tâm trong lĩnh vực phân tích dữ liệu nói chung, xây dựng mô hình phân tích và dự báo nói riêng.

## 2. Phương pháp lựa chọn thuộc tính và chỉ số dẫn báo

### 2.1. Chỉ số dẫn báo

**Chỉ số tác động** là chỉ số mà sự biến động của nó có quan hệ ổn định đến biến động của một số biến (chỉ số) khác. Chỉ số có 3 loại: (1) Chỉ số dẫn báo trước (Leading Indicator): sự biến động của chỉ số này báo trước cho sự biến động của một số chỉ số khác. (2) Chỉ số dẫn báo đồng thời (Coincident Indicator): chính chỉ số này và chỉ số mà nó có quan hệ ổn định xảy ra đồng thời, tức là có tác động đồng thời. (3) Chỉ số dẫn báo sau (Lagging Indicator): cung cấp thông tin biến động trước đó của chỉ số mà chỉ số này có quan hệ.

Trong thực tiễn thường sử dụng chỉ số dẫn báo đồng thời và chỉ số dẫn báo trước. Với khả năng thể hiện được những diễn biến tương tự như các chỉ số kinh tế nhưng ở thời điểm sớm hơn, chỉ số dẫn báo có thể là một cách tiếp cận tương đối hiệu quả và ít tốn kém cho dự báo kinh tế trong ngắn hạn.

*Xây dựng mô hình sử dụng chỉ số dẫn báo đồng thời.* Có hai phương pháp chính xây dựng mô hình sử dụng chỉ số dẫn báo đồng thời là mô hình nhân tố động (dynamic factor models) và mô hình chuyển Markov (Markov switching models).

*Xây dựng mô hình sử dụng chỉ số dẫn báo trước.* Có ba phương pháp chính xây dựng mô hình sử dụng chỉ số dẫn báo trước là mô hình VAR (Vector Autoregression), mô hình nhân tố động (dynamic factor models) và mô hình chuyển Markov (Markov switching models). Ngoài ra, có thể sử dụng các phương pháp khác cho dự báo chỉ số dẫn báo trước như: mạng neural và các phương pháp không tham số

(Neural networks and nonparametric methods), mô hình nhị phân (Binary models), mô hình chuyển quan sát (Observed transition models).

### 2.2. Phương pháp lựa chọn thuộc tính

Lựa chọn thuộc tính là việc chọn ra một tập thuộc tính con từ tập thuộc tính ban đầu sao cho các tập thuộc tính con này thể hiện tốt nhất chức năng của một hệ thống quy nạp. Như vậy, phải xác định tiêu chuẩn lựa chọn thuộc tính.

#### Các mô hình lựa chọn thuộc tính

*Mô hình lọc (Filter)* là phương pháp lựa chọn thuộc tính đơn giản nhất. Đầu vào của mô hình là toàn bộ các thuộc tính của tập dữ liệu, sau khi thực hiện việc đánh giá các thuộc tính sử dụng các độ đo hoặc các tiêu chí nhất định cho trước thì đầu ra của mô hình là danh sách các thuộc tính với điểm số của từng thuộc tính. Việc lựa chọn  $M$  thuộc tính có điểm số cao nhất (hoặc thấp nhất) sẽ cho tập con thuộc tính tốt nhất theo một tiêu chí nhất định.

Ưu điểm của một số phương pháp lọc là: đơn giản, nhanh, dễ mở rộng, không phụ thuộc vào bộ phân lớp, độc lập với bộ phân lớp, độ phức tạp tính toán thấp và sử dụng cho bộ dữ liệu kích cỡ lớn.

*Mô hình đóng gói (Wrapper)* tìm kiếm tập con các thuộc tính tốt bằng cách đánh giá chất lượng của các tập thuộc tính. Việc đánh giá chất lượng thường sử dụng hiệu năng (độ chính xác dự đoán hoặc phân lớp) của thuật toán học. Để đánh giá chất lượng của tập thuộc tính, chúng sử dụng phản hồi từ mô hình dự đoán. Sở dĩ mô hình này được gọi là đóng gói bởi nó luôn “bao quanh” bộ phân lớp. Mô hình đóng gói có thể sử dụng các chiến lược tìm kiếm khác nhau chẳng hạn như tìm kiếm tuần tự, hoặc ngẫu nhiên.

Ưu điểm của mô hình đóng gói là chi phí tính toán thấp, tương tác với bộ phân lớp và hiệu năng cao hơn mô hình lọc.

*Mô hình nhúng (Embedded)* giúp cải tiến hiệu năng phân lớp và tăng tốc độ của quá trình lựa chọn. Mô hình nhúng là sự tích hợp, nhúng kỹ thuật lựa chọn thuộc tính vào mô hình học. Mô hình này kết hợp ưu điểm của mô hình lọc và đóng gói bằng cách sử dụng đồng thời tiêu chí đánh giá độc lập và các thuật toán học để đánh giá tập con các thuộc tính. Mô hình lọc có thể cung cấp một chỉ dẫn thông minh cho mô hình đóng gói, chẳng hạn như: giảm không gian tìm kiếm, một điểm khởi đầu tốt, đường tìm kiếm ngắn và thông minh hơn. Để thực hiện được mô hình nhúng người phát triển cần tìm hiểu cấu trúc của thuật toán học, xác định các tham số có thể sử dụng cho

việc đánh giá mức độ quan trọng của thuộc tính. Nói cách khác, các thuộc tính được xếp hạng ngay trong quá trình thực thi của việc học, không phải sau khi việc học hoàn thành như trong mô hình đóng gói.

### **Một số thuật toán lựa chọn thuộc tính**

Các thuật toán lựa chọn thuộc tính được xét dưới góc độ chiến lược tìm kiếm nào được sử dụng trong giải thuật đó: Tìm kiếm toàn bộ, Tìm kiếm theo kinh nghiệm và Tìm kiếm xác suất. Ngoài ra còn một vài phương pháp lựa chọn thuộc tính khác như: phương pháp trọng số thuộc tính (feature weighting method), phương pháp lai (hybrid method) và phương pháp lớn dần (incremental method).

### **3. Một số mô hình dự báo kinh tế vĩ mô dựa trên phương pháp lựa chọn thuộc tính và chỉ số dẫn báo trên thế giới**

Dự báo các chỉ số kinh tế vĩ mô luôn là vấn đề được nhiều nhà khoa học trên thế giới quan tâm. Ban đầu các nhà khoa học trên thế giới nghiên cứu các mô hình dự báo kinh tế vĩ mô dựa trên chỉ số dẫn báo như Geweke (1977), Sargent và Sims (1977), Stock J.H và Watson M.W (1989)... Đến những năm cuối thế kỷ 20, sang thế kỷ 21 lượng dữ liệu ngày càng lớn nên nhiều nghiên cứu đã kết hợp phương pháp lựa chọn thuộc tính và chỉ số dẫn báo để xây dựng mô hình dự báo kinh tế vĩ mô như nghiên cứu của Armah và Swanson (2010, 2011), Artis et al. (2002), Boivin (2005, 2006) và Stock và Watson (2005, 2006, 2012), Alessandro G. và Tommaso P. (2016),... Mỗi mô hình có những ưu nhược điểm khác nhau và các tác giả đã cố gắng tìm ra mô hình tối ưu nhất cho từng chỉ số dự báo cụ thể.

Từ năm 1977 Geweke, Sargent và Sims đã phát triển mô hình nhân tố dựa trên chỉ số dẫn báo. Tuy nhiên mô hình này chỉ được sử dụng nhiều trong phân tích kinh tế - xã hội sau khi Stock J.H và Watson M.W (1989) công bố nghiên cứu quá trình lựa chọn chỉ số và mô hình dựa trên chỉ số dẫn báo đồng thời và chỉ số dẫn báo sau. Nhóm tác giả này đã giới thiệu nhiều công trình liên quan đến chỉ số dẫn báo trong các năm 1992, 1993, 2002, 2003.

Năm 2002, Stock J.H và Watson M.W (2002a) giới thiệu nghiên cứu về dự báo kinh tế vĩ mô của Mỹ sử dụng chỉ số đồng thời và chỉ số dẫn báo trước. Dữ liệu được thu thập trong khoảng thời gian từ tháng 1/1959 đến tháng 12/1998 gồm 8 biến đo lường hoạt động của nền kinh tế Mỹ; mô hình dự báo cho các biến này trong 6, 12 và 24 tháng và sử dụng 4 phương pháp xây dựng mô hình dự báo: Mô hình dự báo chỉ số khuếch tán (Diffusion Index Forecast), mô hình

tự hồi quy (Autoregressive Forecast), mô hình tự hồi quy vector (VAR - Vector Autoregressive Forecast) và mô hình chỉ số dẫn báo trước (Multivariate Leading Indicator Forecasts).

Tiếp sau đó, Stock J.H và Watson M.W (2003b) giới thiệu nghiên cứu sử dụng chỉ số dẫn báo để dự báo kinh tế vĩ mô trong dự báo thời kỳ suy thoái kinh tế Mỹ năm 2001. Nghiên cứu này kết hợp ba phương pháp: dự báo giá trị trung bình giản đơn (the simple average), giá trị trung vị (the median) và trung bình có trọng số dựa trên sai số (MSFE-weighted average, MSFE - mean squared forecast error). Nghiên cứu chỉ ra rằng kết hợp dự báo là tốt hơn so với sử dụng mô hình tự hồi quy AR.

Năm 2013, Chauvet M. và Potter S. đã tổng kết về các nghiên cứu trong hai thập kỷ về dự báo tăng trưởng của Mỹ. Nhóm tác giả đã xây dựng mô hình dự báo tốc độ tăng trưởng của GDP từ các chỉ số quan trọng. Dữ liệu thực nghiệm từ quý 1 năm 1992 đến quý 4 năm 2010. Mô hình dự báo xây dựng dựa trên mô hình tuyến tính, phi tuyến tính, mô hình cấu trúc và mô hình giảm cấu trúc (Reduced form models). Cuối cùng là kết hợp dự báo (Forecast Combination). Kết quả dự báo của từng mô hình tại từng thời điểm được phân tích và được chỉ rõ về mức độ chính xác dự báo.

Nhiều nghiên cứu tiếp sau đó đã sử dụng chỉ số dẫn báo trong việc xây dựng mô hình dự báo các chỉ số vĩ mô, đặc biệt là dự báo tăng trưởng của Mỹ. Tuy nhiên dữ liệu thu được ngày càng lớn, để nâng cao hiệu suất dự báo một số nghiên cứu đã kết hợp thêm phương pháp lựa chọn thuộc tính, phân tích thành phần chính trong các mô hình dự báo.

Năm 2016, nhóm tác giả Alessandro G. và Tommaso P. đã nghiên cứu mô hình dự báo dựa trên việc kết hợp phương pháp lựa chọn thuộc tính và phương pháp chỉ số khuếch tán để nâng cao độ chính xác trong dự báo các chỉ số kinh tế vĩ mô. Nghiên cứu đưa ra một phương pháp học có giám sát dựa trên việc lựa chọn các thuộc tính theo độ quan trọng của chúng trong mô hình dự đoán hồi quy. Nhóm tác giả đã thực hiện thực nghiệm để so sánh phương pháp của mình với mô hình dự báo dựa trên phương pháp chỉ số khuếch tán (DI) của Stock J.H và Watson M.W (2002a) và cho thấy kết quả dự báo tốt hơn hẳn.

Năm 2015, Hazar A. và Osman T. B. đã thực hiện lựa chọn các nhân tố ảnh hưởng đến nền kinh tế vĩ mô trên thị trường chứng khoán bằng cách sử dụng phương pháp lựa chọn thuộc tính kết hợp với mạng nơron. Nghiên cứu đã sử dụng thuật toán sequential



forward selection (SFS) kết hợp mạng nơron để tìm chỉ số ảnh hưởng trong các chỉ số ứng viên. Sau khi thực hiện nghiên cứu đã chỉ ra mô hình dự đoán có tỉ lệ lỗi còn 9,6%. Với kết quả này, các nhà hoạch định tài chính, nhà đầu tư và nhà phân tích rủi ro có thể làm việc hiệu quả hơn nhiều mà không cần xử lý một lượng lớn dữ liệu, không gặp khó khăn trong việc thu thập và lưu trữ, giảm bớt chi phí dự đoán và hoàn thành phân tích, dự đoán trong một thời gian ngắn.

Năm 2015, tác giả Hyun Hak Kim đã sử dụng mô hình giảm chiều dữ liệu để dự báo các chỉ số kinh tế vĩ mô của Hàn Quốc bao gồm: lạm phát, tăng trưởng GDP, tiêu dùng, xuất khẩu và tổng vốn. Các tác giả đã sử dụng mô hình lai gồm phân tích thành phần chính (PCA), phân tích thành phần chính thưa thớt (SPCA), kết hợp với các phương pháp giảm chiều khác nhau. Nghiên cứu thực hiện 2 thực nghiệm so sánh mô hình lai này với các mô hình chuẩn như mô hình AR cho các biến kinh tế vĩ mô của Hoa Kỳ tại Kim và Swanson (2013b); so sánh với các mô hình tuyến tính và chỉ số khuếch tán trên các biến kinh tế vĩ mô của Hàn Quốc. Kết quả mô hình lai hoạt động tốt và vượt trội hơn hẳn.

Ứng dụng các phương pháp giảm chiều dữ liệu cũng được nhiều nhóm nghiên cứu thực hiện dự báo các chỉ số kinh tế vĩ mô của các nước như: Nhật Bản, Ấn Độ, Đức, Mexico, Brazil, Turkey, Nga, Trung Quốc.

**4. Đề xuất phương pháp xây dựng mô hình dự báo theo cách tiếp cận lựa chọn thuộc tính và chỉ số dẫn báo**

Kế thừa kết quả nghiên cứu trước đây, chúng tôi đề xuất phương pháp xây dựng mô hình theo cách tiếp cận lựa chọn thuộc tính và chỉ số dẫn báo gồm các phần như sau:

**Phần 1:** Phát hiện quan hệ nguyên nhân - kết quả giữa các chỉ tiêu kinh tế vĩ mô cần được dự báo.

Phát hiện quan hệ nguyên nhân - kết quả nhằm xác định tập các chỉ số dẫn báo cốt lõi nằm trong danh mục các chỉ tiêu kinh tế vĩ mô cần được dự báo. Quan hệ nhân quả được đề xuất bởi nhà toán học - giải Nobel kinh tế Granger C.W.J (1969).

Trên cơ sở xác định biến nào là chỉ số dẫn báo trước dựa trên lý thuyết kinh tế và kiểm định nhân quả Granger sẽ thiết lập các phương trình dự báo các chỉ số kinh tế vĩ mô.

**Phần 2:** Thu thập dữ liệu của tất cả các chỉ tiêu kinh tế cả trong và ngoài nước có liên quan và/hoặc có thể tiềm năng là liên quan đến các chỉ số dẫn báo cốt lõi.

Các chỉ tiêu kinh tế trong và ngoài nước có liên quan đến chỉ số dẫn báo nói chung là rất lớn. Ký hiệu S là tập các chỉ tiêu được thu thập dữ liệu này. Số lượng các chỉ tiêu trong S là rất lớn và càng lớn càng tốt, do mô hình dự báo càng chứa nhiều thông tin của các yếu tố có tác động đến sự biến động của biến phụ thuộc (hay biến nội sinh) thì chất lượng dự báo bằng mô hình (hay độ chính xác dự báo) càng cao.

**Phần 3:** Xây dựng phương trình dự báo.

Nhiệm vụ của phần này là xây dựng phương trình dự báo cho từng chỉ số dẫn báo cốt lõi. Để làm việc này cần xác định các nhân tố có ảnh hưởng đến chỉ số này. Đối với các chỉ số kinh tế vĩ mô thì số lượng các nhân tố ảnh hưởng đến nó trong thực tế là khá lớn. Nếu số biến ngoại sinh trong một phương trình dự báo là lớn trong khi kích thước mẫu dữ liệu lại nhỏ thì không thể tiến hành ước lượng mô hình hồi quy được. Vì vậy cần sử dụng các phương pháp nhằm giảm chiều dữ liệu trên mẫu. Cụ thể, với từng chỉ số dẫn báo cốt lõi cần thực hiện những nội dung sau:

(1) Sử dụng các phương pháp lựa chọn thuộc tính để xác định tập con A của S, tập này phải có ích cho việc xây dựng mô hình dự báo chỉ số dẫn báo.

Khi sử dụng các phương pháp lựa chọn thuộc tính, số lượng các thuộc tính được lựa chọn sẽ nhỏ hơn các thuộc tính ban đầu có nghĩa là mô hình sẽ sử dụng ít thông số hơn (tập A sẽ có ít thuộc tính hơn tập S). Việc này cải thiện khả năng khái quát hóa và giảm thời gian thực hiện phức tạp. Các phương pháp lựa chọn thuộc tính có thể sử dụng là: Phương pháp lọc, phương pháp gói và phương pháp nhúng.

(2) Xây dựng tập biến mới B với số lượng nhỏ hơn rất nhiều để thay thế cho tập các chỉ tiêu A.

Trường hợp tập A có số lượng biến lớn cần áp dụng thêm các phương pháp trích xuất thuộc tính. Ngược với phương pháp lựa chọn thuộc tính, phương pháp trích xuất thuộc tính sử dụng tất cả các thuộc tính ban đầu để tính toán một tập các thuộc tính mới. Các thuộc tính mới này sẽ đại diện cho các thuộc tính ban đầu. Các phương pháp thường sử dụng phổ biến là: phương pháp phân tích thành phần chính (PCA); phương pháp phân rã giá trị đơn (SVD) và phân tích yếu tố (FA).

(3) Xây dựng phương trình dự báo chỉ số dẫn báo cốt lõi theo các biến mới trong tập B và thực hiện dự báo kiểm định để đánh giá chất lượng dự báo.

(4) Xây dựng các phương trình dự báo các biến trong tập B (hay phương trình dự báo các biến ngoại

sinh của chỉ số dẫn báo cốt lõi này) bằng sử dụng mô hình trễ phân bố tự hồi quy ARDL (autoregressive distributed lag models).

Mô hình ARDL/EC hữu ích trong việc dự báo và giải thích các mối quan hệ dài hạn từ những biến động trong ngắn hạn. Mô hình đã được sử dụng dự báo trong nhiều vấn đề như: tăng trưởng kinh tế, đầu tư, lạm phát, tỷ lệ thất nghiệp, chi phí chăm sóc sức khỏe, tiêu dùng, thu nhập, ô nhiễm,...

**Phần 4:** Kết hợp các phương trình dự báo để được mô hình dự báo các chỉ tiêu kinh tế vĩ mô.

Từ các phương trình dự báo cho từng biến kinh tế trong Phần 3 sẽ được kết hợp trong mô hình dự báo các chỉ tiêu kinh tế vĩ mô. Mô hình dự báo được sử dụng trong thực tế.

**Phần 5:** Thực hiện và phân tích kết quả dự báo bằng sử dụng mô hình được xây dựng dự báo cho quan sát tiếp theo sau nữa.

## 5. Kết luận

Mô hình dự báo kinh tế vĩ mô vẫn luôn thu hút sự quan tâm của nhiều tổ chức cũng như cá nhân. Nhiều nghiên cứu được thực hiện là sự kết hợp của nhiều nhóm nghiên cứu quốc tế. Các mô hình trong lĩnh vực này cũng ngày càng phát triển. Chỉ số dẫn báo vẫn là một trong những nhân tố đóng vai trò quan trọng trong việc xây dựng mô hình. Để xây dựng các mô hình dự báo kinh tế vĩ mô trên tập dữ liệu rất lớn thì rất cần thiết phải áp dụng các phương pháp giảm chiều dữ liệu. Lựa chọn thuộc tính là một trong các hướng được ứng dụng phổ biến.

Trong bài viết này chúng tôi đã đưa ra phương pháp luận xây dựng mô hình dự báo kinh tế vĩ mô theo cách tiếp cận lựa chọn thuộc tính và chỉ số dẫn báo. Từng kỹ thuật áp dụng đã được chỉ rõ trong từng phần của quy trình xây dựng mô hình. Hiện nay chúng tôi đang tiến hành thực nghiệm phương pháp này nhằm dự báo các chỉ số kinh tế vĩ mô chủ yếu theo quý của Việt Nam. Phần thực nghiệm sẽ được nhóm tác giả công bố trong các bài viết tiếp theo./.

## Tài liệu tham khảo:

- Aruoba, S.B., Diebold, F.X. (2010). Real-time macroeconomic monitoring: real activity, inflation, and interactions. *American Economic Review* 100, pp20–24.
- Banerjee, A.N., Marcellino, M. (2005). Are there any reliable leading indicators for US inflation and GDP growth?. *International Journal of Forecasting*. In press.
- Chauvet M., Potter S., (2013). *Forecasting Output*. In: Elliott, G., Timmermann, A. (Eds.), *Handbook of Economic Forecasting*, Vol. 2. Elsevier.
- Ghysels E, Santa-Clara P, Valkanov R. 2004. *The MIDAS touch: mixed data sampling regression models*. Mimeo, Chapel Hill, NC.

Giannone D, Reichlin L, Small D. 2008. Nowcasting GDP and inflation: the real-time informational content of macroeconomic data releases. *Journal of Monetary Economics* 55: 665–676.

Giannone, D., Reichlin, L., (2013). Now-casting and the real-time data flow. In: Elliott, G., Timmermann, A. (Eds.), *Handbook of Economic Forecasting*, vol. 2. Elsevier.

Massimiliano, M. (2006). Leading indicators. *Handbook of Economic Forecasting*, Vol 1. 879-960.

Michael P. C. and Ana B. G. 2009. Forecasting us output growth using leading indicators: an appraisal using midas models. *Journal of Applied Econometrics J. Appl. Econ.* 24: 1187–1206.

Stock, J.H., Watson, M.W. (1989). New indexes of coincident and leading economic indicators. In: Blanchard, O., Fischer, S. (Eds.), *NBER Macroeconomics Annual*. MIT Press, Cambridge, MA, pp. 351–394.

Stock, J.H., Watson, M.W. (1992). A procedure for predicting recessions with leading indicators: Econometric issues and recent experience. *NBER Working Paper Series No.* 4014.

Stock, J.H., Watson, M.W. (1993). A procedure for predicting recessions with leading indicators: Econometric issues and recent experience. In: Stock, J.H., Watson, M.W. (Eds.), *Business Cycles, Indicators, and Forecasting*. The University of Chicago Press, Chicago, pp. 95–153.

Stock, J.H., Watson, M.W. (2002a). Macroeconomic forecasting using diffusion indexes. *Journal of Business and Economic Statistics* 20, 147–162.

Stock, J.H., Watson, M.W. (2002b). Forecasting using principal components from a large number of predictors. *Journal of the American Statistical Association* 97, 1167–1179.

Stock, J.H., Watson, M.W. (2003a). Forecasting output and inflation: The role of asset prices. *Journal of Economic Literature* 41 (3), 788–829.

Stock, J.H., Watson, M.W. (2003b). How did the leading indicator forecasts perform during the 2001 recession. *Federal Reserve Bank of Richmond Economic Quarterly* 89, 71–90.

Winter, J. (2011). Forecasting GDP Growth in Times of Crisis: Private Sector Forecasts versus Statistical Models. *De Nederlandsche Bank NV Working Paper* 320, November

Wieland, V., Wolters, M.H. (2011). The diversity of forecasts from macroeconomic models of the US economy. *Economic Theory* 47, 247–292.

Lawrence R. Klein (2009). *The Making of National Economic Forecasts*; Edward Elgar Publishing Limited, USA 2009.

Giovanelli, A., & Proietti, T. (2016). On the selection of common factors for macroeconomic forecasting. *Advances in Econometrics*, 35, 595–630. <https://doi.org/10.1108/S0731-9032015000035015>

Altinbas, H., & Biskin, O. T. (2015). Selecting Macroeconomic Influencers on Stock Markets by Using Feature Selection Algorithms. *Procedia Economics and Finance*. [https://doi.org/10.1016/S2212-5671\(15\)01251-4](https://doi.org/10.1016/S2212-5671(15)01251-4)

Kim, H. H. (2015). Forecasting Macroeconomic Variables Using Data Dimension Reduction Methods: The Case of Korea. *Ssrn*. <https://doi.org/10.2139/ssrn.2580591>

Yacine Ait-Sahalia (2016). *Using Principal Component Analysis to Estimate a High Dimensional Factor Model with High-Frequency Data*. Princeton University; October 7, 2016.

Đỗ Văn Thành & Nguyễn Minh Hải (2016A). Phân tích và dự báo thị trường chứng khoán bằng sử dụng chỉ số báo trước. *Kỷ yếu Hội nghị khoa học công nghệ quốc gia lần thứ IX, FAIR, Cần Thơ ngày 4-5/8/2016*, 559-566. DOI: 10.15625/vap.2016.00069.

Đỗ Văn Thành et al, (2016B). Mô hình dự báo sản xuất cao đối với các chỉ số thị trường chứng khoán. *Kỷ yếu Hội nghị khoa học công nghệ quốc gia lần thứ IX, FAIR, Cần Thơ ngày 4-5/8/2016*, 299-308. DOI: 10.15625/vap.2016.00037.

Đỗ Văn Thành (2017A). Mô hình hóa giá cổ phiếu trong ngữ cảnh dữ liệu có số chiều cao. *Kỷ yếu Hội nghị khoa học công nghệ quốc gia lần thứ X, FAIR, Đà Nẵng ngày 17-18/8/2017*, 422-434.

Đỗ Văn Thành & Đỗ Đức Hiếu (2017B). Mô phỏng và phân tích rủi ro dự báo trên tập dữ liệu có số chiều cao. *Kỷ yếu Hội nghị khoa học công nghệ quốc gia lần thứ X, FAIR, Đà Nẵng ngày 17-18/8/2017*, 467-479.