



PHƯƠNG PHÁP DỰ BÁO NGẮN HẠN GIÁ CÁC MẶT HÀNG NÔNG NGHIỆP QUY MÔ LỚN: MINH HỌA BỞI TRƯỜNG HỢP CÁC THỊ TRƯỜNG NÔNG NGHIỆP Ở BẮC KINH

Haoyang Wu, Huaili Wu, Minfeng Zhu, Weifeng Chen và Wei Chen

Tóm tắt:

Để dự báo giá các mặt hàng nông nghiệp tại các thị trường bán buôn khác nhau ở một thành phố, nghiên cứu này đề xuất một mô hình hỗn hợp, kết hợp mô hình dự báo tự hồi quy kết hợp trung bình trượt theo chuỗi thời gian ARIMA và phương pháp hồi quy PLS (Partial Least Squares) là phương pháp được sử dụng trong phân tích đa biến dựa trên các yếu tố thời gian và không gian. Mô hình hỗn hợp này có thể đưa ra được kết quả dự báo hàng tuần về giá của mặt hàng nông nghiệp ở các thị trường khác nhau. Cùng với đó, nghiên cứu này đặt ra các biến số để đo lường xu hướng thay đổi giá dựa trên sự thay đổi của các biến số ngoại sinh và giá các mặt hàng. Do đó, mô hình hỗn hợp này đạt được cảnh báo về thay đổi giá hàng ngày bằng mạng liên kết thông tin thông minh. Mô hình dự báo được thực hiện kiểm tra với dữ liệu của một số mặt hàng nông nghiệp và phân tích sai số. Kết quả cho thấy mô hình hỗn hợp cho độ chính xác cao hơn trong việc dự báo giá hàng tuần và đưa ra các giá trị cảnh báo hàng ngày về giá các mặt hàng nông nghiệp so với từng mô hình, và đến một mức độ nào đó, mô hình này có thể dùng để dự báo hàng ngày sự thay đổi giá của các mặt hàng nông nghiệp.

Từ khoá: Thay đổi cảnh báo, mô hình hỗn hợp, mạng liên kết thông tin thông minh, dự báo giá...

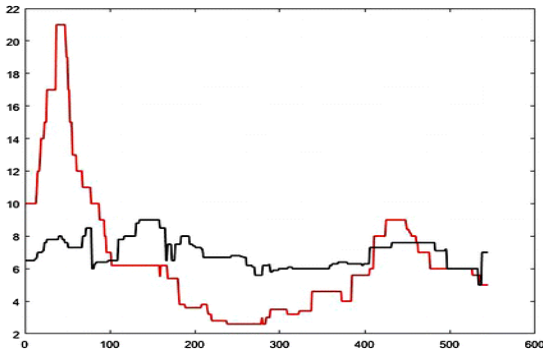
1. Thông tin chung

Có câu nói “Thực phẩm là điều cần thiết tối thiểu nhất của người dân”. Giá mặt hàng nông nghiệp là điều cần thiết quan trọng có liên quan chặt chẽ đến cuộc sống của người dân. Sự biến động của giá cả mặt hàng nông nghiệp bị ảnh hưởng bởi các yếu tố kinh tế xã hội. Do đó, dự báo chính xác xu hướng thay đổi giá có thể định hướng hành vi tiêu dùng của người dân. Đặc biệt điều này có ý nghĩa quan trọng đối với một số vấn đề xã hội nóng bỏng như dự đoán xu hướng kinh tế vĩ mô.

Có nhiều mặt hàng nông nghiệp trên thị trường, trong đó giá các mặt hàng có thể bị ảnh hưởng bởi nhiều yếu tố, thậm chí cùng một mặt hàng có thể được định giá khác nhau ở các thị trường khác nhau. Ví dụ, Hình 1 là giá mặt hàng nông nghiệp hằng ngày ở các thị trường bán buôn chợ nông nghiệp Baliqiao, quận Thông Châu, Bắc Kinh và chợ nông nghiệp Shunxin Shimen, quận Thuận Nghĩa, Bắc Kinh, từ tháng 1/2014 đến tháng 6/2015. Hình 1 cho thấy xu hướng giá cả các mặt hàng nông nghiệp của hai thị trường có sự khác biệt lớn. Người tiêu dùng và các cơ quan hành chính chắc chắn muốn có thông

tin tổng thể về giá dự báo tại một số thị trường nông nghiệp.

Hình 1: Biểu diễn giá bán buôn hàng ngày ở hai thị trường



Đường màu đen là thị trường Baliqiao, quận Thông Châu, Bắc Kinh; đường màu đỏ là thị trường Shunxin Shimen, quận Thuận Nghĩa, Bắc Kinh.

Giá cả mặt hàng nông nghiệp bị ảnh hưởng bởi sự kết hợp của nhiều yếu tố, bao gồm mối quan hệ cung-cầu, thời tiết, chính sách, v.v... Các yếu tố này không thể định lượng theo cùng tiêu chuẩn và có ảnh hưởng không giống nhau đến các mặt hàng nông nghiệp khác nhau ở các thị trường bán buôn đa dạng để dự báo giá cả hàng hóa nông nghiệp.

Dự báo ngắn hạn hiện nay, bao gồm dự báo thay đổi giá hàng tuần và thay đổi giá hàng ngày đang gặp nhiều thách thức, do sự biến động của giá các mặt hàng nông nghiệp bị ảnh hưởng bởi sự kết hợp các yếu tố không chắc chắn. Trong khi đó, điều quan trọng dự báo ngắn hạn là phải dự báo khi có sự thay đổi mạnh mẽ sẽ xảy ra, như trong các trường hợp giá hàng nông nghiệp thay đổi và dao động ổn định. Hiện tại, có ba phương pháp để dự báo ngắn hạn giá mặt hàng nông nghiệp:

1. Phương pháp chuỗi thời gian, bao gồm các phương pháp dự báo ngắn hạn như mô hình ARIMA, mô hình dự báo về phương

sai có điều kiện GARCH (Generalised Autoregressive Conditional Heteroskedasticity là mô hình tổng quát hóa mô hình ARCH; Trong đó, mô hình ARCH: Mô hình phương sai có điều kiện của sai số thay đổi tự hồi quy). Những phương pháp này chỉ dựa trên giá lịch sử của mặt hàng nông nghiệp mà không quan tâm đến các yếu tố khác. Do đó, các mô hình này không hoạt động được khi giá bị ảnh hưởng bởi các yếu tố không theo thời gian.

2. Phương pháp hồi quy, bao gồm mô hình tự hồi quy véc tơ, mô hình trung bình trượt kết hợp tự hồi quy véc tơ. Những phương pháp này có những yếu tố khác nhau để xem xét. Tuy nhiên, do hạn chế của các điều kiện sử dụng, nên không thể áp dụng một mô hình duy nhất để dự báo một số loại mặt hàng nông nghiệp khác nhau trong cùng một thời điểm.

3. Các phương pháp học tập (machine learning, deep learning), bao gồm các mô hình mạng liên kết các thông tin thông minh. Những phương pháp này có phạm vi áp dụng rộng rãi. Tuy nhiên, khi dự báo các mặt hàng nông nghiệp khác nhau, các tác động không thể đảm bảo và có thể xảy ra quá tải. Do đó, các phương pháp này thường được sử dụng để dự báo một số loại mặt hàng nông nghiệp cụ thể.

Các phương pháp hiện tại chủ yếu dựa trên một mô hình duy nhất và nhằm mục tiêu vào một mặt hàng nông nghiệp nhất định trong một thị trường cụ thể. Những phương pháp này không được kiểm tra bằng dữ liệu quy mô lớn và chỉ có thể được sử dụng trong một phạm vi nhỏ. Ngoài ra, các phương pháp hiện tại không xem xét các biến số kinh tế ngoại sinh và sự tương tác giữa các thị trường khác nhau cùng với các yếu tố mùa vụ, làm giảm tính chính xác của dự báo về thời gian biến đổi và biên độ dao động.

Nghiên cứu này đưa ra một mô hình dữ liệu với các thử nghiệm mẫu để giải quyết các vấn đề đã đề cập ở trên và đưa ra một mô hình hỗn hợp mới để dự báo giá hàng nông nghiệp. Chúng ta xem xét lại mô hình ARIMA bằng phương pháp hồi quy PLS, xem xét đến ảnh hưởng của các thị trường nông nghiệp khác nhau trong cùng một thành phố. Chúng ta dự báo sự thay đổi giá hàng tuần của thị trường nông nghiệp bằng cách xem xét các tương tác giữa các thị trường khác nhau và các yếu tố mùa vụ.

Trên cơ sở mô hình hỗn hợp các yếu tố thời gian và không gian, nghiên cứu này cũng đề xuất một mô hình cảnh báo thay đổi giá với một sự thay đổi "khẩn cấp" để định lượng xu hướng thay đổi giá các mặt hàng nông nghiệp. Chúng ta sử dụng các mạng liên kết thông tin thông minh để phân tích "mức độ khẩn cấp" và các biến ngoại sinh khác, và đưa ra dự báo giá trị của hệ số "tính cấp bách" đáp ứng cho sự cảnh báo thay đổi "khẩn cấp". Do đó, ở một mức độ nào đó, chúng ta có thể sử dụng mô hình này dự đoán xu hướng thay đổi giá hàng ngày của các mặt hàng nông nghiệp.

Phương pháp được đề xuất bởi nghiên cứu có những dự báo tốt hơn về 20 loại hàng hoá nông nghiệp tại các thị trường nông nghiệp ở Bắc Kinh. Phân tích sai số và phân tích kết quả hiển thị cho thấy mô hình hỗn hợp của nghiên cứu này đã đạt được kết quả dự báo khả quan. Mô hình hỗn hợp cho thấy sự cải tiến cả về tính chính xác và hiệu quả dự báo so với bất kỳ mô hình đơn lẻ nào khác. Sự đột phá của mô hình được đề xuất bởi nghiên cứu này về cơ bản, bao gồm:

1. Đưa ra một mô hình đề xuất cảnh báo hàng ngày để định lượng và dự báo xu hướng thay đổi hàng ngày của các mặt hàng nông nghiệp.

2. Mô hình này có thể được sử dụng để dự báo rất nhiều loại mặt hàng nông nghiệp có hiệu quả tốt.

3. Mô hình này thực hiện dự báo đồng thời các mặt hàng nông nghiệp ở các thị trường khác nhau trong một thành phố, bằng cách xem xét các yếu tố không gian.

Nghiên cứu này bắt đầu từ các nghiên cứu hiện tại, kết hợp các mô hình dự báo đơn và đề xuất một mô hình dự báo hỗn hợp, mô hình này có thể dự báo đồng thời giá mặt hàng nông nghiệp ở các thị trường khác nhau. Ngoài ra, mô hình hỗn hợp cung cấp kết quả ổn định và chính xác hơn so với các mô hình đơn lẻ hoặc một số mô hình khác. Cùng với đó, nhóm nghiên cứu xây dựng một mô hình cảnh báo giá hàng ngày dựa trên mạng liên kết thông tin thông minh và trong chừng mực nào đó có thể thực hiện dự báo giá hàng ngày của các mặt hàng nông nghiệp, có giá trị ứng dụng cho người dân và các cơ quan hành chính.

2. Tổng quan về nghiên cứu

Mô hình dự báo giá mặt hàng nông nghiệp chủ yếu chia thành hai loại. Một là các mô hình cấu trúc, phân tích các yếu tố giá dựa trên quan điểm kinh tế. Trên quan điểm kinh tế vi mô và kinh tế lượng, Lord đề xuất rằng giá có sự tương tác với nhu cầu, cung cấp và kiểm kê, do đó xây dựng một mô hình dự báo giá các mặt hàng nông nghiệp là một phương trình gắn với thời gian thực.

Một loại khác là các mô hình phi cấu trúc, có nghĩa bỏ qua nguyên tắc dựa trên quan điểm kinh tế và nghiên cứu trực tiếp chuỗi thời gian về giá cả. Box và Jenkins đưa ra mô hình ARIMA. Việc lập mô hình, ước lượng tham số, kiểm tra mô hình và phân tích kết quả dự báo dựa trên giả định: Giá trong tương lai có liên quan đến giá gốc và các biến ngẫu nhiên. Mô hình này bỏ qua ảnh

hưởng của tất cả các yếu tố khác. Rausser và Carter sử dụng mô hình ARIMA để phân tích giá của đậu nành, dầu đậu nành và bột đậu nành trong tương lai, kết luận cho thấy: Đậu nành và bột đậu nành thực hiện tốt trong mô hình ARIMA hơn là trong mô hình ngẫu nhiên. Granger chỉ ra sự khác biệt vượt quá đã xảy ra khi sử dụng mô hình ARIMA để xử lý dữ liệu có lưu trữ dài hạn, do đó đề xuất mô hình tự hồi quy kết hợp mô hình trung bình trượt phân đoạn ARFIMA. Barkoulas và cộng sự tính chênh lệch khác biệt về giá tương lai của hàng hóa nông nghiệp và nhận thấy một số giá tương lai có lưu trữ dài hạn, do đó đáp ứng yêu cầu của mô hình ARFIMA. Mô hình ARIMA dựa trên dữ liệu theo chuỗi thời gian do đó, mô hình này bỏ qua các ảnh hưởng của các yếu tố khác về giá. SIMS đề xuất mô hình tự hồi quy có hướng (VAR) để xây dựng chuỗi thời gian (nhằm khắc phục khiếm khuyết mô hình ARIMA). PARK đã sử dụng các mô hình VAR khác nhau để phân tích giá của thức ăn gia súc và bò, rút ra kết luận: Mô hình tự hồi quy có hướng Bayesian (BVAR) và mô hình không giới hạn hướng (UVAR) (BVAR và UVAR là những mô hình thuộc VAR) đã đưa ra các kết quả dự báo tốt hơn cả về mặt hạn chế của mô hình VAR (RVAR) và mô hình VARMA (tự hồi quy kết hợp trung bình trượt) trong trường hợp này.

Nhưng làm mượt các dữ liệu theo sự khác biệt không thể giải thích được từ quan điểm kinh tế, Engle và Granger phân tích sự kết hợp tuyến tính của các biến dựa trên mối quan hệ đồng cấp. Họ đã đề xuất mô hình hiệu chỉnh sai số vec tơ (VEC), do đó làm mượt dữ liệu theo một cách khác. Do hạn chế của một mô hình duy nhất thường không thể dự báo chính xác giá cả. Yu Le và cộng sự dự báo giá các mặt hàng nông nghiệp tương ứng với ba mô hình san bằng mũ, mô hình hồi quy tuyến tính giản đơn và mô hình

dự báo xám, sau đó tìm ra sự kết hợp tuyến tính tối ưu có sai số (MSE).

Từ lâu các nhà nghiên cứu đã nghiên cứu về xu hướng dài hạn của giá cả mặt hàng nông nghiệp có sự tuần hoàn rõ rệt. Beveridge và Nelson đề xuất một phương pháp tổng quát để làm mượt liên tục các chuỗi thời gian. Phương pháp này chỉ yêu cầu sự thay đổi liên tục của chuỗi thời gian là ổn định. Harvey đã đề xuất cấu trúc mô hình chuỗi thời gian (STS) bao gồm một loạt các mô hình chuỗi thời gian ổn định. Phương pháp này tránh được sự nhận dạng mô hình và tách thành công yếu tố mùa vụ từ sự thay đổi giá các mặt hàng nông nghiệp. Điều này giúp phù hợp với quan điểm kinh tế. Gần đây, một số phương pháp mới đã được đề xuất. Davidson và cộng sự đã sử dụng phương pháp hồi quy có tham số dựa trên phân tích wavelet để ước lượng khoảng thời gian biến đổi và minh họa tiềm năng của phương pháp này. Sự biến động giá là một hướng nghiên cứu quan trọng. Nhiều ngẫu nhiên thường khó có thể quan sát, nhưng là vấn đề quan trọng trong việc dự báo giá cả. Engle đã đề xuất mô hình đặc tả mô hình hóa chuỗi thời gian (ARCH). Mô hình này tin rằng độ biến thiên của nhiễu không phải là hằng số, thay vào đó nó bị ảnh hưởng bởi thông tin trong quá khứ. Bollerslev đã đề xuất mô hình GARCH, đây là mô hình cải tiến mô hình ARCH. Mô hình GARCH thực hiện tốt hơn trong việc mô tả chuỗi thời gian với bộ nhớ lưu trữ dài hạn. Krytsou và cộng sự đề xuất dự báo dài hạn về chuỗi thời gian với nhiễu không còn hoạt động. Thay vào đó, mô hình Mackey-Glass-GARCH có thể được sử dụng. Schroeder đã chia nhiễu về giá thành 4 loại dựa trên độ mạnh của nhân tố quy ước, đặc biệt là nhiễu trắng, nhiễu hồng, nhiễu nâu và nhiễu đen. Các nghiên cứu thực nghiệm sử dụng phương pháp này của Labys

đã đi đến kết luận: Hầu hết các mặt hàng nông nghiệp có nhiều đơn, có nghĩa là dự báo giá hàng nông nghiệp là khá khó khăn. Mạng liên kết thông tin thông minh đã trở thành một phương pháp hữu hiệu để dự báo giá cả. Lapedes và Farber dự báo giá với mạng liên kết thông tin thông minh. Nó có thể phù hợp với một đường cong tùy ý, và có khả năng bao quát tốt. Một hướng dự báo khác là mô hình biến động, Andersen và cộng sự so sánh một số mô hình bao gồm GARCH, biến động ngẫu nhiên và biến động nhiều biến thể. Manfredo và cộng sự sử dụng mô hình biến động dự báo sự biến động của giá ngô và bò. Kroner cộng sự dự báo giá vàng, ngô, bông v.v... với mô hình kỳ vọng - phương sai. Ngày nay, các nhà nghiên cứu đang cân nhắc kết hợp các phương pháp dự báo cấu trúc và phi cấu trúc, làm cho các kết quả dự báo có ý nghĩa kinh tế hơn.

Nghiên cứu dự báo giá hàng tuần của các mặt hàng nông nghiệp bằng cách tích hợp hai mô hình sử dụng phương pháp chuỗi thời gian dựa trên tính tuần hoàn của hàng hoá nông nghiệp, và sử dụng mô hình không gian dựa trên sự liên quan của các thị trường khác nhau. Hơn nữa, nghiên cứu này xử lý các biến ngoại sinh và đạt được cảnh báo về giá hàng ngày bởi các mạng liên kết thông tin thông minh.

3. Xử lý dữ liệu

Nguồn dữ liệu

Dữ liệu giá cả hàng hóa nông nghiệp được khai thác từ trang web của bộ phận thương mại điện tử¹, dữ liệu bao gồm giá hàng ngày của tất cả các mặt hàng nông nghiệp ở các thị trường bán buôn ở Trung Quốc từ ngày 2/1/2014 đến ngày 30/6/2015. Một số dữ liệu bị thiếu do kỳ nghỉ hoặc do

nguyên nhân của mạng lưới thu thập. Nghiên cứu này sử dụng dữ liệu ở Bắc Kinh làm mẫu.

Nghiên cứu này đưa ra dữ liệu thời tiết², tỷ giá hối đoái giữa Trung Quốc và Mỹ³, và giá dầu thô quốc tế⁴ được xem như là các biến kinh tế ngoại sinh. Dữ liệu hàng ngày về thời tiết, tỷ giá hối đoái và giá dầu thô quốc tế bắt đầu được thu thập từ ngày 1/1/2014 đến ngày 30/6/2015, dữ liệu về tỷ giá hối đoái và giá dầu thô quốc tế chỉ có trong những ngày làm việc. Mô hình được xây dựng trong nghiên cứu này dựa trên một bộ dữ liệu lớn. Nghiên cứu này phân tích và đề cập đến giá của tất cả các mặt hàng nông nghiệp ở tất cả các thị trường cũng như số liệu hàng ngày của các biến ngoại sinh khác trong cùng thời điểm và cuối cùng thu được kết quả dự báo.

Xử lý mẫu

Nghiên cứu này sử dụng dữ liệu của 80% ngày trước đó làm mẫu nghiên cứu, và dự báo giá của 20% ngày sau. Giá thực của 20% ngày sau được sử dụng để đánh giá kết quả dự báo.

Mối quan hệ giữa giá hàng hóa nông nghiệp trong ngày t và ngày $t-1$ (ngày trước đó) có thể thay đổi hoặc không thay đổi. Thông thường giá hàng nông nghiệp sẽ luân phiên thay đổi hoặc giữ ổn định. Nhóm nghiên cứu quan sát dữ liệu và thông báo rằng giá thường giữ nguyên trong một thời gian trước khi có sự thay đổi đột ngột. Do đó, chúng ta giả định trong mô hình cảnh báo hàng ngày giá giữ ổn định và thay đổi khi các biến ngoại sinh đạt đến một mức độ nhất định. Vì vậy, nghiên cứu này chỉ định giá ngày cuối cùng cho những ngày thiếu dữ liệu, thay vì sử dụng nội suy tuyến tính theo

² <http://en.tutiempo.net/climate/2014/ws-545110.html>

³ http://www.eia.gov/dnav/pet/pet_pri_spt_s1_d.htm

⁴ <http://www.chinabgao.com/stat/stats/44589.html>

¹ http://nc.mofcom.gov.cn/channel/gxdj/jghq/jg_list.shtml

cách làm thông thường. Điều đó có nghĩa là: Giá (t) = Giá (t-1) nếu Giá (t) bị khuyết. Dữ liệu còn khuyết của các biến ngoại sinh khác cũng được gán theo cùng một cách như trên. Phương pháp tiền xử lý dữ liệu là khác nhau trong các mô hình phụ khác nhau. Phương pháp tiền xử lý trong nghiên cứu này tuân theo nguyên tắc duy trì xu hướng thay đổi giá cả và bỏ qua sự biến động lớn về giá trong khoảng thời gian khá ngắn do người tiêu dùng không thể phản ứng lại với sự biến động lớn về giá cả.

Phương pháp tiền xử lý dữ liệu là khác nhau trong các mô hình phụ khác nhau. Phương pháp tiền xử lý trong nghiên cứu này tuân theo nguyên tắc duy trì xu hướng thay đổi giá cả và bỏ qua sự biến động lớn về giá trong khoảng thời gian khá ngắn do người tiêu dùng không thể phản ứng lại với sự biến động lớn về giá cả.

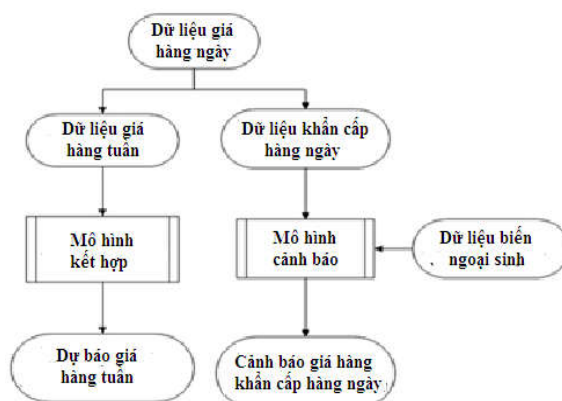
Dự báo giá và mô hình cảnh báo

Nghiên cứu này sử dụng một mô hình hỗn hợp để giải quyết các yếu tố khác nhau, tích hợp các kết quả dự báo của các yếu tố khác nhau và nhận được kết quả dự báo cuối cùng.

Mô hình hỗn hợp có thể được chia thành hai phần: Mô hình dự báo thay đổi giá hàng tuần và mô hình cảnh báo thay đổi giá hàng ngày. Mô hình dự báo thay đổi theo tuần (4.1), mô hình dự báo nhân tố không gian (4.2) và mô hình tích hợp không gian thời gian (4.3), tương ứng với các yếu tố mùa, yếu tố không gian (ảnh hưởng của sự thay đổi giá cả ở các thị trường khác) và sự kết hợp các đầu ra của các mô hình phụ. Mô hình cảnh báo thay đổi giá đề cập đến các biến ngoại sinh (4.4). Nghiên cứu này sử dụng các phương pháp tiền xử lý dữ liệu khác nhau theo các mô hình con khác nhau để có được kết quả dự báo tốt hơn. Khung

của mô hình tổng thể được thể hiện trong Hình 2.

Hình 2: Khung mô hình tổng thể



Mô hình dự báo thời gian

Hầu hết các bảng dữ liệu giá nông nghiệp dựa trên mô hình chuỗi thời gian. Các mô hình này không yêu cầu dữ liệu của bất kỳ biến nào khác và tính khả thi đã được chứng minh. Do đó các mô hình chuỗi thời gian vẫn là một phần quan trọng trong mô hình hỗn hợp của nghiên cứu này.

Xử lý dữ liệu

Các mô hình chuỗi thời gian xử lý tốt trong việc phân tích và dự báo dữ liệu dài hạn, có xu hướng rõ ràng và biến động thường xuyên. Do đó, nghiên cứu này sử dụng giá hàng tuần theo mô hình chuỗi thời gian, bằng cách tính giá trung bình hàng ngày trong một tuần. Mục đích là để nâng cao độ chính xác dự báo, bằng cách tránh ảnh hưởng của biên độ dao động giá bất thường.

Mô hình ARIMA (p, d, q)

Nghiên cứu này dự báo giá hàng tuần, giá các mặt hàng nông nghiệp với mô hình ARIMA làm mô hình dự báo thời gian. Mô hình ARIMA là một mô hình cổ điển và được sử dụng rộng rãi. Các tham số p, d, q tương ứng đại diện cho tự động hồi quy, thời gian khác nhau làm mượt chuỗi thời gian và thứ

tự trung bình trượt. Công thức toán học của mô hình ARIMA là:

$$\theta_i^{(1)}(t) = \mu + \rho(B)\varphi(B)^{-1}\varepsilon(t)$$

$\theta_i^{(1)}(t)$ là các biến ngẫu nhiên, trong mô hình này là sự thay đổi giá cả hàng tuần theo thời gian t

μ là giá trị trung bình

B là toán tử dịch chuyển lùi

$B(W(t)) = W(t-1) \cdot \rho(B)$ là toán tử trung bình trượt

$\rho(B) = 1 - \rho_1^{(B)} - \dots - \rho_p^{(B)}\phi(B)$ là toán tử tự hồi quy

$\phi(B) = 1 - \phi_1^{(B)} - \dots - \phi_p^{(B)} \cdot \varepsilon(t)$ là toán tử xáo trộn độc lập, hoặc sai số ngẫu nhiên.

Trong mô hình này, đầu tiên chúng ta đưa các dữ liệu vào thử nghiệm tính ADF (kiểm tra tăng tính tính DF). Nếu bộ dữ liệu sai trong quá trình kiểm tra sẽ có sự khác biệt của bộ dữ liệu cho đến khi nó có thể vượt qua sự kiểm tra hoặc loại bỏ bộ dữ liệu này. Trong thực tế, hầu hết dữ liệu giá hàng hóa nông nghiệp có thể vượt qua sự kiểm tra ADF, do đó trong mô hình ARIMA (p, d, q) chúng ta gán $d = 1$. Khi đó, giá trị của p và q được lựa chọn bởi AIC (tiêu chí thông tin Akaike) thử nghiệm. Đặt p và q trong khoảng vòng 1 đến 10. Sau đó, đưa bộ dữ liệu nghiên cứu vào kiểm tra AIC, và tìm ra p và q của giá trị AIC thấp nhất. Phải mất một thời gian dài để tìm ra p và q cho mỗi mặt hàng nông nghiệp và một giải pháp thay thế là trực tiếp chọn $p = 10, q = 8$. Các kết quả dự báo là chính xác.

Mô hình dự báo các yếu tố không gian

Nghiên cứu này dự báo giá của tất cả các thị trường nông nghiệp trong một thành phố. Phần này nghiên cứu chủ yếu xem xét

ảnh hưởng của sự thay đổi giá trong các thị trường nông nghiệp khác nhau. Việc xem xét yếu tố này dựa trên hành vi của người tiêu dùng những thay đổi về giá sẽ ảnh hưởng đến hành vi tiêu dùng trong cùng một thành phố.

Xử lý dữ liệu

Người tiêu dùng sẽ không phản ứng với những thay đổi giá trong cùng một ngày. Do đó có một khoảng thời gian trễ trong ảnh hưởng của sự thay đổi giá cả ở các thị trường khác nhau. Nghiên cứu này lấy giá trị trung bình hàng tuần của sự chênh lệch giá, nhằm giữ lại xu hướng thay đổi giá cả, và đủ thời gian cho thời gian phản ứng chậm trễ. Bên cạnh thời gian trễ, sự liên quan giữa các thị trường bán buôn khác nhau là một khó khăn trong thiết kế mô hình dự báo, vì hầu hết các phương pháp trong phân tích hồi quy yêu cầu các biến phải độc lập với nhau. Mục đích của việc thiết kế mô hình trong phần này là đánh giá mức độ ảnh hưởng giữa các thị trường nông nghiệp. Do đó, nghiên cứu này sử dụng phương pháp PLS để dự báo giá dựa trên yếu tố không gian.

Mô hình PLS

Phương pháp phần tử nhỏ nhất bao gồm một thủ tục tương tự như phân tích thành phần chính (PCA), do đó có thể được sử dụng cho các biến có nhiều tương quan. Đối với các mặt hàng nông nghiệp trên thị trường i , nhóm nghiên cứu muốn dự báo thay đổi giá hàng tuần $\theta(t)$ vào thời điểm t . Các biến độc lập là sự thay đổi giá của các thị trường khác $\theta_1^{(t-1)}, \theta_{i-1}^{(t-1)}, \theta_{i+1}^{(t-1)}, \dots, \theta_n^{(t-1)}$ thời gian $t-1$. Nhóm nghiên cứu chuẩn bị quy trình thử nghiệm với các dữ liệu. Sau đó, áp dụng phương pháp PLS xây dựng mô hình và thu được mối quan hệ hồi quy giữa sự thay đổi giá của thị trường mục tiêu và sự thay đổi giá

của các thị trường khác vào thời điểm cuối cùng. Chúng ta có được giá trị dự báo $\theta_i^{(2)}(t)$ của mô hình không gian bằng các mối quan hệ hồi quy.

Thông qua phương pháp PLS, chúng ta có thể lấy được hệ số hồi quy giữa mỗi cặp thị trường nông nghiệp, trong đó một số mức độ phản ánh mối quan hệ có ảnh hưởng giữa các thị trường nông nghiệp.

Hơn nữa, nhóm nghiên cứu sử dụng PLS thay vì trực tiếp sử dụng mô hình ARIMA đa biến vì mô hình ARIMA đa biến đòi hỏi các biến phải được đồng nhất. Tuy nhiên, sự thay đổi giá của các thị trường nông nghiệp ở Trung Quốc cho thấy chuỗi giá thay đổi ở các thị trường nông nghiệp khác nhau có cách tính khác nhau. Do đó, các thị trường khác nhau, không trong phạm vi thử nghiệm đồng nhất. Vì vậy, nhóm nghiên cứu xem xét về cách sử dụng phương pháp PLS, một phương pháp tổng quát hơn, để có thể xử lý được với tất cả các loại đa biến.

Mô hình dự báo hỗn hợp giá hàng tuần

Sau khi xử lý trước hai mô hình ở trên, chúng ta có thể nhận được hai nhóm bộ dữ liệu, dự báo sự khác biệt về mô hình thời gian và không gian của tuần tới (của tuần trước trong quá trình nghiên cứu). Dựa vào phân tích ở trên, chúng ta đã biết sự thay đổi giá hàng tuần chịu ảnh hưởng bởi yếu tố theo thời gian và yếu tố không gian. Tuy nhiên, chúng ta không biết chi tiết hai yếu tố có mối quan hệ với nhau như thế nào. Có hai cách để tìm ra mối quan hệ của hai yếu tố: (i) Phân tích kinh tế; (ii) Kiểm tra một số mô hình có thể với dữ liệu lịch sử và chọn một trong những mô hình tốt nhất.

Mô hình hỗn hợp trong nghiên cứu này là để thấy được sự thú vị trong dự báo của hai mô hình dựa trên bộ dữ liệu nghiên cứu, sau khi đưa kết quả dự báo vào các biến độc

lập và thay đổi giá thực hàng tuần thành hồi quy tuyến tính. Mô hình tuyến tính là một mô hình hiệu quả và tương đối đơn giản, bên cạnh đó nó có thể tiết lộ trọng số của từng yếu tố trong mỗi quan hệ. Ngoài ra, hiện nay chưa có nghiên cứu tương đối về trọng số của hai yếu tố được tổng hợp và số lượng mẫu hiện có là nhỏ. Cuối cùng chúng ta có thể nhận được các mối quan hệ hồi quy của hai yếu tố ảnh hưởng đến sự thay đổi giá theo trọng số khác nhau:

$$\theta_i^{(t)} = \alpha_1 \theta_i^{(1)}(t) + \alpha_2 \theta_i^{(2)}(t)$$

$\theta_i^{(1)}(t)$ là giá trị dự báo của mô hình ARIMA của thị trường i.

$\theta_i^{(2)}(t)$ là giá trị dự báo của mô hình PLA của thị trường i.

Chúng ta có α_1 và α_2 thông qua hồi quy của dữ liệu lịch sử. Chúng ta có thể đưa kết quả dự báo của hai mô hình phụ vào phương trình hồi quy và nhận được các giá trị dự báo cuối cùng về sự thay đổi giá hàng tuần.

Mô hình cảnh báo

Như đã đề cập ở trên, giá hàng hóa nông nghiệp có xu hướng thay đổi sau khi giữ ổn định trong một thời gian. Không có quy tắc rõ ràng nào được quan sát thấy, do đó thời điểm chính xác giá thay đổi là khó dự đoán. Các giải pháp của nghiên cứu này là phương pháp tiền xử lý dữ liệu và có được mức giá hàng tuần. Điều quan trọng là người tiêu dùng phải biết được những thay đổi giá cả có thể xảy ra hàng ngày. Do đó, bài viết này đề xuất một mô hình cảnh báo thay đổi giá để định lượng cường độ thay đổi giá cả có thể bằng giá trị đầu ra.

Giải thuyết biến động giá

Thứ nhất, mô hình này đề xuất một giả thuyết bên cạnh sự biến động xung quanh giá trị trung bình, tất cả các thay đổi

giá đều do sự thay đổi các biến ngoại sinh. Hàng hoá nông nghiệp là một thành phần của nền kinh tế thị trường. Giá của nó không bị ảnh hưởng bởi các biến số kinh tế khác và các biến ngoại sinh, bao gồm: Thời tiết và thay đổi của giá cả. Sự thay đổi này chắc chắn không phải là một dao động xung quanh giá trị trung bình. Vì vậy, đó là một giả thuyết hợp lý.

Sự ảnh hưởng bởi các biến ngoại sinh sẽ tích lũy theo thời gian. Do sự không chắc chắn của ảnh hưởng, phân tích ảnh hưởng tại một thời điểm duy nhất có lỗi lớn. Do đó, phần tiếp theo sẽ đề xuất một số phương pháp để đối phó với các biến ngoại sinh, theo cách này để đồng bộ hóa sự thay đổi giá với sự tích lũy các biến ngoại sinh.

Định nghĩa tính cấp bách và tính mẫu

Việc xử lý dữ liệu giá trong mô hình cảnh báo theo các giả thuyết đã nêu ở trên, trong khi chúng ta hy vọng sẽ đạt được dữ liệu hàng ngày với xu hướng duy trì, điều này có nghĩa là giữ các thông tin liên quan của từng ngày. Vì vậy, chúng ta sử dụng cách sau để xử lý dữ liệu:

1. Làm mượt giá: Dữ liệu giá linh hoạt mỗi ngày, do đó chúng ta có thể giữ lại xu hướng giá và loại bỏ sự biến động không có ý nghĩa. Chúng ta sử dụng phương pháp làm mượt giá trung bình trượt đối với dữ liệu lịch sử của các mặt hàng. Lấy giá trị tham số là 15. Giá của một mặt hàng nông nghiệp trên thị trường i tại thời điểm t là:

$$\begin{aligned} \theta_i^{(1)} &= \theta_i^{(1)} \\ \theta_i^{(2)} &= \frac{\theta_i^{(1)} + \theta_i^{(2)} + \theta_i^{(3)}}{3} \\ &\dots \\ \theta_i^{(8)} &= \frac{\sum_{t=1}^{15} \theta_i^{(t)}}{15} \\ &\dots \end{aligned}$$

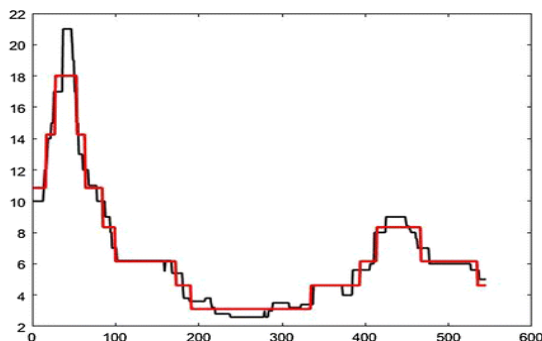
$$\theta_i^{(t)} = \frac{\sum_{i=t-7}^{t+7} \theta_i^{(i)}}{15}$$

2. Phân cụm dữ liệu. Để đồng bộ hóa sự thay đổi giá với sự thay đổi tích lũy của các biến ngoại sinh, trong khi bỏ qua sự dao động nhẹ, nghiên cứu này sử dụng phân tích cụm dữ liệu trong quá trình xử lý. Ở đây chúng ta sử dụng thuật toán phân cụm K-means. Đặt c là số cụm:

$$c = \min\{\text{Số lượng giá khác nhau}, 7\}$$

Chúng ta đặt c nhỏ hơn 7, vì vậy chúng ta có thể chia một cụm là giá trị trung bình, ba cụm cao hơn và ba cụm thấp hơn, để phản ánh sự ổn định và sự dao động lớn của giá, như thể hiện trong Hình 3. Sẽ không có nhiều hơn 7 giá trị khác nhau của giá cả hàng hóa nông nghiệp sau khi phân cụm.

Hình 3: Ảnh hưởng của việc xử lý dữ liệu của mô hình cảnh báo



Đường màu đen để cập đến dữ liệu thô của mô hình cảnh báo; đường màu đỏ để cập đến dữ liệu tiền xử lý của mô hình cảnh báo.

3. Nâng cao chiều hướng. Với $p_i(t)$ là giá mặt hàng nông sản trên thị trường i tại thời điểm t , thiết lập các thay đổi giá trong tương lai gần là $\delta_i(t)$, trong đó xảy ra $N(t)$ ngày. Bây giờ chúng ta có thể nhận được một dữ liệu mới hàng ngày ($\delta_i(t)$, $N_i(t)$).

4. Lấy các biến mới. Sau ba bước thử nghiệm, cuối cùng nhóm nghiên cứu nhận

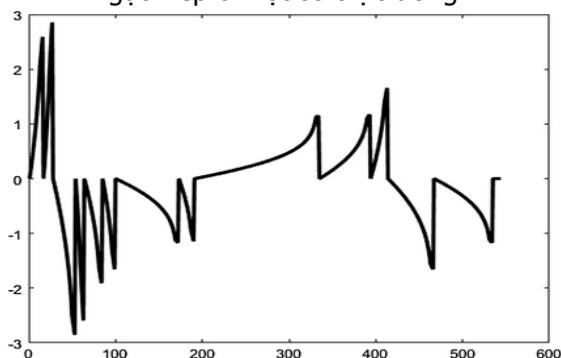
được một số biến mới. Nghiên cứu này ước tính định lượng được khoảng thay đổi giá có thể từ các giá trị của $\delta_i(t)$ và $N_i(t)$. Do đó, xác định một biến số $U_i(t)$ khẩn cấp. Giả sử rằng giá $\theta_i(t)$ kéo dài cho thời gian $T_i(t)$. Dựa vào hiệu quả thực nghiệm và mục đích định lượng, chúng ta định nghĩa $U_i(t)$ như sau:

$$U_i(t) = \frac{(T_i(t) - N_i(t)) \delta_i(t)}{T_i(t) \cdot N_i(t)^{1/4}}$$

Nếu $N_i(t) < 3$, lấy $N_i(t) = 3$, để tránh tình trạng khẩn cấp từ sự thay đổi đột ngột làm cho việc dự báo trở nên khó khăn.

Từ định nghĩa $U_i(t)$, chúng ta thấy sự gia tăng giá lớn hơn hay sớm hơn thì sự thay đổi xảy ra, sự khẩn cấp càng mạnh. Vì vậy, $U_i(t)$ có thể định lượng mức độ thay đổi giá cả khẩn cấp và gửi thông báo cảnh báo. Sự thay đổi cấp bách của giá gạo nếp ở một số thị trường được thể hiện như Hình 4.

Hình 4: Sự thay đổi cấp bách của giá gạo nếp ở một số thị trường



Sự biến đổi các biến ngoại sinh và tính toán mẫu

Một số biến ngoại sinh có xu hướng thay đổi theo cách riêng của chúng. Do đó, không có mối quan hệ rõ ràng với xu hướng thay đổi khẩn cấp của các mặt hàng nông nghiệp. Trong khi đó các biến này là ngẫu nhiên. Vì vậy, không thích hợp và phù hợp để trực tiếp sử dụng dữ liệu hàng ngày của các

biến ngoại sinh này dự báo. Bởi vì chúng ta không thể tránh sự biến động ngẫu nhiên của các biến ngoại sinh và ảnh hưởng do các đặc trưng của chúng.

Do đó chúng ta cần phải tìm ra các yếu tố để phản ánh tốt hơn những thay đổi về giá đang bị ảnh hưởng như thế nào. Nghiên cứu này sẽ xử lý biến ngoại sinh theo 4 bước sau:

1. Trung bình: Chúng ta lấy giá trị trung bình trong 2 tháng qua.
2. Tích lũy: Chúng ta lấy các giá trị tích lũy kể từ lần thay đổi giá cuối cùng. Khi giá thay đổi, tất cả các biến này được đặt thành 0.
3. Lấy giá trị của ngày đó: Chúng ta trực tiếp gán các giá trị thực cho các biến.
4. Ghi lại các giá trị cực đại/cực tiểu: Chúng ta chỉ định các giá trị cực đại/cực tiểu kể từ khi thay đổi giá cuối cùng cho các biến.

Nghiên cứu này đòi hỏi tính cấp bách vì biến số độc lập tương ứng với giá trị tích lũy của thay đổi nhiệt độ cho dù có tuyết, sương mù hoặc bão tố, có các giá trị cực đại tích lũy, giá trị trung bình và giá trị đầu thô và tỷ giá hối đoái của mỗi ngày và do đó có lần 14 biến ngoại sinh.

Mô hình cảnh báo dựa trên mạng liên kết thông tin thông minh

Nghiên cứu này xây dựng một mô hình mạng liên kết thông tin thông minh BP để nghiên cứu các biến ngoại sinh và tính cấp bách. Sự lựa chọn của mạng liên kết thông tin thông minh BP dựa trên 2 yếu tố:

1. Mối quan hệ giữa 14 biến ngoại sinh chưa được biết và không có nghiên cứu nào được tiến hành về định lượng mối quan hệ chính xác của các biến ngoại sinh và sự thay đổi giá cả của hàng hoá nông nghiệp. Các mạng liên kết thông tin thông minh có dạng chức năng linh hoạt bao gồm các mối quan hệ chức năng tuyến tính và phi tuyến tính.

Do đó, có lợi thế duy nhất trong dự báo theo yêu cầu của nghiên cứu này. Phân tích nhiều yếu tố dựa trên mạng liên kết thông tin thông minh sẽ đem lại hiệu quả trong một số ứng dụng.

2. Mỗi quan hệ giữa các biến ngoại sinh và giá cả hàng hóa nông nghiệp có thể dao động theo thời gian. Mô hình mạng liên kết thông tin thông minh có thể được cập nhật theo dữ liệu lịch sử.

Đặt số lớp ẩn là 1: Chúng ta chọn đặt số lớp ẩn theo MSE và chọn phương pháp LM (LM là thuật toán Levenberg-Marquardt hay còn được gọi là phương pháp tối thiểu damped-squares (DLS) dùng để giải quyết các vấn đề không bình thường nhất trong phân tích đa biến). Sau khi xác định được các tham số, chúng ta có thể tính toán bằng mạng liên kết thông tin thông minh.

Trên thực tế, mục đích của sự cấp bách là để phản ánh tác động tích lũy của các biến ngoại sinh. Từ các định nghĩa của 14 biến ngoại sinh, chúng ta có thể thấy một

trong số đó là đơn điệu khi thời gian trôi qua và một số trong đó đang dần tích lũy. Các mạng liên kết thông tin thông minh được tính toán có thể điều chỉnh theo mức độ khẩn cấp hàng ngày. Định nghĩa về tính cấp bách cho thấy mức độ khẩn cấp đo lường xu hướng thay đổi giá cả. Mức độ khẩn cấp cao không có nghĩa là một sự thay đổi giá cả. Thay vào đó, nó cho thấy phạm vi thay đổi giá cả (nếu giá thực sự thay đổi). Xem xét sự không đồng bộ của sự thay đổi giá và tích lũy các biến ngoại sinh, nghiên cứu này là bảo vệ quan điểm với dự báo giá trị cấp bách. Chúng ta xem xét giá trị khẩn cấp của dự báo tuần trước. Với giá trị dự báo $U_i(t)$ tại thời điểm t , giá trị được điều chỉnh $U_i'(t)$ được định nghĩa là:

$$U_i'(t) = \frac{\text{med}\{U_i^{(s)}, s = t-6, \dots, t-1, t\} + U_i^{(t)}}{2}$$

Trong đó: $\text{med}\{U_i^{(s)}, s = t-6, \dots, t-1, t\}$ là điểm giữa của các giá trị khẩn cấp từ ngày $t-6$ đến ngày t .

Còn nữa

Công Hoan (dịch)

Nguồn: <https://journalofbigdata.springeropen.com/track/pdf/10.1186/s40537-016-0062-3?site=journalofbigdata.springeropen.com>