

SO SÁNH MÔ HÌNH ARIMA VÀ VECM TRONG DỰ BÁO CẦU LAO ĐỘNG Ở CÁC DOANH NGHIỆP NGÀNH CÔNG NGHIỆP CHẾ BIẾN TỈNH BÌNH DƯƠNG

Phạm Ngọc Thành

Trường Đại học Lao động – Xã hội (Cơ sở TP. Hồ Chí Minh)

Email: thanhpn@ldxh.edu.vn

Đỗ Thị Hoa Liên

Trường Đại học Lao động – Xã hội (Cơ sở TP. Hồ Chí Minh)

Email: dohoalien@yahoo.com.vn

Hoàng Võ Hằng Phương

Trường Đại học Lao động – Xã hội (Cơ sở TP. Hồ Chí Minh)

Email: hoangvohangphuong@gmail.com;

Ngày nhận: 15/01/2020

Ngày nhận bản sửa: 07/5/2020

Ngày duyệt đăng: 05/12/2020

Tóm tắt:

Bài viết ứng dụng các mô hình trong phương pháp chuỗi thời gian vào dự báo cầu lao động của các doanh nghiệp ngành công nghiệp chế biến ở Bình Dương. Trong đó, hai mô hình phổ biến được sử dụng nhiều trong dự báo cầu lao động trong ngắn hạn là ARIMA và VECM được phân tích và so sánh. Mô hình được xây dựng trên dữ liệu từ năm 1996 đến 2014, sau đó dự báo được thực hiện cho các năm 2015 đến 2017 để kiểm tra mức độ chính xác của các mô hình. Kết quả cho thấy cả hai mô hình đều có năng lực dự báo tốt, tuy nhiên, mô hình ARIMA trong trường hợp này có khả năng dự báo chính xác hơn so với mô hình VECM. Bên cạnh đó, nhóm tác giả tìm thấy vai trò quan trọng của vốn sản xuất đối với sự thay đổi cầu lao động ngành công nghiệp chế biến Bình Dương cả trong ngắn hạn và dài hạn.

Từ khóa: Dự báo, cầu lao động, phương pháp chuỗi thời gian, công nghiệp chế biến, ARIMA, VECM

Mã JEL: C53, E27, J23

A comparative analysis of ARIMA and VECM models in forecasting manpower demand among manufacturing firms in Binh Duong

Abstract:

Time series method was employed to conduct a forecast of manpower demand among manufacturing enterprises in Binh Duong. The two popular models mostly used in forecasting manpower demand in short term, ARIMA and VECM, were used to find out factors influencing manpower demand of the industry in the period of 1997 to 2014. Also, there were a prediction of the manpower demand from 2015 to 2017, a comparison for the quality of forecasting results were then made among the models. The results show that both models performed good forecasts, however, the ARIMA took the higher position in prediction capacity than VECM. Besides, annual average capital takes an important role to the manpower demand in manufacturing enterprises in Binh Duong for short term and long term.

Keywords: Forecasting, manpower demand, time series method, manufacturing, ARIMA, VECM

JEL Codes: C53, E27, J23

1. Giới thiệu

Tại tỉnh Bình Dương, thị trường lao động vẫn tồn những bất cập. Thứ nhất, ở nhiều khu công nghiệp của Bình Dương, nhiều doanh nghiệp thiếu hụt lao động nói chung, đặc biệt là ngành công nghiệp chế biến. Có đến 84,7% lượng lao động làm việc trong ngành công nghiệp chế biến, với đóng góp GDP của ngành trên toàn tỉnh lên đến khoảng 60% vào năm 2018. Bình Dương cũng là một trong ba tỉnh thu hút vốn đầu tư nước ngoài lớn nhất trên toàn quốc. Hơn 90% doanh nghiệp FDI hoạt động trong ngành công nghiệp chế biến và tổng vốn đăng ký các doanh nghiệp này lên đến 81%. Tuy nhiên, kết quả khảo sát nhu cầu lao động của trung tâm dịch vụ việc làm tỉnh Bình Dương trong năm 2019 cho thấy các doanh nghiệp trong ngành thường xuyên phải đối mặt với tình trạng mất ổn định về lao động, thiếu hụt lao động và phải liên tục tuyển dụng. Sự thiếu ổn định về lao động được thể hiện rõ rệt nhất ở các doanh nghiệp da giày, may mặc, điện tử, tức là các lĩnh vực trong ngành công nghiệp chế biến. Nguyên nhân dẫn đến thực trạng này chủ yếu là do doanh nghiệp mở rộng sản xuất kinh doanh và do một bộ phận lao động thay đổi việc làm, nghỉ việc để tìm việc khác. Thứ hai, có sự mất cân bằng giữa cung và cầu lao động do không tương đồng về số lượng và chất lượng giữa hai bên cung, cầu và mong muốn, kỳ vọng của người lao động không phù hợp với người sử dụng lao động. Thứ ba, vẫn có một tỷ lệ nhất định những người trong độ tuổi có khả năng lao động chưa có việc làm (8,7% năm 2015). Như vậy, có thể thấy rằng cả doanh nghiệp lẫn các nhà quản lý ở tỉnh Bình Dương đều cần thiết phải biết được chính xác nhu cầu lao động các ngành nghề để có các kế hoạch chiến lược tuyển dụng nhân sự cho phù hợp.

Hơn nữa, yêu cầu đặt ra hiện nay là để thực hiện quy hoạch, theo dõi, đánh giá, giám sát cũng như điều chỉnh quy hoạch phát triển nguồn nhân lực, việc làm, cần phải dựa trên một nền tảng công tác dự báo có chất lượng. Tuy nhiên, đánh giá một cách khách quan, công tác dự báo thị trường lao động trong thời gian qua ở Việt Nam nói chung và tỉnh Bình Dương nói riêng vẫn chưa được thực hiện một cách đầy đủ và khoa học. Phương pháp/mô hình dự báo nguồn nhân lực dựa trên lý luận khoa học và phù hợp với điều kiện thực tiễn ở Bình Dương còn nhiều bất cập. Những khó khăn về cơ sở dữ liệu khi áp dụng các mô hình dự báo trong điều kiện Việt Nam hiện nay đã được đề cập, song các phương án thay thế giữa các biến hoặc hoàn thiện cơ sở dữ liệu cũng cần tiếp tục đề xuất có sức thuyết phục hơn. Ngoài ra, việc lựa chọn phương pháp dự báo sẽ khó khăn hơn vì một phương pháp dự báo riêng lẻ thường có tính thuyết phục không cao so với kết hợp nhiều phương pháp dự báo được đặt trong bối cảnh so sánh.

Trong những năm gần đây, Bình Dương đã nỗ lực cố gắng nâng cao chất lượng nguồn nhân lực bằng cách đổi mới phương thức đào tạo, có các chương trình dạy nghề, hướng nghiệp, tư vấn nghề nghiệp và giới thiệu việc làm. Nhưng bên cạnh đó, công tác dự báo cầu lao động theo sự phát triển kinh tế địa phương còn chưa được thực hiện một cách khoa học theo các mô hình dự báo, mà nặng về khảo sát doanh nghiệp và từ đó đưa ra các con số kế hoạch. Vì vậy, nhóm tác giả đề xuất hai mô hình nổi tiếng, phù hợp với điều kiện nghiên cứu tại Bình Dương hiện nay, đó là mô hình ARIMA và mô hình VECM để dự báo cầu lao động ngành công nghiệp chế biến Bình Dương và so sánh đánh giá sự phù hợp các mô hình.

2. Cơ sở lý thuyết

Trong các phương pháp kinh tế lượng dùng cho dự báo thì phương pháp chuỗi thời gian được các nhà học thuật khai thác đầu tiên và phổ biến nhất. Phương pháp này căn cứ vào đặc điểm trong quá khứ để xem xét xu hướng của các quy luật vận động, với kỳ vọng các hành vi trong quá khứ sẽ lặp lại và duy trì trong tương lai. Phương pháp này dựa trên mối quan hệ giữa biến phụ thuộc là cầu lao động với các biến độc lập, từ đó, xem xét và đánh giá các tác động của các biến độc lập, các biến trễ của các biến độc lập và biến trễ của biến cầu lao động lên cầu lao động và dự báo cầu lao động. Một số mô hình kinh tế lượng nổi bật thường được sử dụng như mô hình ARIMA của Box-Jenkins, phân tích hồi quy đa biến, ARDL, mô hình VAR, mô hình vector hiệu chỉnh sai số (VECM).

Những nghiên cứu đã áp dụng các mô hình chuỗi thời gian để dự báo cầu lao động, và đều được Coo (2007) đánh giá cao về chất lượng nghiên cứu khi xét điểm tính từ các tiêu chí về dữ liệu nghiên cứu lẫn mô hình và các biến được lựa chọn gồm có LeSage (1990a và 1990b); Crane & Nourzad (1998); Fauvel & cộng sự (1999); Sarantis & Swales (1999); Puri & Soydemir (2000); Krolzig & cộng sự (2002); Wong & cộng

sự (2005); Rapach & Strauss (2005); Wong & cộng sự (2007); Ho (2010); Wong & cộng sự (2011); Bruha (2011); Ho (2012); Vereen & cộng sự (2016); Rapach & Strauss (2012); Krolzig & cộng sự (2002).

ARIMA là mô hình sử dụng độ trễ của chính biến cần nghiên cứu để xác định mô hình dự báo thích hợp, mô hình không quan tâm đến các biến số vĩ mô và các biến độc lập khác. Box & cộng sự (2015) xây dựng ba bước để ước lượng mô hình ARIMA(p,d,q):

- Bước một, kiểm tra tính dừng và xác định p, q bằng các đồ thị tương quan (ACF) và đồ thị tương quan riêng từng phần (PACF);
- Bước hai, tính toán các tham số của mô hình;
- Bước ba, kiểm tra các giả định của mô hình. Nếu mô hình không phù hợp thì quay lại bước hai để thử một mô hình khác tốt hơn.

Mô hình ARIMA được ưa thích trong dự báo cầu lao động ngắn hạn ở hầu hết các nghiên cứu đi trước, mặc dù mô hình khá đơn giản. Điều này có thể thấy ở các nghiên cứu của Puri & Soydemir (2000), Fauvel & cộng sự (1999), Sarantis & Swales (1999), Wong & cộng sự (2005). Fauvel & cộng sự (1999) cho rằng mô hình này vẫn hữu ích trong dự báo vì kết quả của ARIMA có thể làm chuẩn để so sánh với kết quả của các phương pháp khác phức tạp hơn. Đôi khi, ARIMA có thể vượt trội mô hình VAR như ở nghiên cứu của Edlund và Karlsson (1993) đã khẳng định. Minh chứng thực nghiệm ở nhiều nghiên cứu đi trước, ARIMA cung cấp kết quả dự báo có độ chính xác gần bằng với các mô hình khác.

Mô hình VECM là một dạng đặc biệt của VAR, với điều kiện ràng buộc là phải có sự đồng kết hợp giữa các biến. Vì vậy, VECM còn được gọi là mô hình VAR của các chuỗi không dừng, là sự kết hợp giữa VAR và mô hình hiệu chỉnh sai số (ECM) với nhiều biến cùng đồng kết hợp với nhau. Các bước ước lượng mô hình như sau: đầu tiên các chuỗi dữ liệu được kiểm định tính dừng thông qua kiểm định ADF; Sau đó, độ trễ tối ưu được xác định trong mô hình VAR dựa trên một số tiêu chí như AIC, SC, HQ; Tiếp đến là kiểm tra sự đồng kết hợp giữa các chuỗi có sai phân bậc một thông qua kiểm định của Johansen và kiểm định Trace; Cuối cùng, thực hiện dự báo mô hình và kiểm định phần dư của mô hình. So sánh giữa các mô hình trong dự báo cầu lao động, Crane & Nourzad (1998) cho rằng VECM có kết quả chính xác hơn so với mô hình VAR trong dự báo việc làm của ngành sản xuất tại địa phương. Bên cạnh đó, LeSage (1990a và 1990b) đã chứng minh rằng mô hình ECM cũng cho kết quả tốt gần bằng với mô hình VAR, vì vậy, sự kết hợp giữa ECM và VAR là cần thiết. Các nghiên cứu đi trước có xu hướng đề cao tính ưu việt của mô hình VECM đối với các dự báo trong khoảng thời gian dài hơn, mặc dù vậy, trong ngắn hạn tính ưu việt chưa được thể hiện rõ rệt (Fauvel & cộng sự, 1999).

Tại Việt Nam, có ít các nghiên cứu về dự báo cầu lao động theo địa phương, đặc biệt là một ngành lớn có đóng góp cho sự phát triển của địa phương. Hơn nữa, sự so sánh về chất lượng dự báo giữa các phương pháp khác nhau trong dự báo cầu lao động cũng chưa được quan tâm. Đặc biệt phương pháp VECM và ARIMA chưa được ứng dụng vào dự báo cầu lao động của một ngành tại một địa phương nào ở Việt Nam. Với những ưu điểm của hai mô hình cùng với khoảng trống trong nghiên cứu cầu lao động tại Việt Nam ở trên, nhóm tác giả muốn ứng dụng hai mô hình ARIMA và VECM để so sánh và đánh giá lựa chọn mô hình phù hợp hơn khi dự báo cầu lao động của một ngành cho địa phương.

3. Phương pháp nghiên cứu

3.1. Dữ liệu nghiên cứu

Nhược điểm gây cản trở lớn trong quá trình nghiên cứu đó là các số liệu tại Việt Nam trong khoảng thời gian dài về trước đa phần không có trong niên giám thống kê hoặc nếu có thì không đồng bộ với số liệu trong những năm gần đây. Nguyên nhân đến từ việc thay đổi phương pháp khảo sát, cũng như phương pháp tính toán các tiêu chí thống kê trong các thời kì nhất định tại Việt Nam. Điều này khiến cho một số nhóm chỉ tiêu, số liệu không hiển thị hoặc hiển thị khác nhau trong nhiều năm, gây khó khăn cho việc thu thập đầy đủ thông tin của các biến độc lập.

Bình Dương mới được tái thiết lập tỉnh vào ngày 1 tháng 1 năm 1997, nên dữ liệu quá khứ trước năm 1996 không tồn tại. Nhiều chỉ tiêu trong niên giám thống kê có sự khác biệt trong một vài thời kỳ. Vì vậy,

trên cơ sở dữ liệu thứ cấp có sẵn từ niên giám thống kê các năm của Bình Dương, nhóm tác giả sẽ chọn lọc lại những chỉ tiêu có tính nhất quán từ năm 1996 đến 2017 để xác định các yếu tố có khả năng tác động đến cầu lao động, từ đó đưa vào mô hình để kiểm chứng và cuối cùng là dự báo. Những biến phù hợp với những nhược điểm trên trong ngành công nghiệp chế biến thu thập từ niên giám thống kê Bình Dương các năm gồm có biến *Số lao động đang làm việc trong ngành (L)*, *Doanh thu của các doanh nghiệp trong ngành (DT)*, *Vốn sản xuất của các doanh nghiệp (VSX)*. Bên cạnh đó, một yếu tố nữa cũng có khả năng đóng góp vào sự biến động của cầu lao động đó là Lãi suất cho vay (*LS*), thu thập từ IMF. Các biến số lao động (*L*), doanh thu (*DT*) và vốn sản xuất (*VSX*) sẽ được biến đổi lại theo công thức: $\frac{Z_i}{Z_{i-1}}$ với Z_i và Z_{i-1} theo thứ tự là giá trị của các biến ở năm i và năm $(i-1)$. $i=1997, \dots, 2014$. Các giá trị từ năm 2015 -2017 được giữ lại để dự báo và đánh giá năng lực dự báo của mô hình. Như vậy, các biến mới sử dụng trong mô hình được mã hóa lại thành *L1*, *DT1* và *VSX1*.

Một vấn đề khác trong nghiên cứu theo phương pháp chuỗi thời gian đó là dữ liệu thường được thu thập theo tháng hoặc quý. Tuy nhiên, việc có một bộ dữ liệu đầy đủ theo tháng hoặc quý trong một khoảng thời gian dài đối với một địa phương là điều không khả thi. Vì vậy, nghiên cứu chỉ thực hiện dự báo trên bộ số liệu có sẵn và đầy đủ nhất theo năm từ niên giám thống kê Bình Dương qua các năm.

Như vậy, các dữ liệu nghiên cứu trong bài viết được trích xuất từ niên giám thống kê Bình Dương theo năm của ngành công nghiệp chế biến và dữ liệu lãi suất cho vay trung bình hằng năm tại Việt Nam từ IMF trong khoảng thời gian từ 1996 đến 2017.

3.2. Mô hình nghiên cứu

3.2.1. Mô hình ARIMA

Theo Box & cộng sự (2015) biến phụ thuộc $L1_t$ được giải thích theo các giá trị trong quá khứ hoặc giá trị trễ của bản thân biến phụ thuộc và các sai số ngẫu nhiên.

Hầu hết các chuỗi thời gian kinh tế không có tính dừng, tức là chúng sẽ dần kết hợp sau một khoảng thời gian dài. Nếu chuỗi thời gian mà ở sai phân bậc d thì chuỗi có tính dừng, lúc này chuỗi sai phân bậc d là $I(0)$. Lúc này chuỗi thời gian ban đầu là ARIMA(p, d, q), tức là nó là một chuỗi thời gian trung bình trượt kết hợp tự hồi quy. Mô hình ARIMA có dạng:

$$\Delta^d L1_t = \lambda + \eta_1 \Delta^d L1_{t-1} + \eta_2 \Delta^d L1_{t-2} + \dots + \eta_p \Delta^d L1_{t-p} + u_t - \phi_1 u_{t-1} - \dots - \phi_q u_{t-q}$$

Với $\Delta^d L1_t$ là sai phân bậc d ($\Delta^d L1_t = \Delta^{d-1} L1_{t+1} - \Delta^{d-1} L1_t$ với $d \geq 2$, $\Delta^1 L1_t = L1_t - L1_{t-1}$)

Gọi B là phép toán dịch chuyển lùi có dạng: $BL1_t = L1_{t-1}$; tổng quát: $L1_{t-k} = B^k L1_t$; Đặt $L1_t = \Delta L1_t$

$p(B) = (1 - \eta_1 B - \eta_2 B^2 - \dots - \eta_p B^p)$ là phép toán tự hồi quy;

$q(B) = (1 - \phi_1 B - \phi_2 B^2 - \dots - \phi_p B^p)$ là phép toán trung bình trượt.

Nếu chuỗi $L1_t$ có tính dừng với sai phân bậc 1 của nó là $I(0)$, mô hình có thể được viết lại như sau:

$$DL1_t = \lambda + \frac{q(B)}{p(B)} u_t$$

Trong đó, $DL1_t = \Delta L1_t$, $L1_t$ là số lao động ở năm t , $DL1_t$ là sai phân bậc 1 số lao động tại năm t . $B \times DL1_t = DL1_{t-1}$,

3.2.2. Mô hình VECM

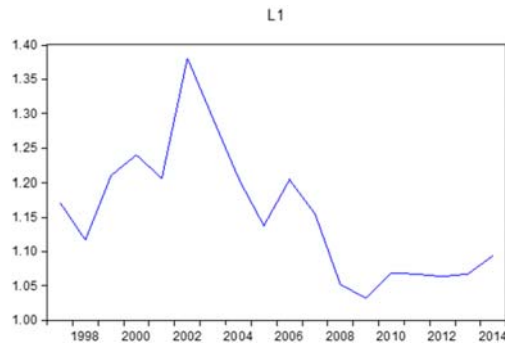
Mô hình VECM là một dạng đặc biệt của mô hình VAR, nó là sự kết hợp giữa VAR và ECM

Mô hình nghiên cứu cầu lao động ngành công nghiệp chế biến ở Bình Dương theo mô hình VECM như sau:

$$\Delta L1_t = \delta + \alpha(\beta' L1_{t-1} + \rho_0) + \sum_{i=1}^p \eta_{1,i} \Delta L1_{t-i} + \sum_{i=1}^p \eta_{2,i} \Delta DT1_{t-i} + \sum_{i=1}^p \eta_{3,i} \Delta VSX1_{t-i} + \sum_{i=1}^p \eta_{4,i} \Delta LS_{t-i} + u_t \quad (2)$$

Trong đó, α là ma trận hệ số điều chỉnh, thể hiện tốc độ mà các biến độc lập hội tụ trở lại giá trị cân bằng; β là ma trận hệ số dài hạn thể hiện mối quan hệ đồng liên kết giữa các biến;

Hình 1: Đồ thị tương quan chuỗi L1



Nguồn: nghiên cứu của nhóm tác giả

Hình 2: Biểu đồ tương quan chuỗi L1

Sample: 1997 2014
Included observations: 18

Autocorrelation	Partial Correlation	AC	PAC	Q-Sta...	Prob
		1 0.707	0.707	10.587	0.001
		2 0.493	-0.01...	16.065	0.000
		3 0.387	0.085	19.651	0.000
		4 0.217	-0.16...	20.860	0.000
		5 -0.02	-0.28...	20.883	0.001
		6 -0.25	-0.25...	22.771	0.001
		7 -0.31	0.021	26.070	0.000
		8 -0.36	-0.02...	30.850	0.000
		9 -0.36	0.079	36.307	0.000
		1... -0.33	-0.00...	41.367	0.000
		1... -0.28	-0.07...	45.398	0.000
		1... -0.19	-0.06...	47.723	0.000

Nguồn: nghiên cứu của nhóm tác giả

$\eta_{j,i}$ là mối quan hệ ngắn hạn giữa các biến với các biến trễ của nó;

p là số trễ, được lựa chọn sao không có sự tự tương quan giữa các sai số ước lượng. Số trễ, p , được xác định bằng phép thử, bắt đầu từ giá trị cao nhất có thể và độ trễ này sẽ được kiểm định ý nghĩa thống kê bằng tỷ số t . AIC và SBC là các tiêu chí để lựa chọn p tối ưu cho mô hình ADF, p được lựa chọn sao cho AIC và SBC nhỏ nhất;

u_t là nhiễu trắng.

Như vậy, các biến mới được mã hóa lại như sau: lao động ($L1$), doanh thu (DTI) và vốn sản xuất ($V SX1$).

Để ứng dụng mô hình VECM các nhà dự báo luôn kỳ vọng các biến có mối quan hệ với nhau trong dài hạn bởi tính đồng kết hợp thể hiện mối quan hệ dài hạn và ổn định giữa các biến với nhau. Phương pháp hợp lý cực đại của Johansen với kiểm định trace được sử dụng để kiểm định tính đồng kết hợp. Các kiểm định

Hình 3: Biểu đồ tương quan chuỗi L1 sai phân bậc 1

Autocorrelation	Partial Correlation	AC	PAC	Q-Sta...	Prob
		1 -0.15	-0.15...	0.4702	0.493
		2 -0.22	-0.25...	1.5303	0.465
		3 0.098	0.019	1.7537	0.625
		4 0.136	0.112	2.2148	0.696
		5 -0.12	-0.05...	2.6355	0.756
		6 -0.26	-0.27...	4.6335	0.592
		7 0.051	-0.11...	4.7189	0.694
		8 -0.02	-0.16...	4.7397	0.785
		9 -0.13	-0.16...	5.4974	0.789
		1... 0.021	-0.04...	5.5184	0.854
		1... 0.043	-0.07...	5.6197	0.897
		1... 0.044	-0.03...	5.7428	0.928

Nguồn: nghiên cứu của nhóm tác giả

Bảng 1: Kiểm định ADF

Biến	Giá trị thống kê t	Giá trị tới hạn	Biến	Giá trị thống kê t	Giá trị tới hạn
L1	-1,509007	-2,666593	$\Delta L1$	-4,411069***	-3,92035

***: mức ý nghĩa 1%

Nguồn: nghiên cứu của nhóm tác giả

Bảng 2: Các mô hình ARIMA dự kiến

Chuỗi thời gian	ARIMA1	ARIMA2	ARIMA3	ARIMA4	ARIMA5	ARIMA6
Mô hình	AR(2), MA(2)	AR(2), MA(6)	AR(6), MA(2)	AR(2), AR(6), MA(2)	AR(2), AR(6), MA(6)	AR(6), MA(2), MA(6)

Nguồn: nghiên cứu của nhóm tác giả.

phần dư từ mô hình VECM bao gồm: Tính dừng phần dư, hiện tượng tự tương quan, phương sai thay đổi giữa các biến, và phân phối chuẩn của phần dư được kiểm tra để xem xét mô hình có vi phạm các giả thiết về phần dư hay không.

4. Kết quả và thảo luận

4.1. Mô hình ARIMA

Bảng 3: Các tiêu chí lựa chọn mô hình

Tiêu chí	ARIMA1	ARIMA2	ARIMA3	ARIMA4	ARIMA5	ARIMA6
R ² điều chỉnh	0,076043	0,172738	0,17823	0,187844	0,177985	0,304381
AIC	-2,071509	-2,133623	-2,14022	-2,026430	-2,016134	-2,041676
SIC	-1,875459	-1,937573	-1,94417	-1,781367	-1,771071	-1,796613

Nguồn: nghiên cứu của nhóm tác giả.

Hình 4: Biểu đồ tương quan của phần dư

Q-statistic probabilities adjusted for 2 ARMA terms

Autocorrelation	Partial Correlation	AC	PAC	Q-Sta...	Prob
1	-0.25...	-0.25...		1.2837	
2	-0.09...	-0.17...		1.4788	
3	0.017	-0.05...		1.4854	0.223
4	0.097	0.077		1.7215	0.423
5	-0.20...	-0.17...		2.8323	0.418
6	0.111	0.035		3.1937	0.526
7	-0.04...	-0.05...		3.2475	0.662
8	-0.09...	-0.12...		3.5419	0.738
9	-0.10...	-0.16...		3.9669	0.784
1...	0.045	-0.11...		4.0586	0.852
1...	0.005	-0.03...		4.0602	0.907
1...	-0.02...	-0.06...		4.0910	0.943

Nguồn: nghiên cứu của nhóm tác giả.

Bảng 4: Kiểm định ADF phần dư của chuỗi

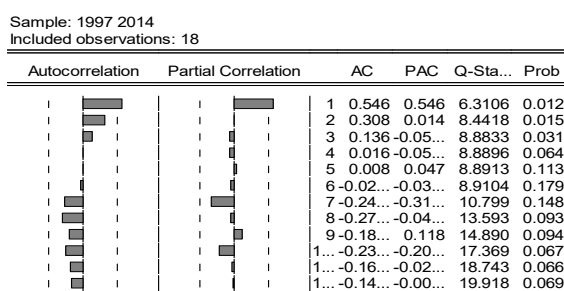
Biến	Giá trị thống kê t	Giá trị tới hạn
DL1	-4,925090***	-3,920350

Nguồn: Nghiên cứu của nhóm tác giả.

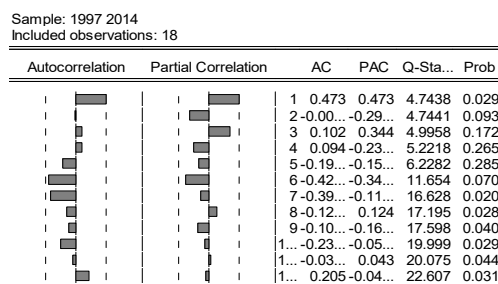
Bảng 5: Kết quả mô hình ARIMA3

Biến	Hệ số	Sai số chuẩn	Thống kê t	Giá trị p
C	-0,009003	0,016520	-0,544983	0,5950
AR(6)	-0,380556	0,508301	-0,748684	0,4674
MA(2)	-0,147264	0,341244	-0,431549	0,6731
SIGMASQ	0,004059	0,001737	2,337183	0,0361

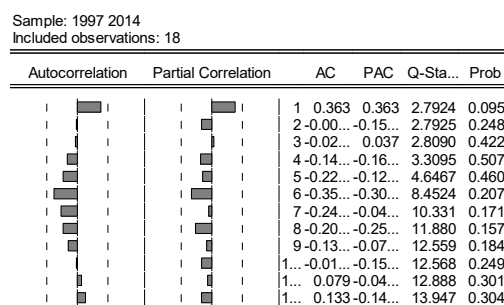
Nguồn: nghiên cứu của nhóm tác giả.

Hình 5: Biểu đồ tương quan chuỗi DT1

Nguồn: nghiên cứu của nhóm tác giả.

Hình 6: Biểu đồ tương quan chuỗi VSX1

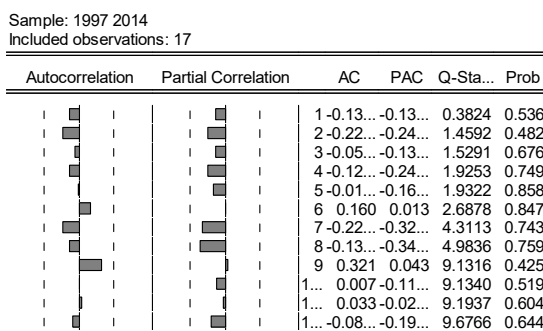
Nguồn: nghiên cứu của nhóm tác giả.

Hình 7: Biểu đồ tương quan chuỗi LS

Nguồn: nghiên cứu của nhóm tác giả.

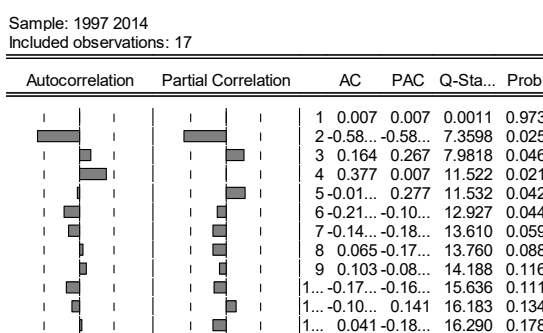
Đồ thị chuỗi L1 của ngành có xu hướng tăng qua các năm nên chắc chắn dữ liệu không thể dao động quanh giá trị trung bình của chuỗi, vi phạm một trong ba điều kiện dừng của chuỗi nên các chuỗi không có tính dừng (Hình 1). Thêm vào đó, ở biểu đồ tương quan, biểu đồ PACF ở tất cả các độ trễ có hình sin và nhiều độ trễ nằm ngoài khoảng tin cậy, hoặc có thể thấy kiểm định toàn bộ giá trị ρ_k của ACF tại 12 độ trễ đều

Hình 8: Biểu đồ tương quan sau khi lấy sai phân bậc 1 Chuỗi DT1



Nguồn: nghiên cứu của nhóm tác giả.

Hình 9: Biểu đồ tương quan sau khi lấy sai phân bậc 1 Chuỗi VSX1



Nguồn: nghiên cứu của nhóm tác giả.

khác 0 có ý nghĩa nên chuỗi không có tính dừng (Hình 2). Cuối cùng, là kiểm định nghiệm đơn vị (ADF), các chuỗi của ngành không dừng do giá trị thống kê t lớn hơn giá trị tới hạn (Bảng 1).

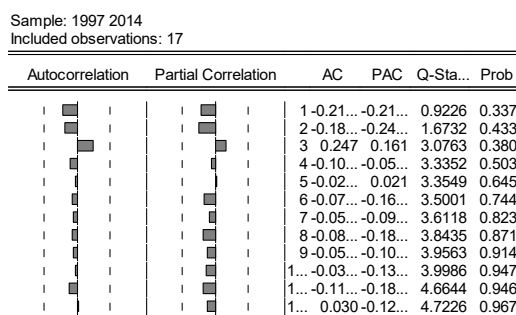
Sau khi chuỗi L1 được lấy sai phân một lần thì được chuỗi dừng với giá trị p ($0,0039 < 0,05$), đồ thị và biểu đồ tương quan cũng cho thấy chuỗi L1 dừng ở sai phân bậc một.

Từ Hình 3, các giá trị ACF và PACF cùng giảm đột ngột tại độ trễ thứ 2, thứ 6 của chuỗi L1. Do vậy, ta có AR(2,6) và MA(2,6) đối với chuỗi L1 ở Bảng 2.

Thực hiện các mô hình theo thứ tự để xem xét lựa chọn mô hình phù hợp nhất. Kết quả của các mô hình được thể hiện ở Bảng 3.

Từ các tiêu chí trên, mô hình có giá trị AIC và SIC nhỏ nhất, R^2 điều chỉnh cao nhất sẽ được lựa chọn. Chuỗi công nghiệp chế biến có mô hình ARMA3 phù hợp nhất.

Hình 10: Biểu đồ tương quan sau khi lấy sai phân bậc 1 Chuỗi LS



Nguồn: nghiên cứu của nhóm tác giả.

Bảng 6: Kiểm định ADF

Biến	Giá trị thống kê t	Giá trị tới hạn	Biến	Giá trị thống kê t	Giá trị tới hạn
L1	-1,509007	-2,666593	$\Delta L1$	-4,411069***	-3,92035
DT1	-2,149403	-2,666593	$\Delta DT1$	-4,739109***	-3,920350
VSX1	-1,496044	-2,666593	$\Delta VSX1$	-4,932805***	-3,959148
LS	-2,468770	-3,886751	ΔLS	-4,625157***	-3,920350

***: mức ý nghĩa 1%

Nguồn: nghiên cứu của nhóm tác giả.

Để lựa chọn mô hình phù hợp nhất thì mô hình đó phải đáp ứng các kiểm định về nhiễu trắng. Kiểm định nghiệm đơn vị cho phần dư bằng biểu đồ tương quan thống kê Q và ADF đều cho thấy phần dư của cả hai chuỗi đều có tính dừng và các giá trị ACF và PACF của phần dư đều nằm trong khoảng tin cậy 95% ở tất cả các độ trễ, nên phần dư của mô hình là nhiễu trắng. Mô hình ARIMA3 của chuỗi công nghiệp chế biến đáp ứng điều kiện các nhiễu không tự tương quan nên là mô hình phù hợp nhất (Hình 4, Bảng 4).

Như vậy, trong số các mô hình ARIMA ở Bảng 3 thì mô hình ARIMA3 là mô hình phù hợp nhất. Kết quả hồi quy của các mô hình được thể hiện ở Bảng 5.

Mặc dù ARIMA3 được xem là phù hợp nhất, nhưng mô hình chỉ hoàn hảo khi các hệ số có ý nghĩa thống kê trong việc giải thích sự thay đổi của số lao động của các doanh nghiệp tại Bình Dương. Bảng 5 cho thấy tất cả các hệ số đều không có ý nghĩa thống kê trong việc giải thích mức độ ảnh hưởng của các biến AR và MA lên sự thay đổi cầu lao động. Tuy nhiên, do mục tiêu của bài nghiên cứu là minh họa việc dự báo, nên mô hình vẫn được chấp nhận với kết quả dự báo.

4.2. Mô hình VECM

Để áp dụng mô hình VECM, trước tiên, tính dừng của các chuỗi phải được đánh giá qua đồ thị và biểu đồ tương quan ở Hình 5, Hình 6, Hình 7 với các chuỗi DT1, VSX1, LS. Đồ thị và biểu đồ tương quan ở tất cả ba chuỗi đều cho thấy các chuỗi không có tính dừng, với các giá trị ACF và PACF ở một số độ trễ khác không có ý nghĩa thống kê. Tương tự, kết quả kiểm định ADF ở Bảng 6 càng khẳng định chắc chắn hơn các

Bảng 7: Kiểm định tính đồng liên kết giữa các biến

H_0	H_A	Kiểm định Trace			Kiểm định giá trị riêng nhỏ nhất		
		Giá trị thống kê Trace	Giá trị tới hạn mức tin cậy 95%	Giá trị p	Giá trị thống kê cực đại Eigen	Giá trị tới hạn mức tin cậy 95%	Giá trị p
$r=0^*$	$r=1$	55,11029	47,85613	0,0090	37,80106	27,58434	0,0017
$r \leq 1$	$r=2$	17,30923	29,79707	0,6172	12,60618	21,13162	0,4891
$r \leq 2$	$r=3$	4,703051	15,49471	0,8395	4,456207	14,26460	0,8083
$r \leq 3$	$r=4$	0,246844	3,841466	0,6193	0,246844	3,841466	0,6193

Nguồn: nghiên cứu của nhóm tác giả.

Bảng 8: Xác định độ trễ của mô hình

Độ trễ	LR	FPE	AIC	SC	HQ
0	NA	0,000	-9,909093	-9,715946*	-9,899202
1	28,49063*	0,000*	-10,49915	-9,533414	-10,44970
2	17,08578	0,000	-10,93998*	-9,201652	-10,85096*

Nguồn: nghiên cứu của nhóm tác giả.

Bảng 9: Kết quả hồi quy mô hình VECM về cầu lao động

Các biến	$\Delta L1_t$	Thống kê t
δ	-0,011229	-0,68261
α	0,276609	0,95269
$L1_{t-1}$	1	
$DT1_{t-1}$	-0,186827	-4,08857***
$VSX1_{t-1}$	-0,662096	-4,32150***
LS_{t-1}	4,018987	11,6671***
ρ_0	-0,572466	
Các biến	Hệ số	Thống kê t
$\Delta L1_{t-1}$	-0,426332	-1,36600
$\Delta DT1_{t-1}$	-0,017101	-0,15175
$\Delta VSX1_{t-1}$	-0,645374	-2,47204**
ΔLS_{t-1}	-1,041357	-1,25718
R^2	0,5025	
Sum sq. resids	0,0404	
S.E. equation	0,06359	
Log likelihood	25,1413	

***: Mức ý nghĩa 1%

**: Mức ý nghĩa 5%

Nguồn: nghiên cứu của nhóm tác giả.

chuỗi có nghiệm đơn vị, vì vậy, chuỗi không dừng.

Tiếp tục xem xét sai phân bậc 1, lúc này, tất cả các chuỗi đều dừng. Có thể thấy rõ điều này qua biểu đồ tương quan và kiểm định ADF. Tất cả các chuỗi sau khi lấy sai phân bậc một đều có các giá trị ACF và PACF nằm trong khoảng tin cậy 95% (Hình 8, Hình 9, Hình 10) và giá trị thống kê t bé hơn giá trị tới hạn (Bảng 6).

Sau khi các chuỗi thời gian đều cùng dừng tại bậc tích hợp thứ 1. Sự đồng liên kết được kiểm tra bằng các kiểm định Trace và kiểm định giá trị riêng nhỏ nhất để xem xét liệu rằng có tồn tại mối quan hệ trong dài hạn giữa các biến hay không.

Bảng 7 cho thấy với cả hai kiểm định đều chấp nhận giả thuyết chuỗi có tính đồng kết hợp với mức ý nghĩa 5%. Hay nói cách khác, có 1 phương trình đồng kết hợp nên các chuỗi có thể chệch khỏi nhau trong

Bảng 10: Kiểm định mối quan hệ nhân quả Granger của các biến độc lập với biến phụ thuộc

Giả thuyết H_0	Kiểm định F		Kiểm định χ^2	
	F	Giá trị p	χ^2	Giá trị p
DT1 không gây nhân quả Granger đối với L1	3,62021	0,0779	0,023029	0,8794
VSX1 không gây nhân quả Granger đối với L1	5,77942	0,0306	6,110997	0,0134
LS không gây nhân quả Granger đối với L1	0,46489	0,5065	1,580496	0,2087

Nguồn: nghiên cứu của nhóm tác giả.

Bảng 11: Các kiểm định mô hình VECM

Kiểm định	Giá trị thống kê	Giá trị p
LM(1)	14,20148	0,5837
H	105,4964	0,3341
NORM	2,709871	0,6075

Nguồn: nghiên cứu của nhóm tác giả

ngắn hạn nhưng lại hội tụ kết hợp với nhau trong dài hạn. Vì vậy, mô hình VECM phù hợp để dự báo cho ngành. Mô hình VECM phụ thuộc không chỉ vào các biến độc lập mà còn có thể được giải thích bởi biến trễ của chính nó và các biến trễ của các biến giải thích. Vì vậy, cần xác định những biến trễ nào sẽ làm cho mô hình tối ưu nhất.

Theo Bảng 8, các biến ở độ trễ ở 1 hoặc 2 đều có cùng số lượng các giá trị LR, FPE, AIC, SC và HQ nhỏ nhất. Tuy nhiên, do số lượng quan sát của các biến không đủ đảm bảo để thực hiện hồi quy có độ trễ là 2, sẽ làm giảm bậc tự do, nên thực hiện mô hình với độ trễ 1 vẫn hợp lí. Như vậy, mô hình có độ trễ bằng 1 và một phương trình đồng liên kết, kết quả hồi quy như Bảng 9.

Mô hình hồi quy cầu lao động trong dài hạn với các biến như sau:

$$L1 = - 0,572466 + 0,186827DT1 + 0,662096VSX1 - 4,018987LS$$

(4,08857) (4,32150) (11,6671)

Kết quả hồi quy cho thấy phương trình cân bằng trong dài hạn phù hợp do các biến độc lập đều có ý nghĩa thống kê trong việc giải thích cầu lao động dài hạn ngành công nghiệp chế biến ở mức ý nghĩa 1%. Kết quả cũng cho thấy hướng tác động của mô hình cũng phù hợp với kỳ vọng ban đầu, khi doanh thu và vốn sản xuất tăng thì cầu lao động tăng, trong khi đó, lãi suất tăng sẽ khiến cho chi phí vay vốn tăng và làm giảm nhu cầu gia tăng lao động.

Bảng 9 cũng cho thấy sự tác động của các biến đến cầu lao động trong ngắn hạn. Trong ngắn hạn chỉ có sự tác động của biến trễ VSX1 cho thấy có ý nghĩa thống kê mức 5%, trong khi đó sự giải thích của các biến trễ của chính cầu lao động (L1), DT1 và LS lại không có ý nghĩa thống kê trong mô hình này. Như vậy, chỉ có vốn sản xuất có sự tác động và ảnh hưởng trong cả dài hạn và ngắn hạn tới biến động của cầu lao động, trong khi đó các nhân tố doanh thu và lãi suất đều thất bại trong việc kết hợp với nhau và với vốn sản xuất để tác động lên cầu lao động trong dài hạn. Tuy nhiên, DT1, VSX1 và LS lại không có sự hiệu chỉnh về cân bằng trong dài hạn.

Để xem xét rõ hơn tác động của các biến lên cầu lao động Bình Dương, kiểm định nhân quả Granger được thực hiện ở Bảng 10. Đối với kiểm định F, bác bỏ giả thuyết H_0 đối với tác động của doanh thu và vốn sản xuất lên cầu lao động. Tăng trưởng doanh thu và vốn sản xuất có tác động nhân quả lên tăng trưởng cầu lao động ngành với mức ý nghĩa 10% và 5% theo thứ tự. Ngược lại, biến động lãi suất không cho thấy có tác động nhân quả Granger đối với cầu lao động khi chấp nhận giả thuyết H_0 rằng biến lãi suất không có mối quan hệ nhân quả với biến cầu lao động ngành. Kết quả kiểm định χ^2 có sự khác biệt với kiểm định F ở yếu tố doanh thu, lúc này doanh thu không có ý nghĩa tác động lên cầu lao động của ngành, còn các nhân tố khác

Bảng 12: Dự báo

Năm	Giá trị thực	ARIMA		VECM	
		Dự báo	Sai số	Dự báo	Sai số
2015	838402	864984	3,171%	873180	4,15%
2016	893570	919655	2,919%	935079	4,65%
2017	937208	967102	3,190%	994376	6,10%
MAPE			3,093%		4,964%
Theil			0,03002		0,04859

Nguồn: nghiên cứu của nhóm tác giả

đều thống nhất kết quả với kiểm định F. Thêm vào đó, theo kết quả kiểm định Granger của tất cả các biến thì sự kết hợp giữa các biến DT1, VSX1 và LS đều thực sự là nguyên nhân gây ra sự biến động của cầu lao động với giá trị $p < 0,05$ (0,0221).

Cuối cùng, các kiểm định được thực hiện để kiểm tra việc thỏa mãn các giả định của mô hình đối với phần dư. Kết quả kiểm định cho thấy mô hình không có tương quan chuỗi giữa các biến trong mô hình với các biến trễ của nó, mô hình có hiện tượng phương sai không đổi, phần dư là nhiễu trắng.

Cả mô hình ARIMA và VECM đều có khả năng dự báo với độ chính xác cao với tỷ lệ sai số trong dự báo của các mô hình đều nằm trong giới hạn chấp nhận được là 10% với sai số của mô hình ARIMA (3,09%) thấp hơn so với mô hình VECM (4,96%). Bên cạnh đó hệ số U của cả hai mô hình đều thấp hơn nhiều so với 1 và gần bằng 0, trong đó, mô hình ARIMA có độ chính xác cao hơn VECM với hệ số 0,03 so với 0,048. Mô hình ARIMA và VECM đều có hệ số U gần bằng 0, có nghĩa là sai số dự báo gần như là 0 và điều này cho thấy các mô hình lựa chọn khá hoàn hảo.

Kết quả nghiên cứu có sự thống nhất về chất lượng dự báo của hai mô hình ARIMA và VECM trong dự báo cầu lao động với các nghiên cứu đi trước của Wong & cộng sự (2007), Wong & cộng sự (2011), Ho (2012). Tuy nhiên, kết quả nghiên cứu của nhóm tác giả lại cho thấy ARIMA là mô hình tối ưu trong dự báo cầu lao động tại Bình Dương hơn so với mô hình VECM. Song, lựa chọn một mô hình phù hợp nhất còn tùy thuộc vào nhiều nguyên nhân. Nguyên nhân chính có thể thấy rõ, đó chính là không xuất hiện mối quan hệ trong dài hạn của các biến ở mô hình VECM nên kết quả dự báo của VECM có độ chính xác thấp hơn so với mô hình ARIMA. Ngoài ra, những nguyên nhân khác gián tiếp khiến cho kết quả khác với mong đợi có thể kể đến như chất lượng và số lượng dữ liệu nghiên cứu. Cỡ mẫu nghiên cứu của nhóm tác giả khá nhỏ so với cỡ mẫu ở các nghiên cứu đi trước. Thứ hai, là do các nhân tố được lựa chọn trong mô hình quá nghèo nàn về số lượng. Thứ ba, vì khoảng thời gian thu thập dữ liệu quá ngắn nên không thể thử nghiệm với các biến trễ lớn hơn (từ hai biến trễ trở lên). Những điều này có thể là nguyên nhân khiến cho mô hình dự báo không đạt kết quả như mong muốn.

5. Kết luận

Dự báo cầu lao động có vai trò quan trọng đối với hoạt động quản lý kinh tế - xã hội nói chung, nhất là ở quá trình lập kế hoạch, cung cấp thông tin và hoạch định chính sách. Cầu lao động, đặc biệt là lao động có trình độ chịu ảnh hưởng bởi qui mô, cơ cấu và trình độ phát triển của nền kinh tế và tổng lực lượng lao động; tính chất và trình độ phát triển của khoa học kỹ thuật.

Một số mô hình dự báo cầu lao động đã được thực hiện trong một số công trình nghiên cứu ở các nước trên thế giới và ứng dụng để dự báo cầu lao động ở Việt Nam. Bài viết trình bày về các phương pháp dự báo cầu lao động đã được sử dụng trong các công trình này, đặc biệt ứng dụng của phương pháp dự báo cho kết quả cụ thể. Các điều kiện để thực hiện dự báo cầu lao động là điều kiện về số liệu, điều kiện về kinh phí thực hiện, về kỹ thuật thực hiện. Tuy nhiên, do khả năng thu thập dữ liệu và trình độ thực hiện các mô hình dự báo còn nhiều hạn chế nên kết quả dự báo chắc chắn không tránh khỏi thiếu sót.

Công tác dự báo cầu lao động tại địa phương là khâu tiên quyết, có vai trò quan trọng trong sự phát triển kinh tế - xã hội địa phương. Đây là căn cứ khoa học để định hướng xây dựng chiến lược phát triển thị trường lao động địa phương. Do đó, đòi hỏi các nhà hoạch định chính sách địa phương phải đầu tư nguồn lực lâu dài cho công tác dự báo; đồng thời người làm công tác dự báo cũng phải tích cực, chủ động xây dựng mô hình, quy trình dự báo cầu lao động phù hợp với điều kiện thực tế của địa phương, có như vậy, kết quả dự báo mới mang lại độ tin cậy cao và có ý nghĩa thiết thực.

Tài liệu tham khảo

- Box, G.E., Jenkins, G.M., Reinsel, G.C. & Ljung, G.M. (2015), *Time series analysis: forecasting and control*, John Wiley & Sons, USA.
- Bruha, J. (2011), 'An Empirical Small Labor Market Model for the Czech Economy', *Finance a Uver: Czech Journal of Economics & Finance*, 61(5), 434-449.
- CCo, L. (2007), *Is it possible to accurately forecast labour market needs?*, Canadian Council on Learning, Ottawa.
- Crane, S.E. & Nourzad, F. (1998), 'Improving local manufacturing employment forecasts using cointegration analysis', *Growth and Change*, 29(2), 175-195.
- Edlund, P.O. & Karlsson, S. (1993), 'Forecasting the Swedish unemployment rate VAR vs. transfer function modelling', *International Journal of Forecasting*, 9(1), 61-76.
- Fauvel, Y., Paquet, A. & Zimmerman, C. (1999), *Short-term forecasting of national and provincial employment in Canada*, Human Resources Development Canada, Canada.
- Ho, P.H. (2010), 'Forecasting construction manpower demand by gray model', *Journal of Construction Engineering and Management*, 136(12), 1299-1305.
- Ho, P.H. (2012) 'Comparison of the grey model and the Box-Jenkins model in forecasting manpower in the UK construction industry', *Twenty-eighth annual conference September 2012 3-5*, Association of Researchers in Construction Management, UK, 369-379.
- Krolzig, H.M., Marcellino, M. & Mizon, G.E. (2002), 'A Markov-switching vector equilibrium correction model of the UK labour market' in Hamilton J.D. & Raj B.(eds), *Advances in Markov-Switching Models*, Physica, Heidelberg, 91-112.
- LeSage, J.P. (1990a), 'Forecasting metropolitan employment using an export-base error-correction model', *Journal of Regional Science*, 30(3), 307-323.
- LeSage, J.P. (1990b), 'A Comparison of the Forecasting Ability of ECM and VAR Models', *The review of Economics and Statistics*, 72(4), 664-671.
- Puri, A. & Soydemir, G. (2000), 'Forecasting industrial employment figures in Southern California: A Bayesian vector autoregressive model', *The Annals of Regional Science*, 34(4), 503-514.
- Rapach, D. E., & Strauss, J. K. (2005), 'Forecasting employment growth in Missouri with many potentially relevant predictors: an analysis of forecast combining methods', *Federal Reserve Bank of St. Louis Regional Economic Development*, 1(1), 97-112.
- Rapach, D. E., & Strauss, J. K. (2012), 'Forecasting US state-level employment growth: An amalgamation approach', *International Journal of Forecasting*, 28(2), 315-327.
- Sarantis, N., & Swales, C. (1999), 'Modelling and forecasting regional service employment in Great Britain', *Economic Modelling*, 16(3), 429-453.
- Vereen, S. C., Rasdorf, W., & Hummer, J. E. (2016), 'Application and Results of a Skilled Labor Demand Forecast Model for the US Construction Industry', *International Journal of engineering science invention*, 5(10), 37-48.
- Wong, J.M., Chan, A.P. & Chiang, Y.H. (2005), 'Time series forecasts of the construction labour market in Hong Kong: the Box-Jenkins approach', *Construction Management and Economics*, 23(9), 979-991.
- Wong, J.M., Chan, A.P., & Chiang, Y.H. (2007), 'Forecasting construction manpower demand: A vector error correction model', *Building and Environment*, 42(8), 3030-3041.
- Wong, J.M., Chan, A.P. & Chiang, Y.H. (2011), 'Construction manpower demand forecasting: A comparative study of univariate time series, multiple regression and econometric modelling techniques', *Engineering, Construction and Architectural Management*, 18(1), 7-29.