

## LỜI CAM ĐOAN

Tôi xin cam đoan luận văn thạc sĩ công nghệ thông tin “**Hệ thống quản lý dân số và dự báo biến động dân số trên địa bàn tỉnh Tây Ninh**” là do tôi nghiên cứu, tổng hợp và thực hiện.

Toàn bộ nội dung luận văn, những điều được trình bày là của chính cá nhân tôi hoặc là được tham khảo, tổng hợp đều được trích xuất với nguồn gốc rõ ràng. Các số liệu, kết quả nêu trong luận văn là trung thực và chưa từng được ai công bố trong bất kỳ công trình nào khác.

Tp. Hồ Chí Minh, ngày 12 tháng 07 năm 2022

**Học viên thực hiện luận văn**

**Huỳnh Tuấn Kiệt**

## LỜI CẢM ƠN

Lời đầu tiên, học viên xin bày tỏ lòng biết ơn sâu sắc đến Thầy **PGS.TS Vũ Đức Lung**, người đã trực tiếp định hướng và hướng dẫn tận tình học viên trong suốt quá trình hoàn thành luận văn. Những kinh nghiệm của Thầy là tiền đề để giúp học viên mở rộng kiến thức và hoàn thành luận văn tốt nghiệp.

Đặc biệt, học viên bày tỏ lòng biết ơn tới các Thầy Cô trong Học viện Công Nghệ Bru Chính Viễn Thông cơ sở Tp. Hồ Chí Minh. Các Thầy Cô đã dạy bảo và luôn tạo điều kiện tốt nhất cho học viên trong suốt quá trình tham gia lớp cao học.

Xin trân trọng cảm ơn các anh/chị cán bộ ở Chi cục dân số tỉnh Tây Ninh, đã hỗ trợ trong việc lấy yêu cầu, cung cấp số liệu cần thiết và các kinh nghiệm trong công tác quản lý dân số trên địa bàn tỉnh Tây Ninh. Cảm ơn các bạn đồng nghiệp cơ quan, đã tạo điều kiện về thời gian và quan tâm động viên tinh thần trong thời gian học viên đi học và hoàn thành luận văn.

Cuối cùng, xin cảm ơn gia đình và bạn bè đã luôn bên tôi, cổ vũ và động viên tôi trong suốt quá trình hoàn thành luận văn này.

Tp. Hồ Chí Minh, ngày 12 tháng 07 năm 2022

**Học viên thực hiện luận văn**

**Huỳnh Tuấn Kiệt**

## MỤC LỤC

LỜI CAM ĐOAN .....	i
LỜI CẢM ƠN .....	ii
DANH SÁCH BẢNG .....	vi
DANH SÁCH HÌNH VẼ.....	vii
MỞ ĐẦU .....	1
CHƯƠNG 1: TỔNG QUAN.....	3
1.1 Tổng quan quản lý dân số .....	3
1.2 Dự báo .....	5
CHƯƠNG 2: NGHIÊN CỨU CÁC YÊU CẦU QUẢN LÝ DÂN SỐ VÀ CÁC MÔ HÌNH DỰ BÁO .....	11
2.1 Các yêu cầu quản lý dân số .....	12
2.2 Các mô hình dự báo .....	13
2.2.1 <i>Chuỗi thời gian và các đại lượng đặc trưng</i> .....	13
2.2.1.1 Giới thiệu chuỗi thời gian (time series) .....	13
2.2.1.2 Các đại lượng đặc trưng của chuỗi thời gian .....	14
2.2.1.3 Các công cụ đo lường hiệu suất cho dự báo .....	17
2.2.2 <i>Mô hình ARIMA thường</i> .....	19
2.2.2.1 Toán tử trễ .....	19
2.2.2.2 Chuỗi thời gian dừng.....	19
2.2.2.3 Quá trình tuyến tính.....	21
2.2.2.4 Quá trình tự hồi qui AR (Auto-regressive) – AR(p) .....	22
2.2.2.5 Quá trình trung bình trượt MA (Moving Average)-MA(q) .....	27
2.2.2.6 Quá trình trung bình trượt tự hồi qui ARMA (p, q) .....	29
2.2.2.7 Mô hình tích hợp trung bình trượt tự hồi qui ARIMA (p, d, q) .....	30
2.2.2.8 Qui trình xây dựng mô hình ARIMA (p, d, q) .....	33
CHƯƠNG 3: XÂY DỰNG HỆ THỐNG QUẢN LÝ VÀ DỰ BÁO DÂN SỐ.....	44
3.1 Xây dựng hệ thống quản lý dân số .....	44
3.1.1 <i>Mô tả dữ liệu</i> .....	44
3.1.2 <i>Xây dựng phần mềm quản lý</i> .....	53
3.2 Xây dựng mô hình dự báo dân số.....	61
Bước 1: nhận dạng mô hình:.....	63

Bước 2: Ước lượng mô hình: .....	67
Bước 3: Kiểm định mô hình: .....	68
Bước 4: Dự báo: .....	69
<b>CHƯƠNG 4: KẾT LUẬN VÀ KIẾN NGHỊ.....</b>	<b>71</b>
<b>4.1 Kết quả đạt được .....</b>	<b>71</b>
<b>4.2 Hạn chế và hướng phát triển.....</b>	<b>71</b>
<b>DANH MỤC CÁC TÀI LIỆU THAM KHẢO .....</b>	<b>72</b>
<b>PHỤ LỤC.....</b>	<b>74</b>

## DANH MỤC CÁC THUẬT NGỮ, CHỮ VIẾT TẮT

<b>Viết tắt</b>	<b>Tiếng Anh</b>	<b>Tiếng Việt</b>
AR	Auto-Regressive	Mô hình tự hồi quy
MA	Moving Average	Mô hình trung bình trượt
ARIMA	Auto-Regressive Integrated Moving Average	Mô hình tự hồi quy trung bình trượt
ACF	Auto Correlation Function	Hàm tự tương quan
PACF	Partial AutoCorrelation Function	Hàm tự tương quan từng phần
MAPE	Mean Absolute Percentage Error	Sai số tương đối phần trăm trung bình
RMSE	Root Mean Squared Error	Sai số căn bậc hai bình phương trung bình
CSDL		Cơ sở dữ liệu
DS-KHHGD		Dân số-kế hoạch hóa gia đình

## DANH SÁCH BẢNG

Bảng 1.1: Dân số Tây Ninh phân theo giới tính và phân theo thành thị, nông thôn (Đơn vị: người) [4].....	3
Bảng 1.2: Tỷ lệ tăng dân số Tây Ninh phân theo giới tính và phân theo thành thị, nông thôn (Đơn vị: %) [4].....	4
Bảng 1.3: Cơ cấu dân số Tây Ninh phân theo giới tính và phân theo thành thị, nông thôn (Đơn vị: %) [4].....	4
Bảng 1.4: So sánh kết quả hiệu suất của 4 thuật toán .....	7
Bảng 1.5: Dự đoán dân số Thổ Nhĩ Kỳ qua so sánh các thuật toán.....	8
Bảng 1.6: Kết quả dự báo dân số Bangladesh đến năm 2030.....	9
Bảng 2.1: Đặc trưng ACF và PACF trong các mô hình tham số.....	30
Bảng 3.1: Lưu thông tin hộ khẩu .....	44
Bảng 3.2: Lưu thông tin nhân khẩu.....	45
Bảng 3.3: Lưu thông tin biến động nhân khẩu.....	48
Bảng 3.4: Lưu thông tin kế hoạch hóa gia đình .....	49
Bảng 3.5: Lưu thông tin lịch sử kế hoạch hóa gia đình .....	50
Bảng 3.6: Lưu thông tin sức khỏe sinh sản.....	51
Bảng 3.7: Lưu thông tin cộng tác viên.....	52
Bảng 3.8: Lưu thông tin địa bàn.....	52

## DANH SÁCH HÌNH VẼ

Hình 1.1: Biểu đồ dân số Việt Nam được dự báo giai đoạn 2017-2026 [2] .....	6
Hình 1.2: Thẻ Summary của Dashboard .....	6
Hình 1.3: Dự báo khuynh hướng số ca nhiễm mới Covid-19 .....	7
Hình 2.1: Các bước xây dựng mô hình ARIMA .....	33
Hình 3.1: Giao diện quản lý hộ - nhân khẩu .....	53
Hình 3.2: Giao diện in PTT (phiếu thu tin).....	54
Hình 3.3: Giao diện quản lý thông tin KHHGD-SKSS. ....	54
Hình 3.4: Giao diện xem thông tin biến động.....	55
Hình 3.5: Giao diện chuyên hộ khẩu.....	55
Hình 3.6: Giao diện tách cá nhân sang hộ khẩu mới. ....	56
Hình 3.7: Giao diện tìm kiếm các thông tin.....	57
Hình 3.8: Giao diện quản lý thông tin công tác viên dân số.....	57
Hình 3.9: Giao diện quản lý địa bàn và gán công tác viên vào địa bàn.....	58
Hình 3.10: Giao diện quản lý thông tin biến động của hộ - nhân khẩu .....	59
Hình 3.11: Giao diện báo cáo dân số kế hoạch hóa gia đình .....	59
Hình 3.12: Giao diện báo cáo dân số kế hoạch hóa gia đình .....	60
Hình 3.13: dân số Tây Ninh từ năm 2006 đến 2020 .....	63
Hình 3.14: Đồ thị chuỗi số liệu dân số Tây Ninh sau khi sử dụng hàm biến đổi Log .....	63
Hình 3.15: ACF và PACF của chuỗi LDANSO .....	64
Hình 3.16: Kiểm định DF của chuỗi LDANSO.....	64
Hình 3.17: ACF và PACF của chuỗi DLDANSO .....	65
Hình 3.18: ACF và PACF của chuỗi DLDANSO .....	66
Hình 3.19: Đồ thị chuỗi số liệu DLDANSO .....	66
Hình 3.20: Mô hình ARIMA (1, 1, 1).....	67
Hình 3.21: Kiểm định sự tương quan ACF .....	68
Hình 3.22: Kiểm định tính dừng .....	68
Hình 3.23: Đồ thị dự báo chuỗi DLDANSO .....	69
Hình 3.24: Đồ thị so sánh chuỗi ban đầu(danso) và chuỗi dự báo (danso_fore) .....	69

## MỞ ĐẦU

Tây Ninh là một tỉnh thuộc vùng Đông Nam Bộ, nằm ở vị trí cầu nối giữa Thành phố Hồ Chí Minh và thủ đô Phnom Pênh, vương quốc Campuchia và là một trong những tỉnh nằm trong vùng kinh tế trọng điểm phía Nam [16]. Nhiều năm qua, Tây Ninh cùng với cả nước luôn quan tâm đến chương trình quản lý Dân số-kế hoạch hóa gia đình, chăm sóc sức khỏe sinh sản nhân dân, quản lý biến động dân số với mục tiêu điều chỉnh tỉ lệ sinh, nâng cao chất lượng cuộc sống, tăng tuổi thọ người dân, nắm bắt được tình hình chuyển đi hay chuyển đến của người dân. Tuy nhiên, qua thời gian, dân số có những thay đổi và có nhiều hạn chế như: mức sinh giữa các địa bàn chênh lệch, mất cân bằng giới tính, người có điều kiện chăm sóc và nuôi dưỡng con còn sinh nhiều, ảnh hưởng đến chất lượng cuộc sống, chỉ số phát triển con người (HDI) còn thấp, tỷ lệ tử vong còn cao, tuổi thọ bình quân tăng trưởng nhưng vẫn còn thấp...điều này làm ảnh hưởng đến công tác quản lý dân số, đồng thời ảnh hưởng đến tình hình phát triển kinh tế xã hội của tỉnh. Vì vậy, ngày nay công tác quản lý và dự báo dân số luôn luôn có ý nghĩa quan trọng.

Hiện nay, việc ứng dụng Công nghệ thông tin (CNTT) trong quá trình quản lý dân số giúp cho việc nhập và cập nhật dữ liệu được nhanh và chính xác hơn. Giúp cho cán bộ dân số có thể thao tác, quản lý và tổng hợp báo cáo được dễ dàng và nhanh chóng. Nhờ vậy công tác quản lý dân số trên địa bàn tỉnh Tây Ninh càng chặt chẽ. Đồng thời, nâng cao ứng dụng CNTT vào việc dự báo xu thế biến động dân số giúp ích trong việc xây dựng chất lượng, phục vụ kịp thời nhu cầu cung cấp thông tin cho ngành Dân số-kế hoạch hóa gia đình và các ban, ngành, đoàn thể khác. Điều này giúp cho các cấp, ngành, địa phương, đơn vị đánh giá, dự báo tình hình, hoạch định chính sách chiến lược về Dân số theo từng giai đoạn.

Từ các thông tin trên cho thấy cần thiết có một hệ thống CNTT quản lý và dự báo dân số nhằm hỗ trợ cơ quan nhà nước Tỉnh hiệu quả hơn trong các hoạch định chính sách. Từ đó đề tài “**Hệ thống quản lý dân số và dự báo biến động dân số trên địa bàn tỉnh Tây Ninh**” được lựa chọn thực hiện.



Đề tài hướng đến xây dựng một hệ thống quản lý và dự báo dân số, và áp dụng có hiệu quả vào thực tiễn. Đề tài giúp cho cán bộ dân số quản lý về tình hình dân số và dự báo xu thế biến động dân số trong những năm tiếp theo; nhằm mục đích nâng cao chất lượng quản lý của cán bộ dân số và đồng thời góp phần cho sự phát triển kinh tế xã hội.

Để thực hiện được mục tiêu trên, cần tiến hành nghiên cứu các nội dung sau: tìm hiểu, thu thập các dữ liệu liên quan, tiếp xúc với cán bộ quản lý dân số để nắm bắt tình hình thực tế; đề ra giải pháp hợp lý cho việc xây dựng và phát triển hệ thống. Đề tài hướng đến nghiên cứu các thuật toán máy học (Machine learning algorithms) để áp dụng trong việc dự báo biến động dân số.

Mục đích nghiên cứu là xây dựng hệ thống quản lý dân số; đồng thời áp dụng các thuật toán dự báo (cụ thể là mô hình ARIMA) cho việc dự báo biến động dân số tỉnh Tây Ninh cho các quý tiếp theo.

Đối tượng nghiên cứu của đề tài tập trung vào phân tích chuỗi thời gian (time series), các mô hình dự báo chuỗi thời gian: AR, MA, ARMA, ARIMA, các công cụ đo lường hiệu suất cho dự báo, các yêu cầu công việc của cán bộ quản lý dân số.

Phạm vi nghiên cứu: Dân số trên địa bàn tỉnh Tây Ninh

Ngoài phần mở bài, mục lục, danh mục hình vẽ, danh mục bảng biểu, tài liệu tham khảo, phụ lục, phần chính của luận văn gồm 4 chương như sau:

*Chương 1. Tổng quan:* Giới thiệu tổng quan về các vấn đề quản lý dân số. Trình bày các nghiên cứu trong nước và ngoài nước liên quan đến dự báo dân số.

*Chương 2. Nghiên cứu các yêu cầu quản lý dân số và các mô hình dự báo:* Các yêu cầu cụ thể về công tác quản lý dân số, các văn bản liên quan. Nghiên cứu các mô hình dự báo ARIMA thường.

*Chương 3. Xây dựng hệ thống quản lý và dự báo dân số:* Mô tả về dữ liệu và hệ thống quản lý dân số. Ứng dụng mô hình ARIMA thường để dự báo biến động dân số Tây Ninh.

*Chương 4. Kết luận và kiến nghị:* Đánh giá kết quả đạt được và hướng phát triển tiếp theo.

## CHƯƠNG 1: TỔNG QUAN

### 1.1 Tổng quan quản lý dân số

Trong ngành khoa học máy tính, các thuật toán máy học nói chung và các mô hình chuỗi thời gian (time series) nói riêng đã đóng góp cho mục đích dự báo xu thế cho các lĩnh vực khác nhau, trong đó có dự báo biến động dân số. Trong công tác dân số-kế hoạch hóa gia đình, thì ngoài việc quản lý thông tin, còn có việc dự báo xu thế các biến động như: tổng dân, trẻ em sinh ra, theo độ tuổi, theo giới tính, hôn nhân, ly hôn, số dân chuyển đi hay chuyển đến... Những biến động đó đều ảnh hưởng quan trọng đến chiến lược, kế hoạch phát triển kinh tế xã hội của tỉnh.

Vấn đề quản lý dân số của Việt Nam nói chung, cũng như tỉnh Tây Ninh nói riêng được xác định là mục tiêu chiến lược. Trong đó có các vấn đề trọng tâm như nâng cao chất lượng nguồn nhân lực, phát triển việc làm có chất lượng cao để tận dụng cơ cấu “dân số vàng”; phát triển hệ thống y tế và giáo dục phù hợp với cơ cấu dân số thay đổi; dân số già và an sinh xã hội; mất cân bằng giới tính khi sinh và hệ lụy; di dân, đô thị hóa ngày càng mạnh mẽ, yêu cầu tư vấn và khám sức khỏe tiền hôn nhân và chất lượng cuộc sống; tác động kinh tế - xã hội trong thời kỳ mức sinh thấp, gia đình nhỏ, gia đình hạt nhân ... Vì vậy cần phải thông tin, giáo dục, tuyên truyền không chỉ cho người dân mà đặc biệt cần thiết đối với cán bộ, công chức trong hệ thống chính trị và các nhà hoạch định chính sách, lãnh đạo và quản lý các cấp, các ngành.

Các bảng thống kê dân số của tỉnh Tây Ninh giai đoạn 2010 – 2020.

**Bảng 1.1: Dân số Tây Ninh phân theo giới tính và phân theo thành thị, nông thôn (Đơn vị: người) [4]**

Năm	Tổng số	Phân theo giới tính		Phân theo thành thị, nông thôn	
		Nam	Nữ	Thành thị	Nông thôn
2010	1.076.486	535.841	540.645	170.004	906.482
2011	1.090.055	542.830	547.225	174.426	915.629
2012	1.103.410	549.718	553.692	178.902	924.508

2013	1.111.396	553.936	557.460	182.583	928.813
2014	1.120.726	558.828	561.898	206.545	914.181
2015	1.129.913	563.652	566.261	206.863	923.050
2016	1.139.654	568.757	570.897	207.181	932.473
2017	1.151.074	574.704	576.370	207.499	943.575
2018	1.160.681	579.751	580.930	207.817	952.864
2019	1.171.683	585.500	586.183	208.300	963.383
Sơ bộ 2020	1.178.329	589.501	588.828	381.106	797.223

**Bảng 1.2: Tỷ lệ tăng dân số Tây Ninh phân theo giới tính và phân theo thành thị, nông thôn (Đơn vị: %) [4]**

Năm	Tổng số	Phân theo giới tính		Phân theo thành thị, nông thôn	
		Nam	Nữ	Thành thị	Nông thôn
2010	0.87	0.92	0.83	2.21	0.63
2011	1.26	1.30	1.22	2.60	1.01
2012	1.23	1.27	1.18	2.57	0.97
2013	0.72	0.77	0.68	2.06	0.47
2014	0.84	0.88	0.80	13.12	-1.58
2015	0.82	0.86	0.78	0.15	0.97
2016	0.86	0.91	0.82	0.15	1.02
2017	1.00	1.05	0.96	0.15	1.19
2018	0.83	0.88	0.79	0.15	0.98
2019	0.95	0.99	0.90	0.23	1.10
Sơ bộ 2020	0.57	0.68	0.45	82.96	-17.25

**Bảng 1.3: Cơ cấu dân số Tây Ninh phân theo giới tính và phân theo thành thị, nông thôn (Đơn vị: %) [4]**

Năm	Tổng số	Phân theo giới tính		Phân theo thành thị, nông thôn	
		Nam	Nữ	Thành thị	Nông thôn
2010	100.00	49.78	50.22	15.79	84.21
2011	100.00	49.80	50.20	16.00	84.00
2012	100.00	49.82	50.18	16.21	83.79
2013	100.00	49.84	50.16	16.43	83.57

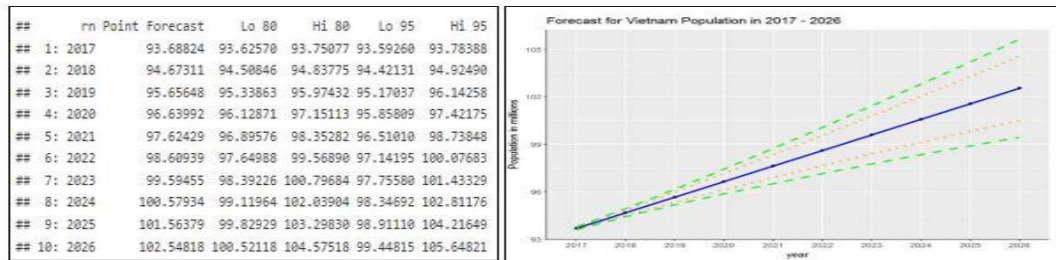
2014	100.00	49.86	50.14	18.43	81.57
2015	100.00	49.88	50.12	18.31	81.69
2016	100.00	49.91	50.09	18.18	81.82
2017	100.00	49.93	50.07	18.03	81.97
2018	100.00	49.95	50.05	17.90	82.10
2019	100.00	49.97	50.03	17.78	82.22
Sơ bộ 2020	100.00	50.03	49.97	32.34	67.66

## 1.2 Dự báo

Quy mô, cơ cấu phân bố dân số tỉnh Tây Ninh đã và đang có xu thế biến đổi nhanh. Để đảm bảo nguyên tắc con người là trung tâm của phát triển, để kế hoạch có tính thực tiễn, hiệu quả cao thì phải tiến hành dự báo dân số. Trong đó, trọng tâm là tốc độ gia tăng dân số, kế hoạch hóa lao động – việc làm, tận dụng cơ cấu “dân số vàng”, kế hoạch y tế và giáo dục, nâng cao chất lượng nguồn nhân lực, cơ cấu dân số theo độ tuổi, yếu tố di dân... Do đó, việc xây dựng hệ thống số liệu quản lý dân số đầy đủ, kịp thời, chính xác và dự báo đáng tin cậy là cơ sở không thể thiếu cho việc kế hoạch phát triển kinh tế - xã hội của tỉnh.

Các công trình nghiên cứu về dự báo dân số có liên quan:

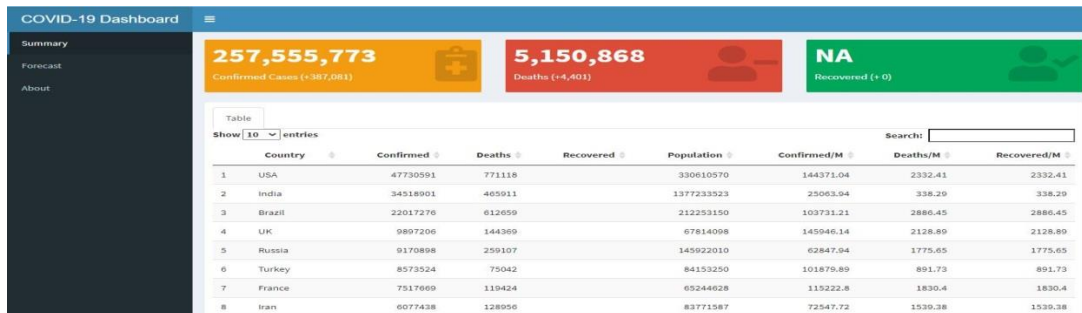
- Năm 2017, tác giả Nguyễn Ngọc Thiều có công trình nghiên cứu: “Dự báo phát triển dân số Việt Nam 2017-2026” [2]. Đề tài nghiên cứu dự đoán phát triển dân số Việt Nam đến 2026, lấy dữ liệu từ <https://www.worldometers.info/> trong khoảng thời gian từ năm 1960 đến 2016 và lưu trữ dưới dạng excel rồi phân tích và dự đoán dân số trong 10 năm tới. Bằng cách xác định các tham số p, d, q trong mô hình ARIMA, từ đó chọn được mô hình ARIMA (2,2,2) là mô hình dự báo cho phát triển dân số Việt Nam trong 10 năm tiếp theo từ 2017 đến 2026. Theo kết quả của dự báo này, đến năm 2024 dân số nước ta vượt mốc 100 triệu người.



**Hình 1.1: Biểu đồ dân số Việt Nam được dự báo giai đoạn 2017-2026 [2]**

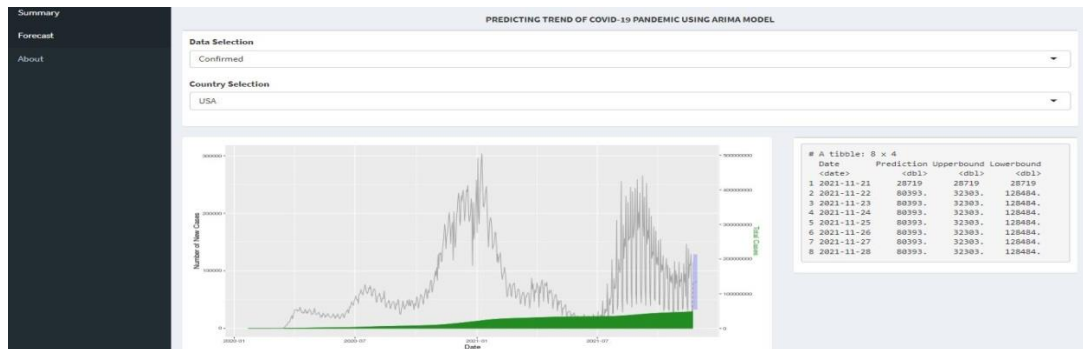
- Năm 2020, nhóm tác giả Nguyễn Quốc Dương và các cộng sự đã có công trình nghiên cứu: “Ứng dụng Shiny kết hợp với mô hình ARIMA để dự báo đại dịch COVID-19” [3]. Trong bài viết này, nhóm tác giả đã sử dụng gói lệnh Shiny (gói lệnh có sẵn trong R) và mô hình ARIMA để xây dựng website nhằm dự báo khuynh hướng của đại dịch Covid-19 cho mỗi quốc gia trên thế giới. Ứng dụng Dashboard cập nhật dữ liệu phân tích hàng ngày của đại dịch Covid-19. Ứng dụng này gồm có 2 thẻ chính là “Summary” và “Forecast”.

+ Thẻ “Summary”, tổng quan về tình hình dịch bệnh Covid-19 bằng cách hiển thị các con số thống kê dịch bệnh. Ba hộp thoại trên đầu trang web với 3 màu: vàng, đỏ, xanh lá cây lần lượt là số ca nhiễm tích lũy (confirmed), số ca tử vong (deaths), số ca phục hồi tích lũy (recovered) trên toàn thế giới. Đồng thời, bên dưới là danh sách hiển thị đầy đủ thông tin các quốc gia có dịch.



**Hình 1.2: Thẻ Summary của Dashboard**

+ Thẻ “Forecast”, gói lệnh Shiny và mô hình ARIMA được sử dụng để dự báo khuynh hướng dịch bệnh Covid-19 cho từng quốc gia cụ thể. Khuynh hướng có 3 tùy chọn: số ca nhiễm tích lũy (confirmed), số ca tử vong (deaths), số ca phục hồi tích lũy (recovered).



**Hình 1.3: Dự báo khuynh hướng số ca nhiễm mới Covid-19**

- Năm 2019, Ahmet Tezcan Tekin đã có công trình nghiên cứu: “Machine learning algorithms to forecast population: Turkey example” [10], nghiên cứu đã sử dụng các thuật toán máy học: LightGBM, Holt-Winters, ARIMA và Prophet để dự đoán dân số: ví dụ cụ thể ở Thổ Nhĩ Kỳ. Các mô hình thuật toán đã đào tạo với dữ liệu từ năm 1960 đến 2017 và được thử nghiệm so sánh với tổng số dân vào năm 2017. Trong quá trình nghiên cứu, 4 thuật toán đã được so sánh và đánh giá dựa vào tiêu chí công cụ đo lường RMSE và MAPE.

+ Khi so sánh hiệu suất của 4 mô hình, thấy rằng thuật toán: Holt-Winters, ARIMA và Prophet cho kết quả chính xác hơn thuật toán LightGBM.

**Bảng 1.4: So sánh kết quả hiệu suất của 4 thuật toán**

Model	Version	RMSE	MAPE	CPU Times (sec)
LightGBM	All Data Set	25098273,89	7,490%	650
LightGBM	Filtered Data Set	30830492,06	7,480%	450
Holt Winters	Population Data	157461,34	0,080%	42
ARIMA	Population Data	136570,57	0,100%	43
Prophet	Population Data	2427166,07	1,430%	86
Last Year Population	Population Data	11555505,55	1,530%	0

+ Kết quả cũng chỉ ra rằng, mô hình ARIMA và Holt-Winters cho kết quả tốt hơn các thuật toán còn lại, đặc biệt là ARIMA.

**Bảng 1.5: Dự đoán dân số Thổ Nhĩ Kỳ qua so sánh các thuật toán**

<b>Models</b>	<b>RMSE</b>	<b>MAPE %</b>	<b>Population Prediction for 2017</b>
LightGBM	6718648,60	8,3208%	87463669
Holt Winters	8360,00	0,0104%	80753380
ARIMA	6434,02	0,0080%	80751454
Prophet	813871,04	1,0100%	79931150
Cohort Component	427708,00	0,5297%	80317312
Turkstat's Prediction	979008,00	1,2125%	79766012
Turkey's Population in 2017			80745020

- “Precision, bias, and uncertainty for state population forecasts: an exploratory analysis of time series models” [11]. Bài báo nghiên cứu về độ chính xác, độ lệch, không chắc chắn cho dự báo dân số thông qua phân tích các mô hình chuỗi thời gian. Trong nghiên cứu này, tác giả sử dụng 6 mô hình của thuật toán ARIMA bao gồm: (1,1,0), (0,1,1), (2,2,0), (0,2,1), (0,2,2) <sup>a</sup>, ln (0,1,1) để dự báo cho 4 tiểu bang ở Hoa Kỳ, dữ liệu thu thập từ năm 1900 đến 2000 và đánh giá kết quả độ chính xác của các mô hình dự báo. Đối với, 2 mô hình tăng trưởng tuyến tính đó là mô hình 1 (ARIMA (1,1,0)) và mô hình 2 (ARIMA (0,1,1)), thì 10 năm dữ liệu là khoảng thời gian đủ đủ để đưa ra dự báo chính xác (MAPE nhỏ nhất). Đối với, 2 mô hình tăng trưởng phi tuyến đó là mô hình 3 (ARIMA (2,2,0)) và mô hình 4 (ARIMA (0,2,1)) thì 10 năm dữ liệu là chưa đủ để dự báo chính xác, mà cần khoảng thời gian dài hơn mới thể hiện được tính ưu việt của 2 mô hình này. Đối với, mô hình 6 (ARIMA (0,2,2)) và mô hình 7 (ARIMA ln (0,1,1)) cũng là mô hình phi

tuyến tính, nhưng gần giống với mô hình 1 và 2. Kết quả, để giải quyết sai sót có thể xảy ra, tác giả đã phát triển thêm mô hình thứ 7 dựa trên mô hình riêng lẻ tốt nhất – xác định bằng cách phân tích các hàm tự tương quan và tự tương quan một phần.

- Năm 2019, Md. Ashek Al Naim và các cộng sự đã có công trình nghiên cứu: “Time series model building and forecasting population of Bangladesh” [12]. Đề tài nghiên cứu xây dựng mô hình ARIMA với GARCH, tập dữ liệu từ năm 1960 đến 2018 từ nguồn (<https://data.worldbank.org/country/bangladesh>) và dự báo dân số của nước Bangladesk đến năm 2030. ARIMA (1,0,1) là mô hình cuối cùng để dự báo, và tác giả cho thấy rằng đây là mô hình ổn định cho dự báo tương lai.

**Bảng 1.6: Kết quả dự báo dân số Bangladesh đến năm 2030**

Year	Population	Year	Population
2020	165792485.6	2026	175334880.2
2021	167448690.8	2027	176833150.8
2022	169078573.5	2028	178305099.0
2023	170682133.8	2029	179750724.8
2024	172259371.7	2030	181170028.2
2025	173810287.2		

Trong nhiều nghiên cứu, đặc biệt các nghiên cứu [2],[10],[11],[12] có liên quan đến dữ liệu dự báo về dân số, thì nhận thấy mô hình ARIMA đã được nhiều bài báo nghiên cứu và cho thấy mô hình này tương đối ổn định cho dự báo tương lai. Mặc dù ARIMA là thuật toán chưa đủ mạnh như các thuật toán LSTM sau này tuy nhiên với bài toán dự báo dân số cũng đã đủ.

Như vậy, trong chương 1 đã giới thiệu tổng quan về công tác dân số và các công trình nghiên cứu liên quan tới dự báo dân số. Trong nhiều công trình nghiên



cứu về dự báo, cho thấy mô hình ARIMA cho kết quả ổn định. Nên chương tiếp theo (chương 2) tiếp tục nghiên cứu các yêu cầu về quản lý dân số, các mô hình dự báo về chuỗi thời gian và mô hình dự báo ARIMA.

## **CHƯƠNG 2: NGHIÊN CỨU CÁC YÊU CẦU QUẢN LÝ DÂN SỐ VÀ CÁC MÔ HÌNH DỰ BÁO**

Hiện nay, việc quản lý dân số ở tỉnh Tây Ninh là rất quan trọng, có nhiều ý nghĩa đến tình hình phát triển kinh tế - xã hội. Trọng tâm là tình hình phát triển tổng dân số, chênh lệch giới tính, độ tuổi, di cư... nhằm nâng cao chất lượng cuộc sống, cũng như các vấn đề về y tế và giáo dục.

Thực tế, việc quản lý dân số tại tỉnh Tây Ninh, đa phần phải quản lý bằng tay, nhập trên excel, dữ liệu nhập phân tán ở mỗi huyện. Điều đó, gây khó khăn cho cán bộ quản lý dân số tỉnh tổng hợp dữ liệu.

Từ đó, rất cần một hệ thống phần mềm quản lý dân số ở tỉnh Tây Ninh. Hệ thống có thể đáp ứng những yêu cầu về mặt quản lý nhằm giúp ích cho cán bộ dân số có thể nhập thông tin một cách dễ dàng, nhanh chóng; đồng thời có cơ sở dữ liệu tập trung, giúp tổng hợp báo cáo một cách nhanh và chính xác.

Các văn bản công văn và quyết định về công tác quản lý dân số có liên quan:

Theo công văn số 18/QĐ-TCDS ngày 17 tháng 3 năm 2016 về việc ban hành Quy định tạm thời về mẫu sổ ghi chép ban đầu, mẫu phiếu thu tin của công tác viên dân số và mẫu biểu báo cáo thống kê chuyên ngành của kho dữ liệu điện tử phục vụ quản lý và điều hành công tác dân số - kế hoạch hóa gia đình.

Theo công văn số 96/TCDS-KHTC ngày 18 tháng 3 năm 2016 về việc hướng dẫn ghi chép ban đầu vào sổ A0, ghi phiếu thu tin và lập báo cáo thống kê chuyên ngành DS-KHHGD. Trong công văn nêu rất rõ các khái niệm liên quan về sổ A0, phiếu thu tin của công tác viên dân số, phạm vi theo dõi DS-KHHGD như: hộ khẩu, nhân khẩu, tạm trú, thường trú thực tế, nhân khẩu chuyển đi – chuyển đến...; đồng thời cũng quy định thời gian thu thập thông tin và thời gian lập các biểu báo cáo thống kê.

Theo văn bản quyết định số 3532/QĐ-BYT ngày 12 tháng 8 năm 2020 về việc quy định về xây dựng và triển khai hệ thống thông tin quản lý Trạm y tế xã, phường, thị trấn. Quyết định này quy định về xây dựng và triển khai Hệ thống

thông tin quản lý Trạm y tế bao gồm: chức năng Hệ thống, hạ tầng kỹ thuật và nhân lực công nghệ thông tin. Trong quyết định quy định rõ cụ thể nhiều phân hệ chức năng mà trạm y tế quản lý, trong đó có quản lý công tác dân số kế hoạch hóa gia đình. Đồng thời cũng quy định rõ các thông tin cần phải quản lý: họ tên, quan hệ với chủ hộ, ngày sinh, giới tính, dân tộc, quốc tịch, nghề nghiệp...; các thông tin biến động: trẻ sinh, người mất, chuyển đi, chuyển đến; quản lý thông tin về biện pháp tránh thai: ngày thực hiện, phương pháp...; quản lý thông tin trẻ em mới sinh: họ tên người mẹ, ngày sinh trẻ, nơi sinh...

Theo quyết định số 198/QĐ-BYT ngày 13/01/2021 về việc quy định danh mục chuẩn định dạng dữ liệu kết nối liên thông y tế cơ sở với nền tảng quản lý thông tin y tế cơ sở V20. Trong quyết định này, có quy định kết nối, chia sẻ của nhiều phân hệ mà trạm y tế quản lý, trong đó có phân hệ Dân số.

## **2.1 Các yêu cầu quản lý dân số**

Khảo sát, phỏng vấn và thu thập các yêu cầu của công việc quản lý dân số cũng như các thông số hỗ trợ hoạch định chính sách của Tỉnh.

Các thông tin cần quản lý dân số được khảo sát như:

- Quản lý thông tin về các địa bàn trong tỉnh: tên địa bàn (ghi rõ: tổ, ấp/khu phố), ấp/khu phố, ghi chú tổng số địa bàn.
- Quản lý thông tin các hộ dân trong từng địa bàn: mã hộ, chủ hộ, địa chỉ.
- Quản lý thông tin các nhân khẩu trong từng hộ dân: họ tên, ngày sinh, giới tính, tình trạng hôn nhân, tình trạng cư trú, trình độ học vấn, trình độ chuyên môn, quan hệ với chủ hộ, địa chủ, ghi chú...
- Quản lý thông tin kế hoạch hóa gia đình: họ tên nhân khẩu, ngày thu thập, các biện pháp tránh thai...
- Quản lý thông tin sức khỏe sinh sản: họ tên người mẹ, ngày thu thập, ngày sinh, là con thứ mấy, nơi sinh, cân nặng trẻ lúc sinh, các thông tin của trẻ (họ tên, ngày sinh, giới tính, quan hệ với chủ hộ...)

- Quản lý các thông tin biến động: sinh, tử, chuyển đi, chuyển đến. Đồng thời có thể phục hồi lại trạng thái ban đầu, khi cán bộ dân số nhập biến động sai.
- Quản lý thông tin của công tác viên dân số: mã công tác viên, họ tên, ngày tham gia, giới tính, địa bàn quản lý, ghi chú...
- Các thông kê báo cáo theo mẫu được quy định trong công văn 18/TCDS-KHTC và 96/TCDS-KHTC.

## 2.2 Các mô hình dự báo

### 2.2.1 Chuỗi thời gian và các đại lượng đặc trưng

#### 2.2.1.1 Giới thiệu chuỗi thời gian (time series)

Time có nghĩa là thời gian, series là chuỗi, thể hiện sự có trình tự, là tập hợp tuần từ các dữ liệu điểm, một dãy các giá trị, thông tin ghi nhận được từ đối tượng nghiên cứu và được sắp xếp theo thứ tự của thời gian hay các mốc của thời gian [17]. Về mặt toán học là một tập các vectors  $z(t)$ ,  $t = 0, 1, 2, \dots$  trong đó  $t$  là thời gian,  $z(t)$  được coi là một biến ngẫu nhiên. Các phép đo được thực hiện trong một sự kiện chuỗi thời gian được sắp xếp theo thứ tự thời gian thích hợp [9].

Chuỗi dữ liệu phụ thuộc vào thời gian chia làm hai loại:

- Chuỗi dữ liệu liên tục theo thời gian: Các quan sát được đo trong khoảng thời gian liên tục. Ví dụ: dữ liệu chứng khoán được cập nhật liên tục; dữ liệu thống kê ca nhiễm covid-19; dữ liệu đo nhiệt độ trong ngày...
- Chuỗi dữ liệu theo theo thời gian được quan sát, đo đạc trong khoảng thời gian rời rạc và cách đều nhau: Các quan sát được thực hiện tách biệt, thường là các quan sát được đo tại các mốc thời gian cách đều nhau. Ví dụ: Chuỗi dữ liệu theo thời gian được đo, thống kê theo tuần, quý, tháng, năm...

Trong luận văn này, giới hạn chỉ đề cập đến chuỗi dữ liệu theo thời gian được quan sát, đo trong khoảng thời gian rời rạc và cách đều nhau.

Dự báo chuỗi dữ liệu phụ thuộc vào thời gian: Là chúng ta đi ước lượng các giá trị của biến ngẫu nhiên chuỗi thời gian  $z_{t+h}$  ( $h \geq 1$ ), dựa trên sự tương quan với các giá trị của biến ngẫu nhiên  $\{z_t\}$  đã được quan sát trong quá khứ.

Chất lượng của dự báo phụ thuộc vào nhiều yếu tố: phụ thuộc vào khoảng cách xa hay gần của dự báo; sự phức tạp của chuỗi thời gian; sự bất thường của chuỗi thời gian mà không thể lường trước được khi tiến hành dự báo...

Ứng dụng của chuỗi thời gian được áp dụng vô nhiều lĩnh vực khác nhau, như:

- Môi trường: dựa vào dữ liệu của sự thay đổi áp suất không khí và nhiệt độ từ các năm trước, có thể dự báo được sự thay đổi thời tiết trong những năm tiếp theo.
- Dân số: từ dữ liệu những năm 1960 đến 2016 về dân số ở Thổ Nhĩ Kỳ, người ta đã cho dự báo khá chính xác những năm tiếp theo [10].

Và còn nhiều lĩnh vực khác nữa như thị trường mua bán, kinh tế - tài chính, kinh doanh sản xuất...

### 2.2.1.2 Các đại lượng đặc trưng của chuỗi thời gian

Giả sử ta có chuỗi thời gian  $\{z_t\}$  gồm  $n$  các quan sát,  $t=1, 2, 3, \dots, n$ .

a. Các đại lượng thống kê đặc trưng cho chuỗi thời gian

**Kỳ vọng:** của biến ngẫu nhiên  $z_t$  là giá trị trung bình của biến ngẫu nhiên đó.

Ký hiệu [8]:

$$E(z_t) = \mu \quad (2.1)$$

Trong phần lớn trường hợp, ta không thể nghiên cứu toàn bộ tổng thể, mà chỉ kỳ vọng (mong muốn) nghiên cứu một số phần tử nào đó của tổng thể (hay còn gọi là mẫu ngẫu nhiên). Kỳ vọng đó gọi là kỳ vọng mẫu [8]:

$$\bar{\mu} = \bar{z} = \frac{1}{n} \sum_{t=1}^n z_t \quad (2.2)$$

**Phương sai:** là phép đo mức chênh lệch giữa các số liệu trong một tập dữ liệu thống kê. Phép đo đại diện cho mức độ phân tán các giá trị xung quanh kỳ vọng của nó [8].

$$\hat{\sigma}_z^2 = \sum_{t=1}^n (z_t - \bar{z})^2 \quad (2.3)$$

**Độ lệch chuẩn:** là căn bậc hai của phương sai [8].

$$\sigma_z = \sqrt{\hat{\sigma}_z^2} \quad (2.4)$$

b. Các đại lượng mô tả mối quan hệ giữa các phần tử trong chuỗi

**Tự hiệp phương sai:** sử dụng để đo mức độ tương quan tuyến tính của hai biến ngẫu nhiên trong cùng một chuỗi thời gian. Nó phản ánh sự phụ thuộc hay độc lập tuyến tính của các biến ngẫu nhiên trong chuỗi.

Tự hiệp phương sai giữa hai biến ngẫu nhiên trong cùng một chuỗi thời gian tại thời điểm  $t$  ký hiệu là  $z_t$  và tại thời điểm  $t+k$  ký hiệu  $z_{t+k}$ , giữa chúng có  $k-1$  quan sát gọi là  $k$  độ trễ, được xác định như sau [8]:

$$\gamma_z(k) = cov[z_t, z_{t+k}] = E[(z_t - \mu)(z_{t+k} - \mu)] \quad (2.5)$$

Trong đó:

+  $\mu$ : là kỳ vọng chung của  $z_t$  và  $z_{t+k}$ . Tự hiệp phương sai khi độ trễ  $k=0$  chính là phương sai của  $z_t$ :  $\gamma_z(0) = cov[z_t - z_t] = \sigma_z^2$

Tương tự, tự hiệp phương sai mẫu được tính [8]:

$$\hat{\gamma}_z(k) = \frac{1}{n} \sum_{t=1}^{n-k} (z_t - \bar{z})(z_{t+k} - \bar{z}) \quad (2.6)$$

$$k=1, 2, \dots, n-1$$

Trong đó:

+  $\bar{z}$  là kỳ vọng mẫu của  $z_t$  và  $z_{t+k}$

**Hàm tự tương quan (ACF):**

ACF: hàm tuần tự tương quan đo lường phụ thuộc tuyến tính giữa các cặp quan sát  $x_z(t)$  và  $x_z(t+k)$ , ứng với thời đoạn  $k=1, 2, \dots$  ( $k$  còn gọi là độ trễ). Với mỗi

độ trễ  $k$ , hàm tự tương quan tại độ trễ  $k$  được xác định qua độ lệch giữa các biến ngẫu nhiên  $x_z(t)$ ,  $x_z(t+k)$  so với giá trị trung bình.

Là đại lượng mô tả tương quan tại trễ  $k$  giữa các giá trị trong chuỗi thời gian, được xác định [8]:

$$\rho_k = \frac{\text{cov}(z_t, z_{t+k})}{\sigma_{z_t} \sigma_{z_{t+k}}} = \frac{\gamma_z(k)}{\sigma_{z_t} \sigma_{z_{t+k}}} = \frac{E[(z_t - \mu)(z_{t+k} - \mu)]}{\sqrt{E[(z_t - \mu)^2]E[(z_{t+k} - \mu)^2]}} \quad (2.7)$$

Trong đó:

+  $\gamma_z(k)$ : là tự hiệp phương sai.

+  $\sigma_{z_t}, \sigma_{z_{t+k}}$ : lần lượt là độ lệch chuẩn của  $z_t$  và  $z_{t+k}$ . Thông thường thì  $x_t$  và  $x_{t+k}$  đều có cùng phương sai là  $\sigma_x^2 = \gamma_x(0)$ , nên tự tương quan tại trễ  $k$  được tính:  $\rho_k = \frac{\gamma_x(k)}{\gamma_x(0)}$ . Khảo sát tự tương quan như một hàm với tham số biến thiên theo trễ  $k$  ( $k=1, 2, \dots$ ) gọi là hàm tự tương quan.

Tự tương quan mẫu được tính theo công thức [8]:

$$\hat{\rho}_k = \frac{\hat{\gamma}_x(k)}{\hat{\gamma}_x(0)} \quad (2.8)$$

Một vài tính chất của tự tương quan mẫu:

+ Tính chất 1:  $-1 \leq \hat{\rho}_k \leq 1$

+ Tính chất 2:  $k = 0 \Rightarrow \hat{\rho}_0 = 1$

+ Tính chất 3:  $\hat{\rho}_k = \hat{\rho}_{-k}$

Nếu như  $z_t$  và  $z_{t+k}$  không tương quan với nhau thì tự tương quan  $\hat{\rho}_k = 0$ , do đó khi  $\text{cov}(z_t, z_{t+k}) = 0$ . Nhưng điều ngược lại chưa hẳn đã đúng.

Dựa trên mối quan hệ tự tương quan giữa các phần tử trong chuỗi và có thể xây dựng được các mô hình dự báo chuỗi thời gian.

### **Hàm tự tương quan từng phần (PACF):**

PACF: song song với việc xác định hàm tương quan giữa các cặp  $z_t$  và  $z_{t+k}$ , ta xác định hàm tự tương quan riêng.

Tự tương quan mẫu  $\hat{\rho}_k$  phản ánh mức độ tương quan giữa hai biến ngẫu nhiên  $z_t$  và  $z_{t+k}$  trong chuỗi thời gian. Tuy nhiên, sự tương quan giữa chúng có thể chịu sự tác động của các biến khác, trong trường hợp này là  $k-1$  biến trung gian  $z_{t+1}$ ,

$z_{t+2} \dots z_{t+k-1}$  ảnh hưởng đến sự tương quan giữa biến  $z_t$  và  $z_{t+k}$ . Do đó tự tương quan từng phần được đưa vào nhằm mục đích mô tả mức độ tương quan trực tiếp giữa hai biến  $z_t$  và  $z_{t+k}$  (không bị ảnh hưởng ràng buộc bởi mối quan hệ với các biến trung gian), Được tính theo công thức [8]:

$$\phi_{kk} = \frac{\rho_k - \sum_{j=1}^{k-1} \phi_{k-1,j} \rho_{k-j}}{1 - \sum_{j=1}^{k-1} \phi_{k-1,j} \rho_j} \quad (2.9)$$

độ trễ  $k=2, 3 \dots$

$$\phi_{kj} = \phi_{k-1,j} - \phi_{kk} \phi_{k-1,k-j} \quad (2.10)$$

$j = 1, 2, \dots, k-1$

Giá trị ban đầu  $\phi_{11} = \rho_1$

Bằng tính toán tương tự trên các quan sát mẫu có được tự tương quan từng phần mẫu  $\hat{\phi}_{kk}$ . Khảo sát tự tương quan từng phần như một hàm với tham số biến thiên theo độ trễ  $k$  được gọi là hàm tự tương quan từng phần.

Tự tương quan từng phần có thể được hiểu theo quan điểm của bài toán dự báo đó là giả định muốn dự báo giá trị của  $x_{t+h}$  từ các giá trị  $z_{t+h-1}, \dots, z_t$ , dựa trên sự kết hợp tuyến tính các giá trị giá khứ này. Sau đó xem xét sai số dự báo bình phương trung bình tối thiểu (MSE – Mean Square Error) [9]:

$$MSE = E \left[ \left( z_{t+h} - \sum_{k=1}^h a_k z_{t+h-k} \right)^2 \right] \quad (2.11)$$

Ta thu được các giá trị có thể của trọng số  $a_1, \dots, a_h$ . Nếu xem xét kết quả này tại một trễ cụ thể, ví dụ là  $h$ , thì khi đó tự tương quan từng phần  $\phi_{kk}$  được coi như là giá trị của hệ số  $a_h$  ( $\phi_{hh} \cong a_h$ ).

### 2.2.1.3 Các công cụ đo lường hiệu suất cho dự báo

Trong các định nghĩa dưới đây, ta gọi  $y_t$  là giá trị thực,  $f_t$  là giá trị dự báo,  $e_t = y_t - f_t$  là giá trị sai số dự báo và  $n$  là kích thước của tập kiểm tra [9]. Và,



Giá trị trung bình:

$$\bar{y} = \frac{1}{n} \sum_{t=1}^n y_t \quad (2.12)$$

Giá trị phương sai:

$$\sigma^2 = \frac{1}{n-1} \sum_{t=1}^n (y_t - \bar{y})^2 \quad (2.13)$$

- **MAPE (Mean Absolutr Percent Error) sai số tương đối phần trăm trung bình**

Được định nghĩa: [9]

$$MAPE = \frac{1}{n} \sum_{t=1}^n \left| \frac{e_t}{y_t} \right| \times 100 \quad (2.14)$$

- Độ đo này thể hiện phần trăm sai số tuyệt đối trung bình đã xảy ra.
- Độc lập với quy mô đo lường, nhưng bị ảnh hưởng bởi quá trình chuyển đổi dữ liệu.

- Không hiển thị hướng của sai số.
- Không liệt kê các sai số cực độ.
- Độ đo này, sai số đối nghịch nhau không bù trừ cho nhau.

- **RMSE (Root Mean Squared Error) sai số căn bậc hai bình phương trung bình**

Được định nghĩa: [9]

$$RMSE = \sqrt{MSE} = \sqrt{\frac{1}{n} \sum_{t=1}^n e_t^2} \quad (2.15)$$

- RMSE là sai số căn bậc hai của MSE.
- RMSE có tất cả thuộc tính tương tự như MSE.

### 2.2.2 Mô hình ARIMA thường

Mô hình ARIMA là viết tắt của tích hợp trung bình trượt tự hồi qui thường (Autoregressive Intergrated Moving Average).

Mô hình ARIMA là một lớp mô hình tuyến tính có khả năng biểu diễn cả chuỗi thời gian tĩnh lẫn không tĩnh. Mô hình ARIMA dựa vào các mẫu tự tương quan trong bản thân của chuỗi thời gian để sinh ra dự đoán [13].

Mô hình ARIMA được sử dụng trong phân tích, dự báo chuỗi thời gian (time series) bằng cách kết hợp các hành vi quan sát được trong quá khứ, trong hiện tại và nhiều hiện tại cộng nhiều trong quá khứ để dự báo dữ liệu trong tương lai.

#### 2.2.2.1 Toán tử trễ

Giả sử có chuỗi các quan sát  $\{z_t\}$ ,  $t=1, 2, \dots, n$

Toán tử trễ, ký hiệu  $B$ , là một toán tử thao tác trên chuỗi thời gian với tính chất là làm dịch chuyển quan sát tại thời gian  $t$  sang quan sát tại thời gian  $t-1$ . Như vậy về mặt toán học, toán tử  $B$  được gọi là toán tử trễ nếu nó thực hiện phép biến đổi [8]:

$$Bz_t = z_{t-1} \quad (2.16)$$

Toán tử trễ có các tính chất điển hình sau:

$$+ B^k z_t = z_{t-k}$$

$$+ B^0 z_t = z_t$$

Bên cạnh toán tử trễ, cũng có thêm toán tử sai phân dùng để thao tác trên chuỗi thời gian được định nghĩa:

$$+ \text{Sai phân bậc 1: } \Delta = z_t - z_{t-1} = (1 - B)z_t$$

$$+ \text{Sai phân bậc 2: } \Delta^2 z_t = \Delta(\Delta z_t) = z_t - 2z_{t-1} + z_{t-2}$$

$$+ \text{Sai phân bậc d: } \Delta^d z_t = \Delta(\Delta^{d-1} z_t)$$

#### 2.2.2.2 Chuỗi thời gian dừng

##### A. Định nghĩa chuỗi thời gian dừng

Mô hình hóa chuỗi thời gian đưa ra dự báo từ mô hình phải luôn giả định rằng chuỗi thời gian phải dừng, bởi vì chuỗi dừng thì các đại lượng đặc trưng chẳng

hạn như: phương sai, kỳ vọng của nó mới có nghĩa đồng thời chỉ khi xây dựng mô hình trên chuỗi thời gian dừng thì dự báo đưa ra mới đáng tin cậy.

Chuỗi thời gian  $z_t$  với ( $t=1, 2, \dots, n$ ) được gọi là dừng nếu kỳ vọng, phương sai không đổi theo thời gian và hiệp phương sai giữa hai quan sát bất kỳ chỉ phụ thuộc vào khoảng cách (độ trễ về thời gian) giữa  $t$  và  $t-k$ , không phụ thuộc vào thời điểm hiệp phương sai được tính, tức là về mặt toán học chuỗi thời gian  $x_t$ , được gọi là dừng nếu:

$$+ \text{Kỳ vọng: } E(z_t) = \mu = \text{const} \quad \forall t$$

$$+ \text{Phương sai: } \text{var}(z_t) = \sigma_x^2 \quad \forall t$$

$$+ \text{Hiệp phương sai: } \gamma_z(k) = \text{cov}(z_q z_{q-k}) \quad \forall t, q | t \neq q$$

Nếu chuỗi dừng không thỏa mãn các điều kiện trên thì được gọi là chuỗi thời gian không dừng. Do vậy tính dừng rất quan trọng khi phân tích chuỗi thời gian, có một số phương pháp để kiểm định chuỗi thời gian dừng như sau:

#### ❖ Phương pháp 1

Kiểm định dựa trên tương quan đồ của hàm tự tương quan ACF. Barlett đã chỉ ra rằng, nếu chuỗi là ngẫu nhiên và dừng thì hàm tự tương quan sẽ có phân bố xấp xỉ với phân bố chuẩn  $N(0, 1/n)$  ( $n$  là số quan sát). Do vậy, nếu chuỗi dừng thì 95% tự tương quan mẫu sẽ nằm trong khoảng giới hạn  $\pm 1.96\sqrt{n}$ . Còn ngược lại thì chuỗi không phải là dừng khi có nhiều tự tương quan mẫu nằm ngoài khoảng giới hạn này.

Tính chất đặc trưng hàm ACF với tham số trễ  $k$  của chuỗi không dừng là nó giảm rất chậm khi  $k$  tăng và PACF thì có xu thế đạt cực đại tại độ trễ 1.

#### ❖ Phương pháp 2

Có thể kiểm tra chuỗi dừng bằng kiểm định nghiệm đơn vị DF. Nếu có tồn tại nghiệm đơn vị trong chuỗi thì kết luận đó là chuỗi không dừng.

### **B. Phương pháp biến đổi chuỗi thời gian dừng**

Thực hiện bằng một trong các phương pháp sau để biến đổi chuỗi không dừng thành chuỗi dừng.

### ❖ Phương pháp khử xu thế

Tính khử xu thế trong chuỗi thời gian là một nguyên nhân chủ yếu làm cho chuỗi không dừng. Giả sử mô hình ước lượng biểu diễn chuỗi thời gian không dừng  $z_t$  theo xu thế tuyến tính như sau:

$$\hat{z}_t = \hat{\beta}_1 + \hat{\beta}_2 t$$

Trong đó:  $\hat{\beta}_1, \hat{\beta}_2$  lần lượt là ước lượng của hệ số chặn và hệ số góc. Để biến đổi chuỗi không dừng  $z_t$  thành chuỗi dừng, một cách tự nhiên là kiểm định chuỗi phần dư thu được  $\hat{a}_t = z_t - \hat{\beta}_1 - \hat{\beta}_2 t$  có phải là chuỗi dừng không.

Ngoài ra xu thế cũng được biểu diễn mô tả bằng hàm đa thức, đường cong tuyến tính, hàm mũ... Và phương pháp loại trừ những xu thế này trong chuỗi để biến đổi thành chuỗi dừng cũng được thực hiện theo cách tương tự.

### ❖ Phương pháp sai phân

Đây là phương pháp rất phổ biến thực hiện biến đổi chuỗi không dừng thành chuỗi dừng.

Loại bỏ thành phần xu thế trong chuỗi: Áp dụng toán tử sai phân bậc  $d$  ( $d \geq 1$ ):

$$\Delta^d z_t = (1 - B)^d z_t$$

### ❖ Phương pháp hàm biến đổi

Chuỗi không dừng có thể có nguyên nhân bởi các dao động trong chuỗi không ổn định. Do đó hàm biến đổi được sử dụng để tác động khiến dao động trong chuỗi trở nên ổn định hơn. Một tập hợp các hàm biến đổi được Box-Cox đưa ra, trong số đó hàm  $z_t = \log(z_t)$  hoặc  $z_t = \sqrt{z_t}$  được sử dụng chủ yếu.

#### 2.2.2.3 Quá trình tuyến tính

Chuỗi thời gian  $\{z_t\}$  là một quá trình tuyến tính, nếu biểu diễn toán học có dạng [8]:

$$z_t = \sum_{j=-\infty}^{\infty} \phi_j a_{t-j} = \phi(B)a_t \quad \forall t \quad (2.17)$$

Trong đó:

+  $a_t$  là nhiễu trắng

+  $\{\phi_j\}$  là các hằng số thỏa điều kiện

$$\sum_{j=-\infty}^{\infty} |\phi_j| < \infty$$

+ B là toán tử trễ với

$$\sum_{j=-\infty}^{\infty} \phi_j B^j$$

Chuỗi thời gian là nhiễu trắng nếu nó hầu như không thể hiện một cấu trúc, hình mẫu nào rõ rệt, cũng như không có bất kỳ sự tự tương quan nào trong chuỗi.

Một quá trình tuyến tính được gọi là quá trình trung bình mượt MA ( $\infty$ ), nếu  $\phi_j = 0$  khi  $j < 0$  và được viết:

$$z_t = \sum_{j=0}^{\infty} \phi_j a_{t-j} \quad (2.18)$$

Trong trường hợp này, toán tử  $\phi\{B\}$  được xem xét như một bộ lọc tuyến tính, mà đầu ra (Output) là chuỗi các quan sát  $\{z_t\}$  khi áp dụng cho chuỗi đầu vào (Input) là nhiễu trắng  $\{a_t\}$ . Điều đó chứng tỏ rằng bộ lọc tuyến tính tại đầu ra là chuỗi dừng khi áp dụng với bất kỳ một chuỗi đầu vào dừng.

#### 2.2.2.4 Quá trình tự hồi qui AR (Auto-regressive) – AR(p)

Quá trình tự hồi qui là quá trình trong đó giá trị được dự báo dựa trên các giá trị trong quá khứ của nó. Mô hình bước ngẫu nhiên (là mô hình mà giá trị sinh ra từ nó được xác định bằng giá trị của quan sát ngày trước nó cộng thêm nhiễu trắng) là một trường hợp đặc biệt của quá trình tự hồi qui.

##### 1. Quá trình tự hồi qui bậc một – AR (1)

Giả sử  $\{z_t\}$  là một chuỗi dừng và các phân tử trong chuỗi có tồn tại mối quan hệ phụ thuộc tuyến tính vào phần tử ngày trước đó của nó cộng với nhiễu trắng, và được biểu diễn tuyến tính đơn giản như sau [8]:

$$z_t = \phi_1 z_{t-1} + a_t \quad (2.19)$$

Trong đó,  $a_t$  là nhiễu trắng và không tương quan với  $z_t$ ,  $\text{cov}(z_t, a_t) = 0$ . Quá trình được mô tả trên được gọi là quá trình tự hồi qui bậc 1 – AR (1).

Có thể viết công thức trên dưới dạng truy hồi sau:

$$z_t = \phi_1 z_{t-1} + a_t$$

$$z_t = \phi_1(\phi_1 z_{t-2} + a_{t-1}) + a_t = \phi_1^2 z_{t-2} + \phi_1 a_{t-1} + a_t$$

$$\begin{aligned} \dots z_t &= a_t + \phi_1 a_{t-1} + \dots + \phi_1^{k-1} z_{t-k-1} = \dots \\ &= a_t + \phi_1 a_{t-1} + \dots + \phi_1^{t-1} a_1 + \phi_1^t z_0 \end{aligned} \quad (2.20)$$

Các đại lượng đặc trưng của quá trình AR (1):

- Kỳ vọng:  $E(z_t) = E(\phi_1 z_{t-1} + a_t) = \phi_1 E(z_{t-1})$

- Phương sai:  $var(z_t) = \frac{\sigma_a^2}{1-\phi_1^2} = \gamma_z(0)$

$$\begin{aligned} var(z_t) &= var(a_t + \phi_1 a_{t-1} + \dots + \phi_1^{t-1} a_1 + \phi_1^t z_0) \\ &= \sigma_a^2 (1 + \phi_1^2 + \dots + \phi_1^{2(t-2)} + \phi_1^{2(t-1)}) \end{aligned}$$

Do  $z_0$  được coi là hằng số nên  $var(z_0) = 0$ , hơn nữa theo định lý chuỗi, nếu  $|\phi_1| < 1$  thì:

$$\sum_{i=1}^{\infty} \phi_1^i = 1 + \phi_1 + \phi_1^2 + \dots \phi_1^i \dots \quad (2.21)$$

là chuỗi hội tụ:

$$\sum_{i=1}^{\infty} \phi_1^i = 1 + \phi_1 + \phi_1^2 + \dots \phi_1^i \dots = \frac{1}{1 - \phi_1^2} \Big|_{\phi_0=1} \quad (2.22)$$

Khi đó phương sai của AR (1) được tính:

- Tự hiệp phương sai:

$$\begin{aligned} cov(z_{t+1} z_t) &= cov((\phi_1 z_t + a_t) z_t) \\ &= \phi_1 cov(z_t z_t) + cov(z_t a_t) = \phi_1 var(z_t) \end{aligned} \quad (2.23)$$

$$\text{suy ra: } ACF(1) = \rho_1 = \frac{cov(z_{t+1} z_t)}{var(z_t)} = \phi_1$$

$$\begin{aligned} cov(z_{t+2} z_t) &= cov((\phi_1 z_{t+1} + a_{t+1}) z_t) \\ &= cov(\phi_1 (\phi_1 z_t + a_t) + a_{t+1}) z_t \\ &= \phi_1^2 cov(z_t z_t) + \phi_1 cov(z_t a_t) + cov(z_t a_{t+1}) \\ &= \phi_1^2 var(z_t) \end{aligned}$$

$$\text{suy ra: } ACF(2) = \rho_2 = \frac{cov(z_{t+2} z_t)}{var(z_t)} = \phi_1^2$$

...

$$\text{cov}(z_{t+k}z_t) = \phi_1^k \text{var}(z_t), \text{ suy ra } ACF(k) = \rho_k = \phi_1^k$$

Nếu  $|\phi_1| > 1$ , khi đó:

$$\sum_{i=0}^{\infty} \phi_1^i = 1 + \phi_1 + \phi_1^2 + \dots \phi_1^i \dots \quad (2.24)$$

không hội tụ do vậy chuỗi sinh ra từ mô hình AR (1) không dừng.

Nếu  $|\phi_1| = 1$  thì quá trình AR (1) được viết  $z_t = z_{t-1} + a_t$  là quá trình bước ngẫu nhiên, do vậy chuỗi sinh ra từ quá trình này cũng không dừng.

Vậy điều kiện để quá trình dừng là  $|\phi_1| < 1$  và nghiệm của  $z_t$  có thể biểu diễn như sau:

$$z_t = \sum_{j=0}^{\infty} \phi_1^j a_{t-j} \quad (2.25)$$

## 2. Quá trình tự hồi qui bậc p – AR(p)

Mở rộng quá trình AR (1) được quá trình tự hồi qui bậc p, AR(p) như sau [14]:

$$z_t = \phi_1 z_{t-1} + \phi_2 z_{t-2} + \dots + \phi_p z_{t-p} + a_t \quad (2.26)$$

Trong đó:  $a_t$  là nhiễu trắng và không tương quan với  $z_t$  còn  $\phi_1, \phi_2 \dots \phi_p$  là p hệ số tự hồi qui được lựa chọn để sinh ra giá trị dự báo phù hợp nhất cho  $z_t$  qua các giá trị trong quá khứ  $z_{t-1}, z_{t-2} \dots z_{t-p}$ . Để thuận tiện hơn được viết lại bằng toán tử trễ như sau:

$$\phi(B)z_t = a_t \quad (2.27)$$

trong đó:  $\phi(B) = 1 - \phi_1 B - \phi_2 B^2 - \dots - \phi_p B^p$

Điều kiện để quá trình AR (p) là hội tụ là  $-1 \leq \phi_i \leq 1, i = 1 \dots p$ , Fuller đã chứng tỏ rằng điều kiện đối với các  $\phi_i$  trên tương đương với điều kiện nghiệm của phương trình đặc trưng  $\phi(B) = 0$  dưới đây lớn hơn 1,  $|B_k| > 1$ . Trong trường hợp có nghiệm phức, thì các nghiệm đó phải nằm ngoài đường tròn đơn vị.

Phương trình đặc trưng:

Đối với AR (1):  $1 - \phi_1 B = 0$

Đối với AR (p):  $1 - \phi_1 B - \phi_2 B^2 - \dots - \phi_p B^p = 0$

Do đó phương trình bậc p có p nghiệm nên phương trình được viết thành:

$$(1 - \alpha_1 B)(1 - \alpha_2 B) \dots (1 - \alpha_p B) = 0 \quad (2.28)$$

Với phương trình trên điều kiện dừng tương đương với điều kiện: tất cả các  $\alpha_i, i=1,2,\dots, p$  đều nằm trong đường tròn đơn vị.

Điều kiện là điều kiện cần để biểu diễn nghiệm của  $z_t$  dựa trên các giá trị hiện tại và quá khứ của nhiễu trắng  $a_t$  và có dạng:

$$\begin{aligned} z_t &= \frac{1}{\phi(B)} a_t = \psi(B) a_t = \psi_1 a_{t-1} + \psi_2 a_{t-2} + \dots + a_t \\ &= \sum_{k=0}^{\infty} \psi_k a_{t-k} \quad |\psi_0 = 1 \end{aligned} \quad (2.29)$$

Giá trị của  $\psi_1, \psi_2, \dots$  phụ thuộc vào  $\phi_1 \dots \phi_p$ , được xác định qua phương trình:  $\psi(B)\phi(B) = 1$ . Đối với quá trình AR dừng tương ứng với  $\sum_{j=0}^{\infty} |\psi_j| < \infty$  thì giá trị của  $\psi_j$  sẽ giảm khi  $j$  tăng. Nếu  $-1 \leq \phi_i \leq 1, i = 1 \dots p$  thì trọng số đặt trên các phần dư giảm dần theo thời gian. Điều này phù hợp với thực tiễn đó là ảnh hưởng của các yếu tố cũ thì càng ít tác động lên giá trị hiện tại.

Nghiệm của  $z_t$  từ phương trình trên thu được bằng cách xác định các hệ số  $\psi_1, \psi_2 \dots$  thỏa mãn phương trình  $\psi(B)\phi(B) = 1$  và để thu được mô hình ban đầu thì nhân hai vế của phương trình với  $\psi(B)$ .

Với dữ liệu chuỗi thời gian có thể được biểu diễn bằng quá trình tự hồi qui AR (p) như trên, vấn đề chính là quyết định xem cấu trúc tự hồi qui nào phù hợp, sau đó xác định giá trị bậc p cho mô hình, sau đó xác định giá trị bậc p cho mô hình. Điều này trở nên dễ dàng hơn khi dùng hàm tự tương quan từng phần PACF như là một công cụ trợ giúp việc xác định bậc p của quá trình. Để xác định PACF của quá trình tự hồi qui bậc p, phương trình (2.26) được viết thành:

$$z_t - \sum_{k=1}^p \phi_k z_{t-k} = a_t \quad (2.30)$$

Nhân hai vế của phương trình (2.30) với  $z_{t+h}, h=1, 2, \dots$  và lấy kỳ vọng giả định rằng với kỳ vọng của  $z_t$  bằng 0, ta có:

$$E \left[ \left( z_t - \sum_{k=1}^p \phi_k z_{t-k} \right) z_{t-h} \right] = E[a_t z_{t-h}] \quad (2.31)$$



Vế trái:

$$E[z_t z_{t-h}] - E\left[\left(\sum_{k=1}^p \phi_k z_{t-k}\right) z_{t-h}\right] = \gamma_z(h) - \sum_{k=1}^p \phi_k E[z_{t-k} z_{t-h}] = \gamma_z(h) - \sum_{k=1}^p \phi_k \gamma_z(h-k)$$

Vế phải:  $E[a_t(a_{t-h} + \psi_1 a_{t-h-1} + \psi_2 a_{t-h-2} \dots)]$

Trường hợp  $h=0$ , thì  $VP = \sigma_a^2$  do  $a_t$  là nhiễu trắng nên chúng độc lập, không tương quan với nhau nên  $E(a_t, a_{t+i}) = 0$  ( $i \neq 0$ ). Ngược lại với  $h \neq 0$  với cùng lý do trên  $VP = 0$ . Vì thế, phương trình (2.31) được viết gộp lại thành:

$$\begin{cases} \gamma_z(0) - \sum_{k=1}^p \phi_k \gamma_z(h-k) = \sigma_1^2 \text{ khi } h = 0 & (2.32) \end{cases}$$

$$\begin{cases} \gamma_z(h) - \sum_{k=1}^p \phi_k \gamma_z(h-k) = 0 \text{ khi } h = 1, 2, 3 \dots & (2.33) \end{cases}$$

ở đây,  $\gamma_z(h) = \gamma_z(-h)$  phương trình trên được gọi là phương trình Yule-Walker.

PACF được tính trên cơ sở hệ số thu được khi đã tối thiểu hóa sai số bình phương trung bình  $MSE = E[(z_{t+h} - \sum_{k=1}^h a_k z_{t+h-k})^2]$ . Thiết lập đạo hàm  $a_j$  bằng 0 dẫn đến phương trình  $E[(z_{t+h} - \sum_{k=1}^h a_k z_{t+h-k}) z_{t+h-j}] = 0$ , ta có thể viết thành:

$$\gamma_z(j) - \sum_{k=1}^h a_k \gamma_z(j-k) = 0, j = 1, 2, \dots, h$$

Từ phương trình này với (2.33) thì rõ ràng là quá trình AR(p), ta có thể chọn  $a_k = \phi_k$  với  $k \leq p$  và  $a_k = 0$  với  $p < k$  để nhận được nghiệm của phương trình trên.

Tính chất của hàm tự tương quan từng phần cho quá trình tự hồi qui AR(p):

- Hàm tự tương quan từng phần PACF,  $\phi_{hh}$ , với tham biến là  $h$ , của quá trình tự hồi qui AR có giá trị khác 0 tại các trũ  $h \leq p$  và bằng 0 tại các bậc trũ  $h > p$ , bậc của quá trình tự hồi qui. Điều này cho phép ước lượng số bộ bậc  $p$  của quá trình tự hồi qui AR qua việc sử dụng hàm tự tương quan từng phần PACF.

### 2.2.2.5 Quá trình trung bình trượt MA (Moving Average)-MA(q)

Quá trình trung bình trượt là quá trình mà giá trị của nó được xác định dựa trên sự kết hợp tuyến tính giữa nhiều không quan sát được ở hiện tại và các nhiễu trong quá khứ.

#### 1. Quá trình trung bình trượt bậc 1 – MA (1)

Giả sử có chuỗi  $\{z_t\}$  là dừng và giá trị của biến ngẫu nhiên  $z_t$  được xác định bằng nhiễu hiện thời cộng với nhiễu trước đó, được biểu diễn như sau [8]:

$$z_t = a_t + \theta_1 a_{t-1} \quad (2.34)$$

Trong đó, nhiễu trắng  $a_t \sim WN(0, \sigma_a^2)$  và không tương quan với  $z_t$ . Quá trình được mô tả trên được gọi là quá trình trung bình trượt bậc 1 – MA (1).

Các đại lượng đặc trưng của quá trình MA (1):

- Kỳ vọng:  $E(z_t) = E(\theta_1 a_{t-1} + a_t) = 0$
- Phương sai:  $var(z_t) = E[(\theta_1 a_{t-1} + a_t)^2] = \theta_1^2 E(a_{t-1}^2) + 0 + E(a_t^2) = \sigma_a^2(1 + \theta_1^2)$
- Tự hiệp phương sai và tự tương quan:

$$\gamma_z(t+h, t) = \begin{cases} \sigma_a^2(1 + \theta_1^2) & h = 0 \\ \sigma_a^2 \theta_1 & h = 1 \\ 0 & |h| > 1 \end{cases}$$

$$\rho_z(h) = \begin{cases} 1 & h = 0 \\ \frac{\theta_1}{(1 + \theta_1^2)} & h = 1 \\ 0 & |h| > 1 \end{cases}$$

Quá trình MA (1) là quá trình sinh ra chuỗi dừng.

#### 2. Quá trình trung bình trượt bậc q – MA (q)

Mở rộng quá trình MA (1) đến bậc q được gọi là quá trình trung bình trượt bậc q, biểu diễn như sau [14]:

$$z_t = a_t + \theta_1 a_{t-1} + \theta_2 a_{t-2} + \dots + \theta_q a_{t-q} \quad (2.35)$$

Trong đó q là bậc trung bình trượt,  $\theta_1 \theta_2 \dots \theta_q$  là các hệ số trung bình trượt. Dùng toán tử trễ, mô hình trên được viết lại như sau:

$$z_t = \theta(B)a_t \quad (2.36)$$

Trong đó:  $\theta(B) = 1 + \theta_1 B + \theta_2 B^2 + \dots + \theta_q B^q$

Đại lượng đặc trưng của quá trình trung bình trượt [8]:

- Kỳ vọng:  $E(z_t) = 0$
- Phương sai:  $var(z_t) = var(a_t + \theta_1 a_{t-1} + \dots + \theta_q a_{t-q}) = \sigma_a^2 (1 + \theta_1^2 + \dots + \theta_q^2)$

- Tự hiệp phương sai:

$$var(z_t, z_{t-k}) = E[(a_t + \theta_1 a_{t-1} + \dots + \theta_q a_{t-q})(a_{t-k} + \theta_1 a_{t-k-1} + \dots + \theta_q a_{t-k-q})]$$

$$\gamma_k = cov(z_t, z_{t-k}) = \begin{cases} \sigma_a^2 \sum_{i=0}^{q-k} \theta_i \theta_{k+i} & k \leq q \\ 0 & k > q \end{cases}$$

$$\theta_0 = 1$$

Tính chất của hàm tự tương quan ACF cho quá trình trung bình trượt MA (q):

Hàm tự tương quan ACF,  $\rho_h$ , với tham biến là h, của quá trình trung bình trượt MA có giá trị khác 0 ở các trễ  $h \leq q$  và bằng 0 khi tại các trễ  $h > q$ , là bậc của quá trình trung bình trượt

Ngoài tính chất dừng, một tính chất khác cần có của mô hình chuỗi thời gian đó là tính khả nghịch, nó đảm bảo cho mô hình chuỗi thời gian có thể hiểu được dễ dàng và cho kết quả dự báo đáng tin cậy. Một chuỗi thời gian được gọi là khả nghịch nếu có thể tái hiện các nhiễu  $a_t$ , qua các giá trị hiện tại và quá khứ  $z_t, z_{t+1} \dots$

Quá trình tự hồi qui AR(p) là quá trình khả nghịch.

Quá trình MA (1) là quá trình khả nghịch, bởi vì:

$$z_t = a_t + \theta_1 a_{t-1}$$

$$a_t = z_t - \theta_1 a_{t-1}$$

Mà lại có:

$$a_{t-1} = z_{t-1} + \theta_1 a_{t-2}$$

Thay vào trên ta được:

$$a_t = z_t - \theta_1 z_{t-1} + \theta_1^2 a_{t-2}$$

Thực hiện tương tự cuối cùng được:

$$a_t = z_t - \theta_1 z_{t-1} + \theta_1^2 z_{t-2} \dots$$

Điều kiện để quá trình MA(q) khả nghịch là tất cả các nghiệm hoặc nghiệm phức của phương trình  $\theta(B) = 1 + \theta_1 B + \theta_2 B^2 + \dots + \theta_q B^q = 0$  đều nằm ngoài đường tròn đơn vị tương đương với giá trị các hệ số  $|\theta_i| < 1, i = 1, \dots, q$ .

Khi quá trình MA(q) thỏa mãn điều kiện khả nghịch thì nhiễu được thể hiện qua  $z_t$  như sau: từ (2.35) ta có:

$$a_t = \frac{1}{\theta(B)} z_t = \pi(B) z_t = \pi_1 z_{t-1} + \pi_2 z_{t-2} + \dots = \sum_{k=0}^{\infty} \pi_k z_{t-k} \mid \pi_0=1$$

Đây là quá trình tự hồi qui vô hạn AR ( $\infty$ ). Chính thức mà nói thì điều kiện khả nghịch được thỏa mãn nếu như:  $\sum_{i=0}^{\infty} |\pi_i| < \infty$ , các giá trị của  $\pi(B)$  phụ thuộc vào  $\theta_1, \dots, \theta_q$ , được xác định dựa trên phương trình  $\theta(B)\pi(B) = 1$ .

### 2.2.2.6 Quá trình trung bình trượt tự hồi qui ARMA (p, q)

Quá trình trung bình trượt tự hồi qui là quá trình gộp của hai quá trình trung bình trượt MA(q) và AR(p), chuỗi được tạo ra từ quá trình này dựa trên sự kết hợp các giá trị quan sát và các nhiễu trong quá khứ. Quá trình này góp thêm vào lớp các mô hình tham số quan trọng để phân tích chuỗi thời gian, nó cải tiến chất lượng của dự báo tốt hơn so với từng mô hình rẽ nhờ sự kết hợp các ưu điểm của mỗi mô hình trong nó.

Quá trình ARMA (p, q) cho chuỗi dừng  $\{z_t\}$  được biểu diễn như sau [14]:

$$(1 - \phi_1 B - \phi_2 B^2 - \dots - \phi_p B^p) z_t = (1 + \theta_1 B + \theta_2 B^2 + \dots + \theta_q B^q) a_t \quad (2.37)$$

Tương đương:  $\phi(B) z_t = \theta(B) a_t$

Trong đó:  $a_t$  là nhiễu trắng  $\sim WN(0, \sigma_a^2)$ ,  $\phi_1, \dots, \phi_p$  là p hệ số tự hồi qui,  $\theta_1, \dots, \theta_q$  là q hệ số trung bình trượt, đa thức  $\phi(B) = 1 - \phi_1 B - \phi_2 B^2 - \dots - \phi_p B^p$  không có thừa số chung với đa thức  $\theta(B) = 1 + \theta_1 B + \theta_2 B^2 + \dots + \theta_q B^q$ .

Nếu p=0 thì quá trình ARMA (p, q) sẽ trở thành quá trình MA(q).

Nếu q=0 thì quá trình ARMA (p, q) sẽ trở thành quá trình AR(p).

Điều kiện để phương trình (2.37) tồn tại nghiệm dừng và duy nhất khi và chỉ khi:  $\Phi(B) = 1 - \phi_1 B - \phi_2 B^2 - \dots - \phi_p B^p \neq 0$

Điều kiện để ARMA (p, q) là dừng nếu tất cả nghiệm của  $\Phi(B)z_t = 0$  nằm ngoài đường tròn đơn vị, tương ứng với hệ số hồi qui  $|\phi_i| < 1$ .

Điều kiện để ARMA (p, q) là khả nghịch nếu tất cả nghiệm của  $\theta(B) = 0$  nằm ngoài đường tròn đơn vị, ứng với hệ số trung bình trượt  $|\theta_i| < 1$ .

Nhận dạng quá trình ARMA (p, q) qua việc sử dụng tương quan đồ của các ACF và PACF được tóm lược trong bảng dưới đây:

**Bảng 2.1: Đặc trưng ACF và PACF trong các mô hình tham số**

Quá trình	ACF	PACF
Quá trình tự hồi qui AR(p)-ARMA(p,0)	Suy giảm theo đồ thị hàm mũ hoặc hình sin	$\hat{\Phi}_{kk} \neq 0, k \leq p$ $\hat{\Phi}_{kk} = 0, k > p$
Quá trình trung bình trượt MA(q)-ARMA (0, q)	$\hat{\rho}_k \neq 0, k \leq p$ $\hat{\rho}_k = 0, k > p$	Suy giảm đồ thị hàm mũ hoặc hình sin
Quá trình ARMA (p, q)	$\hat{\rho}_k \neq 0, k \leq p$ Sau đó suy giảm theo đồ thị hàm mũ hoặc sin	$\hat{\Phi}_{kk} \neq 0, k \leq p$ Sau đó giảm theo đồ thị hàm mũ hoặc sin

### 2.2.2.7 Mô hình tích hợp trung bình trượt tự hồi qui ARIMA (p, d, q)

Một trong những mô hình phân tích chuỗi thời gian (Time series) phổ biến nhất và thường xuyên sử dụng là ARIMA (Auto-Regressive Integrated Moving Average) – tự hồi qui tích hợp trung bình trượt. Mô hình này là mô hình dự báo định lượng theo chuỗi thời gian, giá trị tương lai của yếu tố dự báo sẽ phụ thuộc vào quy luật vận động của chính yếu tố đó. Mô hình này không chỉ xem xét mối quan hệ trong quá khứ với hiện tại của yếu tố dự báo mà còn xem xét tác động từ các chuỗi thời gian khác đến yếu tố dự báo. Mô hình này cũng tương tự như mô hình hồi qui tuyến tính, nhưng bản chất có sự khác biệt là các hệ số hồi qui của mô hình ARIMA được xác định theo tiêu chuẩn hội tụ, sai số còn lại chính là thành phần ngẫu nhiên.

Mô hình ARIMA cơ bản tích hợp của 3 thành phần AR (thành phần tự hồi qui), I (tính dừng của chuỗi thời gian) và MA (thành phần của trung bình trượt) [1],[14].

### 1. Quá trình tích hợp

Các quá trình nêu trên chỉ được áp dụng khi chuỗi thời gian dừng. Tuy nhiên, chuỗi thời gian trong thực tế rất hiếm khi dừng. Vì thế, sự cần thiết phải biến đổi chuỗi trở thành dừng trước khi xây dựng mô hình trên nó. Phương pháp chủ yếu thực hiện điều đó là sai phân, chuỗi sau khi được biến đổi thành chuỗi dừng được gọi là chuỗi tích hợp (Integration).

Nếu chuỗi  $\{z_t\}$  ban đầu được kiểm định là không dừng, thì tiến hành sai phân bậc 1 để biến đổi về chuỗi dừng và nếu chuỗi sai phân bậc 1 ( $\Delta z_t$ ) là dừng thì được gọi là chuỗi tích hợp bậc 1, ký hiệu là I (1). Nếu chuỗi sai phân bậc 1 vẫn chưa dừng thì tiếp tục sai phân bậc 2 ( $\Delta^2 z_t$ ) gọi là chuỗi tích hợp bậc 2, ký hiệu là I (2). Có thể sai phân đến bậc d để chuỗi cuối cùng thu được là dừng, gọi là chuỗi tích hợp bậc d, I (d).

### 2. Mô hình tích hợp trung bình trượt tự hồi qui ARIMA (p, d, q)

Mô hình tích hợp trung bình trượt tự hồi qui là phương pháp để phân tích chuỗi thời gian được Box-Jenkins đề xuất đã mở ra một trang mới trong các công cụ dự báo. Mô hình ARIMA dựa trên ý tưởng rằng chuỗi thời có thể giải thích bằng cách kết hợp các hành vi hiện tại và trong quá khứ với các yếu tố ngẫu nhiên không giải thích được ở hiện tại và quá khứ, gọi là các nhiễu. Nói một cách khác, mô hình ARIMA là mô hình gộp của các mô hình AR, I và MA.

Mô hình gộp ARIMA (p, d, q) chứa đựng mô hình tự hồi qui AR (p), mô hình trung bình trượt MA (q), mô hình tích hợp I (d). Được đặc tả như sau [8]:

$$\begin{aligned} (1 - \phi_1 B - \phi_2 B^2 - \dots - \phi_p B^p)(1 - B)^d z_t & \quad (2.38) \\ & = (1 + \theta_1 B + \theta_2 B^2 + \dots + \theta_q B^q) a_t \end{aligned}$$

Hoặc viết thu gọn lại là:  $\Phi(B)\Delta^d z_t = \theta(B)a_t$

Trong đó:

- $a_t$ : là nhiễu trắng  $\sim WN(0, \sigma_a^2)$ .

- $d$  là bậc sai phân cần thiết để chuỗi dừng ( $d \geq 0$ ).
- Giá trị  $p$  là bậc tự hồi qui được xác định dựa trên tự tương quan từng phần PACF và  $\phi_1, \dots, \phi_p$  là các hệ số tự hồi qui.
  - Giá trị  $q$  là bậc trung bình trượt được xác định dựa trên sự tương quan ACF và  $\theta_1, \dots, \theta_q$  là các hệ số trung bình trượt.

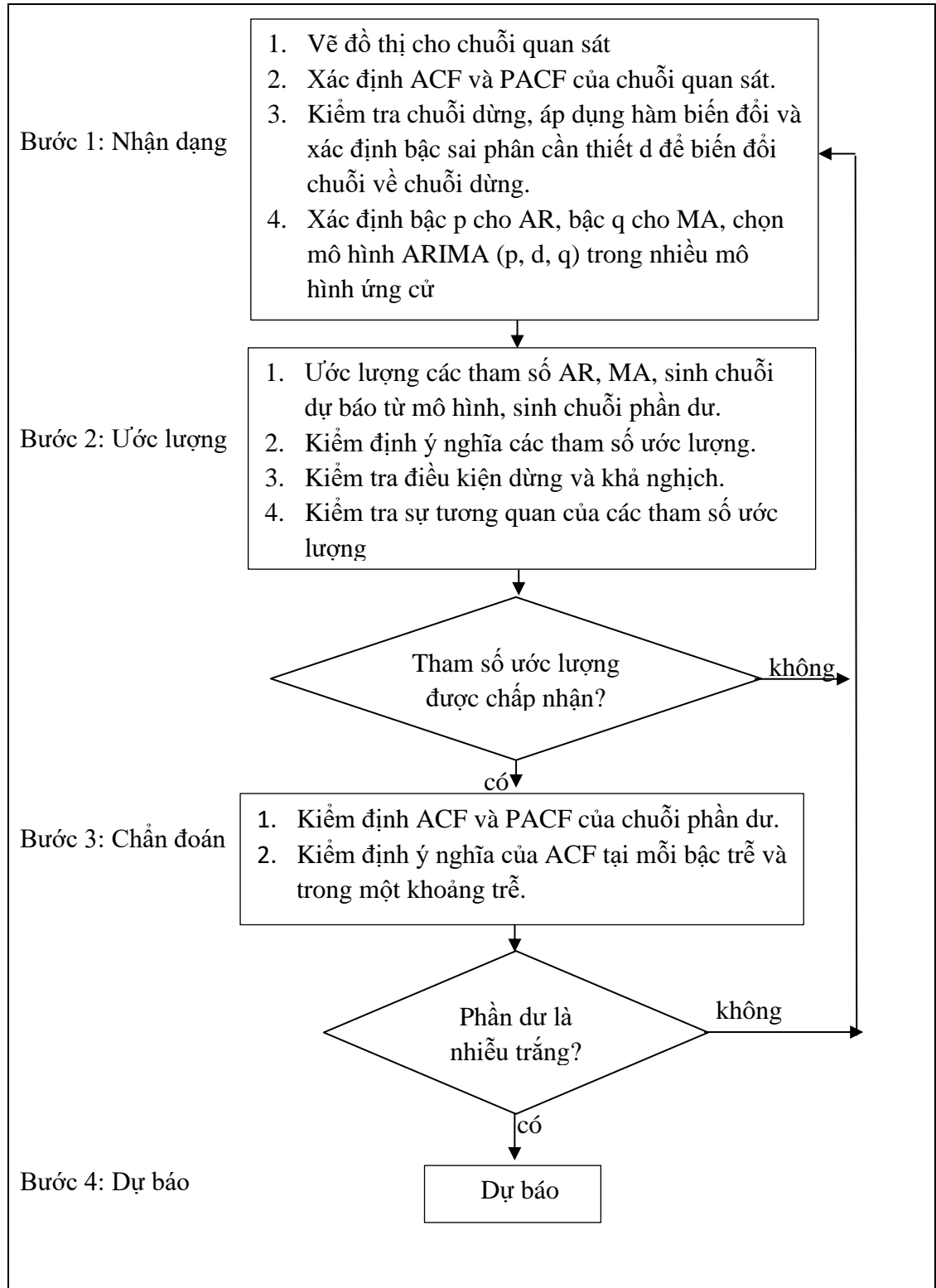
Nếu mô hình ARIMA có dạng ARIMA (0, 0, q) thì đó là quá trình MA(q).

Nếu có dạng ARIMA (p, 0, 0) thì đó là quá trình AR(p).

Nếu có dạng ARIMA (0, d, 0) thì đó là quá trình tích hợp I(d).

Nếu có dạng ARIMA (p, 0, q) thì đó là quá trình ARMA (p, q).

### 2.2.2.8 Quy trình xây dựng mô hình ARIMA (p, d, q)



**Hình 2.1: Các bước xây dựng mô hình ARIMA**



### 1. Nhận dạng từ mô hình (Model Identification)

Nhận dạng mô hình là giai đoạn lựa chọn một mô hình được cho là tốt nhất trong nhiều mô hình ARIMA ( $p, d, q$ ) ứng cử được xây dựng từ chuỗi dữ liệu quan sát. Rõ ràng là có nhiều cách kết hợp bậc  $p, d, q$  tạo ra lớp các mô hình ARIMA ( $p, d, q$ ) khác nhau. Do vậy, pha nhận dạng mô hình bao gồm việc chỉ định bậc tự hồi qui AR( $p$ ), bậc tích hợp I( $d$ ), bậc trung bình trượt MA( $q$ ). Nhưng may mắn là trong thực tế, kinh nghiệm đã chỉ ra rằng các mô hình ARIMA được cho là phù hợp với chuỗi quan sát hiếm khi các giá trị  $p, d, q$  đều đồng thời lớn hơn 2, nên tập các mô hình để lựa chọn cũng không nhiều.

Công cụ chủ yếu để nhận dạng mô hình là tương quan đồ của tự tương quan ACF và tự tương quan từng phần PACF được xây dựng từ chuỗi quan sát. Trong tương quan đồ, khoảng tin cậy ước lượng 95% cũng được vẽ tại mỗi bậc trễ nhằm kiểm định có hay không ACF và PACF tại mỗi bậc trễ khác 0 có nghĩa. Do vậy, có thể biết rằng có hay không sự tương quan trong chuỗi.

Xác định bậc sai phân  $d$  cho chuỗi thì cần phải kiểm tra chuỗi thời gian có phải là dừng không? Nếu không dừng, thì tiến hành sai phân hoặc sử dụng hàm biến đổi để chuyển thành chuỗi dừng trước khi thực hiện phân tích. Bước đầu tiên là xem xét trên đồ thị của chuỗi dữ liệu quan sát. Nó cung cấp những đánh giá sơ lược ban đầu về chuỗi dữ liệu quan sát như: sự biến thiên của chuỗi theo thời gian, tính chu kỳ, mùa vụ trong chuỗi...đồng thời cũng tính giá trị của các đại lượng đặc trưng của chuỗi quan sát như kỳ vọng mẫu, phương sai mẫu, tự hiệp phương sai mẫu. Trên cơ sở đó quyết định thực hiện các phép biến đổi sơ bộ để làm trơn chuỗi dữ liệu, như sử dụng hàm biến đổi (ví dụ hàm log) để khiến cho sự dao động trong chuỗi là ổn định, hoặc là thực hiện phép trừ các giá trị của chuỗi cho kỳ vọng mẫu của nó để chuỗi được phân tích có kỳ vọng mẫu bằng 0. Sau đó, kiểm tra hàm tự tương quan mẫu ACF. Nếu ACF giảm rất chậm theo đồ thị hàm mũ, tức là có sự tự tương quan trong chuỗi, thì tiến hành sai phân để loại bỏ sự tự tương quan. Và một cách cuối cùng rất hữu hiệu để kiểm định chuỗi dừng là dung kiểm định nghiệm đơn vị DF,

nếu giá trị thống kê DF được tính lớn hơn các giá trị tới hạn cho trước thì chúng tỏ trong chuỗi có tồn tại nghiệm đơn vị do vậy chuỗi không dừng.

Xét xét tương quan đồ ACF và PACF của chuỗi dừng để xác định bậc tự hồi qui  $p$  và bậc trung bình trượt  $q$ . Nếu ACF bằng 0 sau trễ  $q$  thì chọn  $q$  là bậc trung bình trượt MA. Nếu PACF bằng 0 sau trễ thì chọn  $p$  là bậc tự hồi qui AR.

Không chỉ dựa vào các đặc điểm nêu trên để lựa chọn giá trị cho các tham số  $p, d, q$  của mô hình, mà nhiều khi việc lựa chọn mô hình phù hợp còn là một nghệ thuật nó phụ thuộc phần lớn vào kinh nghiệm cũng như tri thức sâu rộng trong nhiều lĩnh vực khác nhau của người lập trình. Vì thế, chỉ với một chuỗi quan sát nhưng có nhiều tham số  $p, d, q$  khác nhau được lựa chọn, dẫn đến có một lớp các mô hình ARIMA khác nhau và nhiệm vụ của người lập mô hình là phải tìm được mô hình tốt nhất trong nhiều hình ứng cử.

## 2. Ước lượng mô hình (Model Estimation)

Sau khi chọn được mô hình ứng cử, bước tiếp theo là ước lượng các tham số tự hồi qui  $\phi_1, \phi_2 \dots \phi_p$  cho mô hình tự hồi qui AR( $p$ ) và các tham số trung bình trượt  $\theta_1, \theta_2 \dots \theta_q$  cho mô hình trung bình trượt MA( $q$ ) trong mô hình gộp ARIMA, dựa trên chuỗi dữ liệu quan sát. Sử dụng thuật toán bình phương nhỏ nhất phi tuyến tính (NLS-Nonlinear Least Square) để ước lượng giá trị các tham số này. Chuỗi phần dư  $\hat{a}_t$  thu được tính bằng cách trừ giá trị của chuỗi dữ liệu quan sát  $z_t$  cho giá trị được tính từ mô hình  $\hat{z}_t$ , được gọi là giá trị phù hợp từ mô hình. Thuật toán NLS ước lượng các tham số  $\hat{\phi}_i$  và  $\hat{\theta}_j$  là giá trị thỏa mãn tổng bình phương các phần dư là nhỏ nhất [8].

$$\sum_{t=1}^n \hat{a}_t^2 = \sum_{t=1}^n (z_t - \hat{z}_t)^2 \rightarrow Min \quad (2.39)$$

Việc quan trọng là tìm hiểu xem các giá trị phù hợp  $\hat{z}_t$  được tạo ra như thế nào từ phần dư  $\hat{a}_t$ , biểu diễn trong ngữ cảnh ARIMA là gì? Các giá trị được tính dựa trên mô hình bằng cách gán các ước lượng ban đầu  $\hat{\phi}$  và  $\hat{\theta}$  cho các tham số tự hồi qui  $\phi$  và trung bình trượt  $\theta$ , đặt nhiều ngẫu nhiên  $a_t$  bằng giá trị kỳ vọng của nó

là 0. Sử dụng mô hình ước lượng cùng với các dữ liệu quan sát từng bước sinh ra chuỗi các giá trị phù hợp  $\hat{z}_t$ . Thủ tục này được lặp đi lặp lại và điều chỉnh các tham số ước lượng tại mỗi vòng lặp cho đến khi thu được tổng bình phương phần dư là nhỏ nhất. Sau đó tiến hành trừ giá trị quan sát  $z_t$  cho các giá trị phù hợp  $\hat{z}_t$ , thu được phần dư  $\hat{a}_t$ .

Đây là phương pháp ước lượng thống kê thu được nhiễu ngẫu nhiên, không quan sát được, đó là phần dư  $\hat{a}_t$ . Lưu ý, trong ngữ cảnh của mô hình ARIMA thì các phần dư không thể đều bằng 0 (gọi là “phù hợp tuyệt đối” với số liệu quan sát) do chúng là ước lượng của nhiễu mà có kỳ vọng bằng 0 nhưng cường độ thay đổi ngẫu nhiên. Thay vì thế, các phần dư thu được từ mô hình ARIMA được coi là phù hợp, cường độ của các nhiễu xuất hiện trong các quan sát là không giải thích được bằng mô hình.

### 3. Chẩn đoán mô hình (Model Diagnostic)

Pha ước lượng mô hình và chẩn đoán thường được thực hiện gộp nhau, trong pha ước lượng có thể thực hiện chẩn đoán luôn. Kiểm định tính hợp lý của mô hình là kiểm tra sự phù hợp đối với các tham số ước lượng trong mô hình và phần dư của nó có phải là nhiễu trắng hay không.

#### a. Chẩn đoán tham số ước lượng

Tham số ước lượng được chấp nhận khi thỏa mãn các điều kiện dưới đây:

- Kiểm định T được dùng để kiểm định ý nghĩa thống kê của mỗi tham số ước lượng trong mô hình. Nếu qua kiểm định mà bất kỳ tham số nào bằng 0 có nghĩa thì sẽ bị loại ra khỏi mô hình. Kiểm định F cũng được dùng để kiểm định sự đồng thời bằng 0 của các tham số ước lượng trong mô hình.

- Với mô hình ARIMA, các tham số ước lượng phải thỏa mãn điều kiện dừng cho tham số tự hồi qui và điều kiện khả nghịch cho tham số trung bình trượt. Trong mô hình ước lượng, giá trị tại các tham số ước lượng tự hồi qui, trung bình trượt thường phải giảm dần tại các trễ cách xa điểm dự báo. Điều này thể hiện trong thực tế là ảnh hưởng của các điều kiện, biến động của các hành vi trong chuỗi thời gian càng ở xa trong quá khứ thì càng ít tác động đến giá trị dự báo ở thời điểm hiện tại.

-Kiểm định khoảng tin cậy của các tham số ước lượng. Do các tham số trong mô hình đều là các tham số ước lượng, không phải giá trị tham số thực. Vì thế, cần phải ước lượng khoảng tin cậy mà tham số thực sẽ phụ thuộc vào. Nếu khoảng tin cậy càng nhỏ thì mô hình ước lượng càng đáng tin cậy và ngược lại.

-Ngoài ra sử dụng tiêu chuẩn thông tin Akaike Information Criterion-AIC, Schwarz Information Criterion-SIC, để lựa chọn mô hình hợp lý nhất (theo nghĩa là chúng cùng phù hợp như nhau và đưa ra kết quả dự báo giống nhau, nhưng lại chứa các tham số ít hơn) trong số nhiều mô hình ứng cử.

Nếu các tham số ước lượng của mô hình đều có ý nghĩa thống kê, thỏa mãn điều kiện dừng và khả nghịch hơn nữa mô hình lựa chọn là hợp lý. Khi đó tiếp tục kiểm định sự tương quan trong chuỗi phần dư sinh ra từ mô hình. Nếu không thì quay trở lại giai đoạn đầu, nhận dạng mô hình, để lựa chọn một mô hình khác phù hợp hơn.

### **b.Chẩn đoán phần dư**

Ý tưởng của việc xây dựng mô hình ARIMA là nắm bắt được các cấu trúc tự tương quan trong chuỗi dữ liệu quan sát và trên cơ sở đó xây dựng mô hình phù hợp nhất với chuỗi dữ liệu quan sát. Khi mô hình được cho là phù hợp thì nó phải thể hiện, giải thích được các cấu trúc này và sinh ra chuỗi phần dư không chứa cấu trúc tự tương quan. Do vậy, giai đoạn chẩn đoán phần dư là xem xét trong chuỗi phần dư thu được có còn cấu trúc tự tương quan nào không? Chọn lựa một trong các phương pháp sau để kiểm định sự tự tương quan của chuỗi phần dư ước lượng:

-Theo định nghĩa, chuỗi nhiễu trắng là chuỗi không chứa sự tự tương quan có ý nghĩa tại bất kỳ một trễ nào, tức là hàm tự tương quan ACF của nó bằng 0 tại mọi độ trễ. Do vậy, chỉ cần kiểm định chuỗi phần dư có phải là chuỗi nhiễu trắng hay không? Tức là xem xét các ACF của nó bằng 0 hay không. Trong thực tế chỉ cần 95% ACF bằng 0 một cách có ý nghĩa thì có thể kết luận chuỗi phần dư là nhiễu trắng.

-Dùng phương pháp kiểm định Durbin-Walson kiểm tra tương quan bậc I của chuỗi phần dư bằng mô hình tự hồi qui bậc 1, hoặc dùng kiểm định Breusch-Godfrey LM áp dụng cho tự tương quan bậc cao hơn.

-Kiểm định Q hoặc  $Q_{LB}$  được sử dụng để kiểm định giả thuyết chuỗi phần dư là nhiễu trắng.

Nếu trong chuỗi phần dư có sự tự tương quan một cách có ý nghĩa tại một vài độ trễ thì quay lại bước 1, nhận dạng mô hình, để chọn một mô hình khác phù hợp hơn. Do vậy, xây dựng mô hình ARIMA là một quá trình lặp cho đến khi tìm được mô hình thích hợp nhất.

#### 4. Dự báo (Forecasting)

Khi mô hình ARIMA phù hợp cho chuỗi dữ liệu quan sát được xác định thì có thể dùng để dự báo giá trị của chuỗi và khoảng tin cậy mà giá trị quan sát thực sẽ nằm trong. Mô hình ARIMA được rất nhiều nhà kinh tế sử dụng để đưa ra những dự báo ngắn hạn.

Xét mô hình ARMA (p, q) tổng quát sau [8]:

$$\phi(B)z_t = \theta(B)a_t$$

Viết cụ thể:

$$z_t = \phi_1 z_{t-1} + \dots + \phi_p z_{t-p} + a_t + \theta_1 a_{t-1} + \dots + \theta_q a_{t-q}$$

Dự báo giá trị từ mô hình tại thời điểm t+h, thay thế chỉ số thời gian t bằng chỉ số thời gian t+h, thu được:

$$z_{t+h} = \phi_1 z_{t+h-1} + \dots + \phi_p z_{t+h-p} + a_{t+h} + \theta_1 a_{t+h-1} + \dots + \theta_q a_{t+h-q} \quad (2.40)$$

Phương trình (2.40), nếu thời điểm hiện tại là t khi đó giá trị dự báo cho t+h là  $z_{t+h}$  sẽ được tạo ra với p quan sát trước đó và q nhiễu ngẫu nhiên trước đó, cộng với nhiễu tại thời điểm t+h. Tại giai đoạn ước lượng thu được các tham số ước lượng tự hồi qui  $\hat{\phi}_i$ , trung bình trượt  $\hat{\theta}_j$  và phần dư  $\hat{a}_t$ . Do đó, giá trị của nhiễu tại một số trễ được xác định, ngoài trừ các nhiễu  $a_{t+1}, \dots, a_{t+h}$  là chưa biết. Để đơn giản chúng được gán bằng kỳ vọng của nó là 0. Các giá trị  $z_{t+1}, \dots, z_{t+h+1}$  được thay thế bằng các giá trị dự báo của nó  $\hat{z}_t(1), \dots, \hat{z}_t(h-1)$ . Thay thế tất cả vào phương trình

(2.40), thu được giá trị dự báo theo phương thức phần dư bình phương trung bình nhỏ nhất cho  $z_{t+h}$ , như sau [8]:

$$\hat{z}_t(h) = \hat{\varphi}_1 \hat{z}_{t+h-1} + \dots + \hat{\varphi}_p \hat{z}_{t+h-p} - \hat{\theta}_k \hat{a}_{t+h-k} - \dots - \hat{\theta}_q \hat{a}_{t+h-q} \quad (2.41)$$

Biểu thức (2.41) được gọi là hàm dự báo kết hợp với mô hình ARIMA. Dự báo các giá trị tương lai của chuỗi dữ liệu quan sát được tiến hành theo một trong các cách sau:

**a. Dự báo trước nhiều bước ngoài mẫu (Multi-Step-Ahead Beyond-Sample)**

Mô hình ARMA (p, q) cho ở trên được viết dưới dạng mô hình trung bình trượt MA ( $\infty$ ) [8]:

$$z_t = \frac{\theta(B)}{\varphi(B)} a_t = \psi_\infty(B) a_t \quad (2.42)$$

Hay:  $z_t = a_t + \psi_1 a_{t-1} + \psi_2 a_{t-2} + \psi_3 a_{t-3} + \dots$

Trong đó:  $\psi_1, \psi_2, \psi_3, \dots$  xác định từ phương trình:

$$\psi(B)\varphi(B) = \theta(B)$$

Giả sử rằng chuỗi thời gian  $\{z_t\}$  có n giá trị quan sát  $z_1, z_2, \dots, z_n$ . Dự báo giá trị  $z_{n+h}$  (gọi là dự báo trước h bước ngoài mẫu) từ mô hình sẽ là:

$$z_{t+h} = a_{t+h} + \psi_1 a_{t+h-1} + \psi_2 a_{t+h-2} + \psi_3 a_{t+h-3} + \dots$$

dự báo giá trị  $z_{t+h}$  theo phương thức phần dư bình phương trung bình tối thiểu sẽ là:

$$\hat{z}_t(h) = \psi_1 [a_{t+h-1}] + \psi_2 [a_{t+h-2}] + \psi_3 [a_{t+h-3}] + \dots$$

-Dự báo trước một bước ngoài mẫu (h=1):

$$z_{t+1} = a_{t+1} + \psi_1 a_t + \psi_2 a_{t-1} + \psi_3 a_{t-2} + \dots$$

$$\hat{z}_t(1) = 0 + \psi_1 a_t + \psi_2 a_{t-1} + \psi_3 a_{t-2} + \dots$$

Độ lệch:  $\hat{a}_t(1) = z_{t+1} - \hat{z}_t(1) = a_{t+1}$

Phương sai độ lệch:  $var(\hat{a}_t(1)) = var(a_{t+1}) = \sigma_a^2$

Khoảng tin cậy dự báo:  $\hat{z}_t(1) \pm 1.96 \sqrt{var(\hat{a}_t(1))} = \hat{z}_t(1) \pm 1.96 \sigma_a$

-Dự báo trước hai bước ngoài mẫu (h=2)

$$z_{t+2} = a_{t+2} + \psi_1 a_{t+1} + \psi_2 a_t + \psi_3 a_{t-1} + \dots$$

$$\hat{z}_t(2) = 0 + 0 + \psi_2 a_t + \psi_3 a_{t-1} + \psi_4 a_{t-2} + \dots$$

$$\text{Độ lệch: } \hat{a}_t(2) = z_{t+2} - \hat{z}_t(2) = a_{t+2} + \psi_1 a_{t+1}$$

$$\text{Phương sai độ lệch: } \text{var}(\hat{a}_t(2)) = \text{var}(a_{t+2} + \psi_1 a_{t+1}) = \sigma_a^2(1 + \psi_1^2)$$

$$\text{Khoảng tin cậy dự báo: } \hat{z}_t(2) \pm 1.96\sqrt{\sigma_a^2(1 + \psi_1^2)}$$

-Dự báo trước h bước ngoài mẫu:

$$\hat{z}_t(h) = 0 + 0 + \dots + \psi_k a_t + \psi_{k+1} a_{t-1} \dots$$

$$\text{Độ lệch: } \hat{a}_t(h) = z_{t+h} - \hat{z}_t(h) = a_{t+h} + \psi_1 a_{t+h-1} + \dots + \psi_{h-1} a_{t+1}$$

$$\text{Phương sai độ lệch: } \text{var}(\hat{a}_t(h)) = \sigma_a^2(1 + \psi_1^2 + \dots + \psi_{h-1}^2)$$

Dùng cách dự báo trên để thực hiện dự báo cho các giá trị vượt ra ngoài quan sát cuối cùng trong chuỗi dữ liệu quan sát và giả định rằng các quan sát trong tương lai chưa xuất hiện. Bằng cách lặp lại đệ quy quá trình này, có thể thực hiện bao nhiêu dự báo ngoài mẫu tùy ý. Tuy nhiên, điều đó đồng nghĩa với sai số của dự báo sẽ tăng nhanh chỉ sau khi thực hiện dự báo một vài lần.

### **b. Dự báo trước một bước ngoài mẫu (One-Step-Ahead Beyond-Sample)**

Do sai số của dự báo trước nhiều bước ngoài mẫu tăng nhanh khi dự báo cho giá trị ngoài mẫu ở thời điểm cách xa so với thời điểm cuối cùng của mẫu, nên có lẽ là thích hợp hơn khi chỉ thực hiện dự báo trước một bước ngoài mẫu đồng thời khi có các giá trị quan sát mới xuất hiện thì nó được tích hợp ngay vào mẫu.

Giả sử thời điểm hiện tại là  $t$ , giá trị quan sát gần nhất là  $z_t$ . Thực hiện dự báo trước một bước ngoài mẫu  $\hat{z}_t(1)$  cho giá trị kế tiếp  $z_{t+1}$ . Khi đó, việc phân tích sẽ đợi cho đến thời điểm  $t+1$  để xác định được giá trị quan sát thực sự  $z_{t+1}$  nó sẽ khác đôi chút so với giá trị dự báo  $\hat{z}_t(1)$ . Phần dư chính là ước lượng của ngẫu nhiên  $a_{t+1} = z_{t+1} - \hat{z}_t(1)$ .

Nhập giá trị quan sát mới  $z_{t+1}$  vào mẫu dữ liệu quan sát ban đầu và giá trị nhiễu  $a_{t+1}$  cũng đã xác định. Thực hiện tương tự hoàn toàn có thể dự báo trước một bước khác  $\hat{z}_t(2)$  cho  $z_{t+2}$ . Dự báo theo cách này nhờ việc tích hợp thêm các quan sát thực hiện vào chuỗi ban đầu và các nhiễu được ước lượng, do đó thường cho kết

quả dự báo tốt hơn so với dự báo trước nhiều bước bởi vì không tận dụng thông tin mới bổ sung.

Kiểu dự báo này liên tục cập nhật thêm các giá trị quan sát mới và các nhiễu ngẫu nhiên, nó được gọi là dự báo tích hợp trước một bước ngoài mẫu. Có khả năng phải ước lượng lại các tham số của mô hình sau khi có thêm những giá trị quan sát mới, nhưng điều đó thường không cần thiết bởi vì các giá trị của tham số ước lượng nói chung là không thay đổi một cách có ý nghĩa so với ước lượng từ chuỗi quan sát ban đầu. Chỉ tiến hành ước lượng lại tham số cho mô hình sau khi đã ghép thêm một số đáng kể các quan sát mới vào chuỗi quan sát ban đầu.

Sau khi dự báo được giá trị của chuỗi, phải thực hiện biến đổi ngược giá trị dự báo để đồng nhất với các giá trị quan sát ban đầu. Ví dụ, giả định chuỗi quan sát ban đầu  $z_t$  được biến đổi qua hàm log thành chuỗi dừng, sau đó thực hiện phân tích, xây dựng mô hình ... thì giá trị của dự báo, chẳng hạn  $\hat{z}_t(1)$ , phải được biến đổi ngược lại thành hàm mũ, chẳng hạn  $\widehat{EXP}(z_t(1))$  để đồng nhất với giá trị quan sát ban đầu.

### c. Đánh giá dự báo và khoảng tin cậy của dự báo

Để có cơ sở ra quyết định cần phải đánh giá dự báo. Một cách đánh giá lại dự báo là chúng ta để lại một vài giá trị cuối cùng của mẫu. Sau đó dựa vào một số tiêu chuẩn để đánh giá kết quả dự báo.

Giả sử chúng ta đã dự báo cho các thời kỳ (quan sát)  $n+1, n+2 \dots n+p$

- Sai số dự báo trung bình [8]:

$$\frac{1}{p} \sum_{j=1}^p \hat{a}_{t+j} \quad (2.43)$$

- Tổng bình phương các sai số số dự báo [8]:

$$\sum_{j=1}^p (\hat{a}_{t+j})^2 \quad (2.44)$$

- Căn bậc hai sai số bình phương trung bình [8]:



$$\sqrt{\frac{\sum_{j=1}^p (\hat{a}_{t+j})^2}{p}} \quad (2.45)$$

- Sai số tuyệt đối trung bình [8]:

$$\frac{1}{p} \sum_{j=1}^p |\hat{a}_{t+j}| \quad (2.46)$$

- Phần trăm sai số tuyệt đối [8]:

$$\frac{|\hat{z}_t - z_t|}{z_t} * 100\% \quad (2.47)$$

Nếu mô hình được lựa chọn mà có các sai số nói trên càng nhỏ thì mô hình càng đáng tin cậy, kết quả dự báo có sự chênh lệch rất nhỏ so với giá trị quan sát thực. Nó có thể được sử dụng để dự báo cho giá trị của chuỗi.

Với giá trị  $\alpha$  cho trước, với khoảng tin cậy là  $(1 - \alpha)100\%$  thì giá trị quan sát thực  $z_t$  sẽ nằm trong khoảng [8]:

$$z_{t^*} \in [\hat{z}_{t^*} - T_{\alpha/2}(n - k)\hat{\sigma}_a; \hat{z}_{t^*} + T_{\alpha/2}(n - k)\hat{\sigma}_a] \quad (2.48)$$

Trong đó,  $n$  là số các quan sát,  $\hat{z}_{t^*}$  là giá trị dự báo,  $k$  là số các tham số ước lượng của mô hình,  $\hat{\sigma}_a$  là độ lệch chuẩn của sai số ước lượng.

**Kết luận:** Như vậy, trong chương 2 này chủ yếu giới thiệu các yêu cầu quản lý dân số, và các mô hình ARIMA thường. Một cách tổng quát chuỗi dữ liệu quan sát thực được giả định đầu ra (Output) của một quá trình tuyến tính mà đầu vào (Input). Mô hình ARIMA là một quá trình tuyến tính, nó chứa thành phần thứ nhất là thành phần tích hợp I, thành phần này ám chỉ việc sử dụng phương pháp sai phân theo trễ thường thành chuỗi dừng trước khi tiến hành phân tích; thành phần thứ hai trong mô hình là các mô hình con tự hồi qui không theo mùa vụ AR(p), các mô hình con này dự báo giá trị cho chuỗi dựa trên giá trị trong quá khứ của nó; thành phần cuối cùng trong mô hình là các mô hình con trung bình trượt không theo mùa vụ MA(q), các mô hình này dự báo giá trị của chuỗi dựa trên việc kết hợp các nhiễu trong quá khứ.

Một cách tổng quát mô hình ARIMA dự báo giá trị cho chuỗi mùa vụ hoặc không mùa vụ dựa trên kết hợp các giá trị trong quá khứ và các nhiễu trong quá

khứ. Các kết quả dự báo trên mô hình ARIMA thường có độ chính xác cao do kết hợp nhiều mô hình con trong một mô hình, nó giảm bớt các khiếm khuyết trong dự báo của từng mô hình con và tăng cường ưu việt của từng mô hình con.

Bốn bước cụ thể để xây dựng mô hình ARIMA là nhận dạng mô hình, ước lượng, chẩn đoán và dự báo. Bước nhận dạng mô hình chủ yếu dựa trên ACF và PACF, bước ước lượng là dùng thuật toán NLS ước lượng các tham số và phần dư, sau đó kiểm định và cuối cùng là dự báo giá trị trước nhiều bước hoặc một bước ngoài mẫu. Nói chung là để chọn ra được mô hình ARIMA tốt nhất thì ba bước đầu tiên phải được thực hiện lặp đi lặp lại cho đến khi chọn được mô hình thỏa mãn các kiểm định như kiểm định tham số phải có ý nghĩa, kiểm định phần dư phải là nhiễu trắng và giá trị sai số từ dự báo phải là nhỏ, độ tin cậy lớn.

Trong chương tiếp theo, xây dựng hệ thống quản lý dân số, đồng thời áp dụng mô hình ARIMA thường để thực hiện phân tích và dự báo cho chuỗi thời gian theo mùa vụ về biến động dân số tỉnh Tây Ninh.

## CHƯƠNG 3: XÂY DỰNG HỆ THỐNG QUẢN LÝ VÀ DỰ BÁO DÂN SỐ

### 3.1 Xây dựng hệ thống quản lý dân số

#### 3.1.1 Mô tả dữ liệu

Theo công văn số 96/TCDS-KHTC ngày 18 tháng 3 năm 2016 về việc hướng dẫn ghi chép ban đầu sổ A0, ghi phiếu thu tin và lập báo cáo thống kê chuyên ngành DS-KHHGD. Trong công văn nêu rõ các khái niệm liên quan, phạm vi theo dõi DS-KHHGD như hộ khẩu, nhân khẩu, tạm trú, thường trú thực tế, nhân khẩu chuyển đi – chuyển đến...; thời gian thu thập thông tin. Đồng thời kết hợp với đi khảo sát thực tế tại chi cục dân số tỉnh Tây Ninh, đã phân tích và lưu trữ thành cơ sở dữ liệu như các bảng như sau:

Cơ sở dữ liệu (CSDL) lưu trữ thông tin hộ khẩu: số hộ, mã phường/ xã quản lý, địa bàn, tên chủ hộ/ số hộ, trạng thái.

**Bảng 3.1: Lưu thông tin hộ khẩu**

STT	Tên trường dữ liệu	Kiểu dữ liệu	Kích thước tối đa	Mô tả
1	ID_HO_KHAU	Số	10	Hộ khẩu ID.
2	SO_HO_KHAU	Chuỗi	12	Số hộ khẩu.
3	ID_PHUONG_XA	Số	10	Mã phường/xã – theo danh mục chung toàn quốc.
4	ID_DIA_BAN	Số	10	Mã địa bàn.
5	TEN_GOI	Chuỗi	100	Tên gọi: thường là tên chủ hộ/ số nhà.
6	ID_TRANG_THAI	Chuỗi	5	Mã trạng thái của hộ khẩu: thêm mới (để trống); chuyển đến (ký hiệu: M); chuyển đi (ký hiệu: G); chết/mất (ký hiệu: D).
7	NGAY_BD	Chuỗi	50	Ngày bắt đầu: lưu thông tin ngày bắt đầu khi thêm

				mới/chuyển đến hộ mới.
8	NGAY_KT	Chuỗi	50	Ngày kết thúc: lưu thông tin ngày kết thúc khi hộ chuyển đi/ chết.
9	XAC_NHAN	Số	1	Trạng thái xác nhận: 0-chưa xác nhận; 1-đã xác nhận.

Trong mỗi hộ khẩu, lưu trữ thông tin của các nhân khẩu, bao gồm các thông tin cơ bản: họ tên, ngày sinh, giới tính, quan hệ với chủ hộ, tình trạng cư trú, dân tộc, tình trạng hôn nhân, trình độ học vấn, trình độ chuyên môn, trạng thái (chuyển đi/ chuyển đến...).

**Bảng 3.2: Lưu thông tin nhân khẩu**

STT	Tên trường dữ liệu	Kiểu dữ liệu	Kích thước tối đa	Mô tả
1	ID_DU_LIEU	Số	10	Mã công dân.
2	HO_TEN_DAY_DU	Chuỗi	100	Họ tên đầy đủ của công dân.
3	NGAY_SINH	Ngày	50	Ngày sinh.
4	GIOI_TINH	Số	1	Giới tính: 0-nữ; 1-Nam.
5	ID_QUAN_HE_CHU_HO	Số	5	Mã quan hệ chủ hộ: 0-chủ hộ; 1-vợ/chồng chủ hộ; 3 - bố mẹ; 4 - ông/bà; 5-cháu; 6-quan hệ khác; 21-con đẻ; 22 - con nuôi/con dâu/con rể.
6	TINH_TRANG_CU_TRU	Số	5	Tình trạng cư trú: 1 - thực tế thường trú có mặt; 2 - thực tế thường

				trú vắng mặt; 3 – tạm trú.
7	DAN_TOC	Chuỗi	5	Mã dân tộc: được quy định theo công văn số 96/TCDS-KHTC (mục 3. bảng danh mục các dân tộc Việt Nam).
8	TINH_TRANG_HO N_NHAN	Chuỗi	5	Tình trạng hôn nhân: 1 – chưa có vợ hoặc chồng; 2 – có vợ/có chồng; 4 - ly hôn; 5 – ly thân; 6 – góa.
9	TRINH_DO_HOC_V AN	Chuỗi	5	Trình độ học vấn: 0-chưa rõ; 00-không biết đọc/viết; CS-trung học cơ sở; PT-trung học phổ thông; TH-tiểu học.
10	TRINH_DO_CMKT	Chuỗi	5	Trình độ chuyên môn kỹ thuật: 00-không có chuyên môn kỹ thuật nào; A0-công nhân kỹ thuật (không có bằng hoặc chứng chỉ); A1-công nhân kỹ thuật (có bằng hoặc chứng chỉ); B-sơ cấp chuyên nghiệp; C-trung học chuyên nghiệp; D-cao đẳng; E-đại học; F-thạc sỹ, phó tiến sỹ, tiến sỹ.
11	ID_PHUONG_XA	Số	10	Mã phường/xã – theo danh mục chung toàn quốc.
12	ID_HO_KHAU	Số	10	Mã hộ khẩu

13	ID_TRANG_THAI	Chuỗi	10	Trạng thái: thêm mới (ký hiệu: N); chuyển đến (ký hiệu: M); chuyển đi (ký hiệu: G); trẻ sinh (ký hiệu: S); mất/chết (ký hiệu: D).
14	NGAY_BD	Ngày	50	Ngày bắt đầu: lưu thông tin ngày bắt đầu khi thêm mới/ chuyển đến nhân khẩu mới.
15	NGAY_KT	Ngày	50	Ngày kết thúc: lưu thông tin ngày kết thúc khi nhân khẩu chuyển đi/chết.
16	ID_NGUOI_ME	Số	5	Mã người mẹ. Trong trường hợp thêm mới sức khỏe sinh sản-người con mới được sinh ra, sẽ lưu thông tin mã người mẹ.
17	LA_CON_THU	Số	5	Là con thứ mấy. Trong trường hợp sinh con-lưu lại thông tin số con thứ mấy.
18	DIA_CHI	Chuỗi	500	Địa chỉ của nhân khẩu
19	GHI_CHU	Chuỗi	500	Thông tin ghi chú

Trong mỗi nhân khẩu, có lưu trữ thông tin về biến động nhân khẩu: loại biến động (chuyển đi/ chuyển đến, trẻ mới sinh, hôn nhân...), ngày thu thập.

**Bảng 3.3: Lưu thông tin biến động nhân khẩu**

STT	Tên trường dữ liệu	Kiểu dữ liệu	Kích thước tối đa	Mô tả
1	ID_BIEN_DONG	Số	10	Mã biến động nhân khẩu.
2	ID_DU_LIEU	Số	10	Mã nhân khẩu.
3	LOAI_BIEN_DONG	Chuỗi	10	Loại biến động: 00-sửa sai; 01-xóa do ghi thừa; 02-thêm do ghi thiếu; 13-chuyển đi; 14-chuyển đến; 16-chết; 17-trẻ mới sinh; 18-hôn nhân; 31-kế hoạch hóa gia đình; 32-sức khỏe sinh sản; 34-cá nhân mới cập nhật.
4	ID_PHUONG_XA	Số	10	Mã phường/xã – theo danh mục chung toàn quốc.
5	NGAY_THU_THAP	Ngày	50	Ngày cộng tác viên dân số đi thu thập dữ liệu.
6	NGAY_TAO	Ngày	50	Ngày nhập vô phần mềm.
7	NGAY_CHUYEN_DI	Ngày	50	Ngày nhân khẩu có biến động chuyển đi (nếu có).
8	NGAY_CHUYEN_DEN	Ngày	50	Ngày nhân khẩu có biến động chuyển đến (nếu có).
9	NGAY_CHET	Ngày	50	Ngày nhân khẩu có biến động chết (nếu có).

Trong mỗi nhân khẩu, có thông tin quản lý kế hoạch hóa gia đình cho những nhân khẩu có hôn nhân: ngày bắt đầu sử dụng biện pháp tránh thai, phương thức sử dụng biện pháp tránh thai.

**Bảng 3.4: Lưu thông tin kế hoạch hóa gia đình**

STT	Tên trường dữ liệu	Kiểu dữ liệu	Kích thước tối đa	Mô tả
1	ID_KHHGD	Số	10	Mã kế hoạch hóa gia đình.
2	ID_DU_LIEU	Số	10	Mã nhân khẩu
3	NGAY_BAT_DAU	Ngày	50	Ngày bắt đầu sử dụng biện pháp tránh thai.
4	NGAY_CAP_NHAT	Ngày	50	Ngày cập nhật lên phần mềm.
5	ID_PHUONG_XA	Số	10	Mã phường/xã – theo danh mục chung toàn quốc.
6	ID_BPTT	Chuỗi	50	Mã biện pháp tránh thai: 0-không sử dụng; 1-vòng tránh thai; 1/1-thay vòng tránh thai; 19-thôi sử dụng vòng tránh thai; 2-triệt sản nam; 3-triệt sản nữ; 4-bao cao su; 5-thuốc uống tránh thai; 6-thuốc tiêm tránh thai; 7-thuốc cấy tránh thai; 7/7-thay thuốc cấy tránh thai; 8-biện pháp khác; N-nạo hút thai-không sử dụng; S-diễn giải là sinh-không sử dụng; T-mang thai-không sử dụng.



Bảng 3.4 lưu thông tin kế hoạch hóa gia đình, còn bảng 3.5 lưu thông tin các lần sử dụng kế hoạch hóa gia đình (lịch sử)

**Bảng 3.5: Lưu thông tin lịch sử kế hoạch hóa gia đình**

STT	Tên trường dữ liệu	Kiểu dữ liệu	Kích thước tối đa	Mô tả
1	ID_KHHGD_LS	Số	10	Mã lịch sử kế hoạch hóa gia đình.
2	ID_DU_LIEU	Số	10	Mã nhân khẩu
3	NGAY_BAT_DAU	Ngày	50	Ngày bắt đầu sử dụng biện pháp tránh thai.
4	NGAY_CAP_NHAT	Ngày	50	Ngày cập nhật lên phần mềm.
5	ID_PHUONG_XA	Số	10	Mã phường/xã – theo danh mục chung toàn quốc.
6	ID_BPTT	Chuỗi	50	Mã biện pháp tránh thai được quy định theo công văn số 96 (bảng 2. Mã biện pháp tránh thai và sự kiện thai sản): 0-không sử dụng; 1-vòng tránh thai; 1/1-thay vòng tránh thai; 19-thôi sử dụng vòng tránh thai; 2-triệt sản nam; 3-triệt sản nữ; 4-bao cao su; 5-thuốc uống tránh thai; 6-thuốc tiêm tránh thai; 7-thuốc cấy tránh thai; 7/7-thay thuốc cấy tránh thai; 8-biện pháp khác; N-nạo hút thai-không sử dụng; S-diễn giải là sinh-không sử dụng; T-mang thai-không sử dụng.

Trong những nhân khẩu là nữ trong độ tuổi sinh đẻ, sẽ có thông tin quản lý về sức khỏe sinh sản: ngày thu thập, ngày sinh con, cân nặng trẻ, là con thứ mấy, sinh tại đâu, các thông tin về trẻ sinh (họ tên, giới tính, ngày sinh...).

**Bảng 3.6: Lưu thông tin sức khỏe sinh sản**

STT	Tên trường dữ liệu	Kiểu dữ liệu	Kích thước tối đa	Mô tả
1	ID_SKSS	Số	10	Mã sức khỏe sinh sản.
2	ID_DU_LIEU	Số	10	Mã nhân khẩu.
3	ID_PHUONG_XA	Số	10	Mã phường/xã – theo danh mục chung toàn quốc.
4	NGAY_THU_THAP	Ngày	50	Ngày cộng tác viên dân số đi thu thập.
5	NGAY_SINH	Ngày	50	Ngày sinh của trẻ.
6	NGAY_CAP_NHAT	Ngày	50	Ngày cập nhật lên phần mềm.
7	SINH_CON_TAI	Chuỗi	5	Sinh con ở đâu: S-tại trạm y tế; H-tại nhà; O-đẻ tại nơi khác.
8	LA_CON_THU	Số	5	Là con thứ mấy.
9	CAN_NANG	Chuỗi	10	Cân nặng lúc sinh.

Ngoài ra, cơ sở dữ liệu còn lưu trữ thông tin về cộng tác viên dân số, là những cán bộ hàng tháng đi thu thập các thông tin thay đổi về dân số lên phiếu thu tin.

**Bảng 3.7: Lưu thông tin cộng tác viên**

STT	Tên trường dữ liệu	Kiểu dữ liệu	Kích thước tối đa	Mô tả
1	ID_CONG_TAC_VIEN	Số	10	Mã cộng tác viên
2	TEN_DEM	Chuỗi	100	Tên đệm của công tác viên.
3	TEN	Chuỗi	100	Tên của công tác viên.
4	ID_PHUONG_XA	Số	10	Mã phường/xã – theo danh mục chung toàn quốc.
5	GIOI_TINH	Số	1	Giới tính: 0-nữ; 1-nam.
6	NGAY_THAM_GIA	Ngày	50	Ngày bắt đầu tham gia.

Quản lý danh mục địa bàn. Mỗi xã/ phường quản lý nhiều địa bàn, được đánh số theo thứ tự tổ, ấp/khu phố. Trên mỗi địa bàn có một hoặc nhiều cộng tác viên dân số quản lý.

**Bảng 3.8: Lưu thông tin địa bàn**

STT	Tên trường dữ liệu	Kiểu dữ liệu	Kích thước tối đa	Mô tả
1	ID_DIA_BAN	Số	10	Mã địa bàn.
2	ID_CONG_TAC_VIEN	Số	10	Mã công tác viên.
3	TEN_DIA_BAN	Chuỗi	500	Tên địa bàn.
4	ID_PHUONG_XA	Số	10	Mã phường/xã – theo danh mục chung toàn quốc.
5	DIA_CHI	Chuỗi	500	Địa chỉ của địa bàn.
6	GHI_CHU	Chuỗi	500	Ghi chú thông tin địa bàn.
7	NGAY_CAP_NHAP	Ngày	50	Ngày cập nhật trên phần mềm.

### 3.1.2 Xây dựng phần mềm quản lý

Theo kết quả khảo sát thực tế về nhu cầu sử dụng phần mềm quản lý dân số, để thuận tiện cho người sử dụng thì giao diện đề xuất như hình dưới đây:

Giao diện quản lý hộ - nhân khẩu được phân theo xã/phường và từng địa bàn. Hộ khẩu hiện thị các thông tin gồm: số hộ, SN/tên người; các hộ có gạch màu xanh là cho biết hộ đó chuyển đến; các hộ có gạch màu đỏ là cho biết hộ đó chết/mất; các hộ còn lại là được thêm mới. Các nút chức năng quản lý hộ khẩu: tìm kiếm, thêm hộ, sửa hộ, xóa hộ, in danh sách. Mỗi hộ khẩu thì có những nhân khẩu. Trong nhân khẩu hiển thị các thông tin gồm: họ tên, quan hệ, ngày sinh, dân tộc, giới tính, trình độ học vấn, trình độ chuyên môn, hôn nhân, cư trú; nhân khẩu có gạch màu xanh là cho biết nhân khẩu đó chuyển đến; nhân khẩu có gạch màu đỏ cho biết nhân khẩu đó đã chết/mất. Các nút chức năng quản lý nhân khẩu: thêm, sửa, xóa, in, in PTT (phiếu thu tin).

#### QUẢN LÝ HỘ - NHÂN KHẨU

TT chi tiết
Chuyển hộ khẩu
Tách cá nhân
Tìm kiếm
Báo cáo
Quản lý CTV
QL Biến động

**Hộ khẩu**

Địa phương: Phường Ninh Thành

Địa bàn: Địa Bàn 1, Khu Phố Ninh Phước

Tìm

Số hộ	SN/Tên người	Đã rà soát
00001	Nguyễn Thị Danh	<input type="checkbox"/>
00001.1	Nguyễn Thị Kiều Anh	<input type="checkbox"/>
00001.1	Phạm Đăng Khoa	<input type="checkbox"/>
00002	Đoàn Hữu Lập	<input type="checkbox"/>
00003	Huỳnh Công Tâm	<input type="checkbox"/>
00004	Trang Thị Thủy Hằng	<input checked="" type="checkbox"/>
00005	Trang Thị Hai	<input type="checkbox"/>
00006	Đoàn Minh Chí	<input type="checkbox"/>
00007	Trần Thị Huyền	<input type="checkbox"/>
00007	Nguyễn Văn Liên	<input type="checkbox"/>
00008	Huỳnh Công Luân	<input type="checkbox"/>
00009	Phạm Kim Oanh	<input type="checkbox"/>
00010	Nguyễn Kim Ngọc	<input type="checkbox"/>
00011	Nguyễn Hoa Mai	<input type="checkbox"/>

Số hộ khẩu: 250

Thêm hộ
Sửa hộ
Xóa hộ
In

Thông tin nhân khẩu
Thông tin KHHGD-SKSS
Thông tin biến động

<input type="checkbox"/>	Họ tên	Quan hệ	Ngày sinh	Dân tộc	Giới tính	Trình độ học vấn	tr
<input type="checkbox"/>	ĐOÀN HỮU LẬP	Chủ hộ	03/07/1975	Kinh	Nam	Trung học cơ sở	Không
<input type="checkbox"/>	TRẦN KIM PHƯƠNG	Vợ/Chồng chủ hộ	06/08/1973	Kinh	Nữ	Trung học cơ sở	Không
<input type="checkbox"/>	VÕ THỊ CHẨM	Vợ/Chồng chủ hộ	30/10/1938	Kinh	Nữ	Tiểu học	Không
<input type="checkbox"/>	BÉ TRAI	Con đẻ	27/02/2020	Kinh	Nam		Không
<input type="checkbox"/>	ĐOÀN MINH THY	Con đẻ	05/03/2000	Kinh	Nam	Trung học phổ thông	Không
<input type="checkbox"/>	ĐOÀN MINH THƯ	Con đẻ	05/03/2000	Kinh	Nữ	Trung học phổ thông	Không

Thêm
Sửa
Xóa
In
In PTT

**Hình 3.1: Giao diện quản lý hộ - nhân khẩu**

Mẫu thông tin chi tiết của mẫu in phiếu thu tin. Với mẫu in này, công tác viên dân số sẽ in phiếu này ra, để đi đến từng hộ dân, ghi chép lại thông tin biến động (nếu có thay đổi).

**I- THÔNG TIN CƠ BẢN HỘ SỐ: 00002** Địa chỉ hộ: Đoàn Hữu Lập, Địa Bàn 1, Khu Phố Ninh Phước, Tổ: 1, 2, 3, 4, 5, 15, 16, Phường Ninh Thạnh

Số TT	Họ và tên	Quan hệ với chủ hộ	Giới tính	Ngày sinh	Dân tộc	Trình độ học vấn	Tình trạng hôn nhân	Tình trạng cư
1	ĐOÀN HỮU LẬP	Chủ hộ	Nam	03/07/1973	Kinh	Trung học cơ sở	Có vợ Có chồng	Thực tế thường trú có mặt
2	TRẦN KIM PHƯƠNG	Vợ chồng chủ hộ	Nữ	06/08/1973	Kinh	Trung học cơ sở	Có vợ Có chồng	Thực tế thường trú có mặt
3	ĐOÀN MINH THY	Con đẻ	Nam	05/03/2000	Kinh	Trung học phổ thông	Chưa có vợ hoặc chồng	Thực tế thường trú có mặt
4	BÈ TRẠI	Con đẻ	Nam	27/02/2020	Kinh		Chưa có vợ hoặc chồng	Thực tế thường trú có mặt

**II. THEO DÕI SỬ DỤNG BPTT**  
 Họ và tên: TRẦN KIM PHƯƠNG  
 Năm sinh: 1973 BPTT: Thời số dùng vòng tránh thai  
 Tháng năm bắt đầu SD BPTT: 07/07/2021

Tháng	Năm				
	2016	2017	2018	2019	2020
1					
2					
3					
4					
5					
6					
7					
8					
9					
10					
11					
12					

**III. THEO DÕI CÁC THAY ĐỔI**

**1. Trẻ mới sinh**  
 (1) Họ và tên: .....  
 -Ngày sinh: .....  
 đẻ tại TYT [ ] nhà [ ] nơi khác [ ]  
 là con thứ mấy ..... của ba mẹ  
 (2) Họ và tên: .....  
 -Ngày sinh: .....  
 đẻ tại TYT [ ] nhà [ ] nơi khác [ ]  
 là con thứ mấy ..... của ba mẹ

**2. Người chết**  
 (1) Họ và tên: .....  
 -Ngày chết: .....  
 (2) Họ và tên: .....  
 -Ngày chết: .....

**3. Chuyển đến từ xã khác**  
 (1) Họ và tên: .....  
 Ngày đến: .....  
 (2) Họ và tên: .....  
 Ngày đến: .....  
 (3) Họ và tên: .....  
 Ngày đến: .....

**4. Chuyển đi khỏi xã**  
 (1) Họ và tên: .....  
 Ngày đi: .....  
 (2) Họ và tên: .....  
 Ngày đi: .....  
 (3) Họ và tên: .....  
 Ngày đi: .....

**5. Thay đổi thông tin cơ bản**  
 -Họ, tên; ngày sinh; dân tộc; hôn nhân  
 -Nhân con mới; quan hệ chủ hộ  
 -Sửa sai; xóa do ghi thiếu; thêm do ghi thiếu

Ngày tháng năm	Ghi thay đổi	T

Công tác viên cập nhật thông tin hàng tháng

Hình 3.2: Giao diện in PTT (phiếu thu tin)

Giao diện quản lý thông tin kế hoạch hóa gia đình-sức khỏe sinh sản. Tab này quản lý thông tin kế hoạch hóa gia đình – sức khỏe sinh sản. Kế hoạch hóa gia đình: lưu lại các lần nhập biện pháp tránh thai, ngày tháng, các biện pháp được sử dụng, thông tin nhân khẩu. Sức khỏe sinh sản: thông tin của người mẹ (họ tên, ngày sinh, ngày thu thập, ngày đẻ, con thứ mấy, cân nặng); thông tin của người con.

**Hộ khẩu**  
 Địa phương: Phường Ninh Thạnh  
 Địa bàn: Địa Bàn 1, Khu Phố Ninh Phước

Số hộ	SN/Tên người	Đã rà soát
00001	Nguyễn Thị Danh	<input type="checkbox"/>
00001.1	Nguyễn Thị Kiều Anh	<input type="checkbox"/>
00001.1	Phạm Đăng Khoa	<input type="checkbox"/>
00002	Đoàn Hữu Lập	<input checked="" type="checkbox"/>
00003	Huỳnh Công Tâm	<input type="checkbox"/>
00004	Trang Thị Thủy Hằng	<input checked="" type="checkbox"/>
00005	Trang Thị Hải	<input type="checkbox"/>
00006	Đoàn Minh Chí	<input type="checkbox"/>
00007	Trần Thị Huyền	<input type="checkbox"/>
00007	Nguyễn Văn Liên	<input type="checkbox"/>
00008	Huỳnh Công Luân	<input type="checkbox"/>
00009	Phạm Kim Oanh	<input type="checkbox"/>
00010	Nguyễn Kim Cúc	<input type="checkbox"/>
00011	Nguyễn Hoa Mai	<input type="checkbox"/>
00012	Nguyễn Lạc Thành	<input type="checkbox"/>

Số hộ khẩu: 250

**Thông tin nhân khẩu**

Họ tên	#	#	Ngày tháng	Biện pháp
TRẦN KIM PHƯƠNG	Sửa	Xóa	12/09/2005	Vòng tránh thai
	Sửa	Xóa	25/07/2019	Thay vòng tránh thai
	Sửa	Xóa	01/03/2020	Không sử dụng
	Sửa	Xóa	21/07/2020	Vòng tránh thai
	Sửa	Xóa	07/07/2021	Thôi sử dụng vòng tránh thai

Tổng số: 05

#	#	Ngày thu thập	Ngày đẻ	Thông tin	Là con thứ mấy	Sinh tại
Sửa	Xóa	05/03/2003	05/03/2003	Sinh con	1	Tại cơ sở y tế
Sửa	Xóa	27/03/2020	27/02/2020	Sinh con	2	Tại cơ sở y tế

Số người: 01

Thêm KHHGD Thêm SKSS

Hình 3.3: Giao diện quản lý thông tin KHHGD-SKSS

Quản lý thông tin biến động (hình 3.4) sẽ liệt kê các lần biến động của nhân khẩu. Hình bên dưới là hiển thị 3 lần biến động của nhân khẩu; thông tin về ngày biến động và trạng thái.

Hộ khẩu			Thông tin nhân khẩu		Thông tin KHHGD-SKSS		Thông tin biến động	
Địa phương: Phường Ninh Thành								
Địa bàn: Địa Bàn 1, Khu Phố Ninh Phước								
Số hộ	SN/Tên người	Đã rà soát	Ngày biến động	Trạng thái				
00001	Nguyễn Thị Danh	<input type="checkbox"/>	05-08-1998	Sức khoẻ sinh sản				
00001.1	Nguyễn Thị Kiều Anh	<input type="checkbox"/>	03-02-2003	Sức khoẻ sinh sản				
00001.1	Phạm Đăng Khoa	<input type="checkbox"/>	07-07-2021	Chuyển đi				
00002	Đoàn Hữu Lập	<input type="checkbox"/>						
00003	Huỳnh Công Tâm	<input type="checkbox"/>						
00004	Trang Thị Thúy Hằng	<input checked="" type="checkbox"/>						
00005	Trang Thị Hai	<input type="checkbox"/>						
00006	Đoàn Minh Chí	<input type="checkbox"/>						
00007	Trần Thị Huyền	<input type="checkbox"/>						
00007	Nguyễn Văn Liên	<input type="checkbox"/>						
00008	Huỳnh Công Luân	<input type="checkbox"/>						
00009	Phạm Kim Oanh	<input type="checkbox"/>						
00010	Nguyễn Kim Cúc	<input type="checkbox"/>						
00011	Nguyễn Hoa Mai	<input type="checkbox"/>						
00012	Nguyễn Long Thành	<input type="checkbox"/>						
Số hộ khẩu: 250								

**Hình 3.4: Giao diện xem thông tin biến động**

Giao diện chuyển hộ khẩu (hình 3.5). giao diện này có chức năng khi hộ khẩu cùng xã/phường cần thay đổi/chuyển hộ khẩu từ địa bàn này đến địa bàn khác. Chức năng này thực hiện bằng cách: chọn hộ thuộc địa bàn cần chuyển đi (bên trái) và chọn địa bàn mới cho hộ đó (bên phải); chọn xong, ta nhấn nút “chuyển”.

QUẢN LÝ HỘ - NHÂN KHẨU													
TT chi tiết		Chuyển hộ khẩu		Tách cá nhân		Tìm kiếm		Báo cáo		Quản lý CTV		QL Biến động	
Hộ chuyển đi				Hộ chuyển đến									
Địa phương: Phường Ninh Thành				Địa phương: Phường Ninh Thành									
Địa bàn: Địa Bàn 1, Khu Phố Ninh Phước				Địa bàn: Địa Bàn 1, Khu Phố Ninh Phước									
Số hộ	SN/Tên người			Số hộ	SN/Tên người								
00001	Nguyễn Thị Danh			00001	Nguyễn Thị Danh								
00001.1	Nguyễn Thị Kiều Anh			00001.1	Nguyễn Thị Kiều Anh								
00001.1	Phạm Đăng Khoa			00001.1	Phạm Đăng Khoa								
00002	Đoàn Hữu Lập			00002	Đoàn Hữu Lập								
00005	Trang Thị Hai			00005	Trang Thị Hai								
00007	Trần Thị Huyền			00007	Trần Thị Huyền								
00008	Huỳnh Công Luân			00008	Huỳnh Công Luân								
00009	Phạm Kim Oanh			00009	Phạm Kim Oanh								
00011	Nguyễn Hoa Mai			00011	Nguyễn Hoa Mai								
00012	Nguyễn Long Thành			00012	Nguyễn Long Thành								
00013	Trần Quốc Thái			00013	Trần Quốc Thái								
00014	Nguyễn Thị Thuận			00014	Nguyễn Thị Thuận								
00015	Nguyễn Văn Bé			00015	Nguyễn Văn Bé								
00016	Nguyễn Văn Truyến			00016	Nguyễn Văn Truyến								
00017	Ngô Văn Hưng			00017	Ngô Văn Hưng								
Số hộ khẩu: 230				Số hộ khẩu: 230									

**Hình 3.5: Giao diện chuyển hộ khẩu**

Giao diện tách cá nhân (hình 3.6). Giao diện này có chức năng là tách 1 hay nhiều nhân khẩu từ hộ này (bên trái) sang hộ khác (bên phải); khi chọn nhân khẩu xong, nhấn nút “chuyển đi”; hoặc cần tạo lập thêm hộ mới cho nhân khẩu cần chuyển đi thì nhấn nút “lập hộ mới”.

**QUẢN LÝ HỘ - NHÂN KHẨU**

TT chi tiết | Chuyển hộ khẩu | **Tách cá nhân** | Tìm kiếm | Báo cáo | Quản lý CTV | QL Biến động

**Cá nhân chuyển đi**

Địa phương: Phường Ninh Thạnh

Địa bàn: Địa Bàn 1, Khu Phố Ninh Phước

Hộ khẩu: 00001-Nguyễn Thị Danh

STT	Họ và tên	quan hệ
1	NGUYỄN THỊ DANH	Chủ hộ
2	NGUYỄN MINH QUÂN	Con đẻ
3	NGUYỄN THỊ HỒNG HOA	Con đẻ
4	NGUYỄN THỊ TUYẾT HỒNG	Con đẻ
5	LÊ THỊ NGỌC HÀ	Con nuôi/Con d.
6	NGUYỄN QUÂN ĐẠT	Cháu

Số công dân: 06

**Cá nhân chuyển đến**

Địa phương: Phường Ninh Thạnh

Địa bàn: Địa Bàn 12, Khu Phố Ninh Lợi

Hộ khẩu: 00004-Cao Hạnh Phúc

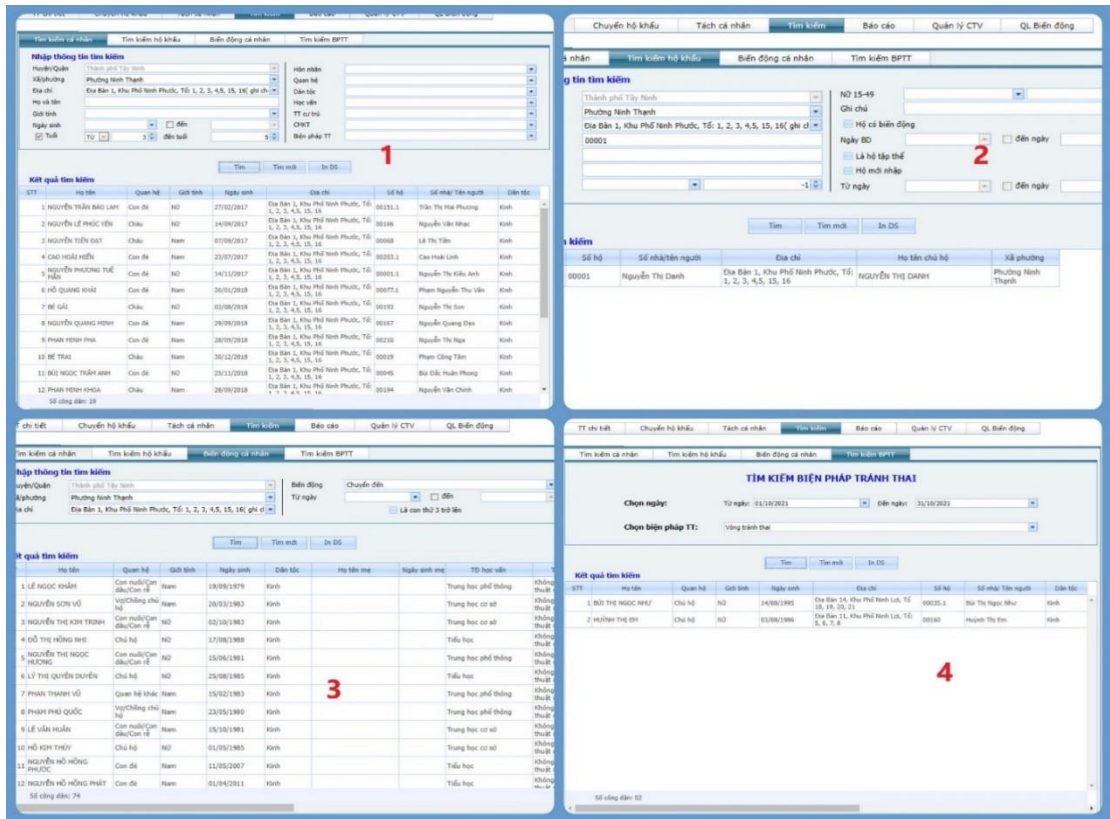
STT	Họ và tên	quan hệ
1	CAO HẠNH PHÚC	Chủ hộ
2	VÕ THỊ QUỐT	Vợ/Chồng chủ h

Số công dân: 02

Chuyển đi | Lập hộ mới

**Hình 3.6: Giao diện tách cá nhân sang hộ khẩu mới**

Giao diện tìm kiếm – mục đích là tìm kiếm các thông tin cần thiết và xuất ra danh sách, phục vụ việc in ấn. Giao diện này chia ra các chức năng tìm kiếm nhỏ (hình 3.7): tìm kiếm cá nhân (số 1), tìm kiếm hộ khẩu (số 2), biến động cá nhân (số 3), tìm kiếm BPTT (số 4). Tìm kiếm cá nhân, chọn các thông tin cần tìm như: địa bàn, họ tên, giới tính, độ tuổi từ - đến, ngày sinh từ - đến, hôn nhân, quan hệ với chủ hộ.... Tìm kiếm hộ khẩu chọn các thông tin tìm: số hộ, họ tên chủ hộ, số người trong hộ, hộ có biến động từ ngày – đến ngày. Biến động cá nhân chọn các thông tin cần tìm: loại biến động (chuyển đi, chuyển đến, trẻ sinh, chết, hôn nhân, kế hoạch hóa gia đình, sức khỏe sinh sản), từ ngày – đến ngày biến động. Tìm kiếm biện pháp tránh thai chọn các biện pháp tránh thai (vòng tránh thai, triệt sản, không sử dụng...). Khi tìm kiếm được các thông tin cần thiết, có thể dụng xuất ra file excel để in ấn và lưu trữ.



**Hình 3.7: Giao diện tìm kiếm các thông tin**

Giao diện quản lý CTV (cộng tác viên). Các thông tin quản lý về cộng tác viên dân số: mã số, họ tên, giới tính, ngày tham gia.



**Hình 3.8: Giao diện quản lý thông tin công tác viên dân số**



Sau khi thêm mới công tác viên, ta thực hiện gán công tác viên đó vô thuộc địa bàn quản lý – bằng cách chọn địa bàn có sẵn hoặc thêm mới địa bàn. Tùy vào mỗi địa bàn, ta có thể gán 1 hay nhiều công tác viên. Công tác viên dân số có nhiệm vụ hàng tháng là đi thu thập thông tin về các hộ - nhân khẩu có biến động trên địa bàn mà họ quản lý.

Địa chỉ đầy đủ	Công tác viên	Tổng số hộ	Ghi chú
Địa Bàn 1, Khu Phố Ninh Phước, Tổ: 1, 2, 3, 4, 5, 15, 16	PHAN VĂN ĐÚ	220	211 hộ
Địa Bàn 10, Khu Phố Ninh Lợi, Tổ: 1, 2, 3, 4	HUỲNH THỊ MÚT	188	152 hộ
Địa Bàn 11, Khu Phố Ninh Lợi, Tổ: 5, 6, 7, 8	NGUYỄN THỊ KIM DÂN	185	137 hộ
Địa Bàn 12, Khu Phố Ninh Lợi, Tổ 8, 9, 10, 11, 12	PHẠM CAO BÍCH THỦY	187	197 hộ
Địa Bàn 13, Khu Phố Ninh Lợi, Tổ 13, 14, 15, 16, 17	LÊ THỊ KIM LIÊN	173	104 hộ
Địa Bàn 14, Khu Phố Ninh Lợi, Tổ 18, 19, 20, 21	NGÔ THỊ MỸ LIÊN	181	110 hộ
Địa Bàn 15, Khu Phố Ninh Hòa, Tổ 2, 3, 4, 5, 6, 7	NGUYỄN VĂN NHAN	181	208 hộ
Địa Bàn 16, Khu Phố Ninh Hòa, Tổ 4, 5, 6, 7, 8, 9, 13	TRẦN THỊ HỒNG	150	
Địa Bàn 17, Khu Phố Ninh Hòa, Tổ 9, 10, 11, 12, 13	HỒ THỊ ĐIẾP	151	217 hộ
Địa Bàn 18, Khu Phố Ninh Phước, Tổ 11, 12, 13, 14, 15, 16, 17, 19	NGUYỄN THÀNH MAI	313	228 hộ
Địa Bàn 19, Khu Phố Ninh Phước, Tổ 1, 2, 3, 4, 5	CAO KIM HOA	185	165 hộ
Địa Bàn 2, Khu Phố Ninh Phước, Tổ 20, 21, 22, 23	NGUYỄN THỊ TRUYỀN	148	123 hộ
Địa Bàn 20, Khu Phố Ninh Phước, Tổ 7, 8, 9, 10	TRIỆU THỊ ÚT	183	164 hộ
Địa Bàn 21, Khu Phố Ninh Phước, Tổ 18, 20, 21, 22, 23, 24, 25, 26	DƯƠNG THỊ TỔ OANH	221	
Địa Bàn 22, Khu Phố Ninh Nghĩa, Tổ 1, 2, 3, 4, 5	TRẦN THỊ THỦY	245	150 hộ
Địa Bàn 3, Khu Phố Ninh Phước, Tổ 25, 27, 28, 29	BÙI KIM NGÂN	171	135 hộ
Địa Bàn 4, Khu Phố Ninh Phước, Tổ 3, 4, 5	PHAN HỒNG NHÃN	190	158 hộ
Địa Bàn 5, Khu Phố Ninh Phước, Tổ 12, 13, 16, 17, 18, 19	ĐỖ THỊ ÁNH NGUYỆT	186	169 hộ
Địa Bàn 6, Khu Phố Ninh Đức, Tổ 1, 2, 3, 4, 5, 6, 7, 8, 9, 10, 11	TRẦN THỊ ANH ĐÀO	239	204 hộ
Địa Bàn 7, Khu Phố Ninh Đức, Tổ 14, 15, 16, 17, 18	NGUYỄN THỊ THU THỦY	219	173 hộ
Địa Bàn 8, Khu Phố Ninh Đức, Tổ 19, 20, 21, 23, 24, 25	THỊ MỸ NA	293	251 hộ
Địa Bàn 9, Khu Phố Ninh Đức, Tổ 7, 8, 9, 10, 11, 12, 13	PHAN VĂN LƯỢNG	252	249 hộ

Tên địa bàn:  Chọn ấp/kp:  Ghi chú:

Danh sách công tác viên:

- BÙI KIM NGÂN
- PHAN VĂN ĐÚ
- THỊ MỸ NA
- NGÔ THỊ MỸ LIÊN
- NGUYỄN VĂN NHAN
- TRẦN THỊ HỒNG
- NGUYỄN THỊ TRUYỀN
- TRẦN THỊ ANH ĐÀO
- LÊ THỊ KIM LIÊN
- CAO KIM HOA
- HỒ THỊ ĐIẾP
- NGUYỄN THỊ KIM DÂN
- PHAN HỒNG NHÃN
- PHAN VĂN LƯỢNG
- HUỲNH THỊ MÚT
- TRIỆU THỊ ÚT
- TRẦN THỊ THỦY
- ĐỖ THỊ ÁNH NGUYỆT
- NGUYỄN THỊ THU THỦY
- PHẠM CAO BÍCH THỦY
- NGUYỄN THÀNH MAI
- DƯƠNG THỊ TỔ OANH

**Hình 3.9: Giao diện quản lý địa bàn và gán công tác viên vào địa bàn**

Giao diện quản lý biến động (hình 3.10). Giao diện này để quản lý các biến động: xem, sửa, phục hồi, in DS (danh sách). Trên giao diện, ta có thể chọn từ ngày – đến ngày, chọn loại biến động (xóa do ghi thừa, thêm do ghi thiếu, chuyển đi, chuyển đến, chết, trẻ mới sinh); ta có thể sửa lại thông tin ngày biến động hoặc phục hồi về trạng thái ban đầu (do thao tác chọn sai) – ví dụ lỡ bấm nhầm nhân khẩu A – chuyển đi, thì ta có thể vô chỗ quản lý này, tìm chọn lại nhân khẩu, và bấm nút “phục hồi”, khi đó nhân khẩu A sẽ quay về trạng thái ban đầu là chưa chuyển đi. Ngoài ra, có thể in danh sách ra, phục vụ cho việc báo cáo.

**QUẢN LÝ HỘ - NHÂN KHẨU**

TT chi tiết | Chuyển hộ khẩu | Tách cá nhân | Tìm kiếm | Báo cáo | Quản lý CTV | **QL Biến động**

---

**QUẢN LÝ THÔNG TIN BIẾN ĐỘNG**  
(*Để bổ sung chức năng phục hồi xóa do ghi thừa!!!*)

Chọn ngày thu thập: Từ ngày: 01/11/2021 Đến ngày: 09/12/2021

Chọn kiểu biến động: Chuyển đến

[Xóa] [Sửa] [Phục hồi] [In DS]

**Danh sách biến động**

STT	Họ Tên	Ngày sinh	Địa chỉ	Số hộ	Số nhà/tên người	Biến động	Ngày thu thập
1	NGUYỄN MINH TÂM	25/05/1967	Địa Bàn 12, Khu Phố Ninh Lợi, T0 8, 9, 10, 11, 12	00159	Nguyễn Minh Tâm	Chuyển đến	11/11/2021
2	NGUYỄN THỊ NGỌC MAI	17/01/1971	Địa Bàn 12, Khu Phố Ninh Lợi, T0 8, 9, 10, 11, 12	00157	Bùi Đình Kim Ngân	Chuyển đến	11/11/2021
3	NGUYỄN THỊ MINH THƯ	02/04/1998	Địa Bàn 12, Khu Phố Ninh Lợi, T0 8, 9, 10, 11, 12	00157	Bùi Đình Kim Ngân	Chuyển đến	11/11/2021
4	NGUYỄN THỊ NGỌC MAI	17/01/1997	Địa Bàn 12, Khu Phố Ninh Lợi, T0 8, 9, 10, 11, 12	00159	Nguyễn Minh Tâm	Chuyển đến	11/11/2021
5	NGUYỄN THỊ MINH THƯ	02/04/1998	Địa Bàn 12, Khu Phố Ninh Lợi, T0 8, 9, 10, 11, 12	00159	Nguyễn Minh Tâm	Chuyển đến	11/11/2021
6	PHẠM HỮU TRÍ	07/05/1991	Địa Bàn 14, Khu Phố Ninh Lợi, T0 18, 19, 20, 21	00166	Võ Thị Kim Chi	Chuyển đến	11/11/2021
7	PHẠM PHÚC AN	14/11/2018	Địa Bàn 14, Khu Phố Ninh Lợi, T0 18, 19, 20, 21	00166	Võ Thị Kim Chi	Chuyển đến	11/11/2021
8	HỒ MINH PHƯƠNG	27/07/1992	Địa Bàn 11, Khu Phố Ninh Lợi, T0 5, 6, 7, 8	00161	Hồ Minh Phương	Chuyển đến	11/11/2021
9	LÊ THỊ BẢO TRẦN	22/07/1993	Địa Bàn 11, Khu Phố Ninh Lợi, T0 5, 6, 7, 8	00161	Hồ Minh Phương	Chuyển đến	11/11/2021
10	HỒ LÊ QUẢN NGHỆ	10/11/2020	Địa Bàn 11, Khu Phố Ninh Lợi, T0 5, 6, 7, 8	00161	Hồ Minh Phương	Chuyển đến	11/11/2021

**Hình 3.10: Giao diện quản lý thông tin biến động của hộ - nhân khẩu**

Giao diện các mẫu báo cáo (hình 3.11). Các mẫu báo cáo theo đúng số quyết định 18/QĐ-TCDS ngày 17 tháng 3 năm 2016. Bao gồm các mẫu biểu cơ bản: báo cáo dân số kế hoạch hóa gia đình theo tháng (cấp xã: biểu 01 – DSX và cấp huyện: biểu 01-DSH); theo quý (cấp xã: biểu 02 – DSX và cấp huyện: biểu 02-DSH); theo năm (cấp xã: biểu 03 – DSX và cấp huyện: biểu 03 – DSH).

**BÁO CÁO DÂN SỐ KẾ HOẠCH HÓA GIA ĐÌNH**

Chọn xã: Phường Ninh Thạnh | Báo cáo: Tháng 10 | Năm: 2021 | Báo cáo dân số, kế hoạch hóa gia đình (Biểu 01)

Tên chỉ tiêu	- TOÀN XÃ	Địa Bàn 1, Khu Phố Ninh Phước, T0 8, 9, 3, 4, 5, 15, 16	Địa Bàn 10, Khu Phố Ninh Lợi, T0 1, 2, 3, 4	Địa Bàn 11, Khu Phố Ninh Lợi, T0 5, 6, 7, 8	Địa Bàn 12, Khu Phố Ninh Lợi, T0 8, 9, 10, 11, 12	Địa Bàn 13, Khu Phố Ninh Lợi, T0 13, 14, 15, 16	Địa Bàn 14, Khu Phố Ninh Lợi, T0 18, 19, 20, 21	Địa Bàn 15, Khu Phố Ninh Hòa, T0 2, 3, 4, 5, 6, 7, 9, 13	Địa Bàn 16, Khu Phố Ninh Hòa, T0 4, 5, 6, 7, 8
1. Tổng số trẻ sinh ra trong tháng	7	0	0	0	2	2	1	0	0
- Số trẻ nam sinh ra	6	0	0	0	2	2	0	0	0
- Số trẻ em nữ sinh ra	1	0	0	0	0	0	1	0	0
- Số trẻ em sinh ra từ con thứ 3 trở lên	2	0	0	0	1	0	1	0	0
- Số trẻ em sinh									
- Số trẻ em sinh									

**BÁO CÁO DÂN SỐ KẾ HOẠCH HÓA GIA ĐÌNH**

Chọn xã: Quý | Báo cáo: Quý 3 | Năm: 2021 | Báo cáo dân số, kế hoạch hóa gia đình (Biểu 02)

Tên chỉ tiêu	- TOÀN XÃ	Địa Bàn 1, Khu Phố Ninh Phước, T0 8, 9, 3, 4, 5, 15, 16	Địa Bàn 10, Khu Phố Ninh Lợi, T0 1, 2, 3, 4	Địa Bàn 11, Khu Phố Ninh Lợi, T0 5, 6, 7, 8	Địa Bàn 12, Khu Phố Ninh Lợi, T0 8, 9, 10, 11, 12	Địa Bàn 13, Khu Phố Ninh Lợi, T0 13, 14, 15, 16	Địa Bàn 14, Khu Phố Ninh Lợi, T0 18, 19, 20, 21	Địa Bàn 15, Khu Phố Ninh Hòa, T0 2, 3, 4, 5, 6, 7, 9, 13	Địa Bàn 16, Khu Phố Ninh Hòa, T0 4, 5, 6, 7, 8
1. Tổng số hộ dân cư tính đến cuối quý (hộ)	4137	206	165	151	159	151	163	175	131
- Trong đó: số hộ gia đình	4104	206	160	151	155	151	160	175	131
2. Tổng số nhân khẩu thực tế thường trú tính đến cuối quý (người)	16510	877	726	700	700	635	714	644	561
3. Số nữ từ 15-49 tuổi sinh con cuối quý (người)	4510	245	195	204	185	188	191	173	153

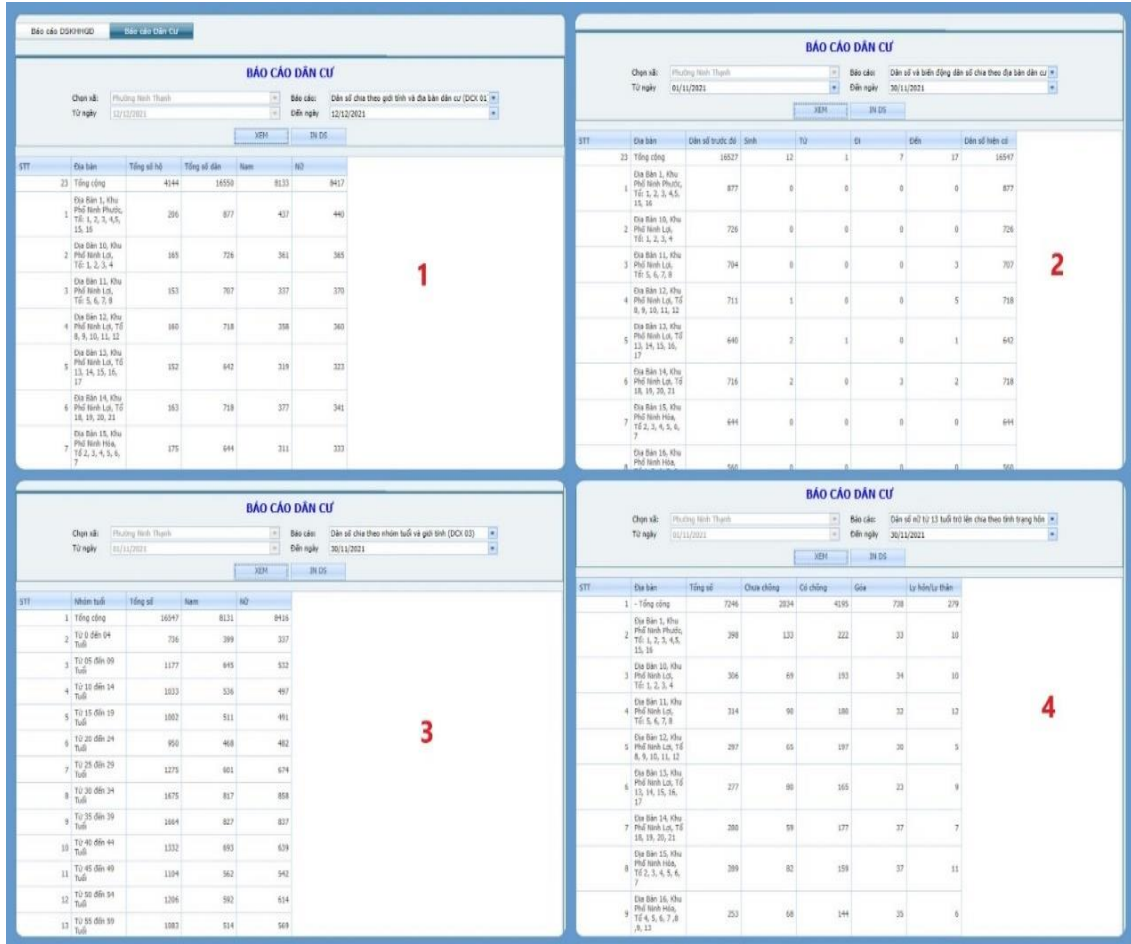
**BÁO CÁO DÂN SỐ KẾ HOẠCH HÓA GIA ĐÌNH**

Chọn xã: Tháng 10 | Báo cáo: Tháng 10 | Năm: 2021 | Báo cáo dân số, kế hoạch hóa gia đình (Biểu 03)

Tên chỉ tiêu	Số lượng
1. Tổng số hộ dân cư hộ tập thể + hộ gia đình) tính đến cuối năm	4144
Trong đó: - số hộ gia đình	4110
2. Tổng số nhân khẩu thực tế thường trú tính đến cuối năm	16550
3. Số phụ nữ có thai đến cuối năm	8117
- Số phụ nữ từ 15 đến 49 tuổi	4526
- Số phụ nữ từ 15 đến 49 tuổi có chồng	2659
4. Tổng số trẻ sinh ra	107
- Số trẻ em nam sinh ra	65
- Số trẻ em nữ sinh ra	42
- Số trẻ em sinh ra là con thứ 3 trở lên	11
- Số trẻ em sinh ra là con của phụ nữ dưới 20 tuổi	9
5. Tổng số người chết trong năm	44

**Hình 3.11: Giao diện báo cáo dân số kế hoạch hóa gia đình**

Các mẫu báo cáo dân cư tiêu biểu: dân số chia theo giới tính và địa bàn dân cư (DCX 01); dân số và biến động dân số theo địa bàn dân cư (DCX 02); dân số chia theo nhóm tuổi và giới tính (DCX 03); dân số nữ từ 13 tuổi trở lên chia theo tình trạng hôn nhân và địa bàn dân cư (DCX 04); dân số chia theo dân tộc và giới tính (DCX 05).



Hình 3.12: Giao diện báo cáo dân số kế hoạch hóa gia đình

### 3.2 Xây dựng mô hình dự báo dân số

Cho chuỗi dữ liệu quan sát về tổng dân số tỉnh Tây Ninh theo quý, từ quý 1 năm 2006 đến quý 4 năm 2020. Ứng dụng mô hình ARIMA, với việc sử dụng phần mềm EViews [6],[7] dự báo biến động dân số Tây Ninh cho quý 1, 2, 3, 4 năm 2021.

Số liệu dân số tỉnh Tây Ninh [5]

**Bảng 3.9: Dân số Tây Ninh từ năm 2006 đến 2020**

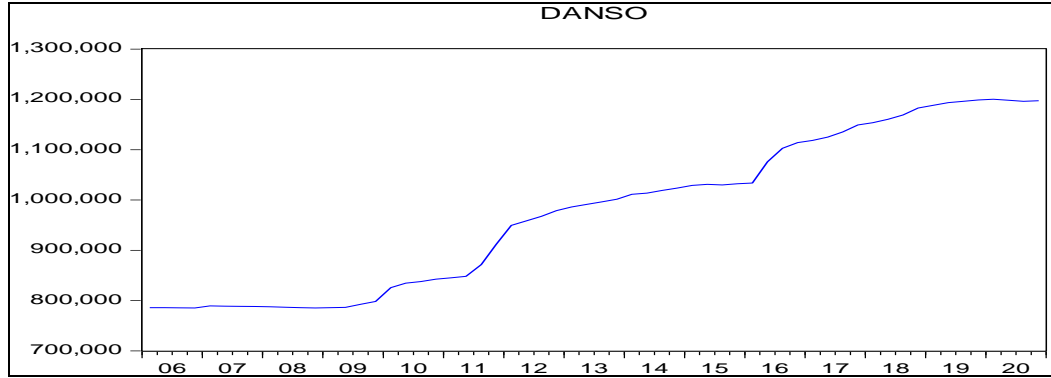
STT	Năm: Quý	Dân số Tây Ninh (người)	STT	Năm: Quý	Dân số Tây Ninh (người)
1	2006:1	785951	31	2013:3	996068
2	2006:2	785682	32	2013:4	1001558
3	2006:3	785511	33	2014:1	1011206
4	2006:4	785234	34	2014:2	1013617
5	2007:1	789224	35	2014:3	1018924
6	2007:2	788836	36	2014:4	1023716
7	2007:3	788343	37	2015:1	1029049
8	2007:4	788214	38	2015:2	1031141
9	2008:1	787470	39	2015:3	1029894
10	2008:2	786707	40	2015:4	1032252
11	2008:3	785816	41	2016:1	1033721
12	2008:4	785228	42	2016:2	1075822
13	2009:1	785792	43	2016:3	1102905

14	2009:2	786348	44	2016:4	1113911
15	2009:3	792504	45	2017:1	1118340
16	2009:4	798314	46	2017:2	1124953
17	2010:1	825559	47	2017:3	1135372
18	2010:2	834657	48	2017:4	1149219
19	2010:3	834880	49	2018:1	1153659
20	2010:4	842424	50	2018:2	1160556
21	2011:1	845123	51	2018:3	1168838
22	2011:2	848188	52	2018:4	1182790
23	2011:3	871622	53	2019:1	1188282
24	2011:4	911750	54	2019:2	1193613
25	2012:1	949429	55	2019:3	1196177
26	2012:2	958346	56	2019:4	1196177
27	2012:3	967384	57	2020:1	1200249
28	2012:4	978596	58	2020:2	1198184
29	2013:1	986015	59	2020:3	1196300
30	2013:2	991203	60	2020:4	1197468

### **Bước 1: nhận dạng mô hình:**

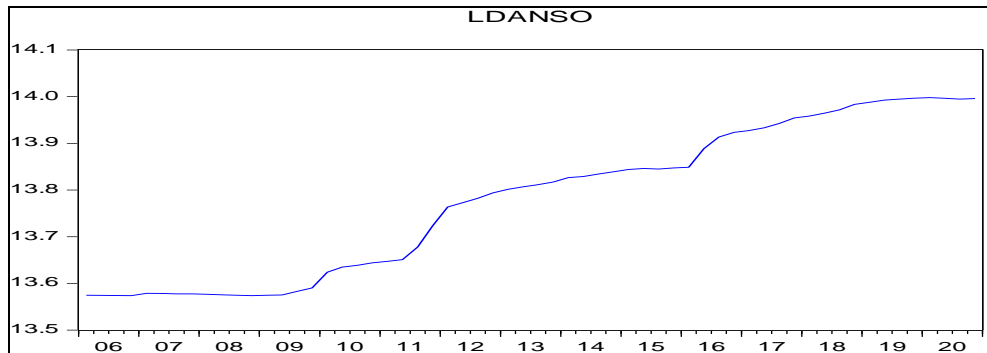
#### *a. Kiểm tra chuỗi dừng*

Đồ thị chuỗi dữ liệu dân số Tây Ninh, chuỗi ký hiệu DANSO



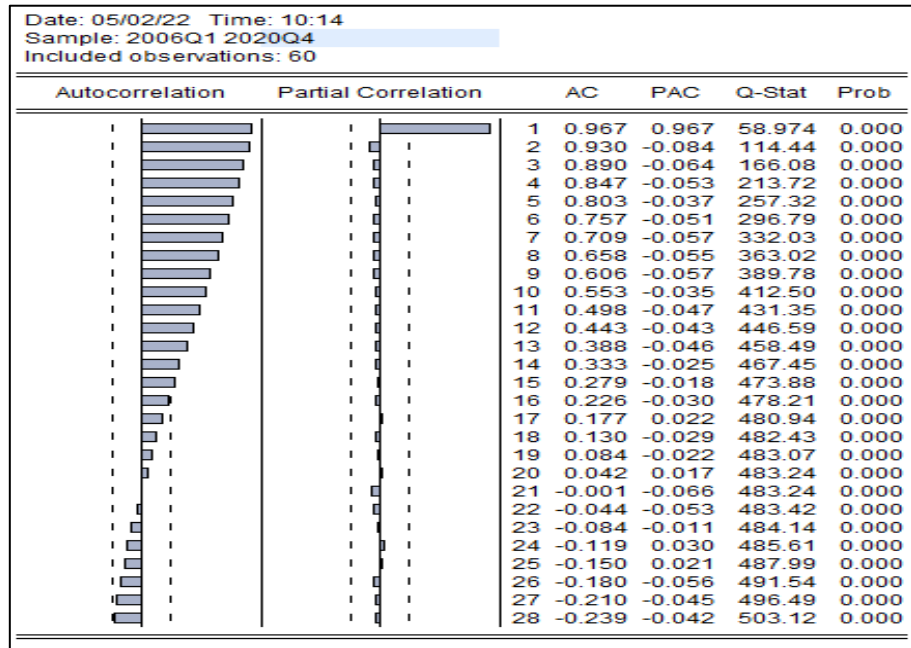
**Hình 3.13: dân số Tây Ninh từ năm 2006 đến 2020**

Phân tích sơ bộ từ đồ thị thì ta thấy chuỗi có xu thế tăng dần. Đồng thời có sự dao động trong đồ thị chuỗi là tương đối không ổn định. Vì thế, áp dụng hàm biến đổi  $LDANSO = \text{Log}(\text{DANSO})$ , để giảm bớt tính bất ổn định của chuỗi trước khi thực hiện phân tích.



**Hình 3.14: Đồ thị chuỗi số liệu dân số Tây Ninh sau khi sử dụng hàm biến đổi Log**

Chuỗi LDANSO thu được sau khi thực hiện hàm biến đổi Log trên DANSO dao động ổn định hơn so với chuỗi ban đầu, nhưng chưa phải là chuỗi dừng, giá trị trung bình của chuỗi vẫn tăng theo thời gian. Điều đó rõ ràng hơn khi xem xét tương quan đồ của hàm tự tương quan – ACF và hàm tự tương quan từng phần – PACF của chuỗi LDANSO:



**Hình 3.15: ACF và PACF của chuỗi LDANSO**

Có rất nhiều giá trị tự tương quan ở cột AC nằm ngoài khoảng tin cậy  $\pm 1.96\sqrt{60} = \pm 0.2530$ .

Trong hình 3.18, tương quan đồ ACF giảm dần khi trễ tăng, PACF tại trễ 1 có giá trị lớn nhất gần với giá trị đơn vị 1, điều đó chứng tỏ chuỗi DANSO chưa phải là chuỗi dừng. Dưới đây là kiểm định DF cho chuỗi LDANSO:

Null Hypothesis: LDANSO has a unit root  
Exogenous: Constant  
Lag Length: 1 (Automatic - based on SIC, maxlag=10)

	t-Statistic	Prob.*
Augmented Dickey-Fuller test statistic	-0.573621	0.8680
Test critical values:		
1% level	-3.548208	
5% level	-2.912631	
10% level	-2.594027	

\*MacKinnon (1996) one-sided p-values.

Augmented Dickey-Fuller Test Equation  
Dependent Variable: D(LDANSO)  
Method: Least Squares  
Date: 05/02/22 Time: 10:16  
Sample (adjusted): 2006Q3 2020Q4  
Included observations: 58 after adjustments

Variable	Coefficient	Std. Error	t-Statistic	Prob.
LDANSO(-1)	-0.004291	0.007481	-0.573621	0.5686
D(LDANSO(-1))	0.567910	0.111234	5.105551	0.0000
C	0.062278	0.103021	0.604521	0.5480

R-squared 0.321789 Mean dependent var 0.007266  
Adjusted R-squared 0.297127 S.D. dependent var 0.010540  
S.E. of regression 0.008837 Akaike info criterion -6.569426  
Sum squared resid 0.004295 Schwarz criterion -6.462851  
Log likelihood 193.5133 Hannan-Quinn criter. -6.527913  
F-statistic 13.04785 Durbin-Watson stat 1.805505  
Prob(F-statistic) 0.000023

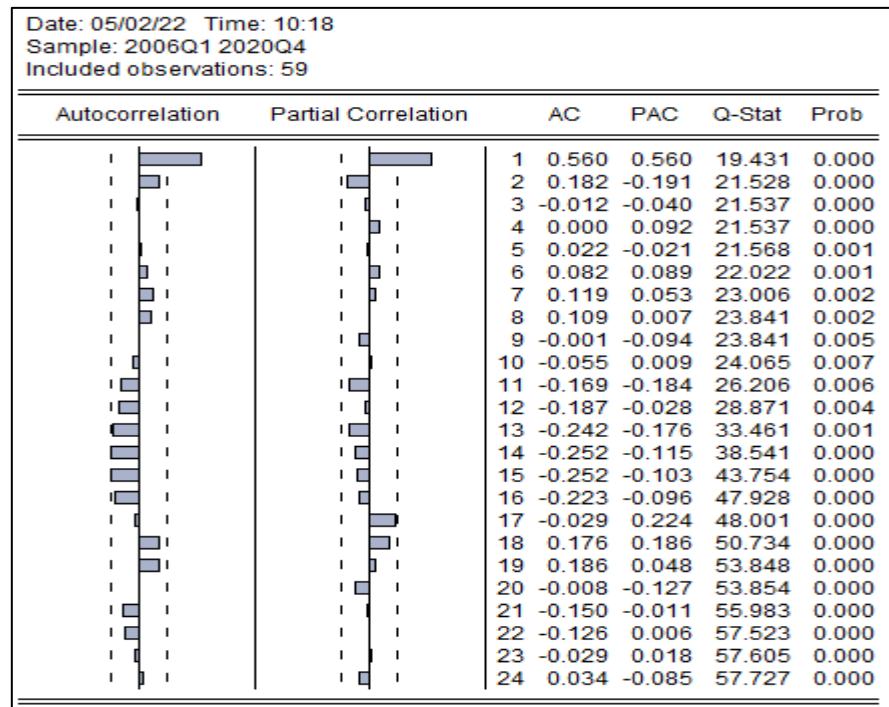
**Hình 3.16: Kiểm định DF của chuỗi LDANSO**

Kiểm định Augmented Dickey-Fuller (DF) cũng chứng tỏ đây không phải là chuỗi dừng. Do giá trị tuyệt đối của thống kê ADF được tính toán qua kiểm định DF là 0.57 nhỏ hơn các giá trị tuyệt đối tới hạn tại các mức ý nghĩa 1% = 3.54; 2% = 2.91; 10% = 2.59

*b. Biến đổi chuỗi không dừng thành chuỗi dừng*

Do chuỗi LDANSO đã được chứng minh qua kiểm tra hàm tự tương quan ACF và kiểm định đơn vị DF ở trên là không dừng, vì thế nó cần phải được biến đổi thành chuỗi dừng trước khi tạo lập mô hình ARIMA.

Thực hiện sai phân bậc 1 trên chuỗi không dừng LDANSO trên, đặt tên là DLDANSO và xem xét tương quan đồ của hàm tự tương quan ACF của nó như sau:



**Hình 3.17: ACF và PACF của chuỗi DLDANSO**

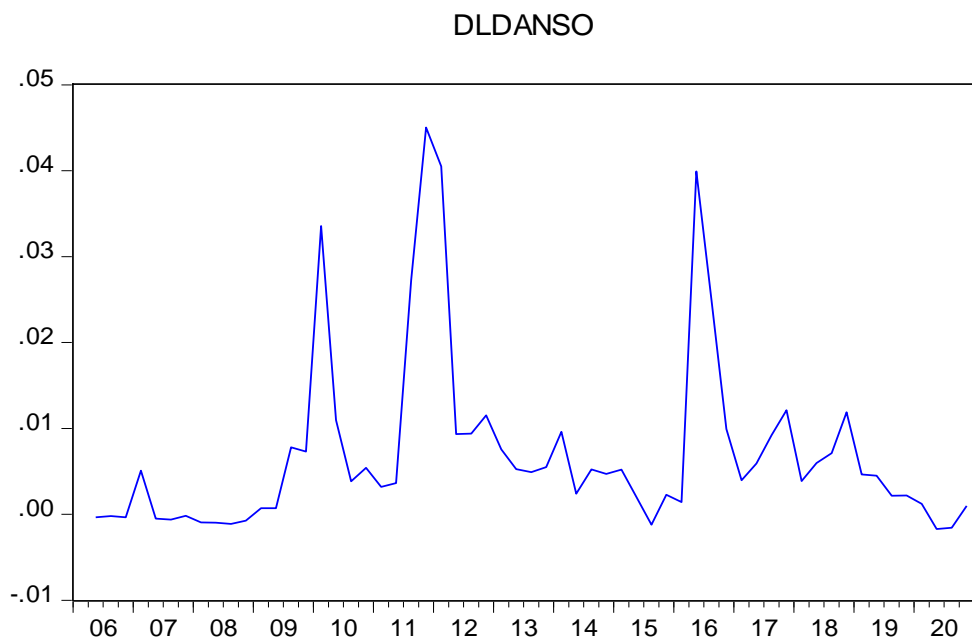
Tương quan đồ ACF và PACF của chuỗi DLDANSO, sau khi đã loại bỏ các thành phần tự tương quan từ chuỗi ban đầu, có đặc điểm ACF có dao động theo hình sin và hầu hết các ACF nằm trong khoảng tới hạn, tức là xấp xỉ bằng 0. Dựa trên điều đó, có thể kết luận chuỗi DLDANSO là chuỗi dừng. Kiểm định nghiệm đơn vị DF, kiểm định chuỗi thời gian dừng, trong hình dưới đây cũng cho kết luận chuỗi DLDANSO là chuỗi thời gian dừng:



Null Hypothesis: D(LDANSO) has a unit root Exogenous: Constant Lag Length: 0 (Automatic - based on SIC, maxlag=10)				
		t-Statistic	Prob.*	
<b>Augmented Dickey-Fuller test statistic</b>				
		-3.966054	0.0030	
Test critical values:				
1% level		-3.548208		
5% level		-2.912631		
10% level		-2.594027		
*Mackinnon (1996) one-sided p-values.				
Augmented Dickey-Fuller Test Equation Dependent Variable: D(LDANSO,2) Method: Least Squares Date: 05/02/22 Time: 10:19 Sample (adjusted): 2006Q3 2020Q4 Included observations: 58 after adjustments				
Variable	Coefficient	Std. Error	t-Statistic	Prob.
D(LDANSO(-1))	-0.437136	0.110219	-3.966054	0.0002
C	0.003189	0.001403	2.273401	0.0269
R-squared	0.219290	Mean dependent var	2.27E-05	
Adjusted R-squared	0.205349	S.D. dependent var	0.009854	
S.E. of regression	0.008784	Akaike info criterion	-6.597944	
Sum squared resid	0.004321	Schwarz criterion	-6.526894	
Log likelihood	193.3404	Hannan-Quinn criter.	-6.570269	
F-statistic	15.72959	Durbin-Watson stat	1.795068	
Prob(F-statistic)	0.000210			

**Hình 3.18: ACF và PACF của chuỗi DLDANSO**

Với giá trị tuyệt đối của thống kê ADF tính bằng kiểm định DF là 3.966054 đều lớn hơn các giá trị tới hạn tại các mức 1% = 3.548208; 5% = 2.912631; 10% = 2.594027, điều đó chứng tỏ chuỗi DLDANSO là dừng. Đồ thị của chuỗi dừng DLDANSO có dạng:



**Hình 3.19: Đồ thị chuỗi số liệu DLDANSO**

c. *Lựa chọn mô hình ARIMA*

Tương quan đồ của chuỗi DLDANSO ở trên thể hiện giá trị ACF và PACF tại trễ 1 đều khác 0, còn tại mọi trễ khác bằng 0. Vì thế mô hình ARIMA (p, d, q) được thiết lập gồm:

- ACF tại trễ 1 khác 0, chọn q=1
- PACF tại trễ 1 khác 0, chọn p=1
- Sai phân bậc 1, chọn d=1

Vậy mô hình được chọn là ARIMA (1, 1, 1)

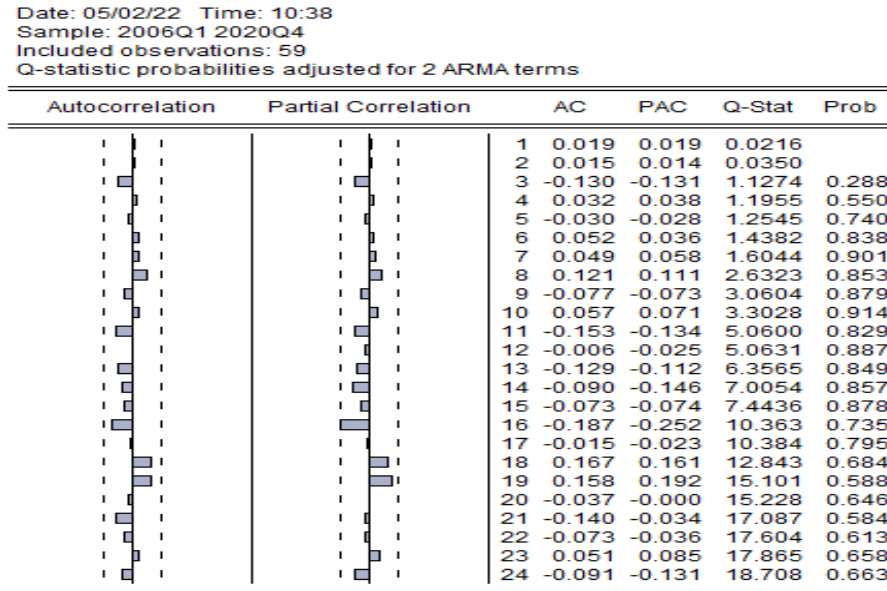
**Bước 2: Ước lượng mô hình:**

Dependent Variable: D(LDANSO)				
Method: ARMA Maximum Likelihood (OPG - BHHH)				
Date: 05/02/22 Time: 10:28				
Sample: 2006Q2 2020Q4				
Included observations: 59				
Convergence achieved after 22 iterations				
Coefficient covariance computed using outer product of gradients				
Variable	Coefficient	Std. Error	t-Statistic	Prob.
C	0.006948	0.005053	1.375100	0.1747
AR(1)	0.403208	0.245326	1.643562	0.1060
MA(1)	0.239792	0.262409	0.913811	0.3648
SIGMASQ	7.19E-05	1.35E-05	5.333223	0.0000
R-squared	0.336351	Mean dependent var		0.007137
Adjusted R-squared	0.300152	S.D. dependent var		0.010496
S.E. of regression	0.008781	Akaike info criterion		-6.559996
Sum squared resid	0.004241	Schwarz criterion		-6.419146
Log likelihood	197.5199	Hannan-Quinn criter.		-6.505013
F-statistic	9.291702	Durbin-Watson stat		1.953311
Prob(F-statistic)	0.000046			
Inverted AR Roots	.40			
Inverted MA Roots	-.24			

**Hình 3.20: Mô hình ARIMA (1, 1, 1)**

### Bước 3: Kiểm định mô hình:

Kiểm định nghiệm đơn vị cho chuỗi phần dư qua việc xem xét tương quan đồ thì ACF như sau:



**Hình 3.21: Kiểm định sự tương quan ACF**

Hình trên với mức ý nghĩa 95% các ACF và PACF của chuỗi phần dư đều bằng 0. Mô hình này chấp nhận để dự báo.

Kiểm nghiệm tính dừng cho chuỗi phần dư

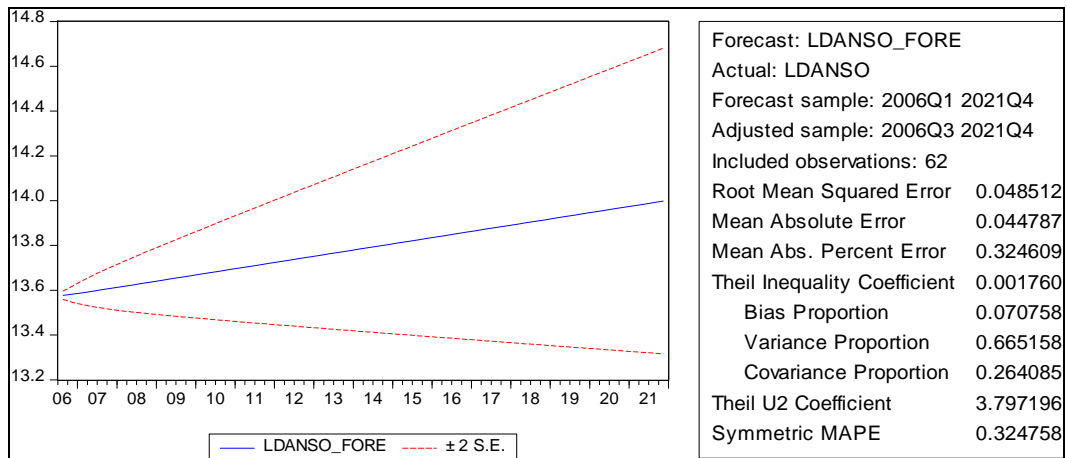
Null Hypothesis: RESID01 has a unit root				
Exogenous: Constant				
Lag Length: 0 (Automatic - based on SIC, maxlag=10)				
		t-Statistic	Prob.*	
<b>Augmented Dickey-Fuller test statistic</b>		<b>-7.374960</b>	<b>0.0000</b>	
Test critical values:	1% level	-3.548208		
	5% level	-2.912631		
	10% level	-2.594027		
*MacKinnon (1996) one-sided p-values.				
Augmented Dickey-Fuller Test Equation				
Dependent Variable: D(RESID01)				
Method: Least Squares				
Date: 05/02/22 Time: 10:41				
Sample (adjusted): 2006Q3 2020Q4				
Included observations: 58 after adjustments				
Variable	Coefficient	Std. Error	t-Statistic	Prob.
RESID01(-1)	-0.981385	0.133070	-7.374960	0.0000
C	0.000170	0.001137	0.149188	0.8819
R-squared	0.492708	Mean dependent var	7.44E-05	
Adjusted R-squared	0.483649	S.D. dependent var	0.012055	
S.E. of regression	0.008662	Akaike info criterion	-6.625849	
Sum squared resid	0.004202	Schwarz criterion	-6.554800	
Log likelihood	194.1496	Hannan-Quinn criter.	-6.598174	
F-statistic	54.39004	Durbin-Watson stat	2.006547	
Prob(F-statistic)	0.000000			

**Hình 3.22: Kiểm định tính dừng**

Với giá trị tuyệt đối của thống kê ADF tính bằng kiểm định DF là 7.374960 đều lớn hơn các giá trị tới hạn tại các mức 1% = 3.548208; 5% = 2.912631; 10% = 2.594027

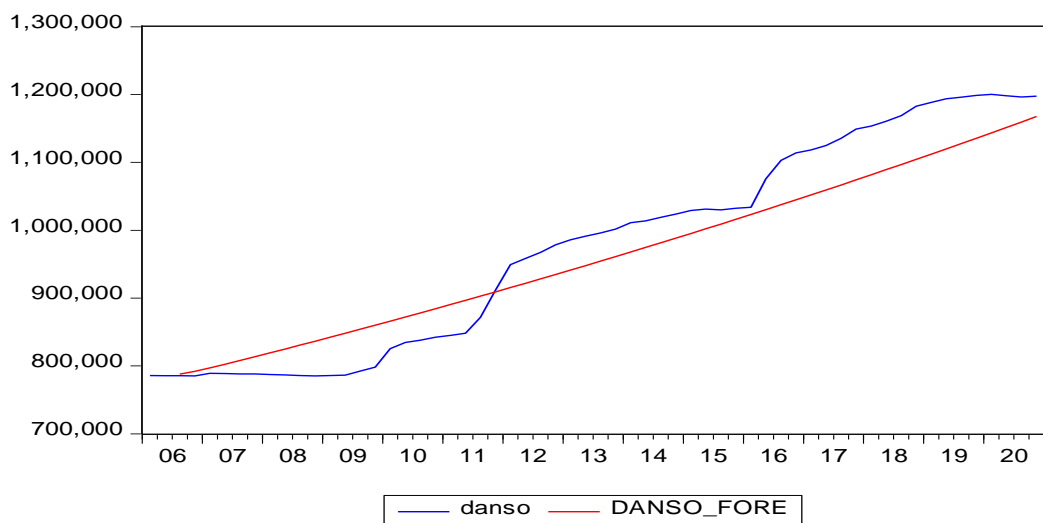
Với 2 cách kiểm định sự tương quan và tính dừng đều cho thấy mô hình ARIMA (1, 1, 1) là tốt, nên ta chấp nhận.

#### **Bước 4: Dự báo:**



**Hình 3.23: Đồ thị dự báo chuỗi DLDANSO**

Giá trị sai số bình phương trung bình (RMSE) và sai số tuyệt đối trung bình (MAPE) là rất nhỏ, đồng thời độ lệch trung bình (Bias Proportion) giữa chuỗi quan sát và chuỗi dự báo là rất nhỏ. Từ đó kết luận là chuỗi dự báo từ mô hình rất phù hợp với chuỗi dự báo.



**Hình 3.24: Đồ thị so sánh chuỗi ban đầu(danso) và chuỗi dự báo (danso\_fore)**

**Dự báo:**

Stt	Năm: quý	Giá trị thật	Giá trị dự báo	Sai số	MAPE
1	2021:1	1196334	1202441	6106.708	0.51045%
2	2021:2	1198155	1209456	11301.11	0.94321%
3	2021:3	1200116	1217333	17217.47	1.43465%
4	2021:4	1205794	1225595	19801.5	1.6422%

Giá trị phần trăm sai số tuyệt đối

$$MAPE = \frac{|giá\ trị\ thật - giá\ trị\ dự\ báo|}{giá\ trị\ thật} \times 100$$

**Đánh giá kết quả**

Các kết quả dự báo về tổng dân số Tây Ninh trong 4 quý tiếp theo cho xấp xỉ với các giá trị quan sát thực, sai số dự báo nhỏ và mô hình ARIMA được lựa chọn khác hợp lý và độ tin cậy cao. Điều đó chứng tỏ mô hình được chọn có thể giải thích được cho sự biến động về dân số tỉnh Tây Ninh.

Kết quả nghiên cứu cho ra kết quả gần như tương đồng với các bài báo cáo khoa học đã nghiên cứu ở chương 1. Cụ thể: ở các bài báo [10],[11],[12].

	Nghiên cứu	Bài báo số [10]
MAPE	0.54045% (2021:1)	0.0080%
RMSE	6106.708 (2021:1)	6434.02

Như vậy, giá trị sai số căn bậc hai bình phương trung bình và phần trăm sai số tuyệt đối của mô hình này tương đối nhỏ và kết quả tương đồng như các bài báo cáo khoa học. Điều đó cho thấy ARIMA là thuật toán được lựa chọn là hợp lý.

## **CHƯƠNG 4: KẾT LUẬN VÀ KIẾN NGHỊ**

### **4.1 Kết quả đạt được**

Thông qua đề tài “**Hệ thống quản lý dân số và dự báo biến động dân số trên địa bàn tỉnh Tây Ninh**” luận văn đã đạt được các kết quả chính sau:

- Trình bày, phân tích tổng quan bài toán quản lý dân số và dự báo phát triển dân số.
- Xây dựng Hệ thống quản lý dân số tỉnh Tây Ninh, giúp ích cho cán bộ dân số có thể nhập những thông tin thu thập lên hệ thống phần mềm: quản lý thông tin về các địa bàn trong tỉnh, các hộ, nhân khẩu, các biến động về dân số như sinh, tử, chuyển đi, chuyển đến; quản lý kế hoạch hóa gia đình, chăm sóc sức khỏe sinh sản; các biểu mẫu báo cáo thống kê liên quan.
- Xây dựng dự báo xu thế dân số nhằm giúp cho các cấp, ngành, đơn vị có bức tranh tổng quan về dân số nhằm hỗ trợ xây dựng chiến lược, hoạch định chính sách về dân số, góp phần vào việc phát triển kinh tế xã hội trên địa bàn tỉnh Tây Ninh.

### **4.2 Hạn chế và hướng phát triển**

Nghiên cứu này chưa so sánh kết quả dự báo với một số mô hình khác như: Mô hình Prophet, mô hình chuỗi thời gian mờ, LSTM, ... Những hạn chế này sẽ được khắc phục trong các nghiên cứu tiếp theo.

## DANH MỤC CÁC TÀI LIỆU THAM KHẢO

- [1] Nguyễn Văn Lượng, Huỳnh Tấn Nguyên (2017), “Ứng dụng mô hình ARIMA để dự báo chỉ số giá tiêu dùng ở Việt Nam”. *Tạp chí khoa học và công nghệ đại học Quảng Bình*, số 12.
- [2] Nguyen Ngoc Thieu, “*Dự báo phát triển dân số Việt Nam 2017-2026*”, Ngày 26/9/2017.
- [3] Nguyễn Quốc Dương, Lê Phương Thảo, Đinh Thị Quỳnh Như, Cao Thị Ái Loan, Phùng Thị Hồng Diễm, (2020), “Ứng dụng Shiny kết hợp mô hình ARIMA để dự báo đại dịch COVID-19”, *Tạp chí Khoa học và Công nghệ VN*.
- [4] Cục thống kê tỉnh Tây Ninh (2020). Niên giám thống kê tỉnh Tây Ninh – năm 2020, tr.59-76
- [5] Cục dân số tỉnh Tây Ninh, số liệu dân số từ quý 1 năm 2006 đến quý 4 năm 2021.
- [6] Phùng Thanh Bình, *Hướng dẫn sử dụng Eviews 6.0*, trường Đại học Kinh tế Tp. Hồ Chí Minh, 53pp
- [7] Nguyễn Duy Tâm (2010), *Phân tích dữ liệu với phần mềm Eviews*, trường Đại học Kinh tế Tp. Hồ Chí Minh, 20 pp
- [8] Đỗ Văn Bình (2007), *Một số mô hình dự báo và áp dụng vào ngành điện*, trường đại học Công Nghệ, đại học Quốc Gia Hà Nội, tr.42-53
- [9] Ratnadip Adhikari, R.K. Agrawal, (2013), “*An introductory study on time series modeling and forecasting*”, LAP Lambert Academic Publishing, Germany, pp.12-45, 2013.
- [10] Ahmet Tezcan Tekin, (2019), “Machine learning algorithms to forecast population: Turkey example”, *International Engineering And Technology Management Summit 2019– ETMS2019*.
- [11] Stanley K. Smith, Jeff Tayman, (2004), “Confidence intervals for Population Forecasts: A case study of time series models for states”, *Paper presented at the annual meeting of the Population Association of America, Boston, 2004*.

- [12] Md. Ashek Al Naim, Md. Abdul khalek and Md. Ayub Ali, (2019), “*Time series model building and forecasting population of Bangladesh*”, 7<sup>th</sup> Int. Conf. on Data Science & SDGs, pp. 421-425, 2019.
- [13] John E. Hanke, Dean W. Wichern (2004). “Business Forecasting” Pearson Prentice Hall, ISBN 0-13-141290-6.
- [14] Kevin Michael Reagan (1984), *An evaluation of ARIMA (BoxJenkins) Models for forecasting wastewater treatment process variables*, University of California, 162 pp
- [15] Shashank Shekhar (2004), *Recursive methods for forecasting short - term traffic flow using seasonal ARIMA time series model*, North Carolina State University.
- [16][https://vi.wikipedia.org/wiki/T%C3%A2y\\_Ninh](https://vi.wikipedia.org/wiki/T%C3%A2y_Ninh), truy cập ngày 03/05/2021
- [17]<https://bigdatauni.com/tin-tuc/tim-hieu-ve-time-series-phan-tich-chuoi-thoi-gian-p-1.html>, truy cập ngày 03/05/2021



## PHỤ LỤC

### ❖ Hệ thống quản lý

Hệ thống quản lý dân số được xây dựng trên nền tảng Web, hỗ trợ .NET Framework version 4.0 trở lên, hệ quản trị cơ sở dữ liệu: SQL server 2008 R2 hoặc cao hơn, ngôn ngữ lập trình C#.

Một số giao diện cần chú ý khi sử dụng hệ thống:

Khi nhấn nút chức năng thêm hộ/sửa hộ (hình 3.1) sẽ hiển thị giao diện để nhập các thông tin quản lý gồm: chọn hộ chuyển đến/thêm mới, ngày thu thập, số hộ, ghi chú; đồng thời, lưu ý là khi thêm mới hộ khẩu thì bắt buộc nhập thông tin chi tiết của chủ hộ. Thông tin quản lý chủ hộ bao gồm: họ tên, ngày sinh, giới tính, quan hệ, cư trú, dân tộc, hôn nhân, TĐ học vấn, CMKT, địa chỉ.

Khi nhấn nút chức năng thêm/sửa (hình 3.1) sẽ hiển thị giao diện để nhập các thông tin quản lý nhân khẩu gồm: chọn nhân khẩu chuyển đến / thêm mới, ngày thu thập, ngày chuyển đến (nếu chọn: chuyển đến), họ tên, ngày sinh, giới tính, quan hệ chủ hộ, cư trú, dân tộc, hôn nhân, học vấn, CMKT, địa chỉ.

Khi nhấn nút thêm KHHGD (hình 3.3) sẽ hiển thị các thông tin quản lý kế hoạch hóa gia đình: họ tên, tuổi, ngày tháng bắt đầu, biện pháp sử dụng. Khi nhấn nút thêm SKSS (hình 3.1.4-5) sẽ hiển thị các thông tin quản lý sức khỏe sinh sản: thông tin của người mẹ (họ tên, ngày sinh, ngày thu thập, ngày đẻ, con thứ mấy, cân nặng); thông tin của người con (họ tên, ngày sinh, giới tính, quan hệ với chủ hộ, cư trú,

### ❖ Dự báo

Trong bài báo cáo này đã sử dụng phần mềm Eviews phiên bản 10.0 để tính toán và dự báo dân số. Một số dòng lệnh chú ý như sau:

`genr ldanso=log(danso) //lấy log. Và gán cho biến ldanso`

`ls ldanso c @trend // xác định chuỗi ldanso có xu thế`

`genr dldanso = d(ldanso) //lấy sai phân bậc 1 cho giá trị ldanso`

`d(ldanso) c ar(1) ma(1) // ước lượng mô hình ARIMA. Với AR là 1; MA là 1`

`genr danso_fore=exp(ldanso_fore) // chuyển đổi giá trị đã chuyển log sang  
giá trị ban đầu, để so sánh.`