

รายงานฉบับสมบูรณ์โครงการวิศวกรรมไฟฟ้า วิชา 2102499 ปีการศึกษา 2558

ชื่อโครงการ การหาโครงสร้างเชิงสาเหตุสำหรับข้อมูลอนุกรมเวลาด้วยวิธีทดสอบทางสถิติบนการประมาณแบบจำลอง

Project Name Exploring Granger causality for time series via statistical tests on estimated models

ชื่อนิพนธ์ นายนันทนซ์ รักษาศรี หมายเลขประจำตัว 5530299121

กลุ่มวิจัย กลุ่มวิจัยการควบคุมขั้นสูงและการหาค่าเหมาะที่สุด

อาจารย์ที่ปรึกษา ผศ.ดร. จิตโกมุท ส่งศิริ ห้องปฏิบัติการวิจัย ระบบควบคุม

ภาควิชาวิศวกรรมไฟฟ้า คณะวิศวกรรมศาสตร์ จุฬาลงกรณ์มหาวิทยาลัย

บทคัดย่อ

โครงการนี้ต้องการที่จะอธิบายความสัมพันธ์ระหว่างตัวแปรของข้อมูลอนุกรมเวลาผ่านเงื่อนไขความสัมพันธ์เชิงเหตุและผล บนแบบจำลองถดถอยตัวเอง และหารูปแบบความสัมพันธ์นั้นผ่านการทดสอบ Wald บนพารามิเตอร์ของแบบจำลอง ผลลัพธ์ของโครงการนี้คือ รูปแบบปัญหาการประมาณแบบจำลองด้วยวิธีกำลังสองต่ำสุด ที่รวมเงื่อนไขความสัมพันธ์เชิงเหตุและผล และเสถียรภาพของแบบจำลอง โดยสำหรับเงื่อนไขเพียงพอของเสถียรภาพ เราบังคับให้เมทริกซ์พลวัตของแบบจำลองมีนอร์มแบบอนันต์น้อยกว่าหนึ่ง ปัญหาที่ได้จะอยู่ในรูปคอนเวกแบบกำลังสอง แบบจำลองที่ประมาณได้จะมีเสถียรภาพแน่นอน อย่างไรก็ตาม แบบจำลองจะมีค่าความคลาดเคลื่อนที่มากเกินไปเนื่องจากเงื่อนไขเสถียรภาพเป็นแค่เงื่อนไขเพียงพอเท่านั้น

คำสำคัญ: ข้อมูลอนุกรมเวลา, ความสัมพันธ์เชิงเหตุและผล, แบบจำลองถดถอยตัวเอง, การทดสอบ Wald, เสถียรภาพของแบบจำลอง

Abstract

This project aims to explain relationship between the time series data by using the Granger causality concept and autoregressive model and estimate the patterns of model parameters by using the Wald test. The result of this study is a least-squares estimation with Granger causality and stability constraints. The stability sufficient condition is added as a condition on the infinity norm of the dynamic matrix and the resulting problem is convex in quadratic form. An estimated model is guaranteed to be stable but the model fitting error is conservatively increased due to the sufficiency of the stability condition.

Keywords: Time series, Granger causality, Autoregressive model, Wald test, System stability

สารบัญ

1	บทนำ	3
2	ระเบียบวิธีที่ใช้ในการทำโครงการ	4
2.1	การประมาณแบบจำลอง AR	4
2.2	การทดสอบ Wald บนค่าประมาณแบบจำลอง AR	6
2.3	การประมาณแบบจำลอง AR ที่รวมเงื่อนไข Granger causality	9
2.4	การรวมเงื่อนไขเสถียรภาพในการประมาณแบบจำลอง	10
2.5	การประมาณแบบจำลองถดถอยตัวเองที่รวมเงื่อนไขความสัมพันธ์ระหว่างเหตุ และผลและเสถียรภาพ	14
2.6	แผนผังการหาแบบจำลองถดถอยตัวเองที่รวมเงื่อนไขความสัมพันธ์ระหว่างเหตุ และผลและเสถียรภาพ	15
3	ผลการทดลอง	17
3.1	การทดสอบ Wald บนตัวประมาณกำลังสองน้อยสุด	17
3.2	การทดสอบ Wald บนแบบจำลองถดถอยตัวเอง	17
3.3	การศึกษาเงื่อนไขความสัมพันธ์ระหว่างเหตุและผลด้วยการทดสอบ Wald	20
3.4	การประมาณแบบจำลองถดถอยตัวเองที่รวมเงื่อนไขความสัมพันธ์ระหว่างเหตุ และผล และเสถียรภาพ	22
4	บทสรุป	25
5	กิตติกรรมประกาศ	25
6	ภาคผนวก	27
6.1	การประมาณแบบ maximum likelihood บนแบบจำลอง AR	27
6.2	ค่า BIC ของการประมาณแบบจำลอง AR	29
6.3	การประมาณแบบ ML และ LS ของแบบจำลอง AR เมื่อรวมเงื่อนไขความสัมพันธ์ ระหว่างเหตุและผล	30
6.4	MATLAB codes	33

1 บทนำ

ในปัจจุบัน ข้อมูลอนุกรมเวลา (time series) นั้นเกี่ยวข้องกับชีวิตประจำวันไม่มากนักน้อย ตัวอย่างเช่น ราคาของหุ้นในแต่ละวัน ปริมาณน้ำฝนในแต่ละเดือน ปริมาณความต้องการใช้ไฟฟ้าในแต่ละวัน หรือ สัญญาณคลื่นไฟฟ้าทางสมอง ถ้าสามารถทราบถึงความสัมพันธ์ระหว่างตัวแปรต่างๆของข้อมูลอนุกรมเวลาก็จะสามารถอธิบายการเปลี่ยนแปลงหรือพลวัตของข้อมูลอนุกรมเวลาว่ามีผลจากอะไรบ้าง ข้อมูลอนุกรมเวลาที่สนใจในโครงการนี้คือ EEG (electroencephalogram) หรือการตรวจคลื่นไฟฟ้าทางสมอง เป็นการบันทึกสัญญาณไฟฟ้าที่เกิดจากผลรวมของกระแสไฟฟ้าของกลุ่มเซลล์ในสมอง โดยเลือกใช้แบบจำลองถดถอยตัวเอง (Autoregressive (AR) model) มาอธิบาย เนื่องจากแบบจำลอง AR นั้นค่าสัญญาณขาออกที่เวลาปัจจุบันจะขึ้นกับค่าของสัญญาณขาออกของเวลาในอดีตและสัญญาณรบกวนในปัจจุบันเท่านั้น ดังสมการ

$$y(t) = c + A_1y(t-1) + A_2y(t-2) + \dots + A_p y(t-p) + v(t) \quad (1)$$

เมื่อ $y(t)$ คือสัญญาณขาออก และ $v(t)$ คือสัญญาณรบกวน ที่เวลา t ดังนั้นแบบจำลอง AR จึงเป็นแบบจำลองที่ดีแบบจำลองหนึ่งสำหรับใช้ในการอธิบายสัญญาณ EEG

สำหรับโครงการนี้ การอธิบายความสัมพันธ์ของตัวแปรในข้อมูลอนุกรมเวลา หรือโครงสร้างเชิงสาเหตุ จะถูกอธิบายด้วยเงื่อนไขความสัมพันธ์ระหว่างเหตุและผล (Granger causality) โดยจาก [6] หลักการพื้นฐานของเงื่อนไขความสัมพันธ์ระหว่างเหตุและผลคือ ถ้าหากตัวแปร x มีผลต่อตัวแปร y แล้วการที่รู้ค่าของตัวแปร x ก่อน ย่อมจะช่วยให้การทำนายค่าของตัวแปร y มีความแม่นยำมากขึ้นกว่าการที่ไม่รู้ค่าของตัวแปร x มาก่อน ดังนั้นเมื่อนำมาประยุกต์ใช้กับแบบจำลอง AR เงื่อนไขความสัมพันธ์ระหว่างเหตุและผลของแบบจำลองนั้นสามารถเขียนได้ในรูปสมการเชิงเส้นของค่าสัมประสิทธิ์หรือค่าพารามิเตอร์ของแบบจำลอง หรือก็คือ y_j ไม่ใช่สาเหตุแบบ Granger ต่อ y_i ก็ต่อเมื่อ

$$(A_k)_{ij} = 0 \quad (2)$$

เมื่อ $(A_k)_{ij}$ คือสมาชิกตัวที่ (i, j) ของเมทริกซ์ A_k ซึ่งคือพารามิเตอร์ของแบบจำลอง AR สำหรับทุกค่า $k = 1, 2, \dots, p$

การประมาณค่าพารามิเตอร์ของแบบจำลองจะเลือกใช้วิธีการประมาณแบบความควรจะเป็นสูงสุด (maximum likelihood (ML) estimation) ถ้าค่าประมาณของพารามิเตอร์มีค่าเท่ากับศูนย์ จะสามารถบอกได้ว่าข้อมูลอนุกรมเวลาไม่สัมพันธ์กัน โดยปรกติแล้วเมทริกซ์ A_k ที่ประมาณได้อาจจะมีบางสมาชิกที่ไม่เท่ากับศูนย์ จึงต้องตั้งสมมติฐานขึ้นมาเพื่อทดสอบสมการ (2) ว่าเป็นจริงหรือไม่ กล่าวคือการทดสอบหารูปแบบโครงสร้างของค่าที่เป็นศูนย์ในพารามิเตอร์ของแบบจำลอง AR การทดสอบสมมติฐานนั้นจะนำหลักการของการทดสอบ Wald มาใช้ โดยจาก [5] การทดสอบ Wald มีหลักการคือ ถ้าสมมติฐานว่าค่าที่แท้จริงของพารามิเตอร์มีค่าเท่ากับศูนย์เป็นจริง แล้วค่าพารามิเตอร์ที่ประมาณได้ควรมีค่าเข้าใกล้ศูนย์อย่างมีนัยยะสำคัญ จากนั้นจะสามารถประมาณค่าพารามิเตอร์ใหม่ได้ภายใต้ข้อกำหนดที่ว่าค่าของสมาชิก

บางตัวของพารามิเตอร์มีค่าเป็นศูนย์ ซึ่งอยู่ในรูปแบบปัญหาการหาค่าที่เหมาะสมที่สุดภายใต้ข้อกำหนดค่าคงที่เป็นศูนย์ (optimization problem with zero constraints)

เนื่องจาก (1) เป็นกระบวนการสุ่ม (random process) ซึ่งต้องกำหนดให้เป็นข้อมูลอนุกรมเวลาที่เป็นกระบวนการคงที่ (stationary process) เนื่องจากกระบวนการที่เราสนใจจะอยู่ในรูปสมการสถานะเชิงเส้นที่มีสัญญาณขาเข้าแบบสุ่ม ดังนั้นกระบวนการนี้จะกลายเป็นกระบวนการคงที่ได้ก็ต่อเมื่อระบบเชิงเส้นมีเสถียรภาพ แต่ค่าพารามิเตอร์ที่ประมาณได้นั้นอาจส่งผลให้ระบบเชิงเส้นไม่มีเสถียรภาพได้ ดังนั้นจึงต้องมีการรวมเงื่อนไขเสถียรภาพเข้าไปในการประมาณพารามิเตอร์ เพื่อให้ระบบเชิงเส้นจากค่าพารามิเตอร์ที่ประมาณได้มีเสถียรภาพ และเป็นกระบวนการที่คงที่

โครงงานนี้มีจุดประสงค์ที่จะศึกษารูปแบบปัญหาการประมาณแบบจำลอง AR ที่รวม 2 คุณลักษณะที่สำคัญคือ 1. รูปแบบโครงสร้างของศูนย์ในพารามิเตอร์ที่จะแสดงถึงความสัมพันธ์ระหว่างเหตุและผลของแบบจำลอง และ 2. แบบจำลองที่ได้จะมีเสถียรภาพ โดยในรายงานนี้จะประกอบไปด้วย ระเบียบวิธีที่ใช้ในการทำโครงงานในหัวข้อที่ 2 ซึ่งจะอธิบายถึงการประมาณแบบจำลองด้วยวิธี ML, การทดสอบ Wald และการรวมเงื่อนไขความสัมพันธ์ระหว่างเหตุและผล และเสถียรภาพ ในหัวข้อที่ 3 จะแสดงผลการทดลองของการทดสอบ Wald, การหารูปแบบเงื่อนไขความสัมพันธ์ระหว่างเหตุและผล และการประมาณพารามิเตอร์ที่รวมเงื่อนไขเสถียรภาพและความสัมพันธ์ระหว่างเหตุและผล และสุดท้ายบทสรุปของโครงงานในหัวข้อที่ 4

2 ระเบียบวิธีที่ใช้ในการทำโครงงาน

2.1 การประมาณแบบจำลอง AR

พิจารณา (1) แบบจำลองถดถอยตัวเอง (Autoregressive (AR) model)

$$y(t) = c + A_1 y(t-1) + A_2 y(t-2) + \dots + A_p y(t-p) + v(t)$$

เมื่อ $y(t) \in \mathbb{R}^n = (y_1(t), y_2(t), \dots, y_n(t))$ เป็นค่าของสัญญาณขาออกของแบบจำลอง และ $y_i(t)$ เป็นสมาชิกตัวที่ i ของ $y(t)$, $A_1, A_2, \dots, A_p \in \mathbb{R}^{n \times n}$ เป็นค่าสัมประสิทธิ์หรือพารามิเตอร์ของแบบจำลอง, $c \in \mathbb{R}^n$ เป็นเวกเตอร์ค่าคงที่, p คืออันดับของแบบจำลอง AR และ $v(t)$ คือสัญญาณรบกวนซึ่งกำหนดให้เป็นสัญญาณรบกวนเกาส์สีขาว (Gaussian white noise) ที่มีค่าเฉลี่ยเป็นศูนย์ และเมทริกซ์ความแปรปรวนเป็น Σ เราสามารถเขียนแบบจำลอง (1) ในรูปแบบสมการเชิงเส้นในพารามิเตอร์ได้ดังนี้

$$y(t) = AH(t) + v(t) \tag{3}$$

เมื่อ $H(t) = [1 \ y(t-1)^T \ y(t-2)^T \ \dots \ y(t-p)^T]^T \in \mathbb{R}^{np+1}$ และ $v(t) \sim \mathcal{N}(0, \Sigma)$ ค่าพารามิเตอร์ของแบบจำลองที่ต้องการประมาณคือ $A = [c \ A_1 \ A_2 \ \dots \ A_p] \in \mathbb{R}^{n \times (np+1)}$ ซึ่งจะนำวิธีการประมาณแบบความควรจะเป็นสูงสุด (ML) มาใช้ประมาณค่าพารามิเตอร์นั้น โดยรูปแบบของ

ปัญหาจะเป็น

$$\underset{A, \Sigma}{\text{maximize}} \quad \frac{N-p}{2} \log \det \Sigma^{-1} - \frac{1}{2} \|L(Y - AH)\|_F^2 \quad (4)$$

โดยที่ L เป็นฟังก์ชันของ Σ^{-1} ด้วยความสัมพันธ์ $L^T L = \Sigma^{-1}$ และ Y กับ H เป็นพารามิเตอร์ของปัญหาที่มีค่าดังนี้

$$Y = [y(p+1) \quad y(p+2) \quad \cdots \quad y(N)]_{n \times (N-p)},$$

$$H = \begin{bmatrix} 1 & 1 & \cdots & 1 \\ y(p) & y(p+1) & \cdots & y(N-1) \\ y(p-1) & y(p) & \cdots & y(N-2) \\ \vdots & \vdots & & \vdots \\ y(1) & y(2) & \cdots & y(N-p) \end{bmatrix}_{(np+1) \times (N-p)}$$

และจะสามารถหาผลเฉลยของ A, Σ ในรูปแบบปิด (closed-form) คือ

$$\hat{A} = YH^T(HH^T)^{-1} \quad (5)$$

$$\hat{\Sigma} = \frac{1}{N-p} \sum_{t=p+1}^N (y(t) - \hat{A}H(t))(y(t) - \hat{A}H(t))^T \quad (6)$$

โดยจะสามารถคำนวณค่า $\hat{\Sigma}$ ได้เมื่อรู้ค่า \hat{A} แล้ว ซึ่งรูปแบบของปัญหา (4) และวิธีการคำนวณผลตอบจะถูกแสดงไว้ในส่วนของภาคผนวกหัวข้อที่ 6.1 และเราสามารถแสดงให้เห็นได้ว่าปัญหา (4) สมมูลกับปัญหากำลังสองน้อยสุด (Least-Squares (LS)) ภายใต้เงื่อนไขที่ว่า $v(t)$ ต้องเป็นสัญญาณรบกวนเกาส์สีขาว (Gaussian white noise)

$$\underset{A}{\text{minimize}} \quad \|Y - AH\|_F^2 \quad (7)$$

ที่มีตัวแปรคือค่าพารามิเตอร์ของแบบจำลอง AR (A)

ปัญหาดังกล่าวมีตัวแปรอยู่ในรูปเมทริกซ์ แต่เพื่อความสะดวกในการวิเคราะห์ปัญหาสำหรับการทดสอบ Wald ที่เราจะกล่าวถึงต่อไป เราสามารถพิจารณา (3) ในรูปแบบสมการที่มีพารามิเตอร์เป็นเวกเตอร์ได้คือ

$$y(t) = \bar{H}(t)\theta + v(t), \quad v(t) \sim \mathcal{N}(0, \Sigma) \quad (8)$$

โดยพารามิเตอร์ที่ต้องการประมาณของแบบจำลองคือ

$$\theta = [B_{11}^T \quad \cdots \quad B_{1n}^T \quad B_{21}^T \quad \cdots \quad B_{2n}^T \quad \cdots \quad B_{nn}^T]^T \in \mathbb{R}^{n^2 p} \quad (9)$$

เมื่อ $B_{ij} = ((A_1)_{ij}, (A_2)_{ij}, \dots, (A_p)_{ij}) \in \mathbb{R}^p$ และ

$$\bar{y}_i = \begin{bmatrix} y_i(t-1) \\ y_i(t-2) \\ \vdots \\ y_i(t-p) \end{bmatrix}_{p \times 1}, \quad \bar{y} = \begin{bmatrix} \bar{y}_1 \\ \bar{y}_2 \\ \vdots \\ \bar{y}_n \end{bmatrix}_{np \times 1}, \quad \bar{H}(t) = \begin{bmatrix} \bar{y}^T & 0 & 0 & \cdots & 0 \\ 0 & \bar{y}^T & 0 & \cdots & 0 \\ 0 & 0 & \bar{y}^T & \cdots & 0 \\ \vdots & \vdots & \vdots & \ddots & \vdots \\ 0 & 0 & 0 & \cdots & \bar{y}^T \end{bmatrix}_{n \times n^2 p}$$

เมื่อนำวิธีการประมาณแบบ ML มาใช้ประมาณค่าพารามิเตอร์นั้น รูปแบบของปัญหา (7) จะสมมูลกับ

$$\underset{\theta}{\text{minimize}} \quad \|F\theta - z\|_2^2 \quad (10)$$

เมื่อ

$$F = [\bar{H}^T(p+1) \quad \bar{H}^T(p+2) \quad \cdots \quad \bar{H}^T(N)]_{n(N-p) \times n^2 p}^T,$$

$$z = [y^T(p+1) \quad y^T(p+2) \quad \cdots \quad y^T(N)]_{n(N-p) \times 1}^T$$

2.2 การทดสอบ Wald บนค่าประมาณแบบจำลอง AR

การทดสอบ Wald

ในหัวข้อนี้ จะกล่าวถึงการทดสอบ Wald ซึ่งเป็นการทดสอบสมมติฐาน (hypothesis test) บนพารามิเตอร์ที่ได้จากการประมาณว่าเป็นไปตามสมมติฐานที่เราตั้งไว้หรือไม่ โดยเป็นเนื้อหาสรุปมาจาก [5]

เริ่มต้นจาก กำหนดให้ $\theta \in \mathbb{R}^p$ เป็นพารามิเตอร์ในแบบจำลอง ที่ถูกประมาณค่าด้วยวิธีการของ ML และให้ $\mathbf{Avar}(\theta)$ คือเมทริกซ์ค่าความแปรปรวนร่วมเชิงเส้นกำกับ (Asymptotic covariance matrix) ของ θ ซึ่งสามารถหาได้จาก

$$\mathbf{Avar}(\theta) = \frac{1}{N} \mathcal{I}(\theta)^{-1}$$

เมื่อ $\mathcal{I}(\theta)$ คือเมทริกซ์ Fisher Information ของ θ ที่คำนวณจาก 1 ตัวอย่างของ y โดย y คือตัวอย่างของข้อมูลและ N คือ จำนวนตัวอย่าง โดยนิยามของ $\mathcal{I}(\theta)$ คือ

$$\mathcal{I}(\theta) = \mathbf{E} [(\nabla_{\theta} \log f(y|\theta))(\nabla_{\theta} \log f(y|\theta))^T]$$

จาก Regularity conditions ค่าคาดหวังของ Hessian มีค่าเท่ากับค่าคาดหวังของ outer product ของเกรเดียนต์ ดังนั้นสามารถคำนวณค่า $\mathcal{I}(\theta)$ ได้อีกแบบจาก

$$\mathcal{I}(\theta) = -\mathbf{E} [\nabla_{\theta}^2 \log f(y|\theta)]$$

การทดสอบสมมติฐานแบบ Wald นั้น จะเริ่มต้นจากการตั้งสมมติฐานว่าง (null hypothesis)

$$H_0 : r(\theta) = 0$$

เมื่อ $r : \mathbb{R}^n \rightarrow \mathbb{R}^m, m \leq n$ คือฟังก์ชันข้อกำหนด (restriction function) บนพารามิเตอร์ θ ตัวอย่าง

เช่น $n = 3$ และเรามีสมมติฐานว่า $\theta_1 = 0, \theta_1 + \theta_3 = 0$ จะได้ว่าฟังก์ชันข้อกำหนด $r(\theta) = \begin{bmatrix} 1 & 0 & 0 \\ 1 & 0 & 1 \end{bmatrix} \begin{bmatrix} \theta_1 \\ \theta_2 \\ \theta_3 \end{bmatrix}$

การทดสอบ Wald ว่า H_0 เป็นจริงหรือไม่ จะใช้ค่าสถิติ

$$W = r(\hat{\theta})^T [D_r(\hat{\theta}) \widehat{\mathbf{Avar}}(\hat{\theta}) D_r(\hat{\theta})^T]^{-1} r(\hat{\theta}) \quad (11)$$

เมื่อ $D_r(\theta)$ คือจาโคเบียนของฟังก์ชัน r เทียบกับ θ และ $\widehat{\mathbf{Avar}}(\hat{\theta})$ คือค่าประมาณของเมทริกซ์ค่าความแปรปรวนช่วงเชิงเส้นกำกับของ $\hat{\theta}$ โดยสำหรับการประมาณแบบ ML ค่าของ $\widehat{\mathbf{Avar}}(\hat{\theta})$ สามารถหาได้จาก

$$\widehat{\mathbf{Avar}}(\hat{\theta}_{\text{ml}}) = \left[- \sum_{i=1}^N \nabla^2 \log f(y_i | \hat{\theta}) \right]^{-1} \quad (12)$$

จาก (11) หาก H_0 เป็นจริง ค่า $r(\hat{\theta})$ ควรจะมีค่าน้อยซึ่งจะทำให้ค่าของ W ควรจะมีค่าน้อยด้วย แต่จะมีค่าน้อยแบบมีนัยสำคัญเท่าใด จึงต้องรู้จากการกระจายตัวของ W โดยค่าทางสถิติของ Wald (W) จะมีการแจกแจงเข้าสู่การแจกแจงแบบไคสแควร์ (Chi-square distribution) โดยมีองศาอิสระเท่ากับ m กำหนดให้ α คือค่าระดับนัยสำคัญ (significance level) ซึ่งนิยามคือ

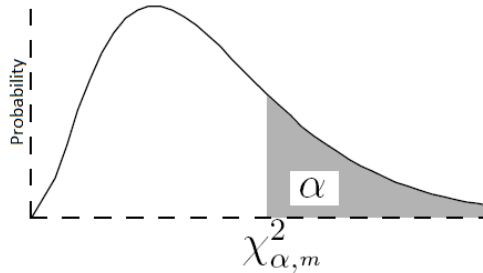
$$\alpha = \mathbf{Prob}(W > \chi_{\alpha, m}^2) = 1 - P(W \leq \chi_{\alpha, m}^2) \approx 1 - F(\chi_{\alpha, m}^2)$$

เมื่อ $F(\chi_{\alpha, m}^2)$ คือฟังก์ชันการแจกแจงแบบไคสแควร์ของ $\chi_{\alpha, m}^2$ และ $\chi_{\alpha, m}^2$ คือค่าวิกฤต (critical value) ดังนั้นจะได้ว่า $\chi_{\alpha, m}^2 = F^{-1}(1 - \alpha)$

การทดสอบสมมติฐานด้วยวิธีการทดสอบ Wald เราจะปฏิเสธ H_0 ถ้า

$$W > \chi_{\alpha, m}^2$$

กล่าวคือ ถ้า W มีค่ามากเกินไปอย่างมีนัยสำคัญ โดยจากรูป 1 พื้นที่แรเงาจะแสดงถึงบริเวณที่จะปฏิเสธสมมติฐานว่าง H_0



รูป 1: รูปแสดงการแจกแจงไคสแควร์

ตัวอย่างการทดสอบ Wald บนตัวประมาณกำลังสองน้อยสุด (Least-Squares)

พิจารณาความสัมพันธ์ระหว่าง y และ x จากสมการเชิงเส้น

$$y = Ax + v$$

โดยที่ $y \in \mathbb{R}^N$, $x \in \mathbb{R}^n$ และ $v \sim \mathcal{N}(0, \Sigma)$ เราสามารถประมาณค่าของ x ได้จากปัญหากำลังสองน้อยสุด (LS)

$$\hat{x} = \arg \min_x \|Ax - y\|^2 = (A^T A)^{-1} A^T y$$

ตัวอย่างการทดสอบที่พบบ่อยคือ การทดสอบว่า $\hat{x}_i = 0$ หรือไม่ สำหรับบาง i

กำหนดสมมติฐานว่าง $\hat{x}_i = 0$ ซึ่งในที่นี้จะได้ฟังก์ชันข้อกำหนดสำหรับการทดสอบ Wald ของ \hat{x}_i ใดๆ คือ $r(\hat{x}) = e_i^T \hat{x}$ และสำหรับประมาณของเมทริกซ์ค่าความแปรปรวนร่วมของ \hat{x} ของการประมาณกำลังสองน้อยสุดนั้น เป็นที่ทราบกันดีว่าอยู่ในรูป

$$\widehat{\text{Avar}}(\hat{x}) = (A^T A)^{-1} A^T \Sigma A (A^T A)^{-1}$$

และค่า W สำหรับการทดสอบ $\hat{x}_i = 0$ จากสมการ (11) จะลดรูปเป็น

$$W_i = \hat{x}_i^2 \left[\widehat{\text{Avar}}(\hat{x})_{ii} \right]^{-1} \quad \text{เมื่อ } i = 1, 2, \dots, n$$

โดยที่ $\widehat{\text{Avar}}(\hat{x})_{ii}$ หมายถึง สมาชิกตำแหน่งที่ (i, i) ของเมทริกซ์ $\widehat{\text{Avar}}(\hat{x})$

การทดสอบ Wald สำหรับเงื่อนไข Granger causality

การทดสอบ Wald สำหรับเงื่อนไข Granger causality นั้นคือการทดสอบสมมติฐานที่ว่า $(A_k)_{ij} = 0$ หรือไม่ เพื่อสังเกตรูปแบบโครงสร้างของศูนย์ในพารามิเตอร์ของแบบจำลองที่ถูกประมาณได้เมื่อใช้ค่าระดับนัยสำคัญต่างๆกัน โดยจะทำการทดสอบ Wald บนแบบจำลอง (8)

การทดสอบ Granger causality จะสอดคล้องกับฟังก์ชันข้อกำหนดสำหรับแต่ละคู่ (i, j) คือ

$$r(\hat{\theta}) = \begin{bmatrix} (\hat{A}_1)_{ij} \\ (\hat{A}_2)_{ij} \\ \vdots \\ (\hat{A}_p)_{ij} \end{bmatrix} = \hat{B}_{ij} \quad (13)$$

และค่าประมาณของเมทริกซ์ค่าความแปรปรวนร่วมของ $\hat{\theta}$ ซึ่งหาได้จาก (12) คือ

$$\widehat{\text{Avar}}(\hat{\theta}) = \left[\sum_{t=p+1}^N \bar{H}(t)^T \hat{\Sigma}^{-1} \bar{H}(t) \right]^{-1}$$

และค่าทางสถิติของ Wald (W) สำหรับ $r(\theta)$ จาก (11) สามารถจัดรูปได้เป็น

$$W_{ij} = \hat{B}_{ij}^T \left[\widehat{\text{Avar}}(\hat{\theta})_{ij} \right]^{-1} \hat{B}_{ij}$$

โดยกำหนดให้ $\widehat{\text{Avar}}(\hat{\theta})_{ij}$ คือเมทริกซ์ย่อยขนาด $p \times p$ ในแนวทแยงมุม (main diagonal block) ของ $\widehat{\text{Avar}}(\hat{\theta})$ ที่สอดคล้องกับพารามิเตอร์ B_{ij} ดังนั้นเราสามารถกำหนดให้ $W = [W_{ij}]$ เป็นเมทริกซ์ทางสถิติของ Wald

จากนั้นจึงใช้การทดสอบ Wald สำหรับทดสอบ $(A_k)_{ij} = 0$ เมื่อ $k = 1, 2, \dots, p$ โดยทดสอบทีละคู่ (i, j) สำหรับ $j \neq i, j = 1, 2, \dots, n$ ทั้งหมด $n^2 - n$ ครั้ง หลังจากทำการทดสอบ Wald จะสามารถหาโครงสร้างเชิงสาเหตุของพารามิเตอร์ A ของแบบจำลอง AR ได้ออกมาในรูปแบบโครงสร้างของศูนย์ ซึ่งผลการทดลองจะถูกกล่าวถึงในหัวข้อที่ 3.2

2.3 การประมาณแบบจำลอง AR ที่รวมเงื่อนไข Granger causality

การทดสอบ Wald บนพารามิเตอร์ของแบบจำลอง AR เมื่อไม่มีข้อกำหนดใดๆ เราจะได้เงื่อนไขความสัมพันธ์ระหว่างเหตุและผลมาในรูปแบบโครงสร้างของศูนย์ ดังนั้นถ้าเราต้องการแบบจำลองที่สอดคล้องกับเงื่อนไขความสัมพันธ์ระหว่างเหตุและผลข้างต้น เราจึงรวมเงื่อนไขในการประมาณพารามิเตอร์ของปัญหา (10) ได้ โดยจะเห็นว่า การประมาณแบบจำลอง AR แบบไม่มีเงื่อนไขใดๆ เราจะใช้ ML ซึ่งสมมูลกับ LS แต่เมื่อรวมเงื่อนไขความสัมพันธ์ระหว่างเหตุและผลแล้วนั้น เราเลือกที่จะใช้ฟังก์ชันจุดประสงค์เป็นแบบกำลังสองน้อยสุด (least-square (LS)) แทน เพราะจะสามารถแก้หาผลเฉลยได้ง่ายกว่า ซึ่งในภาคผนวกหัวข้อที่ 6.3 จะแสดงให้เห็นว่ารูปแบบของปัญหาที่เป็นแบบ ML และ LS ที่รวมเงื่อนไขความสัมพันธ์ระหว่างเหตุและผลนั้น แท้จริงแล้วเป็นปัญหาที่ต่างกัน

ดังนั้น เราสามารถแสดงให้เห็นว่า ปัญหา LS ที่รวมเงื่อนไขความสัมพันธ์ระหว่างเหตุและผล จะจัดรูปได้เป็น

$$\begin{aligned} & \underset{\theta}{\text{minimize}} \quad \|F\theta - z\|_2^2 \\ & \text{subject to} \quad P\theta = 0 \end{aligned} \tag{14}$$

เมื่อ P คือเมทริกซ์การฉาย (projection matrix) ตัวอย่างเช่น $P\theta = \begin{bmatrix} 0 & I & 0 & \cdots & 0 \\ 0 & 0 & I & \cdots & 0 \end{bmatrix} \theta$ จะหมายถึง

ถึงเงื่อนไขว่า $B_{12} = 0$ และ $B_{13} = 0$

ในการหาคำตอบ (14) นั้น เราสามารถพิจารณาปัญหาในรูปแบบพิเศษดังนี้

$$\begin{aligned} & \underset{\theta}{\text{minimize}} \quad \|F\theta - z\|_2^2 \\ & \text{subject to} \quad \theta_i = 0 \end{aligned} \tag{15}$$

กล่าวคือ เราต้องการให้ θ_i บางตัวเป็นศูนย์ เมื่อ $i \in \{1, 2, \dots, n\}$ ซึ่งจะถูกกำหนดมาให้ โดยที่ θ_i คือสมาชิกตัวที่ i ของ θ และ f_i คือคอลัมน์ที่ i ของเมทริกซ์ F จะเห็นว่า

$$F\theta - z = \left(\sum_{k=1}^n f_k \theta_k \right) - z = \left(\sum_{k=1, k \neq i}^n f_k \theta_k \right) - z = \tilde{F}\tilde{\theta} - z$$

ให้ $\tilde{F} = [f_1 \ \cdots \ f_{i-1} \ f_{i+1} \ \cdots \ f_n]$ และ $\tilde{\theta} = (\theta_1, \dots, \theta_{i-1}, \theta_{i+1}, \dots, \theta_n)$

จะได้ผลเฉลยคือ $\tilde{\theta} = (\tilde{F}^T \tilde{F})^{-1} \tilde{F}^T z$ ดังนั้นเราจะหาคำตอบของปัญหากำลังสองน้อยสุด (15) คือ

$$\hat{\theta} = (\hat{\theta}_1, \dots, \hat{\theta}_{i-1}, 0, \hat{\theta}_{i+1}, \dots, \hat{\theta}_n) \quad \text{เมื่อ } \theta_i = 0 \tag{16}$$

จะเห็นว่าเมทริกซ์ \tilde{F} คือเมทริกซ์ F ที่นำเอาคอลัมน์ที่ i ออกไป สำหรับเงื่อนไข $\theta_i = 0$ ใดๆ ดังนั้นสำหรับปัญหา (14) ผลตอบก็จะสามารถหาได้ในทำนองเดียวกันกับปัญหา (15) โดยที่เมทริกซ์ \tilde{F} จะเป็นการนำเอาคอลัมน์ที่สอดคล้องกับเงื่อนไขของปัญหาออกไป เช่น ถ้า $P\theta = 0$ หมายถึงการกำหนดให้ $B_{12} = 0, B_{13} = 0$ จะได้ว่า \tilde{F} คือเมทริกซ์ F ที่เอาเมทริกซ์ย่อย คอลัมน์ที่ 2,3 ออกไป

2.4 การรวมเงื่อนไขเสถียรภาพในการประมาณแบบจำลอง

กระบวนการสุ่ม (1) จะเป็นกระบวนการสุ่มคงที่ (stationary process) ก็ต่อเมื่อระบบเชิงเส้นที่สอดคล้องกันนั้น เป็นระบบที่มีเสถียรภาพ แต่เนื่องจากพารามิเตอร์ที่ประมาณได้อาจส่งผลให้แบบจำลองไม่มีเสถียรภาพได้ เราจึงต้องรวมเงื่อนไขเสถียรภาพเข้าไปในการประมาณพารามิเตอร์ ในหัวข้อนี้จะสรุปเงื่อนไขการมีเสถียรภาพของระบบเชิงเส้น ดังนี้

เงื่อนไขจำเป็นและเพียงพอสำหรับเสถียรภาพ

พิจารณา สมการเชิงเส้นเวลาวิฤต (discrete-time linear system) ในรูปทั่วไป

$$x(t+1) = \mathcal{A}x(t) + \mathcal{B}u(t) \quad (17)$$

เมื่อ x คือตัวแปรสถานะและ \mathcal{A} คือเมทริกซ์พลวัต (dynamic matrix)

เราจะพิจารณาเสถียรภาพของแบบจำลอง (1) โดยเขียนให้อยู่ในรูปสมการเชิงเส้นเวลาวิฤตได้ดังนี้

$$\begin{bmatrix} y(t) \\ y(t-1) \\ \vdots \\ y(t-p+1) \end{bmatrix} = \begin{bmatrix} A_1 & A_2 & \dots & A_{p-1} & A_p \\ I & 0 & \dots & 0 & 0 \\ 0 & I & \dots & 0 & 0 \\ \vdots & \vdots & \ddots & \vdots & \vdots \\ 0 & 0 & \dots & I & 0 \end{bmatrix} \begin{bmatrix} y(t-1) \\ y(t-2) \\ \vdots \\ y(t-p+1) \\ y(t-p) \end{bmatrix} + \begin{bmatrix} v(t) \\ 0 \\ \vdots \\ 0 \end{bmatrix} \quad (18)$$

โดยที่เมทริกซ์พลวัตของ (18) คือ

$$\mathcal{A} = \begin{bmatrix} A_1 & A_2 & \dots & A_{p-1} & A_p \\ I & 0 & \dots & 0 & 0 \\ 0 & I & \dots & 0 & 0 \\ \vdots & \vdots & \ddots & \vdots & \vdots \\ 0 & 0 & \dots & I & 0 \end{bmatrix}_{np \times np} \quad (19)$$

1. เงื่อนไขบนค่าเจาะจงของเมทริกซ์พลวัต

ระบบเชิงเส้นเวลาวิฤตจะมีเสถียรภาพก็ต่อเมื่อค่าเจาะจง (eigenvalue) ทุกตัวของ $\mathcal{A} \in \mathbb{R}^{np \times np}$ มีขนาดน้อยกว่าหนึ่ง นั่นคือ

$$|\lambda_i(\mathcal{A})| < 1, \quad i = 1, 2, \dots, np. \quad (20)$$

เมื่อ $\lambda_i(\mathcal{A})$ คือค่าเจาะจงตัวที่ i ของ \mathcal{A}

สำหรับการประมาณพารามิเตอร์ \mathcal{A} ที่มีเสถียรภาพนั้น คือการหาค่า A_1, A_2, \dots, A_p ที่ทำให้ (20) เป็นจริง ซึ่งออกมาในรูปสมการไม่เชิงเส้น (nonlinear) และยากต่อการนำไปใช้ในการคำนวณจริง

2. ทฤษฎีบทลีสยูนอฟเวลาวิฤต

พิจารณา สมการเลียปูนอฟเวลาวิฤต (discrete-time Lyapunov equation)

$$\mathcal{A}^T P \mathcal{A} - P + Q = 0 \quad (21)$$

สมการข้างต้นจะใช้ในการวิเคราะห์เสถียรภาพของระบบพลวัตเชิงเส้นเวลาวิฤต (17)

ทฤษฎีบทเลียปูนอฟเวลาวิฤต กล่าวว่า

- สำหรับ \mathcal{A} และ $Q \succ 0$ ที่กำหนดให้ ถ้าสามารถหาค่า $P \succ 0$ ที่สอดคล้องกับ (21) จะได้ว่าระบบ (17) มีเสถียรภาพ
- บทกลับ : ถ้าระบบ (17) มีเสถียรภาพแล้ว จะมี $P \succ 0$ และ $Q \succ 0$ ที่สอดคล้องกับ (21) เสมอ

ดังนั้นกล่าวโดยสรุปคือ ระบบ $x(t+1) = \mathcal{A}x(t)$ จะมีเสถียรภาพ ก็ต่อเมื่อมีค่า $P \succ 0$ และ $Q \succ 0$ ซึ่งสอดคล้องกับ (21)

ในปัญหาหาค่าความคุม \mathcal{A} จะถูกกำหนดให้ และเราจะสามารถตรวจสอบการมีเสถียรภาพของ \mathcal{A} ได้จากการหา $P \succ 0$ ที่สอดคล้องกับ (21) แต่ในปัญหาการระบุเอกลักษณ์ของระบบ (system identification) \mathcal{A} จะเป็นพารามิเตอร์ที่เราไม่ทราบค่าและต้องการหา ดังนั้นถ้าเราเลือกที่จะกำหนด $P \succ 0$ มาก่อน เช่น กำหนดให้ $P = I$ และ $Q \succ 0$ สมการ (21) จะสอดคล้องกับ $\mathcal{A}^T \mathcal{A} \prec I$ ซึ่งสมมูลกับ

$$\|\mathcal{A}\|_2 < 1 \quad (22)$$

เมื่อหา \mathcal{A} ที่สอดคล้องกับเงื่อนไข (22) แล้ว ระบบพลวัตเชิงเส้นเวลาวิฤต $x(t+1) = \mathcal{A}x(t)$ จะมีเสถียรภาพ ซึ่งเงื่อนไขข้างต้นนั้นเป็นแค่เงื่อนไขพอเพียงเท่านั้น แต่ไม่ใช่เงื่อนไขจำเป็นสำหรับการมีเสถียรภาพของระบบพลวัตเชิงเส้นเวลาวิฤต

พิจารณาเงื่อนไข $\|\mathcal{A}\|_2 \leq 1$ จาก \mathcal{A} ที่มีโครงสร้างดัง (19) กำหนดให้ $A = [A_1 \ A_2 \ \dots \ A_{p-1}]$ และ $B = A_p$ จะได้

$$\mathcal{A} = \begin{bmatrix} A & B \\ I & 0 \end{bmatrix}$$

เนื่องจากเงื่อนไข $\|\mathcal{A}\|_2 \leq 1$ สมมูลกับ $\mathcal{A}^T \mathcal{A} \preceq I$ จะได้ว่า

$$\mathcal{A}^T \mathcal{A} = \begin{bmatrix} A^T & I \\ B^T & 0 \end{bmatrix} \begin{bmatrix} A & B \\ I & 0 \end{bmatrix} = \begin{bmatrix} A^T A + I & A^T B \\ B^T A & B^T B \end{bmatrix}$$

โดยถ้า $\mathcal{A}^T \mathcal{A} \preceq I$ จะได้ว่า

$$I - \mathcal{A}^T \mathcal{A} = \begin{bmatrix} -A^T A & -A^T B \\ -B^T A & I - B^T B \end{bmatrix} \succeq 0$$

จะเป็นจริงเมื่อ $-A^T A \succeq 0$ ซึ่งเกิดขึ้นได้กรณีเดียวเมื่อ $A = 0$ หรือก็คือ A_1, A_2, \dots, A_{p-1} ต้องมีค่าเป็นศูนย์ทั้งหมด ซึ่งจะกลายเป็นผลตอบที่ไม่มีมีความหมาย เนื่องจากพารามิเตอร์ของแบบจำลอง

A_1, \dots, A_p ต้องมีรูปแบบโครงสร้างของศูนย์แบบเดียวกัน และจะได้ว่า $y(t)$ ขึ้นกับ A_p อย่างเดียว ซึ่งจะเป็นการสรุปที่เฉพาะเจาะจงเกินไป ดังนั้นเราจะไม่พิจารณา $\|A\|_2 \leq 1$ เป็นเงื่อนไขในข้อกำหนดเสถียรภาพ

เงื่อนไขเพียงพอสำหรับเสถียรภาพ

นอกจากเงื่อนไขจำเป็นและเพียงพอสำหรับเสถียรภาพดังที่ได้กล่าวมาข้างต้นแล้ว ยังมีบางเงื่อนไขที่พอเพียงสำหรับการมีเสถียรภาพของระบบเชิงเส้นเวลาวิฤต ดังนี้

1. Gershgorin circle

จาก [7] กำหนดให้ $A = [a_{ij}]$ เป็นเมทริกซ์จัตุรัสขนาด $n \times n$ ใดๆ Gershgorin circles คือ วงกลมบนระนาบเชิงซ้อน n วง ที่มีจุดศูนย์กลางอยู่ที่พิกัด a_{ii} และมีรัศมีเท่ากับ R_i โดยที่ R_i มีนิยามคือ

$$R_i = \sum_{j=1, j \neq i}^n |a_{ij}| \quad i = 1, 2, \dots, n \quad (23)$$

ทฤษฎี Gershgorin กล่าวไว้ว่า ค่าเฉพาะของทุกตัวของ A จะมีค่าอยู่ภายในหรือบนวงกลม Gershgorin อย่างน้อยหนึ่งวง หรือก็คือ

$$|\lambda - a_{ii}| \leq R_i \quad , i = 1, 2, \dots, n \quad (24)$$

เมื่อ λ คือค่าเฉพาะของ A

บทพิสูจน์ทฤษฎี Gershgorin circles สามารถอธิบายได้อย่างกระชับและง่ายดังนี้ กำหนดให้ λ คือค่าเฉพาะของ $A \in \mathbb{R}^{n \times n}$ ซึ่งสอดคล้องกับเวกเตอร์เฉพาะ x และให้ $i \in 1, \dots, n$ ถูกเลือกโดยที่ x_i เป็นเวกเตอร์เฉพาะที่มีขนาดใหญ่มากที่สุด ดังนั้น $|x_i| > 0$ ไม่เช่นนั้น x จะมีค่าเท่ากับศูนย์จะได้ว่า

$$\begin{aligned} \sum_j a_{ij} x_j &= \lambda x_i \\ \sum_{j \neq i} a_{ij} x_j &= \lambda x_i - a_{ii} x_i \\ |\lambda - a_{ii}| &= \left| \frac{\sum_{j \neq i} a_{ij} x_j}{x_i} \right| \leq \sum_{j \neq i} \left| \frac{a_{ij} x_j}{x_i} \right| \leq \sum_{j \neq i} |a_{ij}| = R_i \end{aligned}$$

จากผลลัพธ์ใน (24) นั่นคือ ค่าเฉพาะแต่ละตัวต้องอยู่ภายในหรือบนวงกลม Gershgorin อย่างน้อยหนึ่งวง ดังนั้นการหา A ที่มีเสถียรภาพ คือการบังคับให้วงกลม Gershgorin ทุกวงอยู่ในวงกลมหนึ่งหน่วย ซึ่งจะทำให้ $|\lambda| < 1$ เป็นจริงตาม (24) ด้วย ดังนั้นถ้าเรากำหนดให้

$$\max_i (R_i + |a_{ii}|) \leq 1$$

แทน R_i จาก (23) จะได้ว่า

$$\max_i \sum_{j=1}^n |a_{ij}| \leq 1 \quad i = 1, 2, \dots, n \quad (25)$$

และเนื่องจากเมทริกซ์ A และ A^T มีค่าเงาจะงเหมือนกัน จะได้ว่าเปลี่ยนตำแหน่ง (i, j) ใน (25) จะกลายเป็น

$$\max_j \sum_{i=1}^n |a_{ij}| \leq 1 \quad j = 1, 2, \dots, n \quad (26)$$

เนื่องจาก $\|A\|_\infty$ คือค่าผลรวมของค่าสัมบูรณ์ของค่า a_{ij} ในแถวแต่ละแถวที่มากที่สุดของเมทริกซ์ A (the maximum absolute row sum of the matrix A) และ $\|A\|_1$ คือค่าผลรวมของค่าสัมบูรณ์ของค่า a_{ij} ในคอลัมน์แต่ละคอลัมน์ที่มากที่สุดของเมทริกซ์ A (the maximum absolute column sum of the matrix A) ดังนั้นเงื่อนไข (25) และ (26) จะสอดคล้องกับ $\|A\|_\infty \leq 1$ หรือ $\|A\|_1 \leq 1$

พิจารณาเงื่อนไข $\|A\|_1 \leq 1$ จากโครงสร้างของ A ดัง (19) ถ้าเราต้องการให้เงื่อนไขนี้เป็นจริง เนื่องจากแต่ละคอลัมน์ของ A จะมี 1 อยู่เสมอ ยกเว้นคอลัมน์บล็อกที่มี A_p เราจะได้ว่า A_1, A_2, \dots, A_{p-1} ต้องมีค่าเป็นศูนย์ทั้งหมด ซึ่งจะกลายเป็นผลตอบที่ไม่มี ความหมาย เนื่องจากพารามิเตอร์ของแบบจำลองต้องมีรูปแบบโครงสร้างของศูนย์แบบเดียวกัน ดังนั้นเราจะไม่พิจารณาเงื่อนไขนี้ในข้อกำหนดให้มีเสถียรภาพ

2. ความสัมพันธ์ระหว่าง induced norm และ spectral radius

กำหนดให้ $x \in \mathbb{R}^n$ เป็นเวกเตอร์เงาจะงของ $A \in \mathbb{R}^{n \times n}$ ที่มีค่าเงาจะง λ โดยที่ $\max_i |\lambda_i| = \rho(A)$

คือ spectral radius ดังนั้น A จะมีเสถียรภาพก็ต่อเมื่อ $\rho(A) < 1$

ให้ $v \in \mathbb{R}^n$ เป็นเวกเตอร์ใดๆ เราจะได้ความสัมพันธ์ระหว่าง $\rho(A)$ กับ induced norm ของ A โดยดังนี้

$$\|A\| = \sup_{v \neq 0} \frac{\|Av\|}{\|v\|} \geq \frac{\|Ax\|}{\|x\|} = \frac{|\lambda| \|x\|}{\|x\|} = |\lambda|$$

หรือก็คือ

$$\rho(A) \leq \|A\|$$

ดังนั้นการกำหนดให้ $\|A\| < 1$ จึงเป็นการบังคับให้ $\rho(A) < 1$ หรือก็คือ A จะมีเสถียรภาพนั่นเอง โดยที่ $\|A\|$ เป็น induced norm

จากที่ผ่านมามีผู้ทดลองการรวมเงื่อนไขเสถียรภาพในการประมาณแบบจำลองด้วยวิธีการต่างๆ เช่น การใช้ Jury's test บนระบบเชิงเส้นเวลายืนยงแบบสัญญาณขาเข้าและสัญญาณขาออกเดี่ยว (Single Input Single Output (SISO) Linear-Time-Invariant (LTI) system) จาก [2] โดยจะหาค่าประมาณ

พารามิเตอร์ของระบบเชิงเส้นที่สอดคล้องกับตารางของ Jury เพื่อรับประกันว่าระบบเชิงเส้นจะมีเสถียรภาพแบบ BIBO ซึ่งโครงการนี้จะระบบแบบหลายสัญญาณขาเข้าและหลายสัญญาณขาออก (Multi Input Multi Output (MIMO) Linear system), การใช้ทฤษฎีของ Gershgorin circle บนแบบจำลอง Autoregressive Moving Average with Exogenous Input (ARMAX) จาก [10] โดยเขียนแบบจำลองให้อยู่ในรูปของสมการสถานะ และหาค่าประมาณพารามิเตอร์ของแบบจำลองที่สอดคล้องกับเงื่อนไข (25) ซึ่งจะคล้ายคลึงกันกับโครงการนี้, การนำทฤษฎีบทเลียปูนอฟมาประยุกต์ใช้กับแบบจำลอง Autoregressive Moving Average (ARMA) แบบเวกเตอร์ จาก [8] และ [9] โดยจะประมาณค่าพารามิเตอร์ของแบบจำลอง และจากนั้นหาค่าประมาณพารามิเตอร์ของแบบจำลองใหม่ที่สอดคล้องกันกับสมการเลียปูนอฟ โดยนอร์มของพารามิเตอร์มีขนาดเปลี่ยนแปลงไปจากเดิมน้อยที่สุด ซึ่งจะประมาณ P และ A พร้อมกัน แต่ถ้านำมาประยุกต์ใช้กับโครงการนี้จะมีความซับซ้อนมากขึ้นไปอีกจากเงื่อนไขความสัมพันธ์ระหว่างเหตุและผล ดังนั้นจึงไม่เลือกทำตามวิธีนี้ และจาก [1] จะประยุกต์ใช้ทฤษฎีบทเลียปูนอฟบนแบบจำลองระบบพลวัตเชิงเส้น (linear dynamical systems (LDS) model) โดยจะประมาณค่าพารามิเตอร์ของแบบจำลองที่สอดคล้องกับเงื่อนไข $A^T A < I$ เมื่อ A คือพารามิเตอร์ของแบบจำลอง แต่เนื่องจาก A มีโครงสร้างเหมือน VAR(1) ซึ่งแตกต่างจากโครงสร้างของพารามิเตอร์ในโครงการนี้ อย่างไรก็ตามการทดลองทั้งหมดมีจุดประสงค์เพื่อที่จะบังคับให้ค่าพารามิเตอร์ของแบบจำลองต่างๆที่ประมาณได้ส่งผลให้ระบบมีเสถียรภาพ

2.5 การประมาณแบบจำลองถดถอยตัวเองที่รวมเงื่อนไขความสัมพันธ์ระหว่างเหตุและผลและเสถียรภาพ

ในหัวข้อนี้จะนำเสนอวิธีการประมาณพารามิเตอร์ของแบบจำลองถดถอยตัวเองซึ่งรวมเงื่อนไขความสัมพันธ์ระหว่างเหตุและผลและเงื่อนไขเสถียรภาพ เนื่องจากพารามิเตอร์ที่ถูกประมาณได้หลังจากรวมเงื่อนไขความสัมพันธ์ระหว่างเหตุและผลแล้วอาจจะส่งผลให้ระบบเชิงเส้น (18) ไม่มีเสถียรภาพ

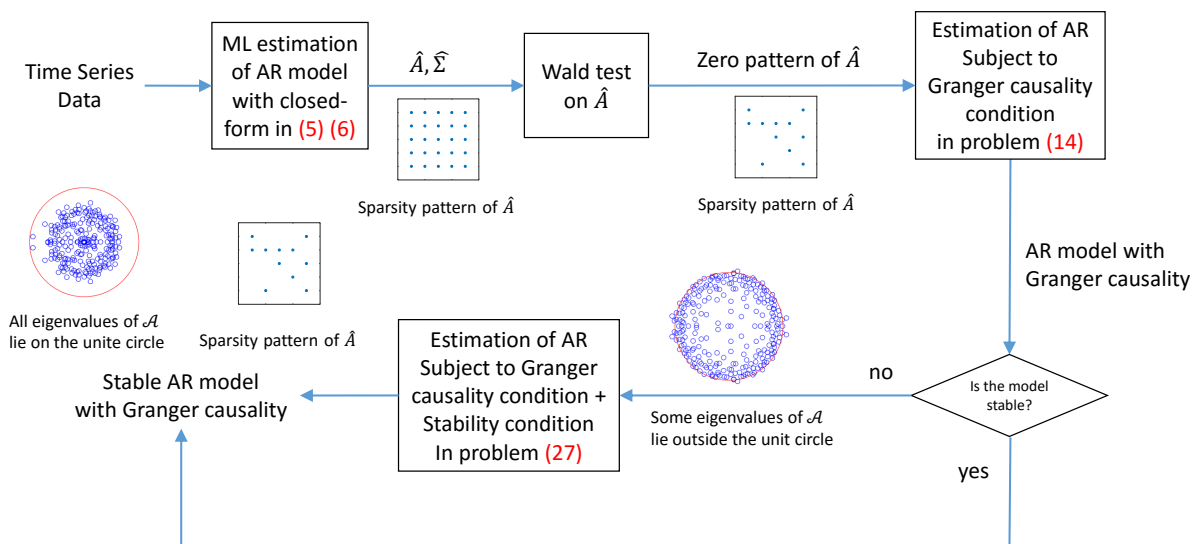
พารามิเตอร์ที่ประมาณได้จะถูกทดสอบด้วยการทดสอบ Wald เพื่อหารูปเงื่อนไขความสัมพันธ์ระหว่างเหตุและผล ซึ่งจะถูกแสดงในรูปแบบโครงสร้างของศูนย์ หลังจากนั้นจะประมาณค่าพารามิเตอร์ใหม่อีกครั้งให้สอดคล้องกับโครงสร้างของศูนย์ที่ทดสอบได้ และรวมเงื่อนไขเสถียรภาพเข้าไป พารามิเตอร์ที่ถูกประมาณได้ในท้ายที่สุดคือจุดประสงค์ของโครงการนี้ รูปแบบของปัญหาจะเป็นการแก้ปัญหา optimization ที่มีเงื่อนไขทั้ง 2 ข้างต้น

$$\begin{aligned} & \underset{A}{\text{minimize}} \quad \|Y - AH\|_F^2 \\ & \text{subject to} \quad \mathcal{A} = \begin{bmatrix} A_1 & A_2 & \dots & A_{p-1} & A_p \\ I & 0 & \dots & 0 & 0 \\ 0 & I & \dots & 0 & 0 \\ \vdots & \vdots & \ddots & \vdots & \vdots \\ 0 & 0 & \dots & I & 0 \end{bmatrix}, \\ & \|\mathcal{A}\|_\infty \leq 1, \\ & (A_k)_{ij} = 0, \quad (i, j) \in I \end{aligned} \tag{27}$$

เมื่อ $I \subset \{1, 2, \dots, n\} \times \{1, 2, \dots, n\}$ คือเซตของคู่อันดับที่เป็นตัวกำหนดโครงสร้างของศูนย์ใน A_k และเงื่อนไข $(A_k)_{ij} = 0$ จะหาจากการทดสอบ Wald ซึ่งจะได้รูปแบบโครงสร้างของศูนย์ออกมา โดยผลลัพธ์จะถูกแสดงไว้ในหัวข้อ 3.4

2.6 แผนผังการหาแบบจำลองถดถอยตัวเองที่รวมเงื่อนไขความสัมพันธ์ระหว่างเหตุและผลและเสถียรภาพ

การหาแบบจำลอง AR ที่มีเสถียรภาพและสอดคล้องกับเงื่อนไขความสัมพันธ์ระหว่างเหตุและผล สามารถสรุปได้ตามแผนผังด้านล่าง



เราสามารถอธิบายขั้นตอนตามแผนผังได้ดังนี้

1. ข้อมูลอนุกรมเวลาสามารถอธิบายได้ด้วยแบบจำลอง AR ซึ่งสามารถหาค่าพารามิเตอร์ของแบบจำลอง \hat{A} และ $\hat{\Sigma}$ ได้ด้วยวิธีการประมาณแบบ ML

2. ทดสอบสมมติฐานแบบ Wald บนพารามิเตอร์ \hat{A} เพื่อหารูปแบบเงื่อนไขความสัมพันธ์ระหว่างเหตุและผล ในรูปของรูปแบบโครงสร้างของศูนย์ ซึ่งสอดคล้องกับค่าระดับนัยสำคัญหนึ่งๆ
3. เมื่อได้รูปแบบโครงสร้างของศูนย์แล้ว เราสามารถประมาณพารามิเตอร์ \hat{A} ใหม่ให้สอดคล้องกับเงื่อนไขความสัมพันธ์ระหว่างเหตุและผล ดังปัญหา (14)
4. เมื่อประมาณพารามิเตอร์ใหม่ได้แล้ว เราต้องทำการตรวจสอบความมีเสถียรภาพของแบบจำลองที่ได้ เนื่องจากถ้าแบบจำลองมีเสถียรภาพอยู่แล้ว การรวมเงื่อนไขเสถียรภาพ $\|A\|_\infty \leq 1$ เข้าไปอีกจะเป็นการจำกัดขอบเขตของแบบจำลองที่เป็นไปได้ให้มีขนาดเล็กลง ส่งผลให้การประมาณอาจมีค่าฟังก์ชันจุดประสงค์ที่สูงขึ้น
5. ถ้าพารามิเตอร์ที่ประมาณได้จากข้อ 3 ส่งผลให้แบบจำลองมีเสถียรภาพ เราก็จะได้แบบจำลอง AR ที่มีเสถียรภาพและสอดคล้องกับเงื่อนไขความสัมพันธ์ระหว่างเหตุและผล
6. แต่ถ้าพารามิเตอร์ที่ประมาณได้จากข้อ 3 ส่งผลให้แบบจำลองไม่มีเสถียรภาพ เราจึงต้องรวมเงื่อนไขการมีเสถียรภาพเข้าไปในการประมาณพารามิเตอร์ใหม่ ดังปัญหา (27) จึงจะได้แบบจำลอง AR ที่มีเสถียรภาพและสอดคล้องกับเงื่อนไขความสัมพันธ์ระหว่างเหตุและผล

3 ผลการทดลอง

3.1 การทดสอบ Wald บนตัวประมาณกำลังสองน้อยสุด

เราจะสร้างข้อมูลขึ้นมาเองจากสมการ

$$y = Ax + v \quad (28)$$

โดย $y \in \mathbb{R}^N$ เมื่อ N คือจำนวนตัวอย่างของข้อมูล ให้ $x \in \mathbb{R}^n, n = 100$ หรือก็คือจำนวนพารามิเตอร์ที่ต้องการประมาณมีทั้งหมด 100 ตัว และ $v \sim \mathcal{N}(0, \sigma^2 I_N)$ โดยให้ $\sigma^2 = \|x\|$ และกำหนดให้ค่าจริงของ x_i บางตัวมีค่าเท่ากับศูนย์และนำไปใช้ในการทดสอบ Wald ต่อไปเพื่อพิจารณาความแม่นยำของการทดสอบ Wald ซึ่งจะพิจารณาจากค่าความผิดพลาดทั้งสองแบบของการทดสอบสมมติฐาน โดยค่าความผิดพลาดแบบที่หนึ่ง (Type I error) คือค่าความน่าจะเป็นที่ค่าพารามิเตอร์ที่ประมาณได้ไม่เท่ากับศูนย์เมื่อกำหนดให้ค่าจริงของพารามิเตอร์เท่ากับศูนย์

$$\text{Type I Error} = \mathbf{Prob}(\hat{x}_i \neq 0 \mid x_i = 0)$$

และค่าความผิดพลาดแบบที่สอง (Type II error) คือค่าความน่าจะเป็นที่ค่าพารามิเตอร์ที่ประมาณได้เท่ากับศูนย์เมื่อกำหนดให้ค่าจริงของพารามิเตอร์ไม่เท่ากับศูนย์

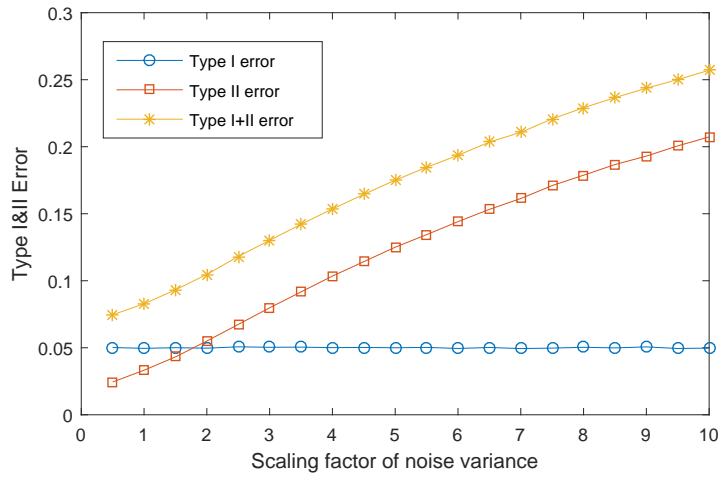
$$\text{Type II Error} = \mathbf{Prob}(\hat{x}_i = 0 \mid x_i \neq 0)$$

จากนั้นจึงทำการทดสอบ Wald โดยทดสอบบนสมการ (28) ที่สร้างขึ้นทั้งหมด 10,000 ครั้ง แต่ละครั้งจะคงค่า A และ x ไว้ ค่าที่เปลี่ยนไปคือสัญญาณรบกวน v เพื่อหาค่าความผิดพลาดทั้งสองแบบและพิจารณาค่าต่างๆที่จะมีผลต่อค่าความผิดพลาดทั้งสองแบบที่เกิดขึ้น ในที่นี้จะพิจารณาค่าระดับนัยสำคัญ, ค่าความแปรปรวนของสัญญาณรบกวน และจำนวนตัวอย่างของข้อมูล

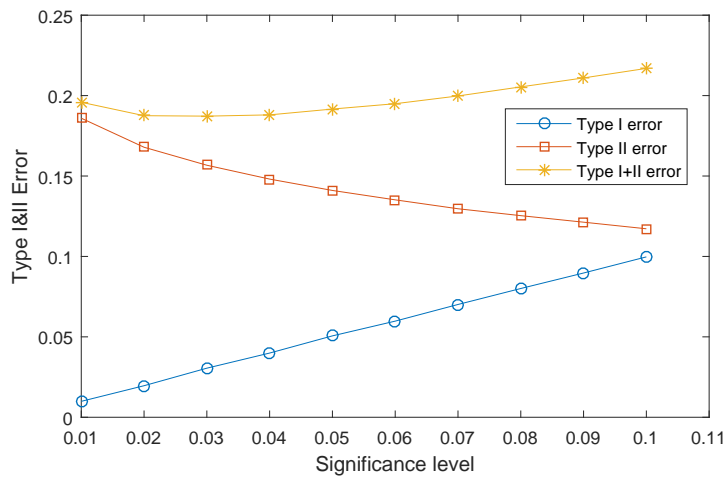
จากผลการทดลอง ดังแสดงในรูป 2, 3 และ 4 จะเห็นว่าค่าความผิดพลาดแบบที่สองจะมีค่าลดลงเมื่อเพิ่มค่าของระดับนัยสำคัญและจำนวนตัวอย่าง แต่จะมีค่าเพิ่มมากขึ้นเมื่อค่าความแปรปรวนของสัญญาณรบกวนมีค่ามากขึ้น ในขณะที่ค่าความผิดพลาดแบบที่หนึ่งจะมีค่าประมาณเท่ากับค่าระดับนัยสำคัญเสมอ เนื่องจากการทดสอบสมมติฐานได้นิยามให้ค่าระดับนัยสำคัญและค่าความผิดพลาดแบบที่หนึ่งเป็นค่าเดียวกัน

3.2 การทดสอบ Wald บนแบบจำลองถดถอยตัวเอง

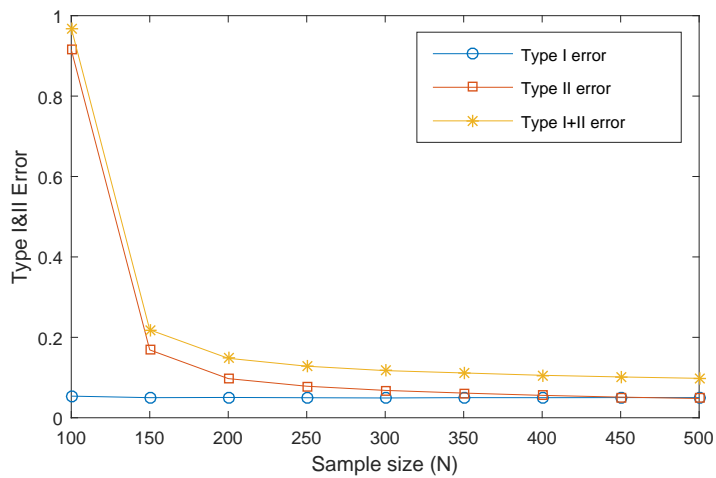
เพื่อสังเกตรูปแบบโครงสร้างของศูนย์ในพารามิเตอร์ของแบบจำลองที่ถูกประมาณได้เมื่อใช้ค่าระดับนัยสำคัญต่างๆกัน เริ่มต้นจากการสร้างแบบจำลองตามสมการ (3) โดยใช้ค่า $y(t) \in \mathbb{R}^n$ อันดับของแบบจำลอง $p = 4$ และจำนวนข้อมูลทั้งหมด 1,000 จุด ($t = 1, 2, \dots, 1000$) พารามิเตอร์ของแบบจำลอง $A_k \in \mathbb{R}^{n \times n}$ สำหรับ $k = 1, 2, 3, 4$ จะถูกสร้างให้มีบางสมาชิก $(A_k)_{ij}$ เป็นศูนย์จำนวนครึ่งหนึ่งของจำนวน



รูป 2: รูปแสดงค่าความผิดพลาดแบบที่หนึ่ง (เส้นสีฟ้า), ค่าความผิดพลาดแบบที่สอง (เส้นสีแดง) และผลรวมค่าความผิดพลาดทั้งสองแบบ (เส้นสีเหลือง) เมื่อเปลี่ยนแปลงค่าความแปรปรวนของสัญญาณรบกวน โดย $\alpha = 0.05$, $N = 150$



รูป 3: รูปแสดงค่าความผิดพลาดแบบที่หนึ่ง (เส้นสีฟ้า), ค่าความผิดพลาดแบบที่สอง (เส้นสีแดง) และผลรวมค่าความผิดพลาดทั้งสองแบบ (เส้นสีเหลือง) เมื่อเปลี่ยนแปลงค่าระดับนัยสำคัญ (α) โดย $\sigma^2 = 10\|x\|$, $N = 150$



รูป 4: รูปแสดงค่าความผิดพลาดแบบที่หนึ่ง (เส้นสีฟ้า), ค่าความผิดพลาดแบบที่สอง (เส้นสีแดง) และผลรวมค่าความผิดพลาดทั้งสองแบบ (เส้นสีเหลือง) เมื่อเปลี่ยนแปลงจำนวนของตัวอย่าง โดย $\alpha = 0.05$, $\sigma^2 = 10\|x\|$

พารามิเตอร์ทั้งหมด หรือก็คือ 50% และ $v(t) \sim \mathcal{N}(0, 10I_n)$ โดยใช้ค่า $n = 20$ แล้วจึงทำการประมาณค่าพารามิเตอร์ A, Σ ซึ่งเป็นผลเฉลยแบบปิด (closed-form) ของปัญหา ML จากสมการ (5), (6) ตามลำดับ

จากนั้นใช้วิธีการจากหัวข้อที่ 2.2 (การทดสอบ Wald สำหรับเงื่อนไขความสัมพันธ์ระหว่างเหตุและผล) เพื่อนำพารามิเตอร์ของแบบจำลองไปทดสอบ Wald ซึ่งจะได้รูปแบบของสมการใหม่ในรูปแบบของสมการ (8) และฟังก์ชันข้อกำหนดสำหรับทดสอบคู่ (i, j) ใดๆ ดัง (13) ซึ่งสามารถเขียนได้เป็น

$$r(\hat{\theta}) = [e_i^T \otimes I_p] \hat{\theta}$$

เมื่อ \otimes คือตัวดำเนินการ kronecker product และ $\hat{\theta} \in \mathbb{R}^{n^2p}$ คือค่าประมาณของ θ ในสมการ (9) ซึ่งมีจำนวนตัวแปรทั้งหมด $n^2p = 1600$ และค่าประมาณของเมทริกซ์ค่าความแปรปรวนร่วมของ $\hat{\theta}$ ซึ่งหาได้จาก (12) คือ

$$\widehat{\mathbf{Avar}}(\hat{\theta}) = \left[\sum_{t=p+1}^N \bar{H}(t)^T \hat{\Sigma}^{-1} \bar{H}(t) \right]^{-1}$$

และค่าทางสถิติของ Wald (W) สำหรับคู่ $(\hat{A}_k)_{ij}$ ใดๆ จาก (11) จะลดรูปเป็น

$$W_{ij} = \hat{B}_{ij}^T \left[\widehat{\mathbf{Avar}}(\hat{\theta})_{ij} \right]^{-1} \hat{B}_{ij}$$

โดยกำหนดให้ $\widehat{\mathbf{Avar}}(\hat{\theta})_{ij}$ คือเมทริกซ์ย่อยขนาด $p \times p$ ในแนวทแยงมุม (main diagonal block) ของ $\widehat{\mathbf{Avar}}(\hat{\theta})$ ที่สอดคล้องกับพารามิเตอร์ B_{ij}

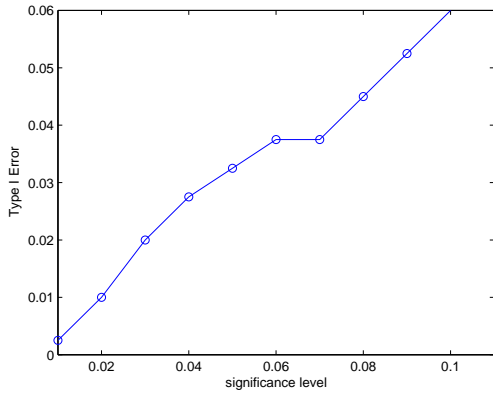
จากนั้นจึงทำการทดสอบ Wald สำหรับ $(A_k)_{ij}$ แต่ละคู่ (i, j) โดยใช้ค่าระดับนัยสำคัญต่างกัน และพิจารณาค่าความผิดพลาดทั้งสองแบบของการทดสอบสมมติฐาน ซึ่งคือค่าความผิดพลาดแบบที่หนึ่ง (Type I error) คือค่าความน่าจะเป็นที่ค่าพารามิเตอร์ที่ประมาณได้ไม่เท่ากับศูนย์เมื่อกำหนดให้ค่าจริงของพารามิเตอร์เท่ากับศูนย์

$$\text{Type I Error} = \mathbf{Prob}((\hat{A}_k)_{ij} \neq 0 | (A_k)_{ij} = 0)$$

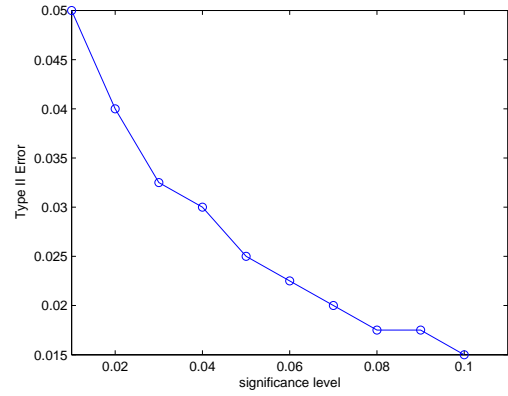
และค่าความผิดพลาดแบบที่สอง (Type II error) คือค่าความน่าจะเป็นที่ค่าพารามิเตอร์ที่ประมาณได้เท่ากับศูนย์เมื่อกำหนดให้ค่าจริงของพารามิเตอร์ไม่เท่ากับศูนย์

$$\text{Type II Error} = \mathbf{Prob}((\hat{A}_k)_{ij} = 0 | (A_k)_{ij} \neq 0)$$

จากการทดลองจะเห็นว่าค่าความผิดพลาดแบบที่สองจะมีค่าลดลงตามค่าระดับนัยสำคัญที่เพิ่มขึ้น ในขณะที่ค่าความผิดพลาดแบบที่หนึ่งจะมีค่าเพิ่มขึ้นตามค่าระดับนัยสำคัญที่เพิ่มขึ้น ดังรูปที่ 5 และแสดงตัวอย่างรูปแบบโครงสร้างของศูนย์ของพารามิเตอร์ที่ประมาณได้เทียบกับค่าจริงในรูปที่ 6

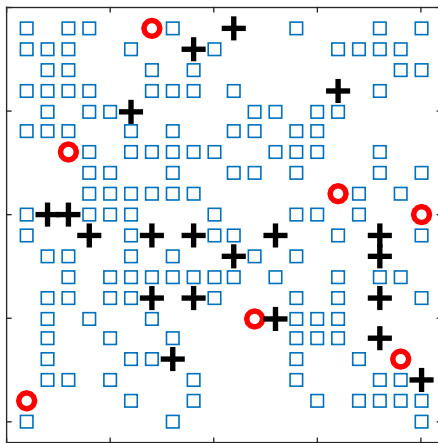


(a) ค่าความผิดพลาดแบบที่หนึ่ง

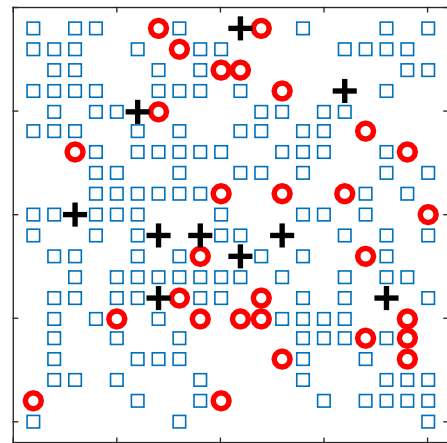


(b) ค่าความผิดพลาดแบบที่สอง

รูป 5: รูปแสดงค่าความผิดพลาดทั้งสองแบบ เมื่อเปลี่ยนแปลงค่าระดับนัยสำคัญ (α)



(a) $\alpha = 0.01$



(b) $\alpha = 0.1$

รูป 6: รูปแสดงตัวอย่างรูปแบบโครงสร้างของศูนย์ของพารามิเตอร์ที่ประมาณได้เมื่อเทียบกับค่าจริง เมื่อ \square แทนจุดข้อมูลที่ไม่เท่ากับศูนย์ทั้งคู่ของค่าจริงและค่าประมาณ \circ แทนจุดที่ค่าจริงเท่ากับศูนย์แต่ค่าประมาณไม่เป็นศูนย์ $+$ แทนจุดที่ค่าจริงไม่เท่ากับศูนย์แต่ค่าประมาณเป็นศูนย์และช่องว่างคือจุดข้อมูลที่เป็นศูนย์ทั้งคู่ของค่าจริงและค่าประมาณ

3.3 การศึกษาเงื่อนไขความสัมพันธ์ระหว่างเหตุและผลด้วยการทดสอบ Wald

เพื่อที่จะเลือกใช้ค่าระดับนัยสำคัญที่เหมาะสม สำหรับการหารูปแบบเงื่อนไขความสัมพันธ์ของพารามิเตอร์ของแบบจำลอง จึงนำเกณฑ์การคัดเลือกตัวแบบโดยข้อสนเทศของเบส์ (Bayesian Information Criterion (BIC)) มาใช้หาค่าที่เหมาะสมนั้น

เริ่มต้นจากการสร้างแบบจำลอง (3) โดยใช้ค่า $y(t) \in \mathbb{R}^n$ อันดับของแบบจำลอง $p = 4$ และจำนวนข้อมูลทั้งหมด 1000 จุด ($t = 1, 2, \dots, 1000$) พารามิเตอร์ของแบบจำลอง $A_k \in \mathbb{R}^{n \times n}$ สำหรับ $k = 1, 2, 3, 4$ จะถูกสร้างให้มีบางสมาชิก $(A_k)_{ij}$ เป็นศูนย์จำนวนครึ่งหนึ่งของจำนวนพารามิเตอร์ทั้งหมด หรือก็คือ 50% และ $v(t) \sim \mathcal{N}(0, 10I_n)$ โดยใช้ค่า $n = 20$ จากนั้นทำการแก้ปัญหา (14) โดยที่รูปแบบ

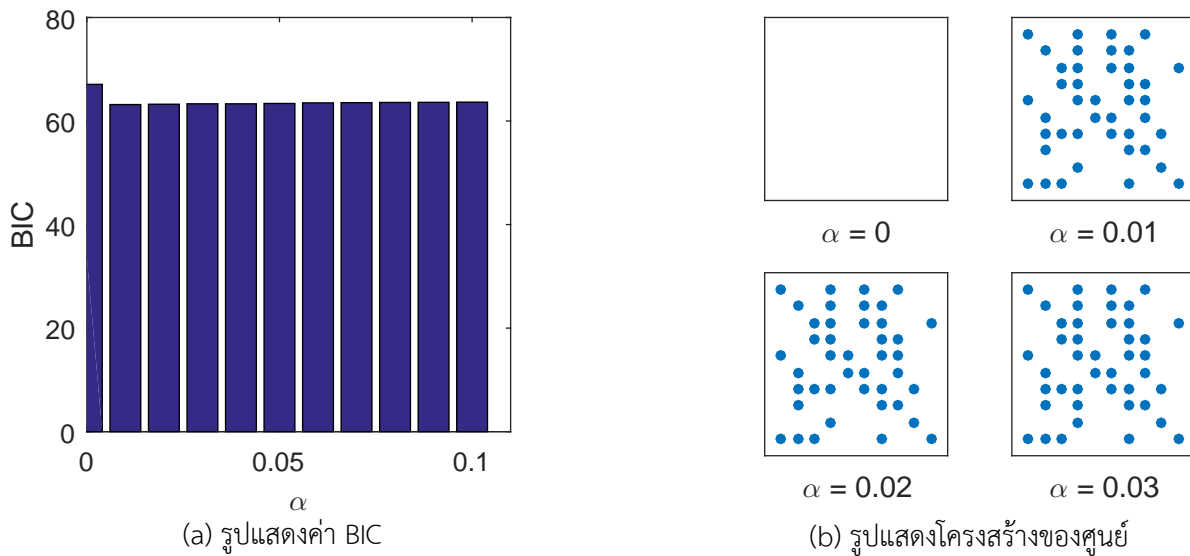
โครงสร้างของศูนย์จะเปลี่ยนไปตามค่าระดับนัยสำคัญทั้งหมด 10 ค่าคือ 0.01, 0.02, ..., 0.1 และหาค่า BIC ได้จาก

$$\text{BIC} = -2\mathcal{L} + d \log N$$

เมื่อ \mathcal{L} คือค่า log-likelihood เป็นฟังก์ชันของ A และ Σ , d คือจำนวนตัวแปรที่ต้องการประมาณค่าทั้งหมด ซึ่งค่า BIC จะสามารถจัดรูปได้เป็น

$$\text{BIC} = \log \det \hat{\Sigma} + \frac{d}{N-p} \log N$$

และจะถูกอธิบายไว้ในภาคผนวกที่ 6.2 โดย $\hat{\Sigma}$ หาได้จาก (6) และค่า d จะเปลี่ยนแปลงไปตามค่าระดับนัยสำคัญที่เปลี่ยนไป เนื่องจากมีผลจากจำนวนของศูนย์ที่หาได้จากการทดสอบ Wald ผลการทดลองที่ได้จะถูกแสดงไว้ในรูป 7



รูป 7: รูปแสดงผลการทดสอบ BIC

จะเห็นว่าผลการทดลองไม่สามารถเลือกค่าระดับนัยสำคัญที่เหมาะสมได้ เนื่องจากเมื่อเพิ่มค่าระดับนัยสำคัญจากศูนย์ขึ้นมาเพียงเล็กน้อยจะส่งผลให้จำนวนตัวแปรที่ต้องการประมาณค่าเพิ่มขึ้นมาอย่างมาก โดยสามารถดูได้จากรูปแบบโครงสร้างของศูนย์ในรูป 8 และเมื่อเพิ่มค่าระดับนัยสำคัญต่อไปเรื่อยๆ จำนวนตัวแปรที่ต้องการประมาณค่าจะค่อยๆ เพิ่มขึ้นเรื่อยๆ ที่ละน้อย เมื่อเทียบกับการเพิ่มขึ้นอย่างมากของจากค่าระดับนัยสำคัญเท่ากับศูนย์เป็น 0.01 ทำให้ผลการทดลองจะเลือกค่าระดับนัยสำคัญที่ต่ำที่สุดเสมอ ซึ่งจากความเป็นจริงแล้วการทดสอบ BIC จะสามารถหาค่าระดับนัยสำคัญที่เหมาะสมได้ถ้าเมื่อเพิ่มค่าระดับนัยสำคัญจากศูนย์ขึ้นมาเล็กน้อยแล้ว จำนวนตัวแปรที่ต้องการประมาณค่าจึงค่อยๆ เพิ่มขึ้นที่ละน้อย

3.4 การประมาณแบบจำลองถดถอยตัวเองที่รวมเงื่อนไขความสัมพันธ์ระหว่างเหตุและผล และเสถียรภาพ

เนื่องจากการประมาณพารามิเตอร์ A จาก (5) และ/หรือการประมาณพารามิเตอร์ A ที่รวมเงื่อนไขความสัมพันธ์ระหว่างเหตุและผลในปัญหา (14) อาจส่งผลให้ระบบเชิงเส้นไม่มีเสถียรภาพได้ เราจึงต้องเพิ่มเงื่อนไขเสถียรภาพเข้าไปเพื่อบังคับให้ A มีเสถียรภาพ

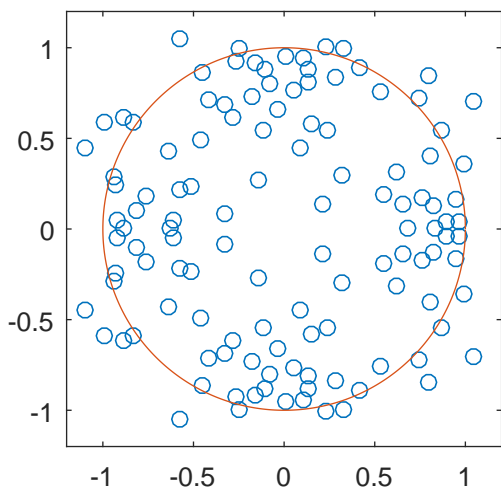
เริ่มต้นจากการสร้างแบบจำลอง (3) โดยใช้ค่า $y(t) \in \mathbb{R}^n$ อันดับของแบบจำลอง $p = 5$ และจำนวนข้อมูลทั้งหมด 200 จุด ($t = 1, 2, \dots, 200$) พารามิเตอร์ของแบบจำลอง $A_k \in \mathbb{R}^{n \times n}$ จะถูกสร้างให้มีบางสมาชิก $(A_k)_{ij}$ เป็นศูนย์จำนวน 50% ของจำนวนพารามิเตอร์ทั้งหมด สำหรับ $k = 1, \dots, 5$ และ $v(t) \sim \mathcal{N}(0, 10I_n)$ โดยใช้ค่า $n = 40$ เราสามารถหาค่าประมาณพารามิเตอร์ A จาก (5) ซึ่งค่า p ที่แท้จริงคือ $p = 5$ แต่ในการประมาณค่าพารามิเตอร์นั้น จะสมมติว่าเราไม่ทราบค่า p ที่แท้จริง และเมื่อใช้ p บางค่า เช่น $p = 3$ ผลปรากฏว่าการประมาณพารามิเตอร์ A นั้นจะทำให้แบบจำลองที่ได้ไม่มีเสถียรภาพ จากนั้นทำการทดสอบ Wald บนพารามิเตอร์ A ที่ประมาณได้ โดยเลือกใช้ค่าระดับนัยสำคัญ $\alpha = 0.05$ เพื่อที่จะหาเงื่อนไขความสัมพันธ์ระหว่างเหตุและผล และแก้ปัญหา

$$\begin{aligned} & \underset{\theta}{\text{minimize}} && \|F\theta - z\|_2^2 \\ & \text{subject to} && P\theta = 0 \end{aligned}$$

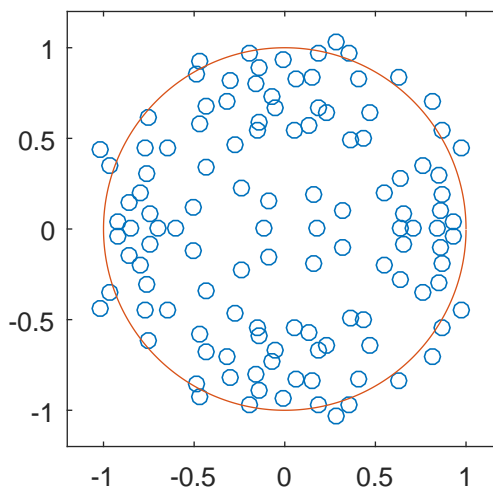
เมื่อ P คือเมทริกซ์การฉาย (projection matrix) เพื่อหาค่าประมาณพารามิเตอร์ของแบบจำลองที่สอดคล้องกับเงื่อนไขความสัมพันธ์ระหว่างเหตุและผล โดยใช้ Granger_cons.m และ ThetaToA.m ในภาคผนวกที่ 6.4 แต่เนื่องจากพารามิเตอร์ที่ประมาณได้อาจไม่มีเสถียรภาพ ดังนั้น เราต้องรวมเงื่อนไขเสถียรภาพเข้าไปในการประมาณพารามิเตอร์ของแบบจำลองที่สอดคล้องกับเงื่อนไขความสัมพันธ์ระหว่างเหตุและผล โดยรูปแบบของปัญหาจะเป็นการแก้ปัญหา optimization ที่มีเงื่อนไขทั้ง 2 ข้างต้น

$$\begin{aligned} & \underset{A}{\text{minimize}} && \|Y - AH\|_F^2 \\ & \text{subject to} && A = \begin{bmatrix} A_1 & A_2 & \dots & A_{p-1} & A_p \\ I & 0 & \dots & 0 & 0 \\ 0 & I & \dots & 0 & 0 \\ \vdots & \vdots & \ddots & \vdots & \vdots \\ 0 & 0 & \dots & I & 0 \end{bmatrix}, \\ & && \|A\|_\infty \leq 1, \\ & && (A_k)_{ij} = 0, \quad (i, j) \in I \end{aligned}$$

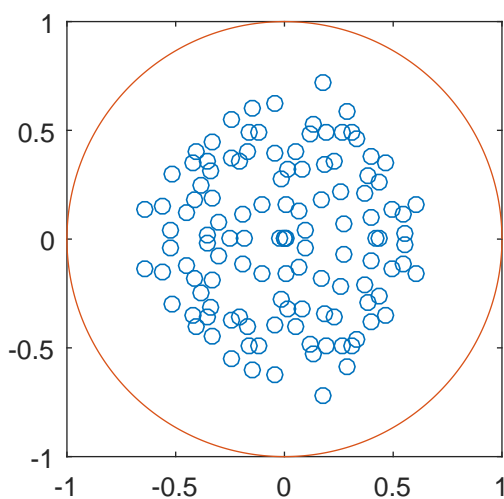
เมื่อ $I \subset \{1, 2, \dots, n\} \times \{1, 2, \dots, n\}$ เป็นเซตของคู่อันดับที่เป็นตัวกำหนดโครงสร้างของศูนย์ใน A_k โดยที่เงื่อนไข $(A_k)_{ij} = 0$ คือรูปแบบโครงสร้างของศูนย์ ที่ทำให้การทดสอบ Wald ซึ่งการแก้ปัญหานี้ เราได้ใช้ซอฟต์แวร์ของ MATLAB ที่ชื่อว่า CVX จาก [3] และ [4] โดยใช้ cvx_granger_stability.m ในภาคผนวกที่ 6.4 และผลลัพธ์ของการทดลองทั้งหมดจะถูกแสดงในรูป 8



(a) เมื่อไม่มีเงื่อนไข (unconstrained LS estimation) หรือปัญหา (7)



(b) เมื่อรวมความสัมพันธ์ระหว่างเหตุและผล หรือปัญหา (14)



(c) เมื่อรวมความสัมพันธ์ระหว่างเหตุและผล และ เสถียรภาพ หรือปัญหา (27)

รูป 8: รูปแสดงตำแหน่งค่าเจาะจงของเมทริกซ์ A บนระนาบเชิงซ้อน ที่คำนวณจากการประมาณแบบกำลังสองต่ำสุด และ รวมเงื่อนไขต่างๆ

จะเห็นว่าตำแหน่งของค่าเจาะจงของ A ที่ประมาณได้จากปัญหาการประมาณที่ไม่มีข้อกำหนดเสถียรภาพ จะมีบางส่วนที่อยู่นอกวงกลม 1 หน่วย แต่เมื่อเพิ่มข้อกำหนดเสถียรภาพเข้าไปตำแหน่งของค่าเจาะจงทั้งหมดจะอยู่ในวงกลม 1 หน่วย ซึ่งแสดงถึงการมีเสถียรภาพของระบบเชิงเส้นเวลาวิฤต โดยจะเห็นว่าตำแหน่งค่าเจาะจงของ (c) จะถูกบีบให้แคบลงมาก เพื่อให้อยู่ในวงกลม 1 หน่วย ซึ่งค่าฟังก์ชันจุดประสงค์จะเพิ่มมากขึ้นเมื่อมีการกำหนดเงื่อนไขเข้าไปมากขึ้น หรือก็คือค่าฟังก์ชันจุดประสงค์ของรูป 8 (c) > (b) > (a) และเป็นบ่งบอกว่า การบังคับใช้เงื่อนไขเสถียรภาพเข้าไป พารามิเตอร์ที่ประมาณได้ จะมีค่าใกล้เคียงกับของจริงน้อยลง

4 บทสรุป

โครงการนี้จะทำการหาเงื่อนไขความสัมพันธ์ระหว่างเหตุและผลของพารามิเตอร์ของแบบจำลอง AR ผ่านรูปแบบโครงสร้างของศูนย์ มีจุดประสงค์เพื่ออธิบายความสัมพันธ์ระหว่างตัวแปรต่างๆ ของข้อมูลอนุกรมเวลาผ่านค่าพารามิเตอร์ของแบบจำลอง ซึ่งรูปแบบโครงสร้างของศูนย์จะสามารถหาได้จากการทดสอบสมมติฐานแบบ Wald แต่เนื่องจากค่าพารามิเตอร์ของแบบจำลองที่ประมาณได้นั้นอาจจะส่งผลให้แบบจำลองไม่มีเสถียรภาพได้ ดังนั้นเราจึงต้องรวมเงื่อนไขเสถียรภาพเข้าไปในการประมาณแบบจำลอง ผลลัพธ์ที่น่าเสนอคือ ปัญหาการประมาณแบบจำลองถดถอยตัวเองในรูปการหาค่าเหมาะสมแบบคอนเวกซ์ ซึ่งสามารถหาคำตอบได้อย่างมีประสิทธิภาพ โดยสุดท้ายจะได้แบบจำลองที่มีเสถียรภาพและสอดคล้องกับเงื่อนไขความสัมพันธ์ระหว่างเหตุและผล

5 กิตติกรรมประกาศ

โครงการวิศวกรรมไฟฟ้า เรื่องการหาโครงสร้างเชิงสาเหตุสำหรับข้อมูลอนุกรมเวลาด้วยวิธีทดสอบทางสถิติบนการประมาณแบบจำลอง จะไม่สามารถประสบความสำเร็จลงได้เลย ถ้าหากขาดความช่วยเหลือ และการสนับสนุนจาก ผู้ช่วยศาสตราจารย์ ดร. จิตโกมุท ส่งศิริ อาจารย์ที่ปรึกษาในวิชาโครงการวิศวกรรมไฟฟ้า โดยการให้คำปรึกษา และคำแนะนำที่เป็นประโยชน์ในการทำโครงการนี้ ตลอดจนให้ความรู้ในการแก้ปัญหา ปรับปรุง และแก้ไขจนรายงานเล่มนี้เสร็จสมบูรณ์ ผู้จัดทำจึงขอขอบพระคุณเป็นอย่างสูง

References

- [1] L. Buesing, J. H. Macke, and M. Sahani. Learning stable, regularised latent models of neural population dynamics. *Network: Computation in Neural Systems*, 23(1-2): 24–47, 2012.
- [2] V. Cerone, D. Piga, and D. Regruto. Bounding the parameters of linear systems with stability constraints. In *American Control Conference (ACC), 2010*, pages 2152–2157. IEEE, 2010.
- [3] Michael Grant and Stephen Boyd. Graph implementations for nonsmooth convex programs. In V. Blondel, S. Boyd, and H. Kimura, editors, *Recent Advances in Learning and Control*, Lecture Notes in Control and Information Sciences, pages 95–110. Springer-Verlag Limited, 2008. http://stanford.edu/~boyd/graph_dcp.html.
- [4] Michael Grant and Stephen Boyd. CVX: Matlab software for disciplined convex programming, version 2.1. <http://cvxr.com/cvx>, March 2014.
- [5] W. H. Greene. *Econometric analysis*, 2008. Upper Saddle River, NJ: Prentice Hall, 24:25–26.
- [6] James Douglas Hamilton. *Time series analysis*, volume 2. Princeton university press Princeton, 1994.
- [7] Roger A Horn and Charles R Johnson. *Matrix analysis*. Cambridge university press, 2012.
- [8] S. L. Lacy and D. S. Bernstein. Subspace identification with guaranteed stability using constrained optimization. *Automatic Control, IEEE Transactions on*, 48(7): 1259–1263, 2003.
- [9] J. Mari, P. Stoica, and T. McKelvey. Vector ARMA estimation: A reliable subspace approach. *Signal Processing, IEEE Transactions on*, 48(7):2092–2104, 2000.
- [10] K. Turksoy, E. S. Bayrak, L. Quinn, E. Littlejohn, and A. Cinar. Guaranteed stability of recursive multi-input-single-output time series models. In *American Control Conference (ACC), 2013*, pages 77–82. IEEE, 2013.

6 ภาคผนวก

6.1 การประมาณแบบ maximum likelihood บนแบบจำลอง AR

พิจารณา $y = (y_1, \dots, y_N)$ เป็นตัวอย่างหรือสัญญาณขาออกที่สังเกตได้จากแบบจำลอง, θ คือค่าพารามิเตอร์ของแบบจำลองที่ต้องการจะประมาณ การประมาณค่า θ จาก y ด้วยหลักการของ ML ทำได้โดยการเลือก θ ที่ทำให้ค่าฟังก์ชันความหนาแน่นของความน่าจะเป็นมีค่าสูงสุด

$$\hat{\theta} = \arg \max_{\theta} f(y|\theta) \quad (29)$$

ซึ่งเป็นปัญหาการหาค่าที่เหมาะสมที่สุดแบบทั่วไป (unconstrained optimization) โดย $f(y|\theta)$ คือฟังก์ชันความหนาแน่นของความน่าจะเป็นแบบมีเงื่อนไข (conditional likelihood function) ของ y สำหรับค่า θ หนึ่งๆ โดยถ้าพิจารณาลอการิทึมของ $f(y|\theta)$ แล้วเราจะสามารถหาค่า $\hat{\theta}$ ได้จากเงื่อนไขเท่ากับศูนย์ของเกรเดียนต์ของลอการิทึมของ $f(y|\theta)$ นั่นคือ

$$\nabla_{\theta} \log f(y|\theta) = 0 \quad (30)$$

เมื่อนำหลักการของ ML มาประยุกต์ใช้กับแบบจำลอง AR จะได้ว่าฟังก์ชันความควรจะเป็นของแบบจำลองคือ

$$\begin{aligned} f(y|\theta) &= f(y|A, \Sigma) = f(y(p+1), y(p+2), \dots, y(N)|y(1), y(2), \dots, y(p), A, \Sigma) \\ &= \frac{1}{(2\pi)^{\frac{(N-p)}{2}}} \cdot \frac{1}{\det \Sigma^{\frac{(N-p)}{2}}} \cdot \exp -\frac{1}{2} \sum_{t=p+1}^N (y(t) - AH(t))^T \Sigma^{-1} (y(t) - AH(t)) \end{aligned} \quad (31)$$

เนื่องจาก (31) อยู่ในรูปของฟังก์ชันเอกซ์โพเนนเชียล (exponential function) เรามักจะพิจารณาลอการิทึมของ (31) แทน เพื่อใช้แก้ปัญหา (29) ได้สะดวกมากขึ้น

$$\begin{aligned} \log(f(y|A, \Sigma)) &\triangleq \mathcal{L}(A, \Sigma) \\ &= \frac{N-p}{2} \log \det \Sigma^{-1} - \frac{1}{2} \sum_{t=p+1}^N (y(t) - AH(t))^T \Sigma^{-1} (y(t) - AH(t)) \end{aligned} \quad (32)$$

เทอมสุดท้ายใน (32) สามารถเขียนใหม่ได้เป็นดังนี้

$$\mathcal{L}(A, \Sigma) = \frac{N-p}{2} \log \det \Sigma^{-1} - \frac{1}{2} \|L(Y - AH)\|_F^2 \quad (33)$$

โดยที่ L เป็นฟังก์ชันของ Σ^{-1} ด้วยความสัมพันธ์ $L^T L = \Sigma^{-1}$

$$Y = [y(p+1) \quad y(p+2) \quad \cdots \quad y(N)]_{n \times (N-p)},$$

$$H = \begin{bmatrix} 1 & 1 & \cdots & 1 \\ y(p) & y(p+1) & \cdots & y(N-1) \\ y(p-1) & y(p) & \cdots & y(N-2) \\ \vdots & \vdots & & \vdots \\ y(1) & y(2) & \cdots & y(N-p) \end{bmatrix}_{(np+1) \times (N-p)}$$

จากนั้นสามารถหาค่าประมาณของ A และ Σ ได้โดยการหาค่าสูงสุดของ $\mathcal{L}(A, \Sigma)$ จะเห็นว่าปัญหา (29) จะเทียบเท่ากับปัญหาค่ากำลังสองน้อยสุด (least-square) (เหตุผลจะแสดงผ่านการคำนวณด้านล่าง) ซึ่งเป็นปัญหาการหาค่าที่เหมาะสมที่สุดแบบทั่วไป (unconstrained optimization) จะได้ว่าผลเฉลยต้องสอดคล้องกับเงื่อนไข (30) ดังนั้น สำหรับการหาค่าสูงสุดของ $\mathcal{L}(A, \Sigma)$ จึงต้องหา A และ Σ ที่ทำให้เกรเดียนต์ของลอการิทึมของฟังก์ชันความน่าจะเป็นแบบมีเงื่อนไขของ y เทียบกับ A และ Σ^{-1} เท่ากับศูนย์ตามลำดับ

$$\nabla_A \mathcal{L}(A, \Sigma) = 0 \text{ และ } \nabla_{\Sigma^{-1}} \mathcal{L}(A, \Sigma) = 0$$

พิจารณา $\nabla_A \mathcal{L}(A, \Sigma) = 0$ โดยที่ $L^T L = \Sigma^{-1}$

$$\begin{aligned} \nabla_A \mathcal{L}(A, \Sigma) &= \frac{\partial}{\partial A} \|L(Y - AH)\|_F^2 \\ &= \frac{\partial}{\partial A} \text{tr}[(L(Y - AH))^T (L(Y - AH))] \\ &= \frac{\partial}{\partial A} \text{tr}[(Y^T - H^T A^T) \Sigma^{-1} (Y - AH)] \\ &= \frac{\partial}{\partial A} \text{tr}[Y^T \Sigma^{-1} Y - Y^T \Sigma^{-1} AH - H^T A^T \Sigma^{-1} Y + H^T A^T \Sigma^{-1} AH] \\ &= \frac{\partial}{\partial A} [\text{tr}(Y^T \Sigma^{-1} Y) - \text{tr}(Y^T \Sigma^{-1} AH) - \text{tr}(H^T A^T \Sigma^{-1} Y) + \text{tr}(H^T A^T \Sigma^{-1} AH)] \end{aligned}$$

ดังนั้น

$$\begin{aligned} \frac{\partial}{\partial A} [\text{tr}(Y^T \Sigma^{-1} Y) - \text{tr}(Y^T \Sigma^{-1} AH) - \text{tr}(H^T A^T \Sigma^{-1} Y) + \text{tr}(H^T A^T \Sigma^{-1} AH)] &= 0 \\ 0 + \frac{\partial}{\partial A} [-\text{tr}(AHY^T \Sigma^{-1}) - \text{tr}(\Sigma^{-1} Y H^T A^T) + \text{tr}(H^T A^T \Sigma^{-1} AH)] &= 0 \\ -(HY^T \Sigma^{-1})^T - \Sigma^{-1} Y H^T + \Sigma^{-1} A H H^T + \Sigma^{-T} A H H^T &= 0 \\ 2\Sigma^{-1} A H H^T - 2\Sigma^{-1} Y H^T &= 0 \end{aligned}$$

จะได้

$$A = Y H^T (H H^T)^{-1} \quad (34)$$

จากการคำนวณข้างต้นจะเห็นว่าพจน์ Σ ไม่มีผลต่อการคำนวณหาค่า A ที่ทำให้ค่า $\mathcal{L}(A, \Sigma)$ มีค่าสูงสุดตาม (34) จึงสอดคล้องกับ \hat{A} ที่เป็นคำตอบของ

$$\hat{A} = \arg \min_A \|Y - AH\|_F^2 \quad (35)$$

เนื่องจากเงื่อนไขเท่ากับศูนย์ของเกรเดียนต์ของ (35) จะออกมาเป็นสมการเดียวกันกับ (34)

พิจารณา $\nabla_{\Sigma^{-1}} \mathcal{L}(A, \Sigma) = 0$

$$\frac{\partial \mathcal{L}(A, \Sigma)}{\partial \Sigma^{-1}} = \frac{N-p}{2} \Sigma - \frac{1}{2} \sum_{t=p+1}^N (y(t) - AH(t))(y(t) - AH(t))^T = 0$$

จะได้

$$\Sigma = \frac{1}{N-p} \sum_{t=p+1}^N (y(t) - AH(t))(y(t) - AH(t))^T$$

ดังนั้นเราจะได้ผลเฉลย A, Σ ในรูปแบบปิด (closed-form) คือ

$$\hat{A} = YH^T(HH^T)^{-1} \quad (36)$$

$$\hat{\Sigma} = \frac{1}{N-p} \sum_{t=p+1}^N (y(t) - \hat{A}H(t))(y(t) - \hat{A}H(t))^T \quad (37)$$

โดยจะสามารถคำนวณค่า $\hat{\Sigma}$ ได้เมื่อรู้ค่า \hat{A} แล้ว

เนื่องจากแบบจำลอง AR สามารถพิจารณาในรูปแบบสมการที่มีพารามิเตอร์เป็นเวกเตอร์ได้ดัง (8)

ดังนั้นลอการิทึมของ (31) จะเปลี่ยนเป็น

$$\begin{aligned} \mathcal{L}(A, \Sigma) &= \frac{N-p}{2} \log \det \Sigma^{-1} - \frac{1}{2} \sum_{t=p+1}^N (y(t) - \bar{H}(t)\theta)^T \Sigma^{-1} (y(t) - \bar{H}(t)\theta) \\ &= \frac{N-p}{2} \log \det \Sigma^{-1} - \frac{1}{2} \|L(F\theta - z)\|_2^2 \end{aligned} \quad (38)$$

เมื่อ $F = [\bar{H}^T(p+1) \quad \bar{H}^T(p+2) \quad \cdots \quad \bar{H}^T(N)]^T$, $z = [y^T(p+1) \quad y^T(p+2) \quad \cdots \quad y^T(N)]^T$

6.2 ค่า BIC ของการประมาณแบบจำลอง AR

หัวข้อนี้จะอธิบายการหาค่า BIC ของแบบจำลอง AR เมื่อรู้ค่า $\hat{A}, \hat{\Sigma}$ โดย $\hat{\Sigma}$ มีรูปแบบตาม (37) จาก

$$\text{BIC} = -2\mathcal{L} + d \log N$$

และจาก (33) $\mathcal{L}(A, \Sigma) = \frac{N-p}{2} \log \det \Sigma^{-1} - \frac{1}{2} \|L(Y - AH)\|_F^2$ จะได้

$$-2\mathcal{L} = (N-p) \log \det \Sigma + \|L(Y - AH)\|_F^2$$

พิจารณา $\|L(Y - AH)\|_F^2$ เมื่อแทนค่า $A = \hat{A}$ และ $\Sigma = \hat{\Sigma}$ จะได้

$$\begin{aligned} \|L(Y - AH)\|_F^2 &= \text{tr}[(L(Y - AH))^T(L(Y - AH))] \\ &= \text{tr}[(Y - \hat{A}H)^T L^T L(Y - \hat{A}H)] \\ &= \text{tr}[(Y - \hat{A}H)(Y - \hat{A}H)^T L^T L] \\ &= \text{tr}[(Y - \hat{A}H)(Y - \hat{A}H)^T \hat{\Sigma}^{-1}] \\ &= \text{tr} \left[\sum_{t=p+1}^N (y(t) - \hat{A}H(t))(y(t) - \hat{A}H(t))^T \hat{\Sigma}^{-1} \right] \\ &= \text{tr}[(N-p)\hat{\Sigma}\hat{\Sigma}^{-1}] \\ &= \text{tr}[(N-p)I_n] \\ &= n(N-p) \end{aligned}$$

ดังนั้น จะได้ว่า

$$\text{BIC} = (N - p) \log \det \hat{\Sigma} + n(N - p) + d \log N$$

เนื่องจากค่า BIC เป็นค่าที่ใช้ในการเปรียบเทียบ เมื่อเปลี่ยนค่าระดับนัยสำคัญไปแต่ละค่าจะส่งผลให้ Σ และ d เปลี่ยนไป (d คือจำนวนพารามิเตอร์ที่ต้องการประมาณค่า) ดังนั้นจะได้

$$\text{BIC} = \log \det \hat{\Sigma} + \frac{d \log N}{N - p}$$

6.3 การประมาณแบบ ML และ LS ของแบบจำลอง AR เมื่อรวมเงื่อนไขความสัมพันธ์ระหว่างเหตุและผล

การประมาณแบบจำลอง AR แบบไม่มีเงื่อนไขใดๆ เราจะใช้ ML ซึ่งสมมูลกับ LS แต่พอเรารวมเงื่อนไขความสัมพันธ์ระหว่างเหตุและผลแล้วนั้น เราเลือกที่จะใช้ฟังก์ชันจุดประสงค์เป็นแบบ LS แทน เพราะจะสามารถแก้หาผลเฉลยได้ง่ายกว่า ซึ่งจะแสดงให้เห็นว่ารูปแบบของปัญหาที่เป็นแบบ ML และ LS ที่รวมเงื่อนไขความสัมพันธ์ระหว่างเหตุและผลนั้นจริงๆ แล้วเป็นปัญหาที่ต่างกัน

พิจารณาลอการิทึมของฟังก์ชันความควรจะเป็นของแบบจำลอง AR

$$\mathcal{L}(A, \Sigma) = \frac{N - p}{2} \log \det \Sigma^{-1} - \frac{1}{2} \text{tr}[(Y - AH)^T \Sigma^{-1} (Y - AH)]$$

กำหนดให้

$$f(A, \Sigma) = -\mathcal{L}(A, \Sigma) = -\frac{N - p}{2} \log \det \Sigma^{-1} + \frac{1}{2} \text{tr}[(Y - AH)^T \Sigma^{-1} (Y - AH)]$$

ให้ \hat{A} คือค่าประมาณของพารามิเตอร์ A และ $\hat{\Sigma}$ ถูกนิยามคือ

$$\hat{\Sigma} = \frac{1}{N - p} (Y - \hat{A}H)(Y - \hat{A}H)^T$$

ดังนั้นจะได้

$$\begin{aligned} \mathcal{L}(\hat{A}, \hat{\Sigma}) &= \frac{N - p}{2} \log \det \hat{\Sigma}^{-1} - \frac{1}{2} \text{tr}[(Y - \hat{A}H)(Y - \hat{A}H)^T \hat{\Sigma}^{-1}] \\ &= \frac{N - p}{2} \log \det \hat{\Sigma}^{-1} - \frac{1}{2} \text{tr}[(N - p)I] \end{aligned}$$

เนื่องจากพจน์ที่สองของ $\mathcal{L}(\hat{A}, \hat{\Sigma})$ เป็นค่าคงที่ ดังนั้นจะได้ว่า $\mathcal{L}(\hat{A}, \hat{\Sigma})$ มีค่าขึ้นอยู่กับ $\hat{\Sigma}$ ซึ่งขึ้นอยู่กับ \hat{A} อย่างเป็นนัย

พิจารณาเงื่อนไขความสัมพันธ์ระหว่างเหตุและผล $(A_k)_{ij} = 0, k = 1, 2, \dots, p, (i, j) \in I$ ซึ่งสามารถเขียนได้ในรูป

$$P(A) = 0$$

เมื่อ $P : \mathbb{R}^{n \times np} \rightarrow \mathbb{R}^{n \times np}$ คือเมทริกซ์การแปลงที่ขึ้นอยู่กับคู่อันดับใน $I \subset \{1, 2, \dots, n\} \times \{1, 2, \dots, n\}$ ตัวอย่างเช่น $I = (1, 2), (2, 3)$, $n = 3, p = 2$ จะได้

$$A = [A_1 \mid A_2]$$

$$P(A) = \begin{bmatrix} 0 & (A_1)_{12} & 0 & | & 0 & (A_2)_{12} & 0 \\ 0 & 0 & (A_1)_{23} & | & 0 & 0 & (A_2)_{23} \\ 0 & 0 & 0 & | & 0 & 0 & 0 \end{bmatrix}$$

เมื่อ 0 คือตำแหน่งที่เท่ากับศูนย์ และ X คือตำแหน่งที่ไม่เท่ากับศูนย์ และกำหนดให้ $Q = I - P$ จะได้

$$A = P(A) + Q(A)$$

พิจารณาปัญหาการประมาณแบบ ML เมื่อรวมเงื่อนไขความสัมพันธ์ระหว่างเหตุและผล

$$\begin{aligned} \text{minimize} \quad & -\frac{N-p}{2} \log \det \Sigma^{-1} + \frac{1}{2} \text{tr}[(Y - AH)^T \Sigma^{-1} (Y - AH)] \\ \text{subject to} \quad & P(A) = 0 \end{aligned} \quad (39)$$

และปัญหาการประมาณแบบ LS เมื่อรวมเงื่อนไขความสัมพันธ์ระหว่างเหตุและผล

$$\begin{aligned} \text{minimize} \quad & \frac{1}{2} \text{tr}[(Y - AH)^T (Y - AH)] \\ \text{subject to} \quad & P(A) = 0 \end{aligned} \quad (40)$$

เราต้องการจะทราบว่าปัญหา (39) กับ (40) สมมูลกันหรือไม่

เนื่องจาก $A = P(A) + Q(A)$ ดังนั้นเมื่อแทนค่า $P(A) = 0$ ใน A บนปัญหา (39) และ (40) ปัญหาทั้งสองจะลดรูปเป็นปัญหาที่ไม่มีข้อกำหนดใดๆ

$$\text{minimize} \quad -\frac{N-p}{2} \log \det \Sigma^{-1} + \frac{1}{2} \text{tr}[(Y - Q(A)H)^T \Sigma^{-1} (Y - Q(A)H)] \quad (41)$$

และ

$$\text{minimize} \quad \frac{1}{2} \text{tr}[(Y - Q(A)H)^T (Y - Q(A)H)] \quad (42)$$

ดังนั้นเราจะสามารถหาคำตอบได้จากเงื่อนไขเท่ากับศูนย์ของเกรเดียน ดังนี้

พิจารณาเงื่อนไขเท่ากับศูนย์ของเกรเดียนของปัญหา (41) เทียบกับ A จะได้

$$\frac{\partial}{\partial A} \text{tr}[(Y - Q(A)H)^T \Sigma^{-1} (Y - Q(A)H)] = Q(\Sigma^{-1}(Y - Q(A)H)H^T) = 0 \quad (43)$$

และ เงื่อนไขเท่ากับศูนย์ของเกรเดียนของปัญหา (41) เทียบกับ Σ^{-1}

$$\frac{\partial}{\partial \Sigma^{-1}} \text{tr}[(Y - Q(A)H)^T \Sigma^{-1} (Y - Q(A)H)] = -\frac{(N-p)}{2} \Sigma + \frac{1}{2} (Y - Q(A)H)(Y - Q(A)H)^T = 0$$

จะได้

$$\Sigma = \frac{1}{(N-p)}(Y - Q(A)H)(Y - Q(A)H)^T \quad (44)$$

แทนค่า (44) ใน (43) จะได้

$$Q \left((N-p) [(Y - Q(A)H)(Y - Q(A)H)^T]^{-1} (Y - Q(A)H)H^T \right) = 0 \quad (45)$$

ซึ่งเป็นสมการไม่เชิงเส้นใน A

พิจารณาเงื่อนไขเท่ากับศูนย์ของเกรเดียนของปัญหา (42) เทียบกับ A

$$\frac{\partial}{\partial A} \text{tr}[(Y - Q(A)H)^T(Y - Q(A)H)] = Q((Y - Q(A)H)H^T) = 0 \quad (46)$$

จะได้

$$Q(Y) = Q(Q(A)HH^T)$$

ซึ่งเป็นสมการเชิงเส้น จะเห็นได้ว่าเงื่อนไข (45) และ (46) นั้นต่างกัน ดังนั้นปัญหา (39) กับ (40) ไม่สมมูลกัน แต่เนื่องจากปัญหา (40) จะเป็นรูปแบบที่สามารถคำนวณได้ง่ายกว่า เราจึงเลือกใช้ปัญหาการประมาณแบบ LS มาใช้เมื่อต้องการรวมเงื่อนไขความสัมพันธ์ระหว่างเหตุและผล ในการประมาณพารามิเตอร์ของแบบจำลอง

6.4 MATLAB codes

ใช้ `genar.m` ในการสร้างข้อมูลแบบจำลองถดถอยตัวเองขึ้นมาในการทดลองที่ 3.2, 3.3 และ 3.4

```
1 %genar(n, p, N, d) generates order-p AR model sparse
   parameter Ai, i = 1,
2 %2, ..., p, with density d and simulates N data
   vectors from the generated
3 %model.
4 %Initial p data vectors are randomly generated and
   the data is corrupted by
5 %noise of variance varn.
6 %Assume that all the model parameters Ai, i = 1, 2,
   ..., p, share the same
7 %sparsity pattern.
8
9 function [Y, A, H] = genar(n, p, N, d, varn)
10
11 Y = zeros(n, N-p);
12 A = zeros(n, n, p);
13 H = zeros(n*p, N-p);
14
15 %create sparsity pattern with nonzero diagonal
   elements
16 sppat = sprandn(n, n, d) | speye(n);
17
18 sdA = sqrt(0.01); %variance of normally
   distributed parameter
19
20 %generate parameters
21 for i = 1:p
22     A(:, :, i) = sdA * sprandn(sppat);
23 end
24
25 %check model stability
26 comp = [speye(n*(p-1)), sparse(n*(p-1), n)]; %
   complementary
27 ITER = 0;
28 while (abs(eigs([A(:, :), comp], 1)) >= 1) && (ITER
   <1000)
29     for i = 1:p
30         A(:, :, i) = sdA * sprandn(sppat);
31     end
32     ITER = ITER + 1;
33 end
34 if ITER == 1000
35     fprintf('ERROR\n')
36     return
37 end
38
39 %generate data
40 H(:, 1) = randn(n*p, 1); %initialize p data
   vectors
41
42 sdn = sqrt(varn);
43 for j = 2:(N-p)
44     ynext = A(:, :)*H(:, j-1) + sdn.*randn(n, 1);
45     H(:, j) = [ynext; H(1:n*(p-1), j-1)];
46 end
```



```
47 Y = [H(1:n, 2:(N-p)), ynext];
```

Listing 1: genar.m

ใช้ `Compute_Aest_Avar.m` ในการคำนวณค่าของ \hat{A} , $\hat{\Sigma}$ และ $\widehat{\text{Avar}}(\hat{\theta})$

```
1 % ML Estimation of AR model and compute the
  parameter for Wald test
2 % INPUT :
3 % Y and H is a data of the AR model
4
5 % OUTPUT :
6 % Aest is model parameter that estimated by ML
  estimation
7 % Avar is parameter the Wald test
8
9 function [Aest, Avar] = Compute_Aest_Avar(Y, H)
10 n = size(Y, 1);
11 p = size(H, 1) / n;
12 N = size(H, 2) + p;
13 Aest = Y * H' / (H * H');
14 Sigma = 0;
15 for t = 1:N-p
16     Sigma = Sigma + (Y(:, t) - Aest * H(:, t)) * (Y(:, t) - Aest * H
  (:, t))';
17 end
18 Sigmahat = Sigma / (N-p);
19 clear Sigma
20 % --- Generate new parameter for Wald test ---;
21 sum = 0;
22 for j = 1:N-p
23     ybar = [];
24     Hbar = [];
25     for a = 1:n
26         for i = a:n:n*p
27             ybar = [ybar H(i, j)];
28         end
29     end
30     for i = 0:n-1
31         Hbar = [Hbar; [zeros(1, i*n*p) ybar zeros(1, n*p
  *(n-i-1))]];
32     end
33     sum = sum + Hbar' / (Sigmahat) * Hbar;
34 end
35 Avar = eye(n*n*p) \ (sum);
36 clear sum;
```

Listing 2: Compute_Aest_Avar.m

ใช้ `funcwaldtest.m` ในการทำ Wald test เพื่อหารูปแบบโครงสร้างของศูนย์ของพารามิเตอร์ของแบบจำลองถดถอยตัวเอง

```
1 % Wald test for AR model with
2 % INPUT :
3 % n is the dimension of the model output
4 % p is the lag-order of the model
5 % Aest is the estimated model parameter with size n-
  by-np
6 % Avar is a consistent estimate of the asymptotic
  covariance matrix of Aest
```

```

7 % alpha is a significance level of the test
8
9 % OUTPUT :
10 % show zero structure of estimated model parameter
11
12 function [W]= funcWaldTest (n , p , Aest , Avar , alpha )
13
14 Ahatt = reshape ( Aest , n , n , p ) ;
15
16 theta = zeros ( n * n * p , 1 ) ; % parameter for testing in
    Wald test
17 num = 1 ;
18 for i = 1 : n
19     for j = 1 : n
20         for k = 1 : p
21             theta ( num ) = Ahatt ( i , j , k ) ;
22             num = num + 1 ;
23         end
24     end
25 end
26 clear num Ahat
27
28 c = chi2inv ( 1 - alpha , p ) ; % a crirical value
29 W = ones ( n , n ) * ( c + 1 ) ;
30 count = 1 ;
31 for i = 1 : n
32     for j = 1 : n
33         if i ~= j
34             r = theta ( ( i - 1 ) * n * p + ( j - 1 ) * p + 1 : ( i - 1 ) * n * p + j *
    p ) ;
35             W ( i , j ) = r ' * ( Avar ( ( i - 1 ) * n * p + ( j - 1 ) * p + 1 : ( i
    - 1 ) * n * p + j * p , ...
36             ( i - 1 ) * n * p + ( j - 1 ) * p + 1 : ( i - 1 ) * n * p + j * p ) \ r
    ) ; %Wald statistic value
37         end
38         count = count + 1 ;
39     end
40 end
41 W = W > c ;
42 figure ;
43 spy ( W ( : , : ) ) ; title ( 'Zero structure of the model
    parameter ' ) ;

```

Listing 3: funcwaldtest.m

ใช้ Parameter.m ในการหา F และ z สำหรับปัญหา (10)

```

1 % find parameter Y,H for AR model and F,z for vector
    form
2 % INPUT :
3
4 % y is a matrix which contain time series data for
    all N time points
5 % matrix y has size n-by-N, column i of y is a time
    series data at time i
6
7 function [Y , H , F , z ] = Parameter ( y , p )
8 a = size ( y ) ;
9 n = a ( 1 ) ;
10 N = a ( 2 ) ;
11 Y = y ( : , p + 1 : N ) ;

```

```

12 H = [];
13 for i = 1:p
14     H = [H; y(:, i:N-p+i-1)];
15 end
16 F = [];
17 for j = 1:N-p
18     ybar = [];
19     Hbar = [];
20     for a = 1:n
21         for i = a:n:n*p
22             ybar = [ybar H(i, j)];
23         end
24     end
25     for i = 0:n-1
26         Hbar = [Hbar; [zeros(1, i*n*p) ybar zeros(1, n*p
27             *(n-i-1))]];
28     end
29 F = [F; Hbar];
30 z = reshape(Y, n*(N-p), 1);

```

Listing 4: Parameter.m

ใช้ `Granger_cons.m` ในการประมาณพารามิเตอร์ของแบบจำลองถดถอยตัวเองที่สอดคล้องกับเงื่อนไขความสัมพันธ์ระหว่างเหตุและผล

```

1 % Estimation of AR model with Granger causality
  constraints
2 % INPUT :
3 % F, z is a data of the AR model which is explained
  in problem (10)
4 % W is a binary matrix that contain a zero pattern
  of AR model parameters
5 % or the Granger causality constraints
6
7 % OUTPUT :
8 % theta is estimated model parameter with Granger
  causality constraints
9 % in problem (14)
10
11 function [theta]= Granger_cons(F, z, W)
12 [row col]= find(W'==0);
13 Fbar=F;
14 for c = 1:length(row)
15     i = col(c);
16     j = row(c);
17     a = size(Fbar);
18     Fbar = [Fbar(:, 1:(i-1)*n*p+(j-1)*p-(c-1)*p) Fbar
19         (:, (i-1)*n*p+j*p+1-(c-1)*p:a(2))];
19 end
20 T = ((Fbar' * Fbar) \ Fbar') * z;
21 theta = ones(n*n*p, 1);
22 c = 1;
23 cc = 0;
24 t = 1;
25 for m = 1:n*n*p
26     i = col(c);
27     j = row(c);
28     if (i-1)*n*p+(j-1)*p+1 <= m && m <= (i-1)*n*p+j*p
29         theta(m) = 0;

```

```

30         cc = cc + 1;
31         if cc == 4
32             c = c + 1;
33             cc = 0;
34             if c > length ( row )
35                 c = 1;
36             end
37         end
38     else
39         theta ( m ) = T ( t ) ;
40         t = t + 1;
41     end
42 end

```

Listing 5: Granger_cons.m

ใช้ ThetaToA.m ในการแปลง θ ไปอยู่ในรูป A

```

1 % To transfrom theta to A
2
3 function [A]= ThetaToA ( theta , n , p )
4 A= zeros ( n , n , p )
5 c = 1;
6 for i = 1 : n
7     for j = 1 : n
8         for k = 1 : p
9             A ( i , j , k ) = theta ( c )
10            c = c + 1;
11        end
12    end
13 end

```

Listing 6: ThetaToA.m

ใช้ CVX_granger_stability.m ในการประมาณพารามิเตอร์ของแบบจำลองถดถอยตัวเองที่มีเสถียรภาพ และสอดคล้องกับเงื่อนไขความสัมพันธ์ระหว่างเหตุและผล

```

1 % Estimation of AR model with Granger causality and
  Stability constraints
2 % INPUT :
3 % Y and H is a data of the AR model which is
  explained in problem (4)
4 % W is a binary matrix that contain a zero pattern
  of AR model parameters or the
5 % Granger causality constraints
6
7 % OUTPUT :
8 % Ags is estimated model parameter with Granger
  causality and
9 % Stability constraints
10
11 function [Ags]= CVX_granger_stability ( Y , H , W )
12 n= size ( Y , 1 ) ;
13 p= size ( H , 1 ) / n ;
14 cvx_precision best
15 cvx_begin sdp
16 variable AA ( n , n * p )
17 variable A2 ( n * p , n * p )
18 minimize norm ( Y - AA * H , 'fro' ) ;
19 subject to

```

```

20     for a = 0:p-1
21         AA( find(W==0)+n*n*a) == 0; % Granger
causality constraints
22     end
23     A2 == [ AA; eye(n*(p-1)) zeros(n*(p-1),n) ];
24     norm(A2, Inf) <= 1; % Stability constraints
25 cvx_end
26 Afs = full(AA);

```

Listing 7: CVX_granger_stability.m