



# โครงการ การเรียนการสอนเพื่อเสริมประสบการณ์

ชื่อโครงการ การประมาณค่าพารามิเตอร์แบบปรับของภาวะความน่าจะเป็น  
สำหรับแบบจำลองพานเนล

Adjusted Likelihood Parameter Estimation for Panel  
Data Models

ชื่อนิสิต นางสาวพัชรสิริ เฟื่องฟู 603 35299 23

ภาควิชา คณิตศาสตร์และวิทยาการคอมพิวเตอร์  
สาขาวิชา คณิตศาสตร์

ปีการศึกษา 2563

คณะวิทยาศาสตร์ จุฬาลงกรณ์มหาวิทยาลัย

การประมาณค่าพารามิเตอร์แบบปรับของภาวะความน่าจะเป็น  
สำหรับแบบจำลองพหุคูณ

นางสาว พัชรสิริ เฟื่องฟู

โครงการนี้เป็นส่วนหนึ่งของการศึกษาตามหลักสูตรวิทยาศาสตรบัณฑิต  
สาขาวิชาคณิตศาสตร์ ภาควิชาคณิตศาสตร์และวิทยาการคอมพิวเตอร์

คณะวิทยาศาสตร์ จุฬาลงกรณ์มหาวิทยาลัย

ปีการศึกษา 2563

ลิขสิทธิ์ของจุฬาลงกรณ์มหาวิทยาลัย

# Adjusted Likelihood Parameter Estimation for Panel Data Models

Patcharasiri Fuangfoo

A Project Submitted in Partial Fulfillment of the Requirements  
for the Degree of Bachelor of Science Program in Mathematics

Department of Mathematics and Computer Science

Faculty of Science

Chulalongkorn University

Academic Year 2020

Copyright of Chulalongkorn University



นางสาว พัชรสิริ เฟื่องฟู: การประมาณค่าพารามิเตอร์แบบปรับของภาวะความน่าจะเป็น สำหรับแบบจำลองพาเนล. (Adjusted Likelihood Parameter Estimation for Panel Data Models) อ.ที่ปรึกษาโครงการหลัก : ผู้ช่วยศาสตราจารย์ ดร. จิราพรรณ สุนทรโชติ, 118 หน้า.

แบบจำลองเฟ-เฮรอตเป็นแบบจำลองที่ใช้กันอย่างแพร่หลาย ดังนั้น จึงมีการศึกษามากมาย เกี่ยวกับแบบจำลองนี้ ทั้งในแง่ของการพัฒนาวิธีการ และการประยุกต์ใช้ งานที่น่าสนใจหนึ่งคือ วิธีการปรับของภาวะควรจะเป็นของค่าความแปรปรวนของ Li และ Lahiri ในปี พ.ศ. 2553 ซึ่งเสนอวิธีการปรับสำหรับแบบจำลองเฟ-เฮรอต

ในโครงการนี้เราได้ขยายแนวคิดของ Li และ Lahiri ไปสู่แบบจำลองที่ทั่วไปมากขึ้น ซึ่งคือแบบจำลองเฟ-เฮรอตแบบพาเนล การขยายมาสู่แบบจำลองพาเนลสามารถรองรับการวิเคราะห์ข้อมูลพร้อมกันหลายปีซึ่งมักเกิดขึ้นในการประยุกต์ สำหรับในการศึกษาของเรา เราศึกษาการประมาณค่าพารามิเตอร์แบบปรับภาวะความน่าจะเป็น ทั้งแบบฟังก์ชันภาวะความน่าจะเป็นโปर्फิลและแบบฟังก์ชันภาวะความน่าจะเป็นเรสซิดัล นอกจากนี้ยังมีการนำไปประยุกต์ใช้กับข้อมูลรายได้เฉลี่ยต่อครัวเรือนของประชากรไทย

ภาควิชา...คณิตศาสตร์และวิทยาการคอมพิวเตอร์...ลายมือชื่อนิสิต... พัชรสิริ เฟื่องฟู

สาขาวิชา...คณิตศาสตร์...ลายมือชื่อ อ.ที่ปรึกษาโครงการหลัก จิราพรรณ สุนทรโชติ  
ปีการศึกษา...2563

# # 6033529923: MAJOR MATHEMATICS

KEYWORDS : PARAMETER ESTIMATION / FAY-HERRIOT MODEL / PANEL DATA / ADJUSTED LIKELIHOOD

MISS PATCHARASIRI FUANGFOO : ADJUSTED LIKELIHOOD PARAMETER ESTIMATION FOR  
PANEL DATA MODELS. ADVISOR : ASST. PROF. JIRAPHAN SUNTORNCHOST Ph.D., 118 pp.

The Fay Herriot model is a widely used model in many applications. Therefore, several studies of the model have been proposed in literature both in the aspect of methodology development and applications. One work is the adjusted likelihood method of variance component of the proposed by Li and Lahiri (2010). Their method was proposed for a classical Fay-Herriot model.

In this study, we extend their technique to a more general model which is the panel version of the Fay Herriot model. The panel extension of the model can accommodate simultaneous modelling of data in multiple years which commonly occur in application. In our study, we study the adjusted likelihood method for both the profile likelihood and residual likelihood. Moreover, application to Thai Socio Economic data is also considered.

Department : Mathematics and Computer Science Student's Signature พัชรวิทย์ ฟุ้งฟู

Field of Study : Mathematics Advisor's Signature จิราภรณ์ สุนทรนชostí

Academic Year : 2020

## กิตติกรรมประกาศ

โครงการฉบับนี้เสร็จสมบูรณ์ได้ เพราะได้รับความอนุเคราะห์ และความช่วยเหลือจากหลาย ๆ ท่าน จึงขอขอบคุณมา ณ ที่นี้

ขอขอบคุณอาจารย์ที่ปรึกษาโครงการ ผู้ช่วยศาสตราจารย์ ดร.จิราพรรณ สุนทรโชติ ที่คอยให้ความรู้ ความช่วยเหลือ ให้คำติชม เป็นที่ปรึกษา ชี้แนะให้เห็นถึงปัญหา และแนะนำแนวทางในการแก้ไขปัญหา ในการทำโครงการตั้งแต่เริ่มต้นจนถึงสิ้นสุด รวมถึงตลอดระยะเวลาในการศึกษาที่จุฬาลงกรณ์มหาวิทยาลัยแห่งนี้

ขอขอบคุณอาจารย์ และรุ่นพี่ทุกท่าน ที่คอยให้ความรู้ คำแนะนำ ชี้แนะแนวทาง ตลอดระยะเวลาในการทำโครงการนี้ และตลอดระยะเวลาในการศึกษาที่จุฬาลงกรณ์มหาวิทยาลัยแห่งนี้

ขอขอบคุณภาควิชาคณิตศาสตร์และวิทยาการคอมพิวเตอร์ที่ให้การสนับสนุนทุนในการทำโครงการนี้

ขอขอบคุณนายอรรถพล อังกุลสิทธิ์ นิสิตปริญญาเอก สาขาคณิตศาสตร์ประยุกต์และวิทยาการคอมพิวเตอร์ ที่ช่วยให้คำแนะนำ และความช่วยเหลือตลอดระยะเวลาในการทำโครงการนี้

ขอขอบคุณครอบครัว และเพื่อน ๆ ทุกคน ที่คอยให้กำลังใจ ช่วยเหลือกันและกัน ให้คำปรึกษาและคำแนะนำ ในการทำโครงการนี้ และขอขอบคุณความอดทนและความพยายามของตนเองที่ทำให้โครงการนี้สำเร็จลุล่วง

สุดท้ายนี้ขอขอบคุณทุนอุดหนุนการศึกษากองทุนสมเด็จพระบรมโอรสาธิราช เจ้าฟ้ามหาวชิราลงกรณ สยามมกุฎราชกุมาร ที่คอยสนับสนุนทุนในการศึกษา

# สารบัญ

	หน้า
บทคัดย่อภาษาไทย .....	ง
บทคัดย่อภาษาอังกฤษ .....	จ
กิตติกรรมประกาศ .....	ฉ
สารบัญ .....	ช
สารบัญตาราง .....	ฌ
สารบัญภาพ .....	ฎ
บทที่ 1 บทนำ.....	1
1.1 ความเป็นมาและเหตุผลการวิจัย .....	1
1.2 วัตถุประสงค์ของการวิจัย .....	2
1.3 ขอบเขตการวิจัย .....	3
1.4 ประโยชน์ที่คาดว่าจะได้รับ.....	3
บทที่ 2 ความรู้พื้นฐานและงานวิจัยที่เกี่ยวข้อง.....	4
2.1 ความรู้พื้นฐาน.....	4
2.1.1 ความรู้พื้นฐานทางด้านทฤษฎีความน่าจะเป็น.....	4
2.1.2 ความรู้พื้นฐานทางสถิติ.....	11
2.1.3 แบบจำลองเชิงเส้น.....	13
2.1.4 แบบจำลองเฟ-เฮรอดท.....	17
2.1.5 ตัวประมาณที่ไม่เอนเอียงเชิงเส้นเชิงประจักษ์ที่ดีที่สุด.....	18
2.1.6 ข้อมูลพานาล.....	20
2.1.7 แผนการสุ่มตัวอย่าง.....	22
2.2 งานวิจัยในอดีต .....	25
บทที่ 3 วิธีดำเนินการและผลการดำเนินการ .....	26
3.1 แบบจำลองเฟ-เฮรอดทแบบจำลองพานาล.....	26



3.2	กรณีความแปรปรวนจากการถดถอยไม่ขึ้นกับเวลา .....	30
3.2.1	กรณีสัมประสิทธิ์การถดถอยไม่ขึ้นกับเวลา.....	31
3.2.2	กรณีสัมประสิทธิ์การถดถอยขึ้นกับเวลา.....	31
3.2.3	การศึกษาเชิงตัวเลข.....	37
3.3	กรณีความแปรปรวนจากการถดถอยขึ้นกับเวลา .....	52
3.3.1	กรณีปัจจัยเกิดจากพื้นที่.....	52
3.3.2	กรณีปัจจัยเกิดจากเวลา.....	53
3.3.3	การศึกษาเชิงตัวเลข.....	58
3.4	บทประยุกต์ของการปรับฟังก์ชันภาวะความน่าจะเป็น .....	79
3.4.1	การคำนวณการประมาณค่าของรายได้ประชากรไทย.....	82
3.4.2	ผลการศึกษาเชิงตัวเลขของวิธีการปรับฟังก์ชันภาวะความน่าจะเป็นกับข้อมูลรายได้เฉลี่ยต่อครัวเรือนของประเทศไทย.....	83
3.5	สรุปผลการศึกษา .....	102
	รายการอ้างอิง.....	103
	ภาคผนวก แบบเสนอหัวข้อโครงการ รายวิชา 2301399 Project Proposal ปีการศึกษา 2563...	105
	ประวัติผู้เขียน .....	109

## สารบัญตาราง

	หน้า
บทที่ 2 ความรู้พื้นฐานและงานวิจัยที่เกี่ยวข้อง .....	4
ตารางที่ 2.1 ตัวอย่างข้อมูลสำหรับแบบจำลองเชิงเส้น .....	17
ตารางที่ 2.2 ข้อมูลรายได้เฉลี่ย สัปดาห์ของคนที่ยื่นจบปริญญาตรี สัปดาห์ของคนทำงานรับจ้าง ของจังหวัดที่ $i$ ณ เวลา $t$ .....	21
บทที่ 3 วิธีดำเนินการและผลการดำเนินการ .....	26
กรณีความแปรปรวนจากการถดถอยไม่ขึ้นกับเวลา	
กรณีสัมประสิทธิ์การถดถอยไม่ขึ้นกับเวลา	
กรณีจำลอง $A$ จากการแจกแจงยูนิฟอร์มในช่วง (1,2)	
ตารางที่ 3.1 จำนวนตัวประมาณความแปรปรวนจากการถดถอย $\hat{A} = 0$ .....	42
ตารางที่ 3.2 ค่า MSE ของ $\hat{A}$ .....	42
ตารางที่ 3.3 ค่า MSE ของ $\hat{\theta}_{it}^{EBLUP}$ กรณี PML และ APML .....	43
ตารางที่ 3.4 ค่า MSE ของ $\hat{\theta}_{it}^{EBLUP}$ กรณี REML และ AREML .....	43
กรณีจำลอง $A$ จากการแจกแจงยูนิฟอร์มในช่วง (4,6)	
ตารางที่ 3.5 จำนวนตัวประมาณความแปรปรวนจากการถดถอย $\hat{A} = 0$ .....	44
ตารางที่ 3.6 ค่า MSE ของ $\hat{A}$ .....	44
ตารางที่ 3.7 ค่า MSE ของ $\hat{\theta}_{it}^{EBLUP}$ กรณี PML และ APML .....	45
ตารางที่ 3.8 ค่า MSE ของ $\hat{\theta}_{it}^{EBLUP}$ กรณี REML และ AREML .....	45
กรณีจำลอง $A$ จากการแจกแจงยูนิฟอร์มในช่วง (9,11)	
ตารางที่ 3.9 จำนวนตัวประมาณความแปรปรวนจากการถดถอย $\hat{A} = 0$ .....	46
ตารางที่ 3.10 ค่า MSE ของ $\hat{A}$ .....	46
ตารางที่ 3.11 ค่า MSE ของ $\hat{\theta}_{it}^{EBLUP}$ กรณี PML และ APML .....	47
ตารางที่ 3.12 ค่า MSE ของ $\hat{\theta}_{it}^{EBLUP}$ กรณี REML และ AREML .....	47
กรณีสัมประสิทธิ์การถดถอยขึ้นกับเวลา	
กรณีจำลอง $A$ จากการแจกแจงยูนิฟอร์มในช่วง (1,2)	
ตารางที่ 3.13 จำนวนตัวประมาณความแปรปรวนจากการถดถอย $\hat{A} = 0$ .....	48
ตารางที่ 3.14 ค่า MSE ของ $\hat{A}$ .....	48
ตารางที่ 3.15 ค่า MSE ของ $\hat{\theta}_{it}^{EBLUP}$ กรณี PML และ APML .....	49

กรณีจำลอง  $A$  จากการแจกแจงยูนิฟอร์มในช่วง (4,6)

ตารางที่ 3.16 จำนวนตัวประมาณความแปรปรวนจากการถดถอย  $\hat{A} = 0$  ..... 49

ตารางที่ 3.17 ค่า MSE ของ  $\hat{A}$  ..... 50

ตารางที่ 3.18 ค่า MSE ของ  $\hat{\theta}_{it}^{EBLUP}$  กรณี PML และ APML..... 50

กรณีจำลอง  $A$  จากการแจกแจงยูนิฟอร์มในช่วง (9,11)

ตารางที่ 3.19 จำนวนตัวประมาณความแปรปรวนจากการถดถอย  $\hat{A} = 0$  ..... 51

ตารางที่ 3.20 ค่า MSE ของ  $\hat{A}$  ..... 51

ตารางที่ 3.21 ค่า MSE ของ  $\hat{\theta}_{it}^{EBLUP}$  กรณี PML และ APML..... 52

กรณีความแปรปรวนจากการถดถอยขึ้นกับเวลา

กรณีปัจจัยเกิดจากพื้นที่

กรณีจำลอง  $A_t$  จากการแจกแจงยูนิฟอร์มในช่วง (1,2)

ตารางที่ 3.22 จำนวนตัวประมาณความแปรปรวนจากการถดถอย  $\hat{A}_t = 0$  ..... 63

ตารางที่ 3.23 ค่า MSE ของ  $\hat{A}_t$ ..... 64

ตารางที่ 3.24 ค่า MSE ของ  $\hat{\theta}_{it}^{EBLUP}$  กรณี PML และ APML..... 65

ตารางที่ 3.25 ค่า MSE ของ  $\hat{\theta}_{it}^{EBLUP}$  กรณี REML และ AREML..... 65

กรณีจำลอง  $A_t$  จากการแจกแจงยูนิฟอร์มในช่วง (4,6)

ตารางที่ 3.26 จำนวนตัวประมาณความแปรปรวนจากการถดถอย  $\hat{A}_t = 0$  ..... 66

ตารางที่ 3.27 ค่า MSE ของ  $\hat{A}_t$ ..... 67

ตารางที่ 3.28 ค่า MSE ของ  $\hat{\theta}_{it}^{EBLUP}$  กรณี PML และ APML..... 67

ตารางที่ 3.29 ค่า MSE ของ  $\hat{\theta}_{it}^{EBLUP}$  กรณี REML และ AREML..... 68

กรณีจำลอง  $A_t$  จากการแจกแจงยูนิฟอร์มในช่วง (9,11)

ตารางที่ 3.30 จำนวนตัวประมาณความแปรปรวนจากการถดถอย  $\hat{A}_t = 0$  ..... 68

ตารางที่ 3.31 ค่า MSE ของ  $\hat{A}_t$ ..... 69

ตารางที่ 3.32 ค่า MSE ของ  $\hat{\theta}_{it}^{EBLUP}$  กรณี PML และ APML..... 70

ตารางที่ 3.33 ค่า MSE ของ  $\hat{\theta}_{it}^{EBLUP}$  กรณี REML และ AREML..... 70

กรณีปัจจัยเกิดจากเวลา

กรณีจำลอง  $A_t$  จากการแจกแจงยูนิฟอร์มในช่วง (1,2)

ตารางที่ 3.34 จำนวนตัวประมาณความแปรปรวนจากการถดถอย  $\hat{A}_t = 0$  ..... 72

ตารางที่ 3.35 ค่า MSE ของ  $\hat{A}_t$ ..... 73

ตารางที่ 3.36 ค่า MSE ของ  $\hat{\theta}_{it}^{EBLUP}$  กรณี PML และ APML..... 73

ตารางที่ 3.37 ค่า MSE ของ  $\hat{\theta}_{it}^{EBLUP}$  กรณี REML และ AREML..... 74

กรณีจำลอง  $A_t$  จากการแจกแจงยูนิฟอร์มในช่วง (4,6)

ตารางที่ 3.38	จำนวนตัวประมาณความแปรปรวนจากการถดถอย $\hat{A}_t = 0$ .....	74
ตารางที่ 3.39	ค่า MSE ของ $\hat{A}_t$ .....	75
ตารางที่ 3.40	ค่า MSE ของ $\hat{\theta}_{it}^{EBLUP}$ กรณี PML และ APML .....	76
ตารางที่ 3.41	ค่า MSE ของ $\hat{\theta}_{it}^{EBLUP}$ กรณี REML และ AREML .....	76

กรณีจำลอง  $A_t$  จากการแจกแจงยูนิฟอร์มในช่วง (9,11)

ตารางที่ 3.42	จำนวนตัวประมาณความแปรปรวนจากการถดถอย $\hat{A}_t = 0$ .....	77
ตารางที่ 3.43	ค่า MSE ของ $\hat{A}_t$ .....	78
ตารางที่ 3.44	ค่า MSE ของ $\hat{\theta}_{it}^{EBLUP}$ กรณี PML และ APML .....	78
ตารางที่ 3.45	ค่า MSE ของ $\hat{\theta}_{it}^{EBLUP}$ กรณี REML และ AREML .....	79
ตารางที่ 3.46	ค่าประมาณความแปรปรวนจากการถดถอย $\hat{A}$ จากข้อมูลรายได้ของประเทศไทย เมื่อ สัมพันธ์การถดถอยไม่ขึ้นกับเวลา .....	84
ตารางที่ 3.47	ค่าประมาณความแปรปรวนจากการถดถอย $\hat{A}$ จากข้อมูลรายได้ของประเทศไทย เมื่อ สัมพันธ์การถดถอยขึ้นกับเวลา .....	84
ตารางที่ 3.48	ค่าประมาณความแปรปรวนจากการถดถอย $\hat{A}_t$ กรณีไม่ปรับ จากข้อมูลรายได้ของ ประเทศไทย เมื่อปัจจัยเกิดจากพื้นที่ .....	91
ตารางที่ 3.49	ค่าประมาณความแปรปรวนจากการถดถอย $\hat{A}_t$ กรณีปรับ จากข้อมูลรายได้ของ ประเทศไทย เมื่อปัจจัยเกิดจากพื้นที่ .....	92
ตารางที่ 3.50	ค่าประมาณความแปรปรวนจากการถดถอย $\hat{A}_t$ กรณีไม่ปรับ (PML) จากข้อมูลรายได้ ของประเทศไทย เมื่อปัจจัยเกิดจากเวลา .....	93
ตารางที่ 3.51	ค่าประมาณความแปรปรวนจากการถดถอย $\hat{A}_t$ กรณีปรับ (APML) จากข้อมูลรายได้ ของประเทศไทย เมื่อปัจจัยเกิดจากเวลา .....	94
ตารางที่ 3.52	ค่าประมาณความแปรปรวนจากการถดถอย $\hat{A}_t$ กรณีไม่ปรับ (REML) จากข้อมูลรายได้ ของประเทศไทย เมื่อปัจจัยเกิดจากเวลา .....	95
ตารางที่ 3.53	ค่าประมาณความแปรปรวนจากการถดถอย $\hat{A}_t$ กรณีปรับ (AREML) จากข้อมูลรายได้ ของประเทศไทย เมื่อปัจจัยเกิดจากเวลา .....	96



# บทที่ 1

## บทนำ

### 1.1 ความเป็นมาและเหตุผลการวิจัย

Small area estimation (SAE) หรือ การประมาณค่าสำหรับพื้นที่ขนาดเล็ก เป็นหนึ่งในวิธีการทางสถิติที่เป็นที่รู้จักสำหรับการประมาณพารามิเตอร์ เมื่อขนาดของตัวอย่างไม่เพียงพอที่จะให้ตัวประมาณค่าที่ได้ มีความน่าเชื่อถือ หลักการของ small area estimation คือ การยืมข้อมูลจากแหล่งข้อมูลที่มีความเกี่ยวข้องกันผ่านแบบจำลอง โดยแบบจำลองที่ใช้อย่างกว้างขวางคือแบบจำลองเฟ-เฮรอต ซึ่งเป็นแบบจำลองที่นิยามผ่าน 2 แบบจำลอง คือ แบบจำลองเชื่อมโยง (Linking Model) และแบบจำลองการสุ่มตัวอย่าง (Sampling Model) ดังนี้

1. แบบจำลองเชื่อมโยง:  $\theta_i = x_i' \beta + v_i$
2. แบบจำลองการสุ่มตัวอย่าง:  $y_i = \theta_i + e_i$

สำหรับ  $i = 1, 2, 3, \dots, n$

เมื่อ

$n$  คือ ขนาดตัวอย่าง

$\theta_i$  คือ ค่าเฉลี่ยสำหรับพื้นที่  $i$

$v_i$  คือ ความคลาดเคลื่อนจากการถดถอย ซึ่งสมมติให้มีการแจกแจงปกติ ที่มีค่าเฉลี่ยเท่ากับ 0 และความแปรปรวนเท่ากับ  $A$

$y_i$  คือ ค่าประมาณของ  $\theta_i$  ที่ได้จากการสำรวจ

$e_i$  คือ ความคลาดเคลื่อนที่เกิดจากการสุ่มตัวอย่าง ซึ่งสมมติให้มีการแจกแจงปกติ ที่มีค่าเฉลี่ยเท่ากับ 0 และความแปรปรวน  $\psi_i$

การประมาณตัวแปรเป้าหมายในแบบจำลองเฟ-เฮรอต คือ ตัวประมาณที่ไม่เอนเอียงเชิงเส้นเชิงประจักษ์ที่ดีที่สุด (Empirical best linear unbiased predictor: EBLUP)

$$\hat{\theta}_i^{EBLUP} = \frac{\hat{A}}{\hat{A} + \psi_i} y_i + \frac{\psi_i}{\hat{A} + \psi_i} x'_i \beta$$

ซึ่งเป็นผลบวกแบบถ่วงน้ำหนักระหว่างค่าที่ได้จากการสำรวจ ( $y_i$ ) และ ค่าที่ได้จากสมการถดถอย ( $x'_i \beta$ ) โดยน้ำหนักจะขึ้นกับความแปรปรวนจากการถดถอย ( $\hat{A}$ ) จะเห็นว่ากรณีที่ความแปรปรวนจากการถดถอย มีค่าเท่ากับ 0 จะทำให้  $\hat{\theta}_i^{EBLUP} = x'_i \beta$  ซึ่งเป็นตัวประมาณที่ได้มาจากการถดถอยเพียงส่วนเดียว ทำให้ตัวประมาณที่ได้ไม่น่าเชื่อถือ เราจึงต้องการหลีกเลี่ยงกรณีความแปรปรวนจากการถดถอยมีค่าเท่ากับ 0 ซึ่งในปี ค.ศ. 2010 Li และ Lahiri ได้เสนอแนวคิดการปรับภาวะความน่าจะเป็นสูงสุด (Adjusted maximum likelihood: ADM) เพื่อไม่ให้เกิดกรณีดังกล่าว นอกจากนี้ต่อมาในปี ค.ศ. 2014 Yoshimori และ Lahiri ได้มีการปรับปรุงวิธีการของ Li และ Lahiri และปี ค.ศ. 2020 Angkunsit และ Suntornchost ได้มีการนำแนวคิดของ Li และ Lahiri ไปปรับสร้างการประมาณค่าความแปรปรวนสำหรับแบบจำลองหลายตัวแปร

ข้อมูลพาเนล เป็นรูปแบบข้อมูลที่มีการบันทึกข้อมูลของหน่วยสำรวจหลายหน่วย โดยบันทึก เป็นค่าตัวแปรที่สนใจจากหลายหน่วยเวลา ทำให้สามารถศึกษาแนวโน้มระหว่างเวลาได้ ซึ่งในปี ค.ศ. 1994 Rao และ Yu ได้ทำการศึกษาเกี่ยวกับข้อมูลพาเนล และการประมาณ empirical best linear unbiased predictor (EBLUP) ของข้อมูลพาเนลไว้แล้ว ซึ่งข้อมูลพาเนลเป็นลักษณะของข้อมูลที่พบในหลายสาขา เช่น ข้อมูลทางด้านเศรษฐศาสตร์

ดังนั้น ในโครงการนี้ เราจะขยายหลักการของ Li และ Lahiri ในการปรับตัวประมาณของความแปรปรวนของแบบจำลองถดถอยให้เป็นแบบจำลองพาเนล และจะใช้แบบจำลองพาเนลในการศึกษาแนวโน้มระหว่างปีของรายได้ของประชากรไทยตั้งแต่ปี พ.ศ. 2552 - พ.ศ. 2560

## 1.2 วัตถุประสงค์ของการวิจัย

เพื่อประยุกต์หลักการการปรับของวิธีภาวะน่าจะเป็นสูงสุดสำหรับแบบจำลองพาเนลกับข้อมูลรายได้ของประชากรไทย ตั้งแต่ปี พ.ศ. 2552 ถึงปี พ.ศ. 2560

## 1.3 ขอบเขตการวิจัย

ข้อมูลที่ใช้ในการศึกษาในโครงการนี้ คือข้อมูลรายได้ของประชากรไทยตั้งแต่ปี พ.ศ. 2552 ถึงปี พ.ศ. 2560 โดยเป็นข้อมูลที่มีการเก็บทุก 2 ปี

## 1.4 ประโยชน์ที่คาดว่าจะได้รับ

### 1.4.1 ในด้านความรู้และประสบการณ์ต่อตัวนิสิตเอง

1. มีความรู้เกี่ยวกับการถดถอยเชิงเส้น
2. มีความรู้เกี่ยวกับแบบจำลองเฟ-เฮรอต
3. มีความรู้เกี่ยวกับแบบจำลองพานเนล
4. มีความรู้ ความสามารถในการใช้โปรแกรม  $R$

### 1.4.2 ในด้านความรู้ ความเข้าใจที่นำไปสู่การแก้ไขปัญหาของสังคมหรือสภาพแวดล้อม

สามารถคาดการณ์แนวโน้มรายได้ของประชากรไทยได้



## บทที่ 2

# ความรู้พื้นฐานและงานวิจัยที่เกี่ยวข้อง

ในส่วนนี้จะกล่าวถึงความรู้พื้นฐานทางด้านทฤษฎีความน่าจะเป็น สถิติ และสถิติสำหรับพื้นที่ขนาดเล็กที่ใช้ในการศึกษาโครงการนี้ รวมถึงงานวิจัยในอดีตที่เกี่ยวข้อง

## 2.1 ความรู้พื้นฐาน

### 2.1.1 ความรู้พื้นฐานทางด้านทฤษฎีความน่าจะเป็น

ในหัวข้อนี้ เราสรุปความรู้พื้นฐานทางด้านทฤษฎีความน่าจะเป็นที่ใช้ในการศึกษาโครงการนี้ โดยหัวข้อที่กล่าวถึงในหัวข้อนี้คือบทนิยาม และคุณสมบัติต่าง ๆ ของตัวแปรสุ่มและเวกเตอร์สุ่ม

#### ปริภูมิตัวอย่าง (Sampling space)

**นิยาม 1 (ปริภูมิตัวอย่าง)** ปริภูมิตัวอย่าง คือ เซตที่มีสมาชิกเป็นผลลัพธ์ที่อาจจะเป็นไปได้ทั้งหมดของการทดลองสุ่ม

#### ตัวแปรสุ่ม (Random variables)

**นิยาม 2 (ตัวแปรสุ่ม)** ตัวแปรสุ่ม คือ ฟังก์ชันที่มีค่าเป็นจำนวนจริง ซึ่งกำหนดโดยแต่ละสมาชิกในปริภูมิตัวอย่าง

**ตัวอย่าง 2.1.** พิจารณาการโยนเหรียญ 1 เหรียญ 3 ครั้ง ดังนั้น ปริภูมิตัวอย่างของการทดลองนี้ คือ

$$S = \{HHH, HHT, HTH, THH, HTT, THT, TTH, TTT\}$$

สมมติให้  $X$  เป็นจำนวนหัวที่ขึ้น จะได้ว่า  $X$  เป็นตัวแปรสุ่ม

### การแจกแจง (Distribution)

**นิยาม 3 (การแจกแจง)** การแจกแจงความน่าจะเป็นของตัวแปรสุ่ม คือ การแจกแจงค่าของตัวแปรสุ่มที่เป็นไปได้ทั้งหมด พร้อมทั้งความน่าจะเป็นที่จะเกิดค่านั้น ๆ

### ฟังก์ชันการแจกแจงสะสม (Cumulative distribution function)

**นิยาม 4 (ฟังก์ชันการแจกแจงสะสม)** เราจะกล่าวว่า  $F_X(x) : \mathbb{R} \rightarrow [0, 1]$  เป็นฟังก์ชันการแจกแจงสะสมของตัวแปรสุ่ม  $X$  ก็ต่อเมื่อ

1.  $\lim_{x \rightarrow -\infty} F_X(x) = 0$  และ  $\lim_{x \rightarrow \infty} F_X(x) = 1$
2.  $F_X(\cdot)$  ไม่เป็นฟังก์ชันลด
3.  $F_X(\cdot)$  เป็นฟังก์ชันต่อเนื่องทางขวา นั่นคือ สำหรับทุกค่า  $x_0$  แล้ว

$$\lim_{x \rightarrow x_0} F_X(x) = F_X(x_0)$$

### การแจกแจงเดียวกัน (Identical distribution)

**นิยาม 5 (การแจกแจงเดียวกัน)** เราจะกล่าวว่า ตัวแปรสุ่ม  $X$  และ  $Y$  มีการแจกแจงเดียวกัน ก็ต่อเมื่อ สำหรับทุก  $x$  แล้ว  $F_X(x) = F_Y(x)$

### ฟังก์ชันความหนาแน่นความน่าจะเป็น (Probability density function)

**นิยาม 6 (ฟังก์ชันความหนาแน่นความน่าจะเป็น)** ฟังก์ชันความหนาแน่นความน่าจะเป็น  $f_X(\cdot)$  ของตัวแปรสุ่ม  $X$  คือ ฟังก์ชันซึ่งสอดคล้อง

$$F_X(x) = \int_{-\infty}^x f_X(t) dt$$

สำหรับทุกค่า  $x$  เมื่อ  $F_X(x)$  คือฟังก์ชันการแจกแจงสะสมของตัวแปรสุ่ม  $X$

ค่าเฉลี่ย ความแปรปรวน และส่วนเบี่ยงเบนมาตรฐาน (Mean, variance, standard deviation)

นิยาม 7 (ค่าเฉลี่ย หรือ ค่าคาดหวัง) ค่าเฉลี่ย หรือ ค่าคาดหวังของตัวแปรสุ่มต่อเนื่อง  $X$  คือ

$$\mu = E(X) = \int_{-\infty}^{\infty} X f_X(x) dx$$

เมื่อ  $f_X(x)$  เป็นฟังก์ชันความหนาแน่นความน่าจะเป็นของตัวแปรสุ่ม  $X$

นิยาม 8 (ความแปรปรวน) ความแปรปรวนของตัวแปรสุ่ม  $X$  คือ

$$\sigma^2 = \text{var}(X) = E(x - \mu)^2$$

นิยาม 9 (ส่วนเบี่ยงเบนมาตรฐาน) รากที่สองของความแปรปรวน เรียกว่า ส่วนเบี่ยงเบนมาตรฐาน นั่นคือ

$$\sigma = \sqrt{\text{var}(X)} = \sqrt{E(x - \mu)^2}$$

การแจกแจงปกติ (Normal distribution)

การแจกแจงปกติเป็นการแจกแจงความน่าจะเป็นของตัวแปรสุ่มที่ส่วนมากจะมีค่าใกล้เคียงค่าเฉลี่ยของตัวแปรเหล่านั้น เป็นส่วนน้อยที่จะมีค่าของตัวแปรที่มากกว่าหรือน้อยกว่าค่าเฉลี่ยอย่างมาก ตัวแปรสุ่มที่มีการแจกแจงปกติ เรียกว่า ตัวแปรสุ่มปกติ (normal random variable)

นิยาม 10 (ฟังก์ชันความหนาแน่นความน่าจะเป็นของการแจกแจงปกติ) ฟังก์ชันความหนาแน่นความน่าจะเป็นของการแจกแจงปกติ ที่มีค่าเฉลี่ย  $\mu$  และความแปรปรวน  $\sigma^2$  คือ

$$f(x|\mu, \sigma^2) = \frac{1}{\sqrt{2\pi\sigma^2}} \exp\left\{-\frac{(x - \mu)^2}{2\sigma^2}\right\}$$

สำหรับ  $-\infty < x < \infty$

เมื่อ  $\mu = 0$  และ  $\sigma^2 = 1$  เราจะได้ฟังก์ชันความหนาแน่นความน่าจะเป็นของการแจกแจงปกติมาตรฐาน ดังนี้

$$g(x) = \frac{1}{\sqrt{2\pi}} \exp\left\{-\frac{x^2}{2}\right\} \quad (2.1)$$

สำหรับ  $-\infty < x < \infty$

## เวกเตอร์สุ่ม (Random vector)

เมื่อเราต้องพิจารณาตัวแปรสุ่มหลาย ๆ ตัวแปร การใช้เวกเตอร์เข้ามาช่วยจะทำให้รูปแบบของข้อมูลกระชับและง่ายต่อการนำไปใช้มากขึ้น เช่น เมื่อเรามีตัวแปรสุ่ม  $X_1, X_2, \dots, X_n$  เราสามารถมองตัวแปรสุ่มเหล่านี้ให้อยู่ในรูปเวกเตอร์  $\mathbf{X}$  ได้ คือ

$$\mathbf{X} = \begin{bmatrix} X_1 \\ X_2 \\ \vdots \\ X_n \end{bmatrix}$$

ทำให้ในทางคณิตศาสตร์สามารถนิยามความหมายของเวกเตอร์สุ่ม ได้เป็นดังนี้

**นิยาม 11 (เวกเตอร์สุ่ม)** เวกเตอร์สุ่มขนาด  $n$  มิติ เป็นฟังก์ชันจากปริภูมิตัวอย่าง (sample space)  $S$  ไปยัง  $\mathbb{R}^n$

**ตัวอย่าง 2.2.** พิจารณาการทอยลูกเต๋า และโยนเหรียญ ซึ่งเป็นอิสระต่อกัน ให้  $X$  เป็นตัวแปรสุ่มแสดงค่าแต้มที่เกิดจากการทอยลูกเต๋า และ ให้  $Y$  เป็นตัวแปรสุ่มโดยที่

$$Y = \begin{cases} 0 & \text{ถ้าเหรียญออกหัว} \\ 1 & \text{ถ้าเหรียญออกก้อย} \end{cases}$$

จะได้ว่า  $(X, Y)$  เป็น เวกเตอร์สุ่มขนาดสองมิติ

## การแจกแจงความน่าจะเป็นร่วมกัน (Joint probability distribution)

**นิยาม 12 (ฟังก์ชันความหนาแน่นความน่าจะเป็นร่วม)** ให้  $f(y_1, \dots, y_p)$  เป็นฟังก์ชันจาก  $\mathbb{R}^p$  ไปยัง  $\mathbb{R}$  จะกล่าวว่าเป็น ฟังก์ชันความหนาแน่นความน่าจะเป็นร่วม (joint probability density function) ของเวกเตอร์สุ่ม  $\mathbf{Y} = (Y_1, \dots, Y_p)$  ก็ต่อเมื่อ สำหรับทุก  $A \subset \mathbb{R}^p$

$$P((Y_1, \dots, Y_p) \in A) = \int_A \dots \int f(y_1, \dots, y_p) dy_1 \dots dy_p$$

## ฟังก์ชันความหนาแน่นความน่าจะเป็นมาร์จินัล (Marginal probability density function)

**นิยาม 13 (ฟังก์ชันความหนาแน่นความน่าจะเป็นมาร์จินัล)** ให้  $f(y_1, \dots, y_p)$  เป็นฟังก์ชันจาก  $\mathbb{R}^p$  ไปยัง  $\mathbb{R}$  ฟังก์ชันความหนาแน่นความน่าจะเป็นมาร์จินัลของตัวแปรสุ่ม  $Y_1, \dots, Y_p$

คือ

$$\begin{aligned} f_{Y_1}(y_1) &= \int_{-\infty}^{\infty} f(y_1, \dots, y_p) dy_2 \dots dy_p && \text{สำหรับทุก } y_1 \\ f_{Y_2}(y_2) &= \int_{-\infty}^{\infty} f(y_1, \dots, y_p) dy_1 dy_3 \dots dy_p && \text{สำหรับทุก } y_2 \\ &\vdots && \\ f_{Y_p}(y_p) &= \int_{-\infty}^{\infty} f(y_1, \dots, y_p) dy_1 \dots dy_{p-1} && \text{สำหรับทุก } y_p \end{aligned}$$

### ความเป็นอิสระต่อกัน (Independent)

**นิยาม 14 (ความเป็นอิสระต่อกัน)** ให้  $Y_1, \dots, Y_p$  เป็นตัวแปรสุ่มต่อเนื่อง ซึ่งมีฟังก์ชันความหนาแน่นความน่าจะเป็นร่วม (joint density function)  $f(y_1, \dots, y_p)$  และ ฟังก์ชันความหนาแน่นความน่าจะเป็นมาร์จินัล  $f_{Y_1}(y_1), \dots, f_{Y_p}(y_p)$  จะกล่าวว่า  $Y_1, \dots, Y_p$  เป็นอิสระต่อกัน ก็ต่อเมื่อ สำหรับทุกค่า  $y_1, \dots, y_p$

$$f(y_1, \dots, y_p) = f_{Y_1}(y_1) \dots f_{Y_p}(y_p) \quad (2.2)$$

### ความแปรปรวนร่วม (Covariance)

**นิยาม 15 (ความแปรปรวนร่วม)** ความแปรปรวนร่วมของตัวแปรสุ่ม  $Y_i$  และ  $Y_j$  คือ

$$\sigma_{ij} = \text{COV}(Y_i, Y_j) = E[(y_i - \mu_i)(y_j - \mu_j)]$$

### เวกเตอร์ค่าเฉลี่ย (Mean vectors)

**นิยาม 16 (เวกเตอร์ค่าเฉลี่ย)** ค่าเฉลี่ยของเวกเตอร์สุ่ม  $\mathbf{Y}$  ที่มีขนาด  $p \times 1$  คือ เวกเตอร์ของค่าเฉลี่ยของตัวแปรสุ่ม  $Y_1, \dots, Y_p$  ในเวกเตอร์  $\mathbf{Y}$  นั่นคือ

$$E(\mathbf{Y}) = E \begin{bmatrix} Y_1 \\ Y_2 \\ \vdots \\ Y_p \end{bmatrix} = \begin{bmatrix} E(Y_1) \\ E(Y_2) \\ \vdots \\ E(Y_p) \end{bmatrix} = \begin{bmatrix} \mu_1 \\ \mu_2 \\ \vdots \\ \mu_p \end{bmatrix} = \boldsymbol{\mu}$$

เมื่อ  $E(Y_i) = \mu_i$  คือค่าเฉลี่ย หรือค่าคาดหวังของตัวแปรสุ่ม  $i$  สำหรับ  $i = 1, \dots, p$

## เมทริกซ์ความแปรปรวนร่วม (Covariance matrix)

**นิยาม 17 (เมทริกซ์ความแปรปรวนร่วม)** กำหนดให้เวกเตอร์สุ่ม  $\mathbf{Y} = (Y_1, Y_2, \dots, Y_p)$  โดยที่  $Y_1, Y_2, \dots, Y_p$  เป็นตัวแปรสุ่มที่มีความแปรปรวน  $\sigma_1^2, \sigma_2^2, \dots, \sigma_p^2$  และความแปรปรวนร่วม  $\sigma_{ij}$  ดังนั้น เมทริกซ์ความแปรปรวนร่วม  $\Sigma$  ของเวกเตอร์สุ่ม  $\mathbf{Y}$  คือ

$$\Sigma = \text{cov}(\mathbf{Y}) = \begin{bmatrix} \sigma_1^2 & \sigma_{12} & \dots & \sigma_{1p} \\ \sigma_{21} & \sigma_2^2 & \dots & \sigma_{2p} \\ \vdots & \vdots & & \vdots \\ \sigma_{p1} & \sigma_{p2} & \dots & \sigma_p^2 \end{bmatrix} = [\sigma_{ij}]$$

### คุณสมบัติของเมทริกซ์

ถ้า  $\mathbf{A}$  เป็นเมทริกซ์ค่าคงตัวขนาด  $k \times p$ ,  $\mathbf{b}$  เป็นเวกเตอร์ค่าคงตัวขนาด  $k \times 1$  และ  $\mathbf{Y}$  เป็นเวกเตอร์สุ่มขนาด  $p \times 1$  ที่มีเมทริกซ์ความแปรปรวนร่วม  $\Sigma$  จะได้ว่า

1.  $E(\mathbf{AY} + \mathbf{b}) = \mathbf{AE}(\mathbf{y}) + \mathbf{b}$
2.  $\text{cov}(\mathbf{Ay} + \mathbf{b}) = \mathbf{A}\Sigma\mathbf{A}'$

### แรงก์เต็ม (Full rank)

**นิยาม 18 (ความไม่อิสระเชิงเส้น)** เซตของเวกเตอร์  $\mathbf{a}_1, \mathbf{a}_2, \dots, \mathbf{a}_n$  จะเรียกว่าไม่เป็นอิสระเชิงเส้นต่อกัน ถ้าสามารถหาค่า  $c_1, c_2, \dots, c_n$  ที่ไม่ใช่ศูนย์พร้อมกัน ซึ่งสอดคล้อง

$$c_1\mathbf{a}_1 + c_2\mathbf{a}_2 + \dots + c_n\mathbf{a}_n = \mathbf{0} \quad (2.3)$$

**นิยาม 19 (ความอิสระเชิงเส้น)** เซตของเวกเตอร์  $\mathbf{a}_1, \mathbf{a}_2, \dots, \mathbf{a}_n$  จะเรียกว่าเป็นอิสระเชิงเส้นต่อกัน ถ้าไม่สามารถหาค่า  $c_1, c_2, \dots, c_n$  ซึ่งสอดคล้องสมการ (2.3) ได้

**นิยาม 20 (แรงก์)** แรงก์ของเมทริกซ์  $\mathbf{A}$  คือ

$$\text{rank}(\mathbf{A}) = \text{จำนวนหลักที่เป็นอิสระเชิงเส้นต่อกันของเมทริกซ์}\mathbf{A}$$

ถ้าเมทริกซ์  $\mathbf{A}$  มีขนาด  $n \times p$  โดยที่  $\text{rank}(\mathbf{A}) = p < n$  แล้ว  $\mathbf{A}$  จะมีแรงก์สูงสุดที่เป็นได้ เรียกว่า แรงก์เต็ม (full rank)

ตัวแปรสุ่มหลายมิติที่ศึกษาในโครงการานนี้คือ ตัวแปรสุ่มแบบปกติหลายตัวแปร ดังนั้น ต่อไปเราจะพิจารณาการแจกแจงปกติหลายตัวแปร (multivariate normal distribution)

### การแจกแจงปกติหลายตัวแปร (multivariate normal distribution)

การแจกแจงปกติหลายตัวแปร เป็นการขยายรูปทั่วไปของการแจกแจงปกติหนึ่งตัวแปร ไปเป็นหลายตัวแปร เริ่มแรกเราจะพิจารณาตัวแปรสุ่มปกติมาตรฐาน  $Z_1, Z_2, \dots, Z_p$  ซึ่งสำหรับตัวแปรสุ่ม  $Z_i$  ( $i = 1, \dots, p$ ) มีค่าเฉลี่ย  $\mu_i = 0$  ความแปรปรวน  $\sigma_i^2 = 1$  และสำหรับ  $i, j = 1, \dots, p$  และ  $i \neq j$  ความแปรปรวนร่วม  $\sigma_{ij} = 0$  หลังจากนั้นเราจะทำการแปลง  $(Z_1, Z_2, \dots, Z_p)$  ที่เป็นอิสระต่อกัน ให้เป็นตัวแปรแบบปกติหลายตัวแปร  $(Y_1, Y_2, \dots, Y_p)$  ที่มีค่าเฉลี่ย ความแปรปรวน และความแปรปรวนร่วมใด ๆ นั่นคือ เราจะให้  $\mathbf{Z} = (Z_1, Z_2, \dots, Z_p)'$  โดยที่  $E(\mathbf{Z}) = \mathbf{0}$ ,  $\text{cov}(\mathbf{Z}) = \mathbf{I}$  เมื่อ  $\mathbf{I}$  คือเมทริกซ์เอกลักษณ์ นั่นคือแต่ละ  $Z_i$  มีการแจกแจงปกติที่มีค่าเฉลี่ยเท่ากับ 0 และความแปรปรวนเท่ากับ 1 และจะแปลง  $\mathbf{Z}$  ให้เป็นเวกเตอร์สุ่มปกติหลายตัวแปร  $\mathbf{Y} = (Y_1, Y_2, \dots, Y_p)'$  โดยที่  $E(\mathbf{Y}) = \boldsymbol{\mu}$  และ  $\text{cov}(\mathbf{Y}) = \boldsymbol{\Sigma}$  เมื่อ  $\boldsymbol{\mu}$  เป็นเวกเตอร์ขนาด  $p \times 1$  และ  $\boldsymbol{\Sigma}$  เป็นเมทริกซ์ขนาด  $p \times p$  เป็นเมทริกซ์สมมาตรซึ่งมีค่าลักษณะเฉพาะ (Eigenvalue) ทุกตัวมากกว่าศูนย์ (positive definite matrix)

เนื่องจากตัวแปรสุ่มแต่ละตัวเป็นอิสระต่อกัน จาก (2.1) และ (2.2) เราจะได้ว่าฟังก์ชันความหนาแน่นร่วม  $g(z_1, z_2, \dots, z_p)$  ของเวกเตอร์สุ่ม  $\mathbf{Z} = (Z_1, \dots, Z_p)$  คือ

$$\begin{aligned} g(z_1, z_2, \dots, z_p) &= g(\mathbf{z}) = g_1(z_1)g_2(z_2) \dots g_p(z_p) \\ &= \frac{1}{\sqrt{2\pi}} \exp\left\{-\frac{z_1^2}{2}\right\} \frac{1}{\sqrt{2\pi}} \exp\left\{-\frac{z_2^2}{2}\right\} \dots \frac{1}{\sqrt{2\pi}} \exp\left\{-\frac{z_p^2}{2}\right\} \\ &= \frac{1}{\sqrt{(2\pi)^p}} \exp\left\{-\sum_{i=1}^p \frac{z_i^2}{2}\right\} \end{aligned}$$

จาก  $\mathbf{z}'\mathbf{z} = z_1^2 + z_2^2 + \dots + z_p^2$  ทำให้ได้ว่า

$$g(\mathbf{z}) = \frac{1}{\sqrt{(2\pi)^p}} \exp\left\{-\frac{\mathbf{z}'\mathbf{z}}{2}\right\} \quad (2.4)$$

นั่นคือ  $\mathbf{Z}$  มีการแจกแจงปกติหลายตัวแปร ที่มีเวกเตอร์ค่าเฉลี่ย  $\mathbf{0}$  และเมทริกซ์ความแปรปรวนร่วม  $\mathbf{I}$

สำหรับการแปลง  $\mathbf{Z}$  ให้เป็น  $\mathbf{Y}$  ที่มีค่าเฉลี่ย  $\boldsymbol{\mu}$  เมทริกซ์ความแปรปรวนร่วม  $\boldsymbol{\Sigma}$  ใด ๆ เรา จะกำหนดให้  $\mathbf{Y} = \boldsymbol{\Sigma}^{1/2}\mathbf{Z} + \boldsymbol{\mu}$  เมื่อ  $\boldsymbol{\Sigma}^{1/2}$  คือ เมทริกซ์ซึ่ง  $\boldsymbol{\Sigma}^{1/2}(\boldsymbol{\Sigma}^{1/2})' = \boldsymbol{\Sigma}$  ทำให้เราได้ว่า

$$E(\mathbf{Y}) = E(\boldsymbol{\Sigma}^{1/2}\mathbf{Z} + \boldsymbol{\mu}) = \boldsymbol{\Sigma}^{1/2}E(\mathbf{Z}) + \boldsymbol{\mu} = \boldsymbol{\mu}$$

$$\text{cov}(\mathbf{Y}) = \text{cov}(\boldsymbol{\Sigma}^{1/2}\mathbf{Z} + \boldsymbol{\mu}) = \boldsymbol{\Sigma}^{1/2}\text{cov}(\mathbf{Z})(\boldsymbol{\Sigma}^{1/2})' = \boldsymbol{\Sigma}^{1/2}\mathbf{I}(\boldsymbol{\Sigma}^{1/2}) = \boldsymbol{\Sigma}$$

ในการหาฟังก์ชันการแจกแจงของ  $\mathbf{Y} = \Sigma^{1/2}\mathbf{Z} + \boldsymbol{\mu}$  จาก  $\mathbf{Z}$  ที่มีการแจกแจงเป็น (2.4) จะใช้เทคนิคการเปลี่ยนตัวแปร ดังนั้น ฟังก์ชันการแจกแจงของ  $\mathbf{Y} = \Sigma^{1/2}\mathbf{Z} + \boldsymbol{\mu}$  คือ

$$f(\mathbf{y}) = g(\mathbf{z}) (|\Sigma^{-1/2}|) \quad (2.5)$$

เมื่อ  $(|\Sigma^{-1/2}|)$  คือค่าสัมบูรณ์ของดีเทอร์มิแนนต์ของ  $\Sigma^{-1/2}$

เนื่องจาก  $\Sigma$  เป็นเมทริกซ์สมมาตรซึ่งมีค่าลักษณะเฉพาะ (Eigenvalue) ทุกตัวมากกว่า ศูนย์ และ จาก  $\mathbf{y} = \Sigma^{1/2}\mathbf{z} + \boldsymbol{\mu}$  ซึ่งทำให้ได้ว่า  $\mathbf{z} = \Sigma^{-1/2}(\mathbf{y} - \boldsymbol{\mu})$  ทำให้เราสามารถเขียน (2.5) ได้เป็น

$$\begin{aligned} f(\mathbf{y}) &= g(\mathbf{z}) |\Sigma^{-1/2}| \\ &= g(\mathbf{z}) |\Sigma|^{-1/2} \\ &= \frac{1}{\sqrt{(2\pi)^p}} \exp\left\{-\frac{\mathbf{z}'\mathbf{z}}{2}\right\} |\Sigma|^{-1/2} \\ &= \frac{1}{\sqrt{(2\pi)^p} |\Sigma|^{1/2}} \exp\left\{-\frac{(\Sigma^{-1/2}(\mathbf{y} - \boldsymbol{\mu}))'(\Sigma^{-1/2}(\mathbf{y} - \boldsymbol{\mu}))}{2}\right\} \\ &= \frac{1}{\sqrt{(2\pi)^p} |\Sigma|^{1/2}} \exp\left\{-\frac{(\mathbf{y} - \boldsymbol{\mu})'\Sigma^{-1}(\mathbf{y} - \boldsymbol{\mu})}{2}\right\} \end{aligned}$$

ซึ่งคือฟังก์ชันการแจกแจงปกติหลายตัวแปร ที่มีค่าเฉลี่ยเท่ากับ  $\boldsymbol{\mu}$  และ เมทริกซ์ความแปรปรวน  $\Sigma$  นั่นคือ  $\mathbf{Y} \sim N(\boldsymbol{\mu}, \Sigma)$

## 2.1.2 ความรู้พื้นฐานทางสถิติ

ในหัวข้อนี้ เราจะกล่าวถึงความรู้พื้นฐานทางสถิติที่ใช้ในการศึกษาโครงการงาน เช่น ตัวอย่างสุ่ม การประมาณค่าพารามิเตอร์ และแบบจำลองเชิงเส้น

ตัวอย่างสุ่ม หมายถึง กลุ่มของตัวอย่างบางหน่วยที่ถูกเลือกโดยวิธีการสุ่ม และนำมาเป็นตัวแทนของประชากรตามขนาดตัวอย่างที่กำหนดไว้ โดยในทางคณิตศาสตร์ได้นิยามไว้ดังนี้

**นิยาม 21 (ตัวอย่างสุ่ม)** ตัวแปรสุ่ม  $X_1, X_2, \dots, X_n$  จะเรียกว่าเป็น ตัวอย่างสุ่มของประชากรขนาด  $n$  ที่มีฟังก์ชันการแจกแจง  $F(x)$  ถ้า  $X_1, X_2, \dots, X_n$  เป็นตัวแปรสุ่มที่เป็นอิสระต่อกัน และฟังก์ชันการแจกแจงมาร์จินัลของแต่ละ  $X_i$  สำหรับ  $i = 1, 2, \dots, n$  เป็นฟังก์ชัน  $F(x)$  เดียวกัน



วิธีการประมาณแบบภาวะความน่าจะเป็นสูงสุด (Maximum likelihood estimation)

**นิยาม 22 (ฟังก์ชันภาวะความน่าจะเป็น)** ให้  $f(\mathbf{x}|\theta)$  เป็นฟังก์ชันการแจกแจงความน่าจะเป็นร่วม (joint pdf) ของกลุ่มตัวอย่าง  $\mathbf{X} = (X_1, X_2, \dots, X_n)$  จะได้ว่า ถ้า  $X = \mathbf{x}$  เป็นค่าสังเกต แล้วฟังก์ชันภาวะความน่าจะเป็น (likelihood function)  $L$  ของ  $\theta$  คือ  $L(\theta|\mathbf{x}) = f(\mathbf{x}|\theta)$

**นิยาม 23 (ตัวประมาณแบบภาวะความน่าจะเป็นสูงสุด)** ตัวประมาณแบบภาวะความน่าจะเป็นสูงสุด (maximum likelihood estimator: MLE) ของพารามิเตอร์  $\theta$  คือ  $\hat{\theta}$  ที่ทำให้ฟังก์ชันภาวะความน่าจะเป็น  $L(\theta|\mathbf{x})$  มีค่าสูงสุด

หมายเหตุ ฟังก์ชันภาวะความน่าจะเป็นส่วนใหญ่มักจะอยู่ในรูปผลคูณของฟังก์ชันความหนาแน่นความน่าจะเป็น ซึ่งจะทำได้ยากต่อการประมาณพารามิเตอร์ ดังนั้น ในด้านการนำไปใช้ จึงมักใช้การแปลงล็อกการิทึม ซึ่งเป็นการเปลี่ยนผลคูณให้อยู่ในรูปผลบวก เพื่อให้ง่ายต่อการคำนวณ จะได้เป็นล็อกฟังก์ชันภาวะความน่าจะเป็น  $\log L(\theta|\mathbf{x})$  และเนื่องจาก ฟังก์ชันล็อกการิทึมเป็นฟังก์ชันเพิ่ม ดังนั้น ตัวประมาณค่าที่ทำให้ล็อกฟังก์ชันภาวะความน่าจะเป็นมีค่าสูงสุด จะเป็นตัวเดียวกันกับตัวประมาณค่าที่ทำให้ฟังก์ชันภาวะความน่าจะเป็นมีค่าสูงสุด

**ตัวอย่าง 2.3.** ให้  $X_1, X_2, \dots, X_n$  เป็นตัวแปรสุ่มที่มีการแจกแจงเดียวกัน คือการแจกแจงปกติที่มีค่าเฉลี่ย  $\mu$  และค่าความแปรปรวน  $\sigma^2$  ที่เป็นอิสระต่อกัน จะได้ว่าตัวประมาณแบบภาวะความน่าจะเป็นสูงสุดของ  $\mu$  และ  $\sigma^2$  มีค่าดังต่อไปนี้

ฟังก์ชันภาวะความน่าจะเป็นของ  $\mu, \sigma^2$  คือ

$$\begin{aligned} L(\mu, \sigma^2|\mathbf{x}) &= f(\mathbf{x}|\mu, \sigma^2) \\ &= \prod_{i=1}^n f(x_i|\mu, \sigma^2) \\ &= \left(\frac{1}{\sqrt{2\pi\sigma^2}}\right)^n \exp\left\{-\sum_{i=1}^n \frac{(x_i - \mu)^2}{2\sigma^2}\right\} \end{aligned}$$

ดังนั้น ล็อกฟังก์ชันภาวะความน่าจะเป็นของ  $\mu, \sigma^2$  คือ

$$l(\mu, \sigma^2|\mathbf{x}) = -\frac{n}{2} \log(2\pi\sigma^2) - \sum_{i=1}^n \frac{(x_i - \mu)^2}{2\sigma^2}$$

หาอนุพันธ์ย่อยของล็อกฟังก์ชันภาวะความน่าจะเป็นเทียบกับพารามิเตอร์ และเทียบกับศูนย์จะได้ว่า

$$0 = \frac{\partial}{\partial \mu} l(\mu, \sigma^2 | \mathbf{x}) = \sum_{i=1}^n \frac{(x_i - \mu)}{\sigma^2}$$

$$0 = \frac{\partial}{\partial \sigma^2} l(\mu, \sigma^2 | \mathbf{x}) = -\frac{n}{2\sigma^2} - \sum_{i=1}^n \frac{(x_i - \mu)^2}{2} \left(-\frac{1}{(\sigma^2)^2}\right)$$

ดังนั้น ตัวประมาณแบบภาวะความน่าจะเป็นสูงสุดของ  $\mu, \sigma^2$  คือ

$$\hat{\mu} = \bar{x}_n \text{ และ}$$

$$\hat{\sigma}^2 = \frac{1}{n} \sum_{i=1}^n (x_i - \hat{\mu})^2 = \frac{1}{n} \sum_{i=1}^n (x_i - \bar{x}_n)^2$$

ตามลำดับ

### ความเอนเอียง (Biased)

**นิยาม 24 (ความเอนเอียง)** ค่าเอนเอียง (Bias) ของตัวประมาณ  $\hat{\theta}$  ของ  $\theta$  คือ

$$B(\hat{\theta}) \equiv E(\hat{\theta}) - \theta$$

เรากล่าวว่า  $\hat{\theta}$  เป็นตัวประมาณที่ไม่เอนเอียงของ  $\theta$  ถ้า  $B(\hat{\theta}) = 0$  นั่นคือ  $E(\hat{\theta}) = \theta$

### ค่าเฉลี่ยกำลังสองของความคลาดเคลื่อน (Mean squared error)

**นิยาม 25 (ค่าเฉลี่ยกำลังสองของความคลาดเคลื่อน)** ค่าเฉลี่ยกำลังสองของความคลาดเคลื่อน (MSE) ของตัวประมาณ  $\hat{\theta}$  ของ  $\theta$  คือ

$$MSE(\hat{\theta}) = E(\hat{\theta} - \theta)^2$$

## 2.1.3 แบบจำลองเชิงเส้น (Linear model)

แบบจำลองเชิงเส้น คือแบบจำลองที่แสดงความสัมพันธ์เชิงเส้นระหว่างตัวแปรต้น (independent variables หรือ predictor variables) และ ตัวแปรตาม (dependent variable หรือ response variable) โดยที่ตัวแปรต้น หมายถึงตัวแปรที่ทำให้เกิดผลตามมา หรือตัวแปรที่เป็นต้นเหตุ และตัวแปรตาม หมายถึง ตัวแปรที่เกิดขึ้นเนื่องจากตัวแปรต้น เช่น หากต้องการศึกษาว่าดินต่างชนิดกันมีผลต่อความสูงของต้นพืชหรือไม่ จากตัวอย่างนี้ ตัวแปรต้น คือ ดินชนิดต่าง ๆ และตัวแปรตาม คือ ความสูงของต้นพืช แบบจำลองเชิงเส้นที่มี  $y$  คือตัวแปรตาม และ  $x_1, x_2, \dots, x_k$  เป็นตัวแปรต้น สามารถเขียนได้ดังนี้

$$y = \beta_0 + \beta_1 x_1 + \beta_2 x_2 + \cdots + \beta_k x_k + \epsilon \quad (2.6)$$

เมื่อ  $\beta_0, \beta_1, \dots, \beta_k$  คือสัมประสิทธิ์การถดถอย และ  $\epsilon$  คือค่าความคลาดเคลื่อน

สำหรับสัมประสิทธิ์การถดถอย  $\beta_0, \beta_1, \dots, \beta_k$  สามารถประมาณได้ด้วยวิธีภาวะความน่าจะเป็นสูงสุด

สำหรับการประมาณสัมประสิทธิ์การถดถอย  $\beta_0, \beta_1, \dots, \beta_k$  ใน (2.6) จะใช้ตัวอย่าง  $n$  ตัวอย่าง (ค่าสังเกต) ของตัวแปรตาม  $y$  และตัวแปรต้น  $x_1, x_2, \dots, x_k$  ดังนั้น แบบจำลองเชิงเส้นสำหรับตัวอย่างที่  $i$  จะเป็นดังนี้

$$y_i = \beta_0 + \beta_1 x_{i1} + \beta_2 x_{i2} + \cdots + \beta_k x_{ik} + \epsilon_i \quad (2.7)$$

เมื่อ  $i = 1, 2, \dots, n$

โดยมีเงื่อนไขดังนี้

1.  $E(\epsilon_i) = 0$  สำหรับ  $i = 1, 2, \dots, n$  นั่นคือ  $E(y_i) = \beta_0 + \beta_1 x_{i1} + \beta_2 x_{i2} + \cdots + \beta_k x_{ik}$
2.  $\text{var}(\epsilon_i) = \sigma^2$  สำหรับ  $i = 1, 2, \dots, n$  นั่นคือ  $\text{var}(y_i) = \sigma^2$
3.  $\text{cov}(\epsilon_i, \epsilon_j) = 0$  สำหรับ  $i \neq j$  นั่นคือ  $\text{cov}(y_i, y_j) = 0$

เราสามารถเขียนกระจาย (2.7) ได้เป็นดังนี้

$$y_1 = \beta_0 + \beta_1 x_{11} + \beta_2 x_{12} + \cdots + \beta_k x_{1k} + \epsilon_1$$

$$y_2 = \beta_0 + \beta_1 x_{21} + \beta_2 x_{22} + \cdots + \beta_k x_{2k} + \epsilon_2$$

⋮

$$y_n = \beta_0 + \beta_1 x_{n1} + \beta_2 x_{n2} + \cdots + \beta_k x_{nk} + \epsilon_n$$

และจาก  $n$  สมการนี้ สามารถเขียนในรูปเมทริกซ์ได้เป็น

$$\begin{bmatrix} y_1 \\ y_2 \\ \vdots \\ y_n \end{bmatrix} = \begin{bmatrix} 1 & x_{11} & x_{12} & \cdots & x_{1k} \\ 1 & x_{21} & x_{22} & \cdots & x_{2k} \\ \vdots & \vdots & \vdots & \ddots & \vdots \\ 1 & x_{n1} & x_{n2} & \cdots & x_{nk} \end{bmatrix} \begin{bmatrix} \beta_0 \\ \beta_1 \\ \vdots \\ \beta_k \end{bmatrix} + \begin{bmatrix} \epsilon_1 \\ \epsilon_2 \\ \vdots \\ \epsilon_n \end{bmatrix}$$

หรือ

$$\mathbf{y} = \mathbf{X}\boldsymbol{\beta} + \boldsymbol{\epsilon} \quad (2.8)$$

และจะได้เงื่อนไขเป็น

$$1. E(\epsilon) = \mathbf{0} \text{ นั่นคือ } E(\mathbf{y}) = \mathbf{X}\boldsymbol{\beta}$$

$$2. \text{cov}(\epsilon) = \sigma^2 \mathbf{I} \text{ นั่นคือ } \text{cov}(\mathbf{y}) = \sigma^2 \mathbf{I}$$

ตัวประมาณของพารามิเตอร์ของแบบจำลองเชิงเส้น สามารถหาได้โดยวิธีกำลังสองน้อยสุด หรือ วิธีการประมาณแบบภาวะความน่าจะเป็นสูงสุด ซึ่งในหัวข้อนี้เราจะนำเสนอการหาค่าประมาณพารามิเตอร์แบบภาวะความน่าจะเป็นสูงสุด

พิจารณาแบบจำลองเชิงเส้น ที่มี  $y$  เป็นตัวแปรตาม และ  $x_1, \dots, x_k$  เป็นตัวแปรต้น ซึ่งแสดงดังนี้ สำหรับ  $i = 1, 2, \dots, n$

$$\text{แบบจำลองเชิงเส้น } y_i = \beta_0 + \beta_1 x_{i1} + \beta_2 x_{i2} + \dots + \beta_k x_{ik} + \epsilon_i$$

ภายใต้สมมติฐานว่า  $\epsilon_i$  มีการแจกแจงปกติที่มีค่าเฉลี่ยเท่ากับ 0 และความแปรปรวน  $\sigma^2$  เราจะได้ฟังก์ชันภาวะความน่าจะเป็น ดังนี้

$$L(\boldsymbol{\beta}|x_1, \dots, x_n) = \frac{1}{(2\pi\sigma^2)^{n/2}} \exp\left\{-\sum_{i=1}^n \frac{(y_i - (\beta_0 + \beta_1 x_{i1} + \beta_2 x_{i2} + \dots + \beta_k x_{ik}))^2}{2\sigma^2}\right\}$$

และได้ล็อกฟังก์ชันภาวะความน่าจะเป็น คือ

$$\log L(\boldsymbol{\beta}|x_1, \dots, x_n) = -\frac{n}{2} \log(2\pi\sigma^2) - \frac{1}{2\sigma^2} \sum_{i=1}^n (y_i - (\beta_0 + \beta_1 x_{i1} + \beta_2 x_{i2} + \dots + \beta_k x_{ik}))^2$$

เมื่อต้องการหาตัวประมาณแบบภาวะความน่าจะเป็นสูงสุดของ  $\beta_j$  สำหรับ  $j = 0, 1, \dots, k$  ทำได้โดยการหาอนุพันธ์ย่อยเทียบกับค่าพารามิเตอร์ และเทียบกับศูนย์ ดังนี้

$$\frac{\partial}{\partial \beta_j} \log L(\boldsymbol{\beta}|x_1, \dots, x_n) = 0$$

ได้ว่า

$$\begin{aligned} \sum_{i=1}^n (y_i - \beta_0 - \beta_1 x_{i1} - \beta_2 x_{i2} - \dots - \beta_k x_{ik}) &= 0 \\ \sum_{i=1}^n x_{i1} (y_i - \beta_0 - \beta_1 x_{i1} - \beta_2 x_{i2} - \dots - \beta_k x_{ik}) &= 0 \\ &\vdots \\ \sum_{i=1}^n x_{ik} (y_i - \beta_0 - \beta_1 x_{i1} - \beta_2 x_{i2} - \dots - \beta_k x_{ik}) &= 0 \end{aligned}$$

ซึ่งเขียนเป็นเมทริกซ์ได้ดังนี้

$$\begin{bmatrix} 1 & 1 & \dots & 1 \\ x_{11} & x_{21} & \dots & x_{n1} \\ x_{12} & x_{22} & \dots & x_{n2} \\ \vdots & \vdots & \ddots & \vdots \\ x_{1k} & x_{2k} & \dots & x_{nk} \end{bmatrix} \begin{bmatrix} (y_1 - \beta_0 - \beta_1 x_{11} - \beta_2 x_{12} - \dots - \beta_k x_{1k}) \\ (y_2 - \beta_0 - \beta_1 x_{21} - \beta_2 x_{22} - \dots - \beta_k x_{2k}) \\ \vdots \\ (y_n - \beta_0 - \beta_1 x_{n1} - \beta_2 x_{n2} - \dots - \beta_k x_{nk}) \end{bmatrix} = \begin{bmatrix} 0 \\ 0 \\ \vdots \\ 0 \end{bmatrix}$$

$$\begin{bmatrix} 1 & 1 & \dots & 1 \\ x_{11} & x_{21} & \dots & x_{n1} \\ x_{12} & x_{22} & \dots & x_{n2} \\ \vdots & \vdots & \ddots & \vdots \\ x_{1k} & x_{2k} & \dots & x_{nk} \end{bmatrix} \begin{bmatrix} y_1 \\ y_2 \\ \vdots \\ y_n \end{bmatrix} - \begin{bmatrix} 1 & x_{11} & x_{12} & \dots & x_{1k} \\ 1 & x_{21} & x_{22} & \dots & x_{2k} \\ \vdots & \vdots & \vdots & \ddots & \vdots \\ 1 & x_{n1} & x_{n2} & \dots & x_{nk} \end{bmatrix} \begin{bmatrix} \beta_0 \\ \beta_1 \\ \vdots \\ \beta_k \end{bmatrix} = \begin{bmatrix} 0 \\ 0 \\ \vdots \\ 0 \end{bmatrix} \quad \text{นั่นคือ}$$

$$\mathbf{X}'[\mathbf{y} - \mathbf{X}\boldsymbol{\beta}] = \mathbf{0}$$

$$\mathbf{X}'\mathbf{y} - \mathbf{X}'\mathbf{X}\boldsymbol{\beta} = \mathbf{0}$$

$$\mathbf{X}'\mathbf{X}\boldsymbol{\beta} = \mathbf{X}'\mathbf{y}$$

ทำให้ได้ว่า ตัวประมาณค่าภาวะความน่าจะเป็นสูงสุด  $\hat{\boldsymbol{\beta}}$  คือ

$$\hat{\boldsymbol{\beta}} = (\mathbf{X}'\mathbf{X})^{-1}\mathbf{X}'\mathbf{y} \quad (2.9)$$

เมื่อ  $(\mathbf{X}'\mathbf{X})$  เป็นเมทริกซ์ที่หาตัวผกผันได้ (invertible matrix)

**ตัวอย่าง 2.4.** พิจารณาข้อมูลในตารางที่ 2.1 แล้วหาสมการพยากรณ์สำหรับตัวแปรตาม  $y$  โดยใช้สมการ (2.9)

ค่าสังเกตที่	$y$	$x_1$	$x_2$
1	2	0	2
2	3	2	6
3	2	2	7
4	7	2	5
5	6	4	9
6	8	4	8
7	10	4	7
8	7	6	10
9	8	6	11
10	12	6	9
11	11	8	15
12	14	8	13

ตารางที่ 2.1: ตัวอย่างข้อมูลสำหรับแบบจำลองเชิงเส้น

จากสมการ (2.9) จะได้ว่า สมการการพยากรณ์สำหรับ ตัวแปรตาม  $y$  เมื่อพิจารณาตัวแปรต้น  $x_1, x_2$  จะเป็นดังนี้  $\hat{y} = 5.37 + 3.01x_1 + 1.29x_2$

### 2.1.4 แบบจำลองเฟ-เฮรียอต (Fay-Herriot model)

แบบจำลองเฟ-เฮรียอต เป็นแบบจำลองที่นำเสนอโดย Fay และ Herriot ในปี 1979 โดยแบบจำลองเฟ-เฮรียอตเป็นแบบจำลองที่มีการนิยามผ่าน 2 แบบจำลอง คือ

1. แบบจำลองเชื่อมโยง (Linking Model) เป็นแบบจำลองที่เชื่อมความสัมพันธ์เชิงเส้นระหว่างตัวแปรอิสระกับตัวแปรตาม
2. แบบจำลองการสุ่มตัวอย่าง (Sampling Model) เป็นแบบจำลองที่แสดงความสัมพันธ์ระหว่างค่าจริงกับค่าประมาณ

สำหรับหน่วยศึกษาที่  $i = 1, 2, \dots, n$  กำหนดให้  $x'_i$  คือเวกเตอร์ของตัวแปรอิสระสำหรับหน่วยที่  $i$  ดังนั้น แบบจำลองเชื่อมโยง สามารถนิยามโดย

$$\theta_i = x'_i \beta + v_i \quad (2.10)$$

เมื่อ  $v_i$  คือ ความคลาดเคลื่อนจากการถดถอย ซึ่งสมมติให้มีการแจกแจงปกติ ที่มีค่าเฉลี่ยเท่ากับ 0 และความแปรปรวนเท่ากับ  $A$

กำหนดให้ ตัวแปรประชากรที่เราสนใจคือ  $\theta_i$  ซึ่งถูกประมาณด้วยค่าประมาณจากการเก็บตัวอย่าง  $y_i$  ดังนั้น แบบจำลองการสุ่มตัวอย่างคือ

$$y_i = \theta_i + e_i \quad (2.11)$$

เมื่อ  $e_i$  คือ ความคลาดเคลื่อนที่เกิดจากการสุ่มตัวอย่าง ซึ่งสมมติให้มีการแจกแจงปกติ ที่มีค่าเฉลี่ยเท่ากับ 0 และความแปรปรวน  $\psi_i$

## 2.1.5 ตัวประมาณที่ไม่เอนเอียงเชิงเส้นเชิงประจักษ์ที่ดีที่สุด (Empirical best linear unbiased predictor: EBLUP)

ตัวประมาณเชิงเส้นที่ไม่เอนเอียงที่ดีที่สุด (Best linear unbiased predictor: BLUP)

ตัวประมาณเชิงเส้นที่ไม่เอนเอียงที่ดีที่สุด หมายถึง ตัวประมาณเชิงเส้นที่ไม่เอนเอียงและให้ค่าความแปรปรวนน้อยที่สุด

ทฤษฎีของเกาส์ มาร์คอฟ กล่าวว่า ถ้า  $E(\mathbf{y}) = \mathbf{X}\beta$  และ  $\text{cov}(\mathbf{y}) = \sigma^2 \mathbf{I}$  แล้วตัวประมาณ  $\hat{\beta}_0, \hat{\beta}_1, \dots, \hat{\beta}_k$  ใน (2.9) จะเป็นตัวประมาณที่ไม่เอนเอียงเชิงเส้นที่ดีที่สุด  
พิจารณาแบบจำลองเชิงเส้นแบบผสมทั่วไป

$$\mathbf{y} = \mathbf{X}\beta + \mathbf{Z}\mathbf{v} + \mathbf{e} \quad (2.12)$$

เมื่อ

$\mathbf{y}$  เป็นเมทริกซ์ขนาด  $n \times 1$

$\mathbf{X}$  และ  $\mathbf{Z}$  เป็นเมทริกซ์ที่มีเรงก์เต็ม (full rank) ขนาด  $n \times k$  และ  $n \times h$  ตามลำดับ

$\mathbf{v}$  และ  $\mathbf{e}$  เป็นเวกเตอร์สุ่มที่มีการแจกแจงที่เป็นอิสระต่อกัน มีค่าเฉลี่ยเท่ากับ  $\mathbf{0}$  และเมทริกซ์ความแปรปรวนร่วมเป็น  $\mathbf{G}$  และ  $\mathbf{R}$  ตามลำดับ ซึ่งขึ้นกับพารามิเตอร์ความแปรปรวน  $\delta = (\delta_1, \dots, \delta_q)'$

ให้  $V$  เป็นเมทริกซ์ของความแปรปรวนร่วม โดยที่  $V = R + ZGZ'$  เป็นเมทริกซ์ไม่เอกฐาน เราสนใจที่จะประมาณ  $\mu = I'\beta + m'v$  ของพารามิเตอร์การถดถอย  $\beta$  และค่าจริง  $v$  สำหรับค่าคงที่  $I$  และ  $m$  ตัวประมาณเชิงเส้นของ  $\mu$  อยู่ในรูป  $\hat{\mu} = a'y + b$  เมื่อทราบค่า  $a, b$  และ  $\hat{\mu}$  จะเป็นตัวประมาณที่ไม่เอนเอียงถ้า

$$E(\hat{\mu}) = \mu$$

และค่าเฉลี่ยกำลังสองของความคลาดเคลื่อน (MSE) ของ  $\hat{\mu}$  คือ

$$MSE(\hat{\mu}) = E(\hat{\mu} - \mu)^2 = \text{var}(\hat{\mu})$$

ในกรณีที่เราทราบค่า  $\delta$  ตัวประมาณ BLUP ของ  $\mu$  คือ

$$\tilde{\mu}^H = t(\delta, y) = I'\tilde{\beta} + m'\tilde{v} = I'\tilde{\beta} + m'GZ'V^{-1}(y - X\tilde{\beta}) \quad (2.13)$$

เมื่อ  $\tilde{\beta} = \tilde{\beta}(\delta) = (X'V^{-1}X)^{-1}X'V^{-1}y$  และ  $\tilde{v} = \tilde{v}(\delta) = GZ'V^{-1}(y - X\tilde{\beta})$

ที่มาของตัวประมาณ BLUP ของ  $\mu$

ตัวประมาณเชิงเส้น  $\hat{\mu} = a'y + b$  เป็นตัวประมาณที่ไม่เอนเอียงสำหรับ  $\mu = I'\beta + m'v$  ภายใต้แบบจำลอง  $y = X\beta + Zv + e$  นั่นคือ  $E(\hat{\mu}) = \mu$  ก็ต่อเมื่อ  $a'X = I'$  และ  $b = 0$  ค่าเฉลี่ยกำลังสองของความคลาดเคลื่อน (MSE) ของ  $\hat{\mu}$  คือ

$$MSE(\hat{\mu}) = \text{var}(\hat{\mu}) = a'Va - 2a'ZGm + m'Gm$$

หลังจากนั้นทำให้  $\text{var}(\hat{\mu})$  น้อยที่สุด โดยขึ้นกับเงื่อนไขความเอนเอียง  $a'X = I'$  และใช้การคูณลากรางจ์ด้วย  $2\lambda$  จะได้

$$Va + X\lambda = ZGm$$

แก้สมการหา  $a$  จะได้

$$a = -V^{-1}X\lambda + V^{-1}ZGm \quad (2.14)$$

แทน (2.14) ใน  $a'X = I'$  และแก้หา  $\lambda$  จะได้

$$\lambda = -(X'V^{-1}X)^{-1}I + (X'V^{-1}X)^{-1}X'V^{-1}ZGm \quad (2.15)$$



แทน (2.15) ใน (2.14) และใช้  $\tilde{\beta} = (X'V^{-1}X)^{-1}X'V^{-1}y$  จะได้

$$a'y = I'\tilde{\beta} + m'GZ'V^{-1}(y - X\tilde{\beta})$$

ซึ่งเป็นตัวเดียวกับสมการ (2.13)

ตัวประมาณที่ไม่เอนเอียงเชิงเส้นที่ดีที่สุดของแบบจำลองเฟ-เฮรอตที่นิยามใน (2.10) และ (2.11) คือ

$$\hat{\theta}_i^{EBLUP} = \frac{\hat{A}}{\hat{A} + \psi_i} y_i + \frac{\psi_i}{\hat{A} + \psi_i} x_i' \beta$$

ซึ่งเป็นผลบวกแบบถ่วงน้ำหนักระหว่างค่าที่ได้จากการสำรวจ ( $y_i$ ) และ ค่าที่ได้จากสมการถดถอย ( $x_i' \beta$ ) โดยน้ำหนักคือ  $\frac{\hat{A}}{\hat{A} + \psi_i}$  และ  $\frac{\psi_i}{\hat{A} + \psi_i}$  โดยที่  $\hat{A}$  คือค่าประมาณความแปรปรวนที่ได้จากวิธีการประมาณแบบภาวะความน่าจะเป็นสูงสุด

## 2.1.6 ข้อมูลพาเนล (Panel Data)

ข้อมูลพาเนลเป็นรูปแบบข้อมูลที่มีการบันทึกข้อมูลของหน่วยสำรวจหลายหน่วย โดยบันทึกเป็นค่าตัวแปรที่สนใจจากหลายหน่วยเวลา ทำให้สามารถศึกษาแนวโน้มระหว่างเวลาได้ โดยแบบจำลองของข้อมูลพาเนล เป็นดังนี้

สำหรับ  $i = 1, 2, \dots, I$  และ  $t = 1, 2, \dots, T$

$$y_{it} = \alpha_i + \beta_i X_{it} + u_{it}$$

เมื่อ

$I$  คือ จำนวนของหน่วยข้อมูล

$T$  คือ ระยะเวลาในการศึกษา

$y_{it}$  คือ ค่าสังเกตของตัวแปรตามจากหน่วยวิเคราะห์ที่  $i$  ณ เวลา  $t$

$X_{it}$  คือ ค่าสังเกตของตัวแปรอิสระจากหน่วยวิเคราะห์ที่  $i$  ณ เวลา  $t$

$u_{it}$  คือ ตัวแปรสุ่มคลาดเคลื่อน

$\alpha_i, \beta_i$  คือ พารามิเตอร์ของแบบจำลอง

**ตัวอย่าง 2.5.** พิจารณาข้อมูลรายได้รายปีเฉลี่ยของจังหวัด ( $y_{it}$ ) โดยมีสัดส่วนของคนที่เรียนจบปริญญาตรี ( $x_{it1}$ ) และ สัดส่วนของคนที่ทำงานรับจ้าง ( $x_{it2}$ ) เป็นตัวแปรอิสระ ซึ่งสามารถแสดงข้อมูลได้ดังตาราง 2.2

จังหวัด ( $i$ )	เวลา ( $t$ )	รายได้เฉลี่ยต่อคน ( $y_{it}$ พันบาท)	สัดส่วนของคนที่จบ ปริญญาตรี ( $x_{it1}$ )	สัดส่วนของคนที่ ทำงานรับจ้าง ( $x_{it2}$ )
1	1	156	0.203	0.443
1	2	124	0.301	0.346
1	3	508	0.498	0.333
1	4	113	0.542	0.232
1	5	429	0.630	0.360
2	1	975	0.493	0.391
2	2	720	0.241	0.498
2	3	844	0.270	0.576
⋮	⋮	⋮	⋮	⋮

ตารางที่ 2.2: ตารางแสดงข้อมูลรายได้เฉลี่ย สัดส่วนของคนที่เรียนจบปริญญาตรี สัดส่วนของคนที่ทำงานรับจ้างของ 10 จังหวัด ระยะเวลา 5 ปี

จากตัวอย่างนี้ เราสมมติให้เป็นการเก็บข้อมูล 5 ปี ของจังหวัด 10 จังหวัด จะเห็นว่าเรามีข้อมูลที่เป็นรายได้เฉลี่ยสัดส่วนของคนที่เรียนจบปริญญาตรี และสัดส่วนของคนที่ทำงานรับจ้างของจังหวัดที่  $i$  เมื่อ  $i = 1, 2, \dots, 10$  ตั้งแต่ปีที่ 1 จนถึงปีที่ 5 ซึ่งสามารถนำมาเขียนเป็นแบบจำลองได้ดังนี้

สำหรับ  $i = 1, 2, \dots, 10$  และ  $t = 1, 2, \dots, 5$

$$y_{it} = \alpha_i + \beta_1 x_{it1} + \beta_2 x_{it2} + u_{it}$$

เมื่อ  $y_{it}$  คือ รายได้เฉลี่ยของจังหวัดที่  $i$  ณ เวลา  $t$

$x_{it1}$  คือ สัดส่วนของคนที่เรียนจบปริญญาตรีของจังหวัดที่  $i$  ณ เวลา  $t$

$x_{it2}$  คือ สัดส่วนของคนที่ทำงานรับจ้างของจังหวัดที่  $i$  ณ เวลา  $t$

## 2.1.7 แผนการสุ่มตัวอย่าง

ในการวางแผนการสำรวจ จะต้องมีการตัดสินใจว่าจะดำเนินการสำมะโนหรือสำรวจตัวอย่าง หากทำการสำรวจตัวอย่าง ก็จะต้องมีการวางแผนการสุ่มตัวอย่างว่าจะมีการเลือกตัวอย่างอย่างไร และเลือกด้วยวิธีใด ซึ่งวิธีการเลือกตัวอย่าง แบ่งออกเป็น 2 ประเภท คือ การเลือกหน่วยตัวอย่างโดยไม่ใช้ความน่าจะเป็น และการเลือกหน่วยตัวอย่างโดยใช้ความน่าจะเป็น ในที่นี้จะกล่าวถึงการเลือกหน่วยตัวอย่างโดยใช้ความน่าจะเป็น ซึ่งวิธีนี้สามารถประมาณค่าประชากรได้ สำหรับแผนการสุ่มตัวอย่าง สามารถแบ่งตามขั้นตอนของการเลือกหน่วยตัวอย่างได้เป็น แผนการสุ่มตัวอย่างขั้นเดียว (single stage sampling) และ แผนการสุ่มตัวอย่างหลายขั้น (multi-stage sampling)

### 1. แผนการสุ่มตัวอย่างขั้นเดียว (single stage sampling)

แผนการสุ่มตัวอย่างขั้นเดียว เป็นแผนการสุ่มตัวอย่างซึ่งกำหนดให้มีการเลือกหน่วยตัวอย่างที่ให้ข้อมูลได้โดยตรงจากประชากร โดยสามารถกำหนดวิธีเลือกหน่วยตัวอย่างได้หลายวิธี เช่น แผนการสุ่มตัวอย่างอย่างง่าย (simple random sampling: SRS) หรือแผนการสุ่มตัวอย่างแบบมีระบบ (systematic sampling: SYS) หรือแผนการสุ่มตัวอย่างแบบแบ่งชั้นภูมิ (stratified sampling) หรือแผนการสุ่มตัวอย่างแบบแบ่งกลุ่ม (cluster sampling) ในที่นี้จะขอกล่าวถึงเพียง 2 แบบ คือ แผนการสุ่มตัวอย่างอย่างง่าย (simple random sampling: SRS) และ แผนการสุ่มตัวอย่างแบบแบ่งชั้นภูมิ (stratified sampling)

#### (a) แผนการสุ่มตัวอย่างอย่างง่าย (simple random sampling: SRS)

แผนการสุ่มตัวอย่างอย่างง่าย เป็นการเลือกหน่วยตัวอย่างขนาด  $n$  หน่วย จากประชากร  $N$  หน่วย โดยหน่วยตัวอย่างแต่ละหน่วยในประชากร มีโอกาสถูกเลือกเท่า ๆ กัน สำหรับในแต่ละครั้งของการเลือก และตัวอย่างแต่ละตัวอย่างที่เป็นไปได้ทั้งหมด มีโอกาสที่จะถูกเลือกเท่า ๆ กัน โดยสูตรการประมาณค่าเฉลี่ยและความแปรปรวนของค่าประมาณ สามารถคำนวณได้ดังต่อไปนี้

##### i. การประมาณค่าเฉลี่ย ( $\bar{x}$ )

$$\bar{x} = \frac{1}{n} \sum_{i=1}^n x_i \quad (2.16)$$

ii. การประมาณค่าความแปรปรวนของค่าประมาณ

$$v(\bar{x}) = \frac{s^2}{n} \left( \frac{N-n}{N} \right) = \left( 1 - \frac{n}{N} \right) \frac{s^2}{n} \quad (2.17)$$

เมื่อ

$N$  คือ จำนวนหน่วยทั้งหมดในประชากร

$n$  คือ จำนวนหน่วยในตัวอย่าง

$x_i$  คือ ข้อมูลของหน่วยที่  $i$  สำหรับ  $i = 1, 2, \dots, n$

$s^2 = \frac{1}{n-1} \sum_{i=1}^n (x_i - \bar{x})^2$  คือ ค่าความแปรปรวนของตัวอย่าง

(b) แผนการสุ่มตัวอย่างแบบแบ่งชั้นภูมิ (stratified sampling)

แผนการสุ่มตัวอย่างแบบแบ่งชั้นภูมิเป็นแผนการสุ่มตัวอย่างที่มีการแบ่งหน่วยต่าง ๆ ในประชากรออกเป็นกลุ่ม ๆ ตามลักษณะบางอย่าง โดยที่ประชากรที่อยู่ในกลุ่มเดียวกันจะมีลักษณะดังกล่าวเหมือนกัน แล้วจึงทำการเลือกหน่วยตัวอย่างจากประชากรแต่ละกลุ่มมาให้ครบทุกกลุ่ม วิธีการแบ่งประชากรออกเป็นกลุ่ม ๆ ตามลักษณะบางอย่าง เรียกว่า การแบ่งชั้นภูมิ (stratification) และเรียกแต่ละกลุ่มของประชากรที่ถูกแบ่งว่า ชั้นภูมิ (stratum) โดยสูตรการประมาณค่าเฉลี่ยและความแปรปรวนของค่าประมาณค่า สามารถคำนวณได้ดังต่อไปนี้

สมมติแบ่งประชากร  $N$  หน่วย ออกเป็น  $H$  ชั้นภูมิ โดยแต่ละชั้นภูมิประกอบด้วย  $N_1, N_2, \dots, N_H$  หน่วยตามลำดับ และ  $\sum_{i=1}^H N_i = N$  และในแต่ละชั้นภูมิได้เลือกตัวอย่างขึ้นมา จำนวน  $n_1, n_2, \dots, n_h$  หน่วย และ  $\sum_{i=1}^H n_i = n$  จะได้ว่า

i. ค่าเฉลี่ยตัวอย่าง ( $\bar{x}_{st}$ ) คือ

$$\bar{x}_{st} = \sum_{h=1}^H W_h \bar{x}_h \quad (2.18)$$

ii. ค่าความแปรปรวนของค่าประมาณค่าเฉลี่ยตัวอย่าง คือ

$$v(\bar{x}_{st}) = \sum_{h=1}^H \frac{W_h^2 s_h^2}{n_h} \left( \frac{N_h - n_h}{N_h} \right) = \sum_{h=1}^H W_h^2 \left( 1 - \frac{n_h}{N_h} \right) \frac{s_h^2}{n_h} \quad (2.19)$$

เมื่อ

สำหรับ  $h = 1, 2, \dots, H$

$N_h$  คือ จำนวนหน่วยทั้งหมดในประชากรในชั้นภูมิที่  $h$

$n_h$  คือ จำนวนหน่วยตัวอย่างในชั้นภูมิที่  $h$

$x_{hi}$  คือ ข้อมูลของหน่วยที่  $i$  ในชั้นภูมิที่  $h$

$W_h = \frac{N_h}{n_h}$  คือ ค่าถ่วงน้ำหนักของตัวอย่างในชั้นภูมิที่  $h$

$\bar{x}_h = \frac{1}{n_h} \sum_{i=1}^{n_h} x_{hi}$  คือ ค่าเฉลี่ยของตัวอย่างในชั้นภูมิที่  $h$

$s_h^2 = \frac{1}{n_h - 1} \sum_{i=1}^{n_h} (x_{hi} - \bar{x}_h)^2$  คือ ความแปรปรวนของตัวอย่างในชั้นภูมิที่  $h$

## 2. แผนการสุ่มตัวอย่างหลายชั้น (multi-stage sampling)

แผนการสุ่มตัวอย่างแบบหลายชั้นเป็นแผนการสุ่มตัวอย่างที่มีกระบวนการเลือกหน่วยตัวอย่างมากกว่า 1 ขั้นตอน ซึ่งหน่วยตัวอย่างแต่ละชั้นก็จะแตกต่างกันไป ในการเลือกหน่วยตัวอย่างแต่ละขั้นตอนนั้นอาจจะใช้แผนการสุ่มตัวอย่างแบบใดก็ได้ ในที่นี้จะกล่าวถึง กรณีแผนการสุ่มตัวอย่างแบบสองขั้นเท่านั้น

แผนการสุ่มตัวอย่างแบบสองขั้น (two stage sampling) เป็นการสุ่มตัวอย่างที่มีขั้นตอนการเลือกหน่วยตัวอย่างจากประชากร 2 ชั้น โดยเลือกจากหน่วยที่ใหญ่ก่อน แล้วจึงทำการเลือกหน่วยย่อยในหน่วยใหญ่ด้วยวิธีใดวิธีหนึ่งตามที่วางแผนไว้

- (a) ขั้นแรก (primary stage) เป็นการเลือกกลุ่มของหน่วยตัวอย่างขึ้นมาจำนวนหนึ่ง กลุ่มของหน่วยตัวอย่างในขั้นแรกเรียกว่า หน่วยตัวอย่างขั้นแรก (primary sampling unit: PSU)
- (b) ขั้นที่สอง (secondary stage) เป็นการเลือกหน่วยตัวอย่างจำนวนหนึ่ง จากกลุ่มของหน่วยตัวอย่างที่สุ่มเลือกมาได้จากขั้นแรก และทำการเก็บรวบรวมข้อมูลจากหน่วยตัวอย่างที่เลือกได้ในขั้นที่สอง กลุ่มของหน่วยตัวอย่างที่สุ่มเลือกมาในขั้นที่สอง เรียกว่า หน่วยตัวอย่างขั้นที่สอง (secondary sampling unit: SSU)

**ตัวอย่าง 2.6.** การสำรวจภาวะเศรษฐกิจและสังคมของครัวเรือนในจังหวัดเชียงราย จำแนกตามเขตการปกครอง จะมีการสุ่มตัวอย่างตามแผนการสุ่มตัวอย่างหลายชั้น โดยที่

- (a) หน่วยตัวอย่างชั้นแรก คือ เขตแจงนับตัวอย่าง (enumeration area: EA)  
 (b) หน่วยตัวอย่างชั้นที่สอง คือ ครัวเรือนตัวอย่าง

## 2.2 งานวิจัยในอดีต

ในปี ค.ศ. 2010 Li และ Lahiri ได้เสนอแนวคิดการปรับภาวะความน่าจะเป็นสูงสุด (Adjusted maximum likelihood: ADM) เพื่อไม่ให้เกิดกรณีที่ความแปรปรวนจากการถดถอยจะมีค่าเท่ากับ 0 ดังนี้

กำหนดให้  $\mathbf{X} = (X_1, X_2, \dots, X_m)'$ ,  $\mathbf{v} = (v_1, v_2, \dots, v_m)'$ ,  $\mathbf{e} = (e_1, e_2, \dots, e_m)'$  และ  $\mathbf{Y} = (Y_1, Y_2, \dots, Y_m)'$  แบบจำลองเฟ-เฮรอตจะสามารถเขียนได้เป็น  $\mathbf{Y} = \mathbf{X}\beta + \mathbf{v} + \mathbf{e}$  ซึ่งเป็นกรณีพิเศษของแบบจำลองเชิงเส้นทั่วไปที่มีความแปรปรวนร่วม  $\Sigma = \mathbf{D} + \mathbf{A}$  เมื่อ  $\mathbf{D} = \text{diag}(D_1, D_2, \dots, D_m)$

Li และ Lahiri ได้เสนอแนวคิดการปรับภาวะความน่าจะเป็นสูงสุด โดย

$$L_{adj}(A) = A \times L(A)$$

เมื่อ  $L(A)$  เป็นฟังก์ชันภาวะความน่าจะเป็น ซึ่งในที่นี้พิจารณา 2 ฟังก์ชัน คือ

1. ฟังก์ชันภาวะความน่าจะเป็นโปรไฟล์  $L_p(A) = c|\Sigma|^{-\frac{1}{2}} \exp\{-\frac{1}{2}\mathbf{y}'\mathbf{P}\mathbf{y}\}$
2. ฟังก์ชันภาวะความน่าจะเป็นเรสซิดัล  $L_{RE}(A) = c|\mathbf{X}'\Sigma^{-1}\mathbf{X}|^{-\frac{1}{2}}|\Sigma|^{-\frac{1}{2}} \exp\{-\frac{1}{2}\mathbf{y}'\mathbf{P}\mathbf{y}\}$

เมื่อ  $c$  เป็นค่าคงที่ และ  $\mathbf{P} = \Sigma^{-1} - \Sigma^{-1}\mathbf{X}(\mathbf{X}'\Sigma^{-1}\mathbf{X})^{-1}\mathbf{X}'\Sigma^{-1}$

Li และ Lahiri (2010) พบว่า วิธีการปรับฟังก์ชันภาวะความน่าจะเป็นสูงสุดของ Li และ Lahiri สามารถลดโอกาสที่  $A \leq 0$  ได้ อีกทั้งยังสามารถลดความเอนเอียง และลดค่าเฉลี่ยของกำลังสองของความคลาดเคลื่อนได้

# บทที่ 3

## วิธีดำเนินการและผลการดำเนินการ

ในบทนี้ เราขยายแนวคิดของ Li และ Lahiri ในการปรับฟังก์ชันภาวะความน่าจะเป็นของแบบจำลองเฟ-เฮรอต มายังแบบจำลองเฟ-เฮรอตแบบพานเนล และศึกษาบทประยุกต์ในข้อมูลรายได้ต่อครัวเรือนของประชากรไทย ในปี พ.ศ. 2552 - 2560 โดยเริ่มแรกจะกล่าวถึงแบบจำลองเฟ-เฮรอตแบบพานเนล การคำนวณการประมาณค่า และมีการจำลองข้อมูลเพื่อนำไปวิเคราะห์ค่าความแปรปรวน  $A$  และแสดงผลจากการจำลองข้อมูล และบทประยุกต์ของการปรับฟังก์ชันภาวะความน่าจะเป็นกับข้อมูลรายได้ของประชากรไทย โดยจะพิจารณารูปแบบของแบบจำลอง 2 รูปแบบ คือ

1. กรณีความแปรปรวนจากการถดถอยไม่ขึ้นกับเวลา
2. กรณีความแปรปรวนจากการถดถอยขึ้นกับเวลา

### 3.1 แบบจำลองเฟ-เฮรอตแบบพานเนล

โครงสร้างแบบจำลองเฟ-เฮรอตแบบพานเนลนั้น เกิดจากการขยายแนวคิดของแบบจำลองเฟ-เฮรอต โดยจะมีการพิจารณาถึงข้อมูลจากการสำรวจประชากรจากหลายหน่วยเวลา ซึ่งเป็นแบบจำลองที่นิยามผ่าน 2 แบบจำลอง คือ แบบจำลองเชื่อมโยง (Linking Model) และแบบจำลองการสุ่มตัวอย่าง (Sampling Model) ดังนี้

สำหรับ  $i = 1, \dots, n$  และ  $t = 1, \dots, T$  กำหนดให้

$\theta_{it}$  คือ ข้อมูลประชากรหน่วยที่  $i$  ณ เวลา  $t$

$X_{it}$  คือ เวกเตอร์ของตัวแปรอิสระสำหรับหน่วยที่  $i$  ณ เวลา  $t$

จะได้ว่าแบบจำลองเชื่อมโยง ซึ่งแสดงความสัมพันธ์ระหว่าง  $\theta_{it}$  และ  $X_{it}$  ดังนี้

$$\theta_{it} = X'_{it}\beta + v_i + \epsilon_{it}$$

เมื่อ

$v_i$  คือ ปัจจัยที่เกิดจากพื้นที่  $i$

$\epsilon_{it}$  คือ ความคลาดเคลื่อนจากการถดถอยของพื้นที่ที่  $i$  ณ เวลา  $t$  ซึ่งสมมติให้มีการแจกแจงปกติที่มีค่าเฉลี่ยเท่ากับ 0 และความแปรปรวนเท่ากับ  $A$

$n$  คือ ขนาดตัวอย่าง

$T$  คือ ระยะเวลาในการศึกษาข้อมูล

ในส่วนของแบบจำลองการสุ่มตัวอย่าง ซึ่งแสดงความสัมพันธ์ระหว่างค่าจริงและตัวประมาณค่า สามารถเขียนได้ด้วย

$$y_{it} = \theta_{it} + e_{it}$$

เมื่อ

$y_{it}$  คือ ค่าประมาณของข้อมูลประชากรหน่วยที่  $i$  ณ เวลา  $t$

$e_{it}$  คือ ความคลาดเคลื่อนที่เกิดจากการสุ่มตัวอย่าง ซึ่งสมมติให้มีการแจกแจงปกติที่มีค่าเฉลี่ยเท่ากับ 0 และความแปรปรวน  $\psi_{it}$

แบบจำลองดังกล่าวสามารถเขียนในรูปสมการเมทริกซ์ได้ดังต่อไปนี้

ให้  $\theta = (\theta_{1t}, \theta_{2t}, \dots, \theta_{nt})'$  เป็นข้อมูลประชากรที่สนใจ โดยแบบจำลองที่ใช้ เราสมมติว่า  $\theta$  มีความเกี่ยวข้องกับตัวแปรช่วย  $X_{it} = (x_{it1}, x_{it2}, \dots, x_{itp})'$  เมื่อ  $p$  คือ จำนวนตัวแปรช่วย  $\beta = (\beta_1, \beta_2, \dots, \beta_p)'$  เป็นเวกเตอร์ของสัมประสิทธิ์การถดถอย  $v = (v_1, v_2, \dots, v_n)'$  เป็นปัจจัยที่เกิดจากพื้นที่ จะได้แบบจำลองเชื่อมโยงเป็นดังนี้



$$\begin{bmatrix} \theta_{11} \\ \theta_{12} \\ \vdots \\ \theta_{1T} \\ \theta_{21} \\ \theta_{22} \\ \vdots \\ \theta_{2T} \\ \vdots \\ \theta_{n1} \\ \theta_{n2} \\ \vdots \\ \theta_{nT} \end{bmatrix} = \begin{bmatrix} x_{111} & x_{112} & \dots & x_{11p} \\ x_{121} & x_{122} & \dots & x_{12p} \\ \vdots & \vdots & \vdots & \vdots \\ x_{1T1} & x_{1T2} & \dots & x_{1Tp} \\ x_{211} & x_{212} & \dots & x_{21p} \\ x_{221} & x_{222} & \dots & x_{22p} \\ \vdots & \vdots & \vdots & \vdots \\ x_{2T1} & x_{2T2} & \dots & x_{2Tp} \\ \vdots & \vdots & \vdots & \vdots \\ x_{n11} & x_{n12} & \dots & x_{n1p} \\ x_{n21} & x_{n22} & \dots & x_{n2p} \\ \vdots & \vdots & \vdots & \vdots \\ x_{nT1} & x_{nT2} & \dots & x_{nTp} \end{bmatrix} \begin{bmatrix} \beta_1 \\ \beta_2 \\ \vdots \\ \beta_p \end{bmatrix} + \begin{bmatrix} v_1 \\ v_1 \\ \vdots \\ v_1 \\ v_2 \\ v_2 \\ \vdots \\ v_2 \\ \vdots \\ v_n \\ v_n \\ \vdots \\ v_n \end{bmatrix} + \begin{bmatrix} \epsilon_{11} \\ \epsilon_{12} \\ \vdots \\ \epsilon_{1T} \\ \epsilon_{21} \\ \epsilon_{22} \\ \vdots \\ \epsilon_{2T} \\ \vdots \\ \epsilon_{n1} \\ \epsilon_{n2} \\ \vdots \\ \epsilon_{nT} \end{bmatrix} \quad (3.1)$$

ซึ่งสามารถเขียนในรูปเมทริกซ์ได้เป็น  $\boldsymbol{\theta} = \mathbf{X}\boldsymbol{\beta} + \mathbf{v} + \boldsymbol{\epsilon}$  และเราได้ทำการแปลงเมทริกซ์ เพื่อจัดให้อยู่ในรูปของการถดถอยเชิงเส้น เป็นดังนี้

$$\begin{bmatrix} \theta_{11} \\ \theta_{12} \\ \vdots \\ \theta_{1T} \\ \theta_{21} \\ \theta_{22} \\ \vdots \\ \theta_{2T} \\ \vdots \\ \theta_{n1} \\ \theta_{n2} \\ \vdots \\ \theta_{nT} \end{bmatrix} = \begin{bmatrix} x_{111} & x_{112} & \dots & x_{11p} & 1 & 0 & 0 & \dots & 0 \\ x_{121} & x_{122} & \dots & x_{12p} & 1 & 0 & 0 & \dots & 0 \\ \vdots & \vdots & \vdots & \vdots & \vdots & \vdots & \vdots & \vdots & \vdots \\ x_{1T1} & x_{1T2} & \dots & x_{1Tp} & 1 & 0 & 0 & \dots & 0 \\ x_{211} & x_{212} & \dots & x_{21p} & 0 & 1 & 0 & \dots & 0 \\ x_{221} & x_{222} & \dots & x_{22p} & 0 & 1 & 0 & \dots & 0 \\ \vdots & \vdots & \vdots & \vdots & \vdots & \vdots & \vdots & \vdots & \vdots \\ x_{2T1} & x_{2T2} & \dots & x_{2Tp} & 0 & 1 & 0 & \dots & 0 \\ \vdots & \vdots & \vdots & \vdots & \vdots & \vdots & \vdots & \vdots & \vdots \\ x_{n11} & x_{n12} & \dots & x_{n1p} & 0 & 0 & \vdots & \dots & 1 \\ x_{n21} & x_{n22} & \dots & x_{n2p} & 0 & 0 & \vdots & \dots & 1 \\ \vdots & \vdots & \vdots & \vdots & \vdots & \vdots & \vdots & \vdots & \vdots \\ x_{nT1} & x_{nT2} & \dots & x_{nTp} & 0 & 0 & \vdots & \dots & 1 \end{bmatrix} \begin{bmatrix} \beta_1 \\ \beta_2 \\ \vdots \\ \beta_p \\ v_1 \\ v_2 \\ \vdots \\ v_n \end{bmatrix} + \begin{bmatrix} \epsilon_{11} \\ \epsilon_{12} \\ \vdots \\ \epsilon_{1T} \\ \epsilon_{21} \\ \epsilon_{22} \\ \vdots \\ \epsilon_{2T} \\ \vdots \\ \epsilon_{n1} \\ \epsilon_{n2} \\ \vdots \\ \epsilon_{nT} \end{bmatrix} \quad (3.2)$$

ซึ่งสามารถเขียนในรูปเมทริกซ์ได้เป็น

$$\theta = Z\gamma + \epsilon \quad (3.3)$$

โดยที่  $Z = [X, B]$  เมื่อ

$$B = \begin{bmatrix} 1 & 0 & 0 & \dots & 0 \\ 1 & 0 & 0 & \dots & 0 \\ & & \vdots & & \\ 1 & 0 & 0 & \dots & 0 \\ 0 & 1 & 0 & \dots & 0 \\ 0 & 1 & 0 & \dots & 0 \\ & & \vdots & & \\ 0 & 1 & 0 & \dots & 0 \\ & & \vdots & & \\ 0 & 0 & \vdots & \dots & 1 \\ 0 & 0 & \vdots & \dots & 1 \\ & & \vdots & & \\ 0 & 0 & \vdots & \dots & 1 \end{bmatrix}$$

เป็นเมทริกซ์ขนาด  $nT \times n$  และ  $\gamma = (\beta_1, \beta_2, \dots, \beta_p, v_1, v_2, \dots, v_n)'$  เป็นเมทริกซ์ขนาด  $(p + n) \times 1$

พิจารณาแบบจำลองการสุ่มตัวอย่าง ให้  $y = (y_{1t}, y_{2t}, \dots, y_{nt})'$  เป็นค่าประมาณของ  $\theta$  จะได้แบบจำลองการสุ่มตัวอย่างเป็น

$$\begin{bmatrix} y_{11} \\ y_{12} \\ \vdots \\ y_{1T} \\ y_{21} \\ y_{22} \\ \vdots \\ y_{2T} \\ \vdots \\ y_{n1} \\ y_{n2} \\ \vdots \\ y_{nT} \end{bmatrix} = \begin{bmatrix} \theta_{11} \\ \theta_{12} \\ \vdots \\ \theta_{1T} \\ \theta_{21} \\ \theta_{22} \\ \vdots \\ \theta_{2T} \\ \vdots \\ \theta_{n1} \\ \theta_{n2} \\ \vdots \\ \theta_{nT} \end{bmatrix} + \begin{bmatrix} e_{11} \\ e_{12} \\ \vdots \\ e_{1T} \\ e_{21} \\ e_{22} \\ \vdots \\ e_{2T} \\ \vdots \\ e_{n1} \\ e_{n2} \\ \vdots \\ e_{nT} \end{bmatrix} \quad (3.4)$$

ซึ่งสามารถเขียนในรูปเมทริกซ์ได้เป็น

$$\mathbf{y} = \boldsymbol{\theta} + \mathbf{e} \quad (3.5)$$

ดังนั้น แบบจำลองเฟ-เฮรอตแบบพานอลสามารถเขียนผ่าน (3.3) และ (3.5) เป็นเมทริกซ์ได้ดังนี้

$$\mathbf{y} = \mathbf{Z}\boldsymbol{\gamma} + \boldsymbol{\epsilon} + \mathbf{e} \quad (3.6)$$

## 3.2 กรณีความแปรปรวนจากการถดถอยไม่ขึ้นกับเวลา

ในหัวข้อนี้ เราจะพิจารณาแบบจำลองเฟ-เฮรอตแบบพานอล ในกรณีที่ความแปรปรวนจากการถดถอยไม่ขึ้นกับเวลา ซึ่งจะพิจารณาใน 2 กรณี คือ

1. กรณีสัมประสิทธิ์การถดถอยไม่ขึ้นกับเวลา
2. กรณีสัมประสิทธิ์การถดถอยขึ้นกับเวลา

### 3.2.1 กรณีสัมประสิทธิ์การถดถอยไม่ขึ้นกับเวลา

ในหัวข้อนี้ เราพิจารณาแบบจำลองพหุคูณกรณีสัมประสิทธิ์การถดถอยไม่ขึ้นกับเวลา ซึ่งสามารถเขียนได้ดังนี้

1. แบบจำลองเชิงมโยง:  $\theta_{it} = x'_{it}\beta + v_i + \epsilon_{it}$

โดยที่  $\epsilon_{it} \sim N(0, A)$  เมื่อ  $A$  คือ ความแปรปรวนที่เกิดจากการถดถอยที่ไม่ขึ้นกับเวลา

2. แบบจำลองการสุ่มตัวอย่าง:  $y_{it} = \theta_{it} + e_{it}$

โดยที่  $e_{it} \sim N(0, \psi_{it})$  เมื่อ  $\psi_{it}$  คือความแปรปรวนที่เกิดจากการสุ่มตัวอย่าง

โดยแบบจำลองเชิงมโยงจะเขียนกระจายได้ดังสมการ (3.1) และทำการแปลงเมทริกซ์เพื่อจัดให้อยู่ในรูปของการถดถอยเชิงเส้นได้ดังสมการ (3.2) และแบบจำลองการสุ่มตัวอย่างจะได้ดังสมการ (3.4) ทำให้ได้แบบจำลองเฟ-เฮรอตแบบพหุคูณกรณีสัมประสิทธิ์ไม่ขึ้นกับเวลาดังสมการ (3.6) จะได้ว่า  $\mathbf{y} \sim N(\mathbf{Z}\boldsymbol{\gamma}, \boldsymbol{\Sigma})$  เมื่อ  $\boldsymbol{\Sigma} = \mathbf{A}\mathbf{I} + \boldsymbol{\psi}$  โดยที่  $\boldsymbol{\psi} = \text{diag}(\psi_{11}, \psi_{12}, \dots, \psi_{nT})$

$$\begin{aligned}
 \boldsymbol{\Sigma} &= A \begin{bmatrix} 1 & 0 & \dots & 0 \\ 0 & 1 & \dots & 0 \\ \vdots & \vdots & \ddots & \vdots \\ 0 & 0 & \dots & 1 \end{bmatrix} + \begin{bmatrix} \psi_{11} & 0 & \dots & 0 \\ 0 & \psi_{12} & \dots & 0 \\ \vdots & \vdots & \ddots & \vdots \\ 0 & 0 & \dots & \psi_{nT} \end{bmatrix} \\
 &= \begin{bmatrix} A & 0 & \dots & 0 \\ 0 & A & \dots & 0 \\ \vdots & \vdots & \ddots & \vdots \\ 0 & 0 & \dots & A \end{bmatrix} + \begin{bmatrix} \psi_{11} & 0 & \dots & 0 \\ 0 & \psi_{12} & \dots & 0 \\ \vdots & \vdots & \ddots & \vdots \\ 0 & 0 & \dots & \psi_{nT} \end{bmatrix} \\
 &= \begin{bmatrix} A + \psi_{11} & 0 & \dots & 0 \\ 0 & A + \psi_{12} & \dots & 0 \\ \vdots & \vdots & \ddots & \vdots \\ 0 & 0 & \dots & A + \psi_{nT} \end{bmatrix} \tag{3.7}
 \end{aligned}$$

### 3.2.2 กรณีสัมประสิทธิ์การถดถอยขึ้นกับเวลา

ในหัวข้อนี้ เราจะพูดถึงการจำลองข้อมูล ในกรณีที่สัมประสิทธิ์การถดถอยขึ้นกับเวลา จะทำให้ได้แบบจำลองดังนี้

1. แบบจำลองเชื่อมโยง:  $\theta_{it} = x'_{it}\beta_t + v_i + \epsilon_{it}$

โดยที่  $\epsilon_{it} \sim N(0, A)$  เมื่อ  $A$  คือ ความแปรปรวนที่เกิดจากการถดถอยที่ไม่ขึ้นกับเวลา

2. แบบจำลองการสุ่มตัวอย่าง:  $y_{it} = \theta_{it} + e_{it}$

โดยที่  $e_{it} \sim N(0, \psi_{it})$  เมื่อ  $\psi_{it}$  คือความแปรปรวนที่เกิดจากการสุ่มตัวอย่าง

จากแบบจำลองเชื่อมโยงจะทำให้เขียนเป็นเมทริกซ์ได้ดังนี้ สำหรับ  $i = 1, 2, \dots, n$  และ  $t = 1, 2, \dots, T$

$$\begin{bmatrix} \theta_{11} \\ \theta_{12} \\ \vdots \\ \theta_{1T} \\ \theta_{21} \\ \theta_{22} \\ \vdots \\ \theta_{2T} \\ \vdots \\ \theta_{n1} \\ \theta_{n2} \\ \vdots \\ \theta_{nT} \end{bmatrix} = \begin{bmatrix} x_{111} & \dots & x_{11p} & 0 & \dots & 0 & \dots & 0 & \dots & 0 \\ 0 & \dots & 0 & x_{121} & \dots & x_{12p} & \dots & 0 & \dots & 0 \\ \vdots & \vdots & \vdots & \vdots & \vdots & \vdots & \ddots & \vdots & \vdots & \vdots \\ 0 & \dots & 0 & 0 & \dots & 0 & \dots & x_{1T1} & \dots & x_{1Tp} \\ x_{211} & \dots & x_{21p} & 0 & \dots & 0 & \dots & 0 & \dots & 0 \\ 0 & \dots & 0 & x_{221} & \dots & x_{22p} & \dots & 0 & \dots & 0 \\ \vdots & \vdots & \vdots & \vdots & \vdots & \vdots & \ddots & \vdots & \vdots & \vdots \\ 0 & \dots & 0 & 0 & \dots & 0 & \dots & x_{2T1} & \dots & x_{2Tp} \\ \vdots & \vdots & \vdots & \vdots & \vdots & \vdots & \vdots & \vdots & \vdots & \vdots \\ x_{n11} & \dots & x_{n1p} & 0 & \dots & 0 & \dots & 0 & \dots & 0 \\ 0 & \dots & 0 & x_{n21} & \dots & x_{n2p} & \dots & 0 & \dots & 0 \\ \vdots & \vdots & \vdots & \vdots & \vdots & \vdots & \ddots & \vdots & \vdots & \vdots \\ 0 & \dots & 0 & 0 & \dots & 0 & \dots & x_{nT1} & \dots & x_{nTp} \end{bmatrix} \begin{bmatrix} \beta_{11} \\ \beta_{12} \\ \vdots \\ \beta_{1p} \\ \beta_{21} \\ \beta_{22} \\ \vdots \\ \beta_{2p} \\ \vdots \\ \beta_{T1} \\ \beta_{T2} \\ \vdots \\ \beta_{Tp} \end{bmatrix} + \begin{bmatrix} v_1 \\ v_1 \\ \vdots \\ v_1 \\ v_2 \\ v_2 \\ \vdots \\ v_2 \\ \vdots \\ v_n \\ v_n \\ \vdots \\ v_n \end{bmatrix} + \begin{bmatrix} \epsilon_{11} \\ \epsilon_{12} \\ \vdots \\ \epsilon_{1T} \\ \epsilon_{21} \\ \epsilon_{22} \\ \vdots \\ \epsilon_{2T} \\ \vdots \\ \epsilon_{n1} \\ \epsilon_{n2} \\ \vdots \\ \epsilon_{nT} \end{bmatrix}$$

และเราจะทำการแปลงเมทริกซ์เพื่อจัดให้อยู่ในรูปของการถดถอยเชิงเส้น เป็นดังนี้

$$\begin{bmatrix} \theta_{11} \\ \vdots \\ \theta_{1T} \\ \theta_{21} \\ \vdots \\ \theta_{2T} \\ \vdots \\ \theta_{n1} \\ \vdots \\ \theta_{nT} \end{bmatrix} = \begin{bmatrix} x_{111} & \dots & x_{11p} & 0 & 0 & \dots & 0 & 1 & 0 & \dots & 0 \\ \vdots & \vdots & \vdots & \ddots & \vdots & \vdots & \vdots & \vdots & \vdots & \vdots & \vdots \\ 0 & \dots & 0 & 0 & x_{1T1} & \dots & x_{1Tp} & 1 & 0 & \dots & 0 \\ x_{211} & \dots & x_{21p} & 0 & 0 & \dots & 0 & 0 & 1 & \dots & 0 \\ \vdots & \vdots & \vdots & \ddots & \vdots & \vdots & \vdots & \vdots & \vdots & \vdots & \vdots \\ 0 & \dots & 0 & 0 & x_{2T1} & \dots & x_{2Tp} & 0 & 1 & \dots & 0 \\ \vdots & \vdots & \vdots & \vdots & \vdots & \vdots & \vdots & \vdots & \vdots & \vdots & \vdots \\ x_{n11} & \dots & x_{n1p} & 0 & 0 & \dots & 0 & 0 & 0 & \dots & 1 \\ \vdots & \vdots & \vdots & \ddots & \vdots & \vdots & \vdots & \vdots & \vdots & \vdots & \vdots \\ 0 & \dots & 0 & 0 & x_{nT1} & \dots & x_{nTp} & 0 & 0 & \dots & 1 \end{bmatrix} \begin{bmatrix} \beta_{11} \\ \vdots \\ \beta_{1p} \\ \vdots \\ \beta_{T1} \\ \vdots \\ \beta_{Tp} \\ \vdots \\ v_1 \\ \vdots \\ v_n \end{bmatrix} + \begin{bmatrix} \epsilon_{11} \\ \vdots \\ \epsilon_{1T} \\ \epsilon_{21} \\ \vdots \\ \epsilon_{2T} \\ \vdots \\ \epsilon_{n1} \\ \vdots \\ \epsilon_{nT} \end{bmatrix}$$

ซึ่งสามารถเขียนในรูปเมทริกซ์ได้เป็น

$$\boldsymbol{\theta} = \mathbf{Z}\boldsymbol{\gamma} + \boldsymbol{\epsilon} \quad (3.8)$$

โดยที่  $\mathbf{Z} = [\mathbf{X}, \mathbf{B}]$  เมื่อ

$$\mathbf{B} = \begin{bmatrix} 1 & 0 & \dots & 0 \\ 1 & 0 & \dots & 0 \\ \vdots & \vdots & \vdots & \vdots \\ 1 & 0 & \dots & 0 \\ 0 & 1 & \dots & 0 \\ 0 & 1 & \dots & 0 \\ \vdots & \vdots & \vdots & \vdots \\ 0 & 1 & \dots & 0 \\ 0 & 0 & \dots & 1 \\ 0 & 0 & \dots & 1 \\ \vdots & \vdots & \vdots & \vdots \\ 0 & 0 & \dots & 1 \end{bmatrix}$$

เป็นเมทริกซ์ขนาด  $nT \times n$  และ  $\boldsymbol{\gamma} = (\beta_{11}, \beta_{12}, \dots, \beta_{1p}, \dots, \beta_{T1}, \beta_{T2}, \dots, \beta_{Tp}, v_1, v_2, \dots, v_n)'$   
เป็นเมทริกซ์ขนาด  $(Tp + n) \times 1$

เมื่อพิจารณาแบบจำลองการสุ่มตัวอย่าง จะสามารถเขียนในรูปเมทริกซ์ได้ดังสมการ (3.5) ดังนั้น แบบจำลองเฟ-เฮรอตแบบพหุคูณสามารถเขียนผ่าน (3.8) และ (3.5) ได้ดังนี้

$$\mathbf{y} = \mathbf{Z}\boldsymbol{\gamma} + \boldsymbol{\epsilon} + \mathbf{e} \quad (3.9)$$

จาก (3.9) จะได้ว่า  $\mathbf{y} \sim N(\mathbf{Z}\boldsymbol{\gamma}, \boldsymbol{\Sigma})$  เมื่อ  $\boldsymbol{\Sigma} = \mathbf{A} + \boldsymbol{\psi}$  ซึ่ง  $\boldsymbol{\Sigma}$  เป็นดังสมการ (3.7)

ในทางปฏิบัติ เราไม่รู้ค่า  $\mathbf{A}$  แต่เราจะสามารถประมาณค่า  $\mathbf{A}$  ได้ โดยค่าความแปรปรวน  $\mathbf{A}$  จะถูกประมาณโดยตัวประมาณแบบภาวะความน่าจะเป็นสูงสุดโปรไฟล์ (Profile maximum likelihood estimator: PML) ตัวประมาณแบบภาวะความน่าจะเป็นสูงสุดโปรไฟล์แบบปรับ (Adjusted profile maximum likelihood estimator: APML) ตัวประมาณแบบภาวะความน่าจะเป็นสูงสุดเรสซิดูอัล (Residual maximum likelihood estimator: REML) และ ตัวประมาณแบบภาวะความน่าจะเป็นสูงสุดเรสซิดูอัลแบบปรับ (Adjusted residual maximum likelihood estimator: AREML) โดยการใช้คำสั่ง `optim` ในโปรแกรม R

1. ตัวประมาณแบบภาวะความน่าจะเป็นสูงสุดโปรไฟล์ (Profile maximum likelihood estimator: PML)

จาก (3.6) จะได้ฟังก์ชันภาวะความน่าจะเป็นโปรไฟล์ เป็นดังนี้

$$\begin{aligned} L_{PML}(A|\mathbf{y}) &= f(\mathbf{y}|A) \\ &= \frac{1}{[(2\pi)^{nT}|\boldsymbol{\Sigma}|]^{1/2}} \exp\left\{-\frac{1}{2}(\mathbf{y} - \mathbf{z}\boldsymbol{\gamma})'\boldsymbol{\Sigma}^{-1}(\mathbf{y} - \mathbf{z}\boldsymbol{\gamma})\right\} \end{aligned}$$

เมื่อแทน  $\boldsymbol{\gamma}$  ด้วย  $\hat{\boldsymbol{\gamma}} = (\mathbf{Z}'\boldsymbol{\Sigma}^{-1}\mathbf{Z})^{-1}\mathbf{Z}'\boldsymbol{\Sigma}^{-1}\mathbf{y}$  จะได้ว่า

$$L_{PML}(A|\mathbf{y}) = \frac{1}{[(2\pi)^{nT}|\boldsymbol{\Sigma}|]^{1/2}} \exp\left\{-\frac{1}{2}\mathbf{y}'\mathbf{P}\mathbf{y}\right\}$$

เมื่อ  $\mathbf{P} = \boldsymbol{\Sigma}^{-1} - \boldsymbol{\Sigma}^{-1}\mathbf{Z}(\mathbf{Z}'\boldsymbol{\Sigma}^{-1}\mathbf{Z})^{-1}\mathbf{Z}'\boldsymbol{\Sigma}^{-1}$

และได้ว่าล็อกฟังก์ชันภาวะความน่าจะเป็นโปรไฟล์ จะเป็น

$$l_{PML}(A|\mathbf{y}) = -\frac{1}{2}(nT \log(2\pi) + \log|\boldsymbol{\Sigma}| + \mathbf{y}'\mathbf{P}\mathbf{y}) \quad (3.10)$$

2. ตัวประมาณแบบภาวะความน่าจะเป็นสูงสุดเรสซิดูอัล (Residual maximum likelihood estimator: REML)

ให้  $\mathbf{F}$  เป็นเมทริกซ์ที่มี  $\text{rank}(\mathbf{F}) = n - p$  ซึ่ง  $\mathbf{F}'\mathbf{Z} = 0$  และ  $\mathbf{W} = \mathbf{F}'\mathbf{y}$  จะได้ว่า  $\mathbf{W} \sim N(\mathbf{F}'\mathbf{Z}\boldsymbol{\gamma}, \mathbf{F}'\boldsymbol{\Sigma}\mathbf{F}) = N(0, \mathbf{F}'\boldsymbol{\Sigma}\mathbf{F})$

ดังนั้น ฟังก์ชันภาวะความน่าจะเป็นเรสซิดัล เป็นดังนี้

$$\begin{aligned}
 L_{REML}(A|\mathbf{y}) &= f(\mathbf{y}|A) \\
 &= \frac{1}{[(2\pi)^{nT}|\mathbf{F}'\boldsymbol{\Sigma}\mathbf{F}]^{1/2}} \exp\left\{-\frac{1}{2}\mathbf{W}'(\mathbf{F}'\boldsymbol{\Sigma}\mathbf{F})^{-1}\mathbf{W}\right\} \\
 &= \frac{1}{[(2\pi)^{nT}|\mathbf{F}'\boldsymbol{\Sigma}\mathbf{F}]^{1/2}} \exp\left\{-\frac{1}{2}\mathbf{y}'\mathbf{F}(\mathbf{F}'\boldsymbol{\Sigma}\mathbf{F})^{-1}\mathbf{F}'\mathbf{y}\right\} \\
 &= \frac{1}{[(2\pi)^{nT}|\mathbf{F}'\boldsymbol{\Sigma}\mathbf{F}]^{1/2}} \exp\left\{-\frac{1}{2}\mathbf{y}'\mathbf{P}\mathbf{y}\right\} \\
 &= \frac{1}{[(2\pi)^{nT}|\mathbf{Z}'\boldsymbol{\Sigma}^{-1}\mathbf{Z}||\boldsymbol{\Sigma}|]^{1/2}} \exp\left\{-\frac{1}{2}\mathbf{y}'\mathbf{P}\mathbf{y}\right\} \\
 &= \frac{1}{[(2\pi)^{nT}|\boldsymbol{\Sigma}|]^{1/2}} (|\mathbf{Z}'\boldsymbol{\Sigma}^{-1}\mathbf{Z}|)^{-1/2} \exp\left\{-\frac{1}{2}\mathbf{y}'\mathbf{P}\mathbf{y}\right\} \\
 &= (|\mathbf{Z}'\boldsymbol{\Sigma}^{-1}\mathbf{Z}|)^{-1/2} L_{PML}(A|y)
 \end{aligned}$$

และล็อกฟังก์ชันภาวะความน่าจะเป็นเรสซิดัล เป็นดังนี้

$$l_{REML}(A|\mathbf{y}) = -\frac{1}{2}(nT \log(2\pi) + \log|\boldsymbol{\Sigma}| + \log(|\mathbf{Z}'\boldsymbol{\Sigma}^{-1}\mathbf{Z}|) + \mathbf{y}'\mathbf{P}\mathbf{y}) \quad (3.11)$$

เมื่อ  $\mathbf{P} = \boldsymbol{\Sigma}^{-1} - \boldsymbol{\Sigma}^{-1}\mathbf{Z}(\mathbf{Z}'\boldsymbol{\Sigma}^{-1}\mathbf{Z})^{-1}\mathbf{Z}'\boldsymbol{\Sigma}^{-1}$

สำหรับตัวประมาณพารามิเตอร์ที่ใช้คือ ตัวประมาณที่ไม่เอนเอียงเชิงเส้นที่ดีที่สุด (best linear unbiased predictor: BLUP)

$$\theta_{it}^{BLUP} = \frac{A}{A + \psi_{it}} y_{it} + \frac{\psi_{it}}{A + \psi_{it}} z'_{it} \gamma$$

เมื่อ  $A$  คือค่าความแปรปรวน

หลังจากได้ตัวประมาณภาวะความน่าจะเป็นสูงสุด  $\hat{A}$  แล้ว จะได้ว่าตัวประมาณที่ไม่เอนเอียงเชิงเส้นเชิงประจักษ์ที่ดีที่สุด ( $\hat{\theta}^{EBLUP}$ ) ดังนี้

$$\hat{\theta}_{it}^{EBLUP} = \frac{\hat{A}}{\hat{A} + \psi_{it}} y_{it} + \frac{\psi_{it}}{\hat{A} + \psi_{it}} z'_{it} \gamma$$

จะเห็นว่ากรณีที่  $\hat{A} < 0$  จะขัดแย้งกับความหมายของ  $A$  ซึ่งเป็นความแปรปรวน ซึ่งจะทำให้ตัวประมาณค่า  $\hat{\theta}_{it}^{EBLUP}$  คลาดเคลื่อนจากค่าจริง และในกรณีที่  $\hat{A} = 0$  เขียนแทนด้วย  $\hat{A}_0$  จะทำให้  $\hat{\theta}_{it}^{EBLUP} = z'_{it} \gamma$  ซึ่งเป็นปัญหาเดียวกันกับที่เกิดขึ้นจากงานของ Li และ Lahiri เมื่อปี ค.ศ. 2010 ดังนั้น ต่อไปเราจะขยายหลักการปรับตัวประมาณ  $\hat{A}$  มาใช้กับแบบจำลองพาดานเพื่อที่จะทำให้  $\hat{A} > 0$  ดังนี้



$$L_{adj}(A|\mathbf{y}) = A \times L(A|\mathbf{y})$$

เมื่อ  $L(A, |\mathbf{y})$  คือฟังก์ชันภาวะความน่าจะเป็นแบบโปรไฟล์และแบบเรสซิดัล ตามลำดับต่อไปจะเป็นการแสดงว่า หลักการปรับตัวประมาณ  $\hat{A}$  ให้ค่า  $\hat{A}$  ที่เป็นบวก

**บทตั้ง 3.2.1** [ Li 2007 ] ให้  $f(\cdot, \cdot, \dots, \cdot)$  เป็นฟังก์ชันของ  $x = (x_1, x_2, \dots, x_n)$  ซึ่งเป็นฟังก์ชันต่อเนื่องและให้ค่าบวก โดยที่  $x_i \geq 0$  สำหรับ  $i = 1, \dots, n$  และ

$$\lim_{x_j \rightarrow \infty} \prod_{i=1}^n x_i^\alpha f(x_1, x_2, \dots, x_n) = \lim_{x_j \rightarrow \infty} g(x_1, x_2, \dots, x_n) = 0$$

สำหรับทุก  $j = 1, \dots, n$  ซึ่ง  $\alpha > 0$  จะได้ว่า มี  $x_{01}, x_{02}, \dots, x_{0n}$  ซึ่ง

$$g(x_{01}, x_{02}, \dots, x_{0n}) = \max_{x_1, x_2, \dots, x_n} g(x_1, x_2, \dots, x_n)$$

และ  $x_{0i} > 0$  สำหรับทุก  $i = 1, 2, \dots, n$

**ทฤษฎีบท 3.2.2** ตัวประมาณแบบภาวะความน่าจะเป็นสูงสุดโปรไฟล์แบบปรับ (APML estimator) และตัวประมาณแบบภาวะความน่าจะเป็นสูงสุดเรสซิดัลแบบปรับ (AREML estimator) มีค่ามากกว่าศูนย์

**บทพิสูจน์** พิจารณาตัวประมาณแบบภาวะความน่าจะเป็นสูงสุดโปรไฟล์แบบปรับ (APML estimator) จะได้ฟังก์ชันภาวะความน่าจะเป็นโปรไฟล์แบบปรับ ดังนี้

$$\begin{aligned} L_{APML}(A|\mathbf{y}) &= A \times L_{PML}(A|\mathbf{y}) \\ &= A \times \frac{1}{[(2\pi)^{nT} |\Sigma|]^{1/2}} \exp\left\{-\frac{1}{2} \mathbf{y}' \mathbf{P} \mathbf{y}\right\} \end{aligned}$$

เมื่อ  $\mathbf{P} = \Sigma^{-1} - \Sigma^{-1} \mathbf{Z} (\mathbf{Z}' \Sigma^{-1} \mathbf{Z})^{-1} \mathbf{Z}' \Sigma^{-1}$

เห็นได้ชัดว่าฟังก์ชันภาวะความน่าจะเป็นโปรไฟล์  $L_{PML}(A|\mathbf{y})$  เป็นฟังก์ชันของ  $A$  ซึ่งเป็นฟังก์ชันต่อเนื่องและให้ค่าบวก จะได้ว่า

$$\begin{aligned} &\lim_{A \rightarrow \infty} A \times \frac{1}{[(2\pi)^{nT} |\Sigma|]^{1/2}} \exp\left\{-\frac{1}{2} \mathbf{y}' \mathbf{P} \mathbf{y}\right\} \\ &= \frac{1}{[(2\pi)^{nT}]^{1/2}} \lim_{A \rightarrow \infty} \frac{A}{\prod_{d=1}^D |\Sigma_d|^{1/2}} \exp\left\{-\frac{1}{2} \mathbf{y}' \mathbf{P} \mathbf{y}\right\} \\ &= \frac{1}{[(2\pi)^{nT}]^{1/2}} \lim_{A \rightarrow \infty} \frac{A}{\prod_{d=1}^D |A + \psi_d|^{1/2}} \exp\left\{-\frac{1}{2} \mathbf{y}' \mathbf{P} \mathbf{y}\right\} \\ &= \frac{1}{[(2\pi)^{nT}]^{1/2}} \left( \lim_{A \rightarrow \infty} \frac{A}{\prod_{d=1}^D |A + \psi_d|^{1/2}} \right) \left( \lim_{A \rightarrow \infty} \exp\left\{-\frac{1}{2} \mathbf{y}' \mathbf{P} \mathbf{y}\right\} \right) \\ &= 0 \end{aligned}$$

เมื่อ  $\lim_{A \rightarrow \infty} \exp\{-\frac{1}{2}\mathbf{y}'\mathbf{P}\mathbf{y}\} = 1$

โดยบทตั้ง 3.2.1 จะได้ว่ามี  $A$  ซึ่ง  $L_{APML}(A|\mathbf{y})$  ให้ค่าสูงสุด โดยที่  $A > 0$  และในทำนองเดียวกัน ตัวประมาณแบบภาวะความน่าจะเป็นสูงสุดเรสซิดัลแบบปรับ (AREML estimator) มีค่ามากกว่าศูนย์

จากหลักการปรับ จะทำให้ได้ล็อกฟังก์ชันภาวะความน่าจะเป็น (log-likelihood function) ดังนี้

1. ตัวประมาณแบบภาวะความน่าจะเป็นสูงสุดโปรไฟล์แบบปรับ (Adjusted profile maximum likelihood estimator: APML)

$$l_{APML}(A|\mathbf{y}) = \log(A) - \frac{1}{2}(nT \log(2\pi) + \log|\Sigma| + \mathbf{y}'\mathbf{P}\mathbf{y})$$

$$\text{เมื่อ } \mathbf{P} = \Sigma^{-1} - \Sigma^{-1}\mathbf{Z}(\mathbf{Z}'\Sigma^{-1}\mathbf{Z})^{-1}\mathbf{Z}'\Sigma^{-1}$$

2. ตัวประมาณแบบภาวะความน่าจะเป็นสูงสุดเรสซิดัลแบบปรับ (Adjusted residual maximum likelihood estimator: AREML)

$$l_{AREML}(A|\mathbf{y}) = \log(A) - \frac{1}{2}(nT \log(2\pi) + \log|\Sigma| + \log(|\mathbf{Z}'\Sigma^{-1}\mathbf{Z}|) + \mathbf{y}'\mathbf{P}\mathbf{y})$$

$$\text{เมื่อ } \mathbf{P} = \Sigma^{-1} - \Sigma^{-1}\mathbf{Z}(\mathbf{Z}'\Sigma^{-1}\mathbf{Z})^{-1}\mathbf{Z}'\Sigma^{-1}$$

### 3.2.3 การศึกษาเชิงตัวเลข

ในหัวข้อนี้ เราจะทำการศึกษาประสิทธิภาพของตัวประมาณค่าแบบภาวะความน่าจะเป็นสูงสุดแบบปรับ ทั้งในแบบภาวะความน่าจะเป็นสูงสุดโปรไฟล์และแบบภาวะความน่าจะเป็นสูงสุดเรสซิดัล โดยพิจารณาแบบจำลองเฟ-เฮรอตแบบพานอล สำหรับพื้นที่ย่อย 10 พื้นที่ และเวลา 5 ปี ดังนี้

กรณีสัมประสิทธิ์การถดถอยไม่ขึ้นกับเวลา

สำหรับ  $i = 1, 2, \dots, 10$  และ  $t = 1, 2, \dots, 5$

พิจารณา แบบจำลองเชื่อมโยง

$$\theta_{it} = x'_{it}\beta + v_i + \epsilon_{it} \quad (3.12)$$

เมื่อ  $\text{var}(\epsilon_{it}) = A$

และแบบจำลองการสุ่มตัวอย่าง

$$y_{it} = \theta_{it} + e_{it} \quad (3.13)$$

เมื่อ  $\text{var}(e_{it}) = \psi_{it}$

โดยมีการตั้งค่าการจำลองข้อมูลดังนี้

1. จำนวนครั้งที่ทำการจำลอง:  $K = 1000$  ครั้ง
2. จำนวนตัวอย่าง: 10
3. สัมประสิทธิ์การถดถอย  $\beta_1 = \beta_2 = 1$
4. จำลอง  $\theta_{it} (i = 1, 2, \dots, n)$  และ  $(t = 1, 2, \dots, T)$  จากสมการ (3.12)  
โดยจำลอง  $x_{it1}, x_{it2}$  จากการแจกแจงปกติ ที่มีค่าเฉลี่ย 10 ความแปรปรวน 1 และ  $1.05^2$  นั่นคือ  $x_{it1} \sim N(10, 1)$  และ  $x_{it2} \sim N(10, 1.05^2)$  ตามลำดับ
5. จำลอง  $v_i$  จากการแจกแจงปกติที่มีค่าเฉลี่ย 10 ความแปรปรวน  $1.65^2$   
นั่นคือ  $v_i \sim N(10, 1.65^2)$
6. จำลอง  $A$  ใน 3 กรณี คือ
  - (a) จำลอง  $A$  จากการแจกแจงยูนิฟอร์มในช่วง (1,2) นั่นคือ  $A \sim Unif(1, 2)$
  - (b) จำลอง  $A$  จากการแจกแจงยูนิฟอร์มในช่วง (4,6) นั่นคือ  $A \sim Unif(4, 6)$
  - (c) จำลอง  $A$  จากการแจกแจงยูนิฟอร์มในช่วง (9,11) นั่นคือ  $A \sim Unif(9, 11)$
7. จำลอง  $\epsilon_{it}$  จากการแจกแจงปกติที่มีค่าเฉลี่ย 0 ความแปรปรวน  $A$   
นั่นคือ  $\epsilon_{it} \sim N(0, A)$
8. จำลอง  $\psi_{it}$  ใน 2 กรณี คือ
  - (a) จำลอง  $\psi_{it}$  จากการแจกแจงยูนิฟอร์มในช่วง  $(\frac{1}{2}A, A)$  นั่นคือ  $\psi_{it} \sim Unif(\frac{1}{2}A, A)$
  - (b) จำลอง  $\psi_{it}$  จากการแจกแจงยูนิฟอร์มในช่วง  $(\frac{3}{2}A, 2A)$  นั่นคือ  $\psi_{it} \sim Unif(\frac{3}{2}A, 2A)$
9. จำลอง  $e_{it}$  จากการแจกแจงปกติที่มีค่าเฉลี่ย 0 ความแปรปรวน  $\psi_{it}$   
นั่นคือ  $e_{it} \sim N(0, \psi_{it})$
10. จำลอง  $y_{it}$  จาก  $y_{it} = \theta_{it} + e_{it}$

### กรณีสัมประสิทธิ์การถดถอยขึ้นกับเวลา

สำหรับ  $i = 1, 2, \dots, 10$  และ  $t = 1, 2, \dots, 5$

พิจารณา แบบจำลองเชิงมโยง

$$\theta_{it} = x'_{it}\beta_t + v_i + \epsilon_{it} \quad (3.14)$$

เมื่อ  $\text{var}(\epsilon_{it}) = A$

และ แบบจำลองการสุ่มตัวอย่าง

$$y_{it} = \theta_{it} + e_{it} \quad (3.15)$$

เมื่อ  $\text{var}(e_{it}) = \psi_{it}$

โดยมีการตั้งค่าการจำลองข้อมูลดังนี้

1. จำนวนครั้งที่ทำการจำลอง:  $K = 1000$  ครั้ง
2. จำนวนตัวอย่าง: 10
3. จำลองสัมประสิทธิ์การถดถอย  $\beta_1$  และ  $\beta_2$  จากการแจกแจงยูนิฟอร์มในช่วง (1,2) นั่นคือ  $\beta_1 \sim Unif(1, 2)$  และ  $\beta_2 \sim Unif(1, 2)$
4. จำลอง  $\theta_{it} (i = 1, 2, \dots, n)$  และ  $(t = 1, 2, \dots, T)$  จากสมการ (3.14) โดยจำลอง  $x_{it1}, x_{it2}$  จากการแจกแจงปกติ ที่มีค่าเฉลี่ย 10 ความแปรปรวน 1 และ  $1.05^2$  นั่นคือ  $x_{it1} \sim N(10, 1)$  และ  $x_{it2} \sim N(10, 1.05^2)$  ตามลำดับ
5. จำลอง  $v_i$  จากการแจกแจงปกติที่มีค่าเฉลี่ย 10 ความแปรปรวน  $1.65^2$  นั่นคือ  $v_i \sim N(10, 1.65^2)$
6. จำลอง  $A$  ใน 3 กรณี คือ
  - (a) จำลอง  $A$  จากการแจกแจงยูนิฟอร์มในช่วง (1,2) นั่นคือ  $A \sim Unif(1, 2)$
  - (b) จำลอง  $A$  จากการแจกแจงยูนิฟอร์มในช่วง (4,6) นั่นคือ  $A \sim Unif(4, 6)$
  - (c) จำลอง  $A$  จากการแจกแจงยูนิฟอร์มในช่วง (9,11) นั่นคือ  $A \sim Unif(9, 11)$
7. จำลอง  $\epsilon_{it}$  จากการแจกแจงปกติที่มีค่าเฉลี่ย 0 ความแปรปรวน  $A$  นั่นคือ  $\epsilon_{it} \sim N(0, A)$

8. จำลอง  $\psi_{it}$  ใน 2 กรณี คือ

(a) จำลอง  $\psi_{it}$  จากการแจกแจงยูนิฟอร์มในช่วง  $(\frac{1}{2}A, A)$  นั่นคือ  $\psi_{it} \sim Unif(\frac{1}{2}A, A)$

(b) จำลอง  $\psi_{it}$  จากการแจกแจงยูนิฟอร์มในช่วง  $(\frac{3}{2}A, 2A)$  นั่นคือ  $\psi_{it} \sim Unif(\frac{3}{2}A, 2A)$

9. จำลอง  $e_{it}$  จากการแจกแจงปกติที่มีค่าเฉลี่ย 0 ความแปรปรวน  $\psi_{it}$

นั่นคือ  $e_{it} \sim N(0, \psi_{it})$

10. จำลอง  $y_{it}$  จาก  $y_{it} = \theta_{it} + e_{it}$

หลังจากได้ข้อมูลจากการจำลองข้อมูลแล้ว จึงนำข้อมูลดังกล่าวมาหาจำนวนของการเกิด  $\hat{A} \leq 0$  และ หาค่าเฉลี่ยของกำลังสองของความคลาดเคลื่อน (Mean squared error) ของตัวประมาณค่าความแปรปรวนจากการถดถอย  $\hat{A}$  และหาค่าเฉลี่ยของกำลังสองของความคลาดเคลื่อน (Mean squared error) ของ  $\hat{\theta}_{it}^{EBLUP}$

การคำนวณค่าเฉลี่ยของกำลังสองของความคลาดเคลื่อน (Mean squared error: MSE) ของตัวประมาณค่า

1. การคำนวณ MSE ของตัวประมาณค่าความแปรปรวนจากการถดถอย  $\hat{A}$

$$MSE = \frac{1}{K} \sum_{k=1}^K (\hat{A}_k - A_k)^2 \quad (3.16)$$

เป็นสูตรคำนวณค่าเฉลี่ยของกำลังสองของความคลาดเคลื่อน (Mean squared error: MSE) ระหว่างตัวประมาณค่าความแปรปรวนจากการถดถอย  $\hat{A}$  และ ค่าจริงของความแปรปรวนจากการถดถอย  $A$

2. การคำนวณ MSE ของตัวประมาณค่าที่ไม่เอนเอียงเชิงเส้นเชิงประจักษ์ที่ดีที่สุด  $\hat{\theta}_{it}^{EBLUP}$

2.1 การคำนวณ MSE แบบที่ 1

$$MSE = \frac{1}{nTK} \sum_{i=1}^n \sum_{t=1}^T \sum_{k=1}^K (\hat{\theta}_{itk}^{EBLUP} - \theta_{itk}^{BLUP})^2 \quad (3.17)$$

เป็นสูตรคำนวณค่าเฉลี่ยของกำลังสองของความคลาดเคลื่อน (Mean squared error: MSE) ระหว่างตัวประมาณค่า  $\hat{\theta}_{it}^{EBLUP}$  ที่ได้จากค่าประมาณความแปรปรวนจากการถดถอย  $\hat{A}$  และตัวประมาณค่า  $\theta_{it}^{BLUP}$  ที่ได้จากค่าจริงของความแปรปรวนจากการถดถอย  $A$

## 2.2 การคำนวณ MSE แบบที่ 2

$$MSE = \frac{1}{nTK} \sum_{i=1}^n \sum_{t=1}^T \sum_{k=1}^K (\hat{\theta}_{itk}^{EBLUP} - \theta_{itk})^2 \quad (3.18)$$

เป็นสูตรคำนวณค่าเฉลี่ยของกำลังสองของความคลาดเคลื่อน (Mean squared error: MSE) ระหว่างตัวประมาณค่า  $\hat{\theta}_{it}^{EBLUP}$  ที่ได้จากค่าประมาณความแปรปรวนจากการถดถอย  $\hat{A}$  และค่าจริง  $\theta_{it}$

### ผลการวิเคราะห์การจำลองข้อมูล กรณีสัมประสิทธิ์การถดถอยไม่ขึ้นกับเวลา

ผลการวิเคราะห์การจำลองข้อมูลแสดงดังตารางที่ 3.1 ถึง ตารางที่ 3.12 โดยตารางที่ 3.1 ตารางที่ 3.5 และตารางที่ 3.9 แสดงจำนวนตัวประมาณความแปรปรวนจากการถดถอย  $\hat{A}_t$  มีค่าน้อยกว่าหรือเท่ากับ 0 จากแบบที่ไม่ปรับ และแบบที่ปรับแล้ว จากการจำลอง  $A$  จากการแจกแจงยูนิฟอร์มในช่วง (1,2), (4,6) และ (9,11) ตามลำดับ

ตารางที่ 3.2 ตารางที่ 3.6 และตารางที่ 3.10 แสดงค่าเฉลี่ยของกำลังสองของความคลาดเคลื่อนของตัวประมาณความแปรปรวนจากการถดถอย ( $\hat{A}$ ) เมื่อเทียบกับความแปรปรวนจากการถดถอยที่เราจำลองไว้ตั้งแต่ต้น ( $A$ ) จากแบบที่ไม่ปรับ และแบบที่ปรับแล้ว จากการจำลอง  $A$  จากการแจกแจงยูนิฟอร์มในช่วง (1,2), (4,6) และ (9,11) ตามลำดับ

ตารางที่ 3.3 ตารางที่ 3.7 และตารางที่ 3.11 แสดงค่าเฉลี่ยของกำลังสองของความคลาดเคลื่อนของตัวประมาณที่ไม่เอนเอียงเชิงเส้นเชิงประจักษ์ที่ดีที่สุดที่คำนวณจากตัวประมาณความแปรปรวนจากการถดถอย  $\hat{A}$  ( $\hat{\theta}_{it}^{EBLUP}$ ) เทียบกับที่คำนวณจากการถดถอยที่เราจำลองไว้ตั้งแต่ต้น ( $\hat{\theta}_{it}^{BLUP}$ ) และเทียบกับค่าจริงที่เราจำลองได้ตั้งแต่ต้น ( $\theta_{it}$ ) กรณี PML และ APML จากการจำลอง  $A$  จากการแจกแจงยูนิฟอร์มในช่วง (1,2), (4,6) และ (9,11) ตามลำดับ โดยที่  $\hat{A}_0$  ที่แสดงในตารางคือ  $\hat{A}_0 = \max(0, \hat{A})$

ตารางที่ 3.4 ตารางที่ 3.8 และตารางที่ 3.12 แสดงค่าเฉลี่ยของกำลังสองของความคลาดเคลื่อนของตัวประมาณที่ไม่เอนเอียงเชิงเส้นเชิงประจักษ์ที่ดีที่สุดที่คำนวณจากตัวประมาณความแปรปรวนจากการถดถอย  $\hat{A}$  ( $\hat{\theta}_{it}^{EBLUP}$ ) เทียบกับที่คำนวณจากการถดถอยที่เราจำลองไว้ตั้งแต่ต้น ( $\hat{\theta}_{it}^{BLUP}$ ) และเทียบกับค่าจริงที่เราจำลองได้ตั้งแต่ต้น ( $\theta_{it}$ ) กรณี REML และ AREML จากการจำลอง  $A$  จากการแจกแจงยูนิฟอร์มในช่วง (1,2), (4,6) และ (9,11) ตามลำดับ โดยที่  $\hat{A}_0$  ที่แสดงในตารางคือ  $\hat{A}_0 = \max(0, \hat{A})$

กรณีจำลอง  $A$  จากการแจกแจงยูนิฟอร์มในช่วง  $(1,2)$  นั่นคือ  $A \sim Unif(1,2)$

ความแปรปรวนจากการสุ่มตัวอย่าง	PML	APML	REML	AREML
$\psi_{it} \sim Unif(\frac{1}{2}A, A)$	1000	0	1000	0
$\psi_{it} \sim Unif(\frac{3}{2}A, 2A)$	1000	0	1000	0

ตารางที่ 3.1: จำนวนตัวประมาณความแปรปรวนจากการถดถอย  $\hat{A} \leq 0$  จากการจำลอง 1000 ครั้ง

จากตารางที่ 3.1 จะเห็นว่า จำนวนตัวประมาณความแปรปรวนจากการถดถอย  $\hat{A} \leq 0$  ของกรณีที่ยังไม่ได้ปรับทั้ง PML และ REML มีจำนวนมาก ในขณะที่ กรณีที่ปรับแล้วทั้ง APML และ AREML ไม่พบตัวประมาณความแปรปรวนจากการถดถอย  $\hat{A} \leq 0$

ความแปรปรวนจากการสุ่มตัวอย่าง	PML	APML	REML	AREML
$\psi_{it} \sim Unif(\frac{1}{2}A, A)$	899.9612	0.0785	698.4005	0.9357
$\psi_{it} \sim Unif(\frac{3}{2}A, 2A)$	1043.2437	0.1127	739.4091	0.9323

ตารางที่ 3.2: ค่าเฉลี่ยของกำลังสองของความคลาดเคลื่อน (Mean squared error) ของตัวประมาณความแปรปรวนจากการถดถอย  $\hat{A}$

จากตารางที่ 3.2 จะเห็นว่า ค่าเฉลี่ยของกำลังสองของความคลาดเคลื่อนของตัวประมาณความแปรปรวนจากการถดถอย  $\hat{A}$  ของกรณีที่ปรับแล้วทั้ง APML และ AREML มีค่าน้อยกว่า ค่าเฉลี่ยของกำลังสองของความคลาดเคลื่อนของกรณีที่ยังไม่ได้ปรับ ทั้ง PML และ REML

ความแปรปรวนจากการ สุ่มตัวอย่าง	ค่าจริง	PML		APML
		$\hat{A}$	$\hat{A}_0$	
$\psi_{it} \sim Unif(\frac{1}{2}A, A)$	$\hat{\theta}_{it}^{BLUP}$	0.0632	0.0839	0.0009
	$\theta_{it}$	0.1222	0.1497	0.0644
$\psi_{it} \sim Unif(\frac{3}{2}A, A)$	$\hat{\theta}_{it}^{BLUP}$	0.2341	0.0543	0.0010
	$\theta_{it}$	0.3234	0.1500	0.0965

ตารางที่ 3.3: ค่าเฉลี่ยของกำลังสองของความคลาดเคลื่อน (Mean squared error) ของ  $\hat{\theta}_{it}^{EBLUP}$  กรณี  $A \sim Unif(1, 2)$  ของ PML และ APML

จากตารางที่ 3.3 จะเห็นว่า ค่าเฉลี่ยของกำลังสองของความคลาดเคลื่อนของ  $\hat{\theta}_{it}^{EBLUP}$  เทียบกับ  $\hat{\theta}_{it}^{BLUP}$  และเทียบกับ  $\theta_{it}$  ของกรณีที่ปรับแล้ว มีค่าน้อยกว่ากรณีที่ยังไม่ปรับ

พิจารณา กรณีความแปรปรวนจากการสุ่มตัวอย่าง มีค่าน้อยกว่า ความแปรปรวนจากการถดถอย จะเห็นว่า เมื่อใช้ค่า  $\hat{A}$  จริง ซึ่งอาจเป็นลบได้ จะให้ค่าเฉลี่ยของกำลังสองของความคลาดเคลื่อน ที่น้อยกว่า เมื่อใช้ค่า  $\hat{A}_0$  ที่ใช้ค่า  $\hat{A}$  ที่มากกว่าศูนย์ เนื่องจากจำนวน  $\hat{A} \leq 0$  มีจำนวนมาก จึงส่งผลให้  $\frac{\hat{A}}{\hat{A} + \psi_{it}}$  มีค่าใกล้ 1

ความแปรปรวนจากการ สุ่มตัวอย่าง	ค่าจริง	REML		AREML
		$\hat{A}$	$\hat{A}_0$	
$\psi_{it} \sim Unif(\frac{1}{2}A, A)$	$\hat{\theta}_{it}^{BLUP}$	0.0643	0.0830	0.0746
	$\theta_{it}$	0.1241	0.1485	0.1401
$\psi_{it} \sim Unif(\frac{3}{2}A, A)$	$\hat{\theta}_{it}^{BLUP}$	0.2461	0.0539	0.0501
	$\theta_{it}$	0.3332	0.1490	0.1452

ตารางที่ 3.4: ค่าเฉลี่ยของกำลังสองของความคลาดเคลื่อน (Mean squared error) ของ  $\hat{\theta}_{it}^{EBLUP}$  กรณี  $A \sim Unif(1, 2)$  ของ REML และ AREML

จากตารางที่ 3.4 จะเห็นว่า ค่าเฉลี่ยของกำลังสองของความคลาดเคลื่อนของ  $\hat{\theta}_{it}^{EBLUP}$  เทียบกับ  $\hat{\theta}_{it}^{BLUP}$  และเทียบกับ  $\theta_{it}$  ของกรณีที่ปรับแล้ว มีค่าน้อยกว่ากรณีที่ยังไม่ปรับ

พิจารณา กรณีความแปรปรวนจากการสุ่มตัวอย่าง มีค่าน้อยกว่า ความแปรปรวนจากการถดถอย จะเห็นว่า เมื่อใช้ค่า  $\hat{A}$  จริง ซึ่งอาจเป็นลบได้ จะให้ค่าเฉลี่ยของกำลังสองของความคลาดเคลื่อน ที่น้อยกว่า เมื่อใช้ค่า  $\hat{A}_0$  ที่ใช้ค่า  $\hat{A}$  ที่มากกว่าศูนย์ เนื่องจากจำนวน  $\hat{A} \leq 0$  มีจำนวนมาก จึงส่งผลให้  $\frac{\hat{A}}{\hat{A} + \psi_{it}}$  มีค่าใกล้ 1



กรณีจำลอง  $A$  จากการแจกแจงยูนิฟอร์มในช่วง  $(4,6)$  นั่นคือ  $A \sim Unif(4,6)$

ความแปรปรวนจากการสุ่มตัวอย่าง	PML	APML	REML	AREML
$\psi_{it} \sim Unif(\frac{1}{2}A, A)$	1000	0	1000	0
$\psi_{it} \sim Unif(\frac{3}{2}A, 2A)$	1000	0	1000	0

ตารางที่ 3.5: จำนวนตัวประมาณความแปรปรวนจากการถดถอย  $\hat{A} \leq 0$  จากการจำลอง 1000 ครั้ง

จากตารางที่ 3.5 จะเห็นว่า จำนวนตัวประมาณความแปรปรวนจากการถดถอย  $\hat{A} \leq 0$  ของกรณีที่ยังไม่ได้ปรับทั้ง PML และ REML มีจำนวนมาก ในขณะที่ กรณีที่ปรับแล้วทั้ง APML และ AREML ไม่พบตัวประมาณความแปรปรวนจากการถดถอย  $\hat{A} \leq 0$

ความแปรปรวนจากการสุ่มตัวอย่าง	PML	APML	REML	AREML
$\psi_{it} \sim Unif(\frac{1}{2}A, A)$	827.2519	0.1386	636.0507	0.9294
$\psi_{it} \sim Unif(\frac{3}{2}A, 2A)$	638.0222	0.3729	485.9815	0.9119

ตารางที่ 3.6: ค่าเฉลี่ยของกำลังสองของความคลาดเคลื่อน (Mean squared error) ของตัวประมาณความแปรปรวนจากการถดถอย  $\hat{A}$

จากตารางที่ 3.6 จะเห็นว่า ค่าเฉลี่ยของกำลังสองของความคลาดเคลื่อนของตัวประมาณความแปรปรวนจากการถดถอย  $\hat{A}$  ของกรณีที่ปรับแล้วทั้ง APML และ AREML มีค่าน้อยกว่า ค่าเฉลี่ยของกำลังสองของความคลาดเคลื่อนของกรณีที่ยังไม่ได้ปรับ ทั้ง PML และ REML

ความแปรปรวนจากการ สุ่มตัวอย่าง	ค่าจริง	PML		APML
		$\hat{A}$	$\hat{A}_0$	
$\psi_{it} \sim Unif(\frac{1}{2}A, A)$	$\hat{\theta}_{it}^{BLUP}$	0.3379	0.2868	0.0794
	$\theta_{it}$	0.5434	0.5011	0.2934
$\psi_{it} \sim Unif(\frac{3}{2}A, A)$	$\hat{\theta}_{it}^{BLUP}$	2.5308	0.1821	0.0597
	$\theta_{it}$	2.8510	0.4978	0.3776

ตารางที่ 3.7: ค่าเฉลี่ยของกำลังสองของความคลาดเคลื่อน (Mean squared error) ของ  $\hat{\theta}_{it}^{EBLUP}$  กรณี  $A \sim Unif(4, 6)$  ของ PML และ APML

จากตารางที่ 3.7 จะเห็นว่า ค่าเฉลี่ยของกำลังสองของความคลาดเคลื่อนของ  $\hat{\theta}_{it}^{EBLUP}$  เทียบกับ  $\hat{\theta}_{it}^{BLUP}$  และเทียบกับ  $\theta_{it}$  ของกรณีที่ปรับแล้ว มีค่าน้อยกว่ากรณีที่ยังไม่ปรับ

ความแปรปรวนจากการ สุ่มตัวอย่าง	ค่าจริง	REML		AREML
		$\hat{A}$	$\hat{A}_0$	
$\psi_{it} \sim Unif(\frac{1}{2}A, A)$	$\hat{\theta}_{it}^{BLUP}$	0.3636	0.2854	0.2756
	$\theta_{it}$	0.5654	0.4984	0.4886
$\psi_{it} \sim Unif(\frac{3}{2}A, A)$	$\hat{\theta}_{it}^{BLUP}$	6.3536	0.1840	0.1787
	$\theta_{it}$	6.6525	0.5047	0.4995

ตารางที่ 3.8: ค่าเฉลี่ยของกำลังสองของความคลาดเคลื่อน (Mean squared error) ของ  $\hat{\theta}_{it}^{EBLUP}$  กรณี  $A \sim Unif(4, 6)$  ของ REML และ AREML

จากตารางที่ 3.8 จะเห็นว่า ค่าเฉลี่ยของกำลังสองของความคลาดเคลื่อนของ  $\hat{\theta}_{it}^{EBLUP}$  เทียบกับ  $\hat{\theta}_{it}^{BLUP}$  และเทียบกับ  $\theta_{it}$  ของกรณีที่ปรับแล้ว มีค่าน้อยกว่ากรณีที่ยังไม่ปรับ

กรณีจำลอง  $A$  จากการแจกแจงยูนิฟอร์มในช่วง  $(9,11)$  นั่นคือ  $A \sim Unif(9, 11)$

$\psi_{it}$	PML	APML	REML	AREML
$\psi_{it} \sim Unif(\frac{1}{2}A, A)$	1000	0	1000	0
$\psi_{it} \sim Unif(\frac{3}{2}A, 2A)$	1000	0	989	0

ตารางที่ 3.9: จำนวนตัวประมาณความแปรปรวนจากการถดถอย  $\hat{A} \leq 0$  จากการจำลอง 1000 ครั้ง

จากตารางที่ 3.9 จะเห็นว่า จำนวนตัวประมาณความแปรปรวนจากการถดถอย  $\hat{A} \leq 0$  ของกรณีที่ยังไม่ได้ปรับทั้ง PML และ REML มีจำนวนมาก ในขณะที่ กรณีที่ปรับแล้วทั้ง APML และ AREML ไม่พบตัวประมาณความแปรปรวนจากการถดถอย  $\hat{A} \leq 0$

ความแปรปรวนจากการสุ่มตัวอย่าง	PML	APML	REML	AREML
$\psi_{it} \sim Unif(\frac{1}{2}A, A)$	649.1229	0.2888	497.6969	0.9172
$\psi_{it} \sim Unif(\frac{3}{2}A, 2A)$	326.7126	1.8035	167.0825	0.8713

ตารางที่ 3.10: ค่าเฉลี่ยของกำลังสองของความคลาดเคลื่อน (Mean squared error) ของตัวประมาณความแปรปรวนจากการถดถอย  $\hat{A}$

จากตารางที่ 3.10 จะเห็นว่า ค่าเฉลี่ยของกำลังสองของความคลาดเคลื่อนของตัวประมาณความแปรปรวนจากการถดถอย  $\hat{A}$  ของกรณีที่ปรับแล้วทั้ง APML และ AREML มีค่าน้อยกว่า ค่าเฉลี่ยของกำลังสองของความคลาดเคลื่อนของกรณีที่ยังไม่ได้ปรับ ทั้ง PML และ REML

ความแปรปรวนจากการ สุ่มตัวอย่าง	ค่าจริง	PML		APML
		$\hat{A}$	$\hat{A}_0$	
$\psi_{it} \sim Unif(\frac{1}{2}A, A)$	$\hat{\theta}_{it}^{BLUP}$	1.8281	0.5752	0.2795
	$\theta_{it}$	2.2469	0.9988	0.7049
$\psi_{it} \sim Unif(\frac{3}{2}A, A)$	$\hat{\theta}_{it}^{BLUP}$	$2 \times 10^6$	0.3609	0.1688
	$\theta_{it}$	$2 \times 10^6$	0.9969	0.8091

ตารางที่ 3.11: ค่าเฉลี่ยของกำลังสองของความคลาดเคลื่อน (Mean squared error) ของ  $\hat{\theta}_{it}^{EBLUP}$  กรณี  $A \sim Unif(9, 11)$  ของ PML และ APML

จากตารางที่ 3.11 จะเห็นว่า ค่าเฉลี่ยของกำลังสองของความคลาดเคลื่อนของ  $\hat{\theta}_{it}^{EBLUP}$  เทียบกับ  $\hat{\theta}_{it}^{BLUP}$  และเทียบกับ  $\theta_{it}$  ของกรณีที่ปรับแล้ว มีค่าน้อยกว่ากรณีที่ยังไม่ปรับ

ความแปรปรวนจากการ สุ่มตัวอย่าง	ค่าจริง	REML		AREML
		$\hat{A}$	$\hat{A}_0$	
$\psi_{it} \sim Unif(\frac{1}{2}A, A)$	$\hat{\theta}_{it}^{BLUP}$	2.3574	0.5687	0.5572
	$\theta_{it}$	2.7796	1.0025	0.9910
$\psi_{it} \sim Unif(\frac{3}{2}A, A)$	$\hat{\theta}_{it}^{BLUP}$	$1 \times 10^5$	0.3638	0.3524
	$\theta_{it}$	$1 \times 10^5$	0.9940	0.9834

ตารางที่ 3.12: ค่าเฉลี่ยของกำลังสองของความคลาดเคลื่อน (Mean squared error) ของ  $\hat{\theta}_{it}^{EBLUP}$  กรณี  $A \sim Unif(9, 11)$  ของ REML และ AREML

จากตารางที่ 3.12 จะเห็นว่า ค่าเฉลี่ยของกำลังสองของความคลาดเคลื่อนของ  $\hat{\theta}_{it}^{EBLUP}$  เทียบกับ  $\hat{\theta}_{it}^{BLUP}$  และเทียบกับ  $\theta_{it}$  ของกรณีที่ปรับแล้ว มีค่าน้อยกว่ากรณีที่ยังไม่ปรับ

### ผลการวิเคราะห์การจำลองข้อมูล กรณีสัมประสิทธิ์การถดถอยขึ้นกับเวลา

ผลการวิเคราะห์การจำลองข้อมูลแสดงดังตารางที่ 3.13 ถึง ตารางที่ 3.21 โดยตารางที่ 3.13 ตารางที่ 3.16 และตารางที่ 3.19 แสดงจำนวนตัวประมาณความแปรปรวนจากการถดถอย  $\hat{A}_t$  มีค่าน้อยกว่าหรือเท่ากับ 0 จากแบบที่ไม่ปรับ และแบบที่ปรับแล้ว จากการจำลอง  $A$  จากการแจกแจงยูนิฟอร์มในช่วง (1,2), (4,6) และ (9,11) ตามลำดับ

ตารางที่ 3.14 ตารางที่ 3.17 และตารางที่ 3.20 แสดงค่าเฉลี่ยของกำลังสองของความคลาดเคลื่อนของตัวประมาณความแปรปรวนจากการถดถอย ( $\hat{A}$ ) เมื่อเทียบกับความแปรปรวนจาก

การถดถอยที่เราจำลองไว้ตั้งแต่ต้น ( $A$ ) จากแบบที่ไม่ปรับ และแบบที่ปรับแล้ว จากการจำลอง  $A$  จากการแจกแจงยูนิฟอร์มในช่วง (1,2), (4,6) และ (9,11) ตามลำดับ

ตารางที่ 3.15 ตารางที่ 3.18 และตารางที่ 3.21 แสดงค่าเฉลี่ยของกำลังสองของความคลาดเคลื่อนของตัวประมาณที่ไม่เอนเอียงเชิงเส้นเชิงประจักษ์ที่ดีที่สุดที่คำนวณจากตัวประมาณความแปรปรวนจากการถดถอย  $\hat{A}$  ( $\hat{\theta}_{it}^{EBLUP}$ ) เทียบกับที่คำนวณจากการถดถอยที่เราจำลองไว้ตั้งแต่ต้น ( $\hat{\theta}_{it}^{BLUP}$ ) และเทียบกับค่าจริงที่เราจำลองได้ตั้งแต่ต้น ( $\theta_{it}$ ) กรณี PML และ APML จากการจำลอง  $A$  จากการแจกแจงยูนิฟอร์มในช่วง (1,2), (4,6) และ (9,11) ตามลำดับ โดยที่  $\hat{A}_0$  ที่แสดงในตารางคือ  $\hat{A}_0 = \max(0, \hat{A})$

กรณีจำลอง  $A$  จากการแจกแจงยูนิฟอร์มในช่วง (1,2) นั่นคือ  $A \sim Unif(1, 2)$

ความแปรปรวนจากการสุ่มตัวอย่าง	PML	APML
$\psi_{it} \sim Unif(\frac{1}{2}A, A)$	862	0
$\psi_{it} \sim Unif(\frac{3}{2}A, 2A)$	760	0

ตารางที่ 3.13: จำนวนตัวประมาณความแปรปรวนจากการถดถอย  $\hat{A} \leq 0$  จากการจำลอง 1000 ครั้ง

จากตารางที่ 3.13 จะเห็นว่า จำนวนตัวประมาณความแปรปรวนจากการถดถอย  $\hat{A} \leq 0$  ของกรณีที่ยังไม่ได้ปรับ PML มีจำนวนมาก ในขณะที่ กรณีที่ปรับแล้ว APML ไม่พบตัวประมาณความแปรปรวนจากการถดถอย  $\hat{A} \leq 0$

ความแปรปรวนจากการสุ่มตัวอย่าง	PML	APML
$\psi_{it} \sim Unif(\frac{1}{2}A, A)$	21.6269	100.3554
$\psi_{it} \sim Unif(\frac{3}{2}A, 2A)$	29.5707	92.0092

ตารางที่ 3.14: ค่าเฉลี่ยของกำลังสองของความคลาดเคลื่อน (Mean squared error) ของตัวประมาณความแปรปรวนจากการถดถอย  $\hat{A}$

จากตารางที่ 3.14 จะเห็นว่า ค่าเฉลี่ยของกำลังสองของความคลาดเคลื่อนของตัวประมาณ

ความแปรปรวนจากการถดถอย  $\hat{A}$  ของกรณีที่ปรับแล้ว APML มีค่ามากกว่า ค่าเฉลี่ยของกำลังสองของความคลาดเคลื่อนของกรณีที่ยังไม่ได้ปรับ PML

ความแปรปรวนจากการ สุ่มตัวอย่าง	ค่าจริง	PML		APML
		$\hat{A}$	$\hat{A}_0$	
$\psi_{it} \sim Unif(\frac{1}{2}A, A)$	$\hat{\theta}_{it}^{BLUP}$	$6 \times 10^{11}$	1.0166	0.3896
	$\theta_{it}$	$6 \times 10^{11}$	2.7955	0.1241
$\psi_{it} \sim Unif(\frac{3}{2}A, A)$	$\hat{\theta}_{it}^{BLUP}$	$6 \times 10^{11}$	0.4867	0.6516
	$\theta_{it}$	$6 \times 10^{11}$	2.4940	0.3070

ตารางที่ 3.15: ค่าเฉลี่ยของกำลังสองของความคลาดเคลื่อน (Mean squared error) ของ  $\hat{\theta}_{it}^{EBLUP}$  กรณี  $A \sim Unif(1, 2)$  ของ PML และ APML

จากตารางที่ 3.15 จะเห็นว่า ค่าเฉลี่ยของกำลังสองของความคลาดเคลื่อนของ  $\hat{\theta}_{it}^{EBLUP}$  เทียบกับ  $\hat{\theta}_{it}^{BLUP}$  และเทียบกับ  $\theta_{it}$  ของกรณีที่ปรับแล้ว APML ส่วนใหญ่ มีค่าน้อยกว่ากรณีที่ยังไม่ปรับ PML

กรณีจำลอง  $A$  จากการแจกแจงยูนิฟอร์มในช่วง (4,6) นั่นคือ  $A \sim Unif(4, 6)$

ความแปรปรวนจากการสุ่มตัวอย่าง	PML	APML
$\psi_{it} \sim Unif(\frac{1}{2}A, A)$	0	0
$\psi_{it} \sim Unif(\frac{3}{2}A, 2A)$	10	0

ตารางที่ 3.16: จำนวนตัวประมาณความแปรปรวนจากการถดถอย  $\hat{A} \leq 0$  จากการจำลอง 1000 ครั้ง

จากตารางที่ 3.16 จะเห็นว่ากรณีที่ความแปรปรวนจากการสุ่มตัวอย่างมีค่าน้อย จำนวนตัวประมาณความแปรปรวนจากการถดถอย  $\hat{A} \leq 0$  ของกรณีที่ยังไม่ได้ปรับ PML และกรณีที่ปรับแล้ว APML ไม่พบตัวประมาณความแปรปรวนจากการถดถอย  $\hat{A} \leq 0$  แต่เมื่อความแปรปรวนจากการสุ่มตัวอย่างมีค่ามากขึ้น จำนวนตัวประมาณความแปรปรวนจากการถดถอย  $\hat{A} \leq 0$  ของกรณีที่ยังไม่ได้ปรับ PML จะมีจำนวนมากขึ้น แต่กรณีที่ปรับแล้ว APML ไม่พบตัวประมาณความแปรปรวนจากการถดถอย  $\hat{A} \leq 0$

ความแปรปรวนจากการสุ่มตัวอย่าง	PML	APML
$\psi_{it} \sim Unif(\frac{1}{2}A, A)$	61.0491	74.3283
$\psi_{it} \sim Unif(\frac{3}{2}A, 2A)$	39.7897	54.7850

ตารางที่ 3.17: ค่าเฉลี่ยของกำลังสองของความคลาดเคลื่อน (Mean squared error) ของตัวประมาณความแปรปรวนจากการถดถอย  $\hat{A}$

จากตารางที่ 3.17 จะเห็นว่า ค่าเฉลี่ยของกำลังสองของความคลาดเคลื่อนของตัวประมาณความแปรปรวนจากการถดถอย  $\hat{A}$  ของกรณีที่ปรับแล้ว APML มีค่ามากกว่า ค่าเฉลี่ยของกำลังสองของความคลาดเคลื่อนของกรณีที่ยังไม่ได้ปรับ PML

ความแปรปรวนจากการสุ่มตัวอย่าง	ค่าจริง	PML		APML
		$\hat{A}$	$\hat{A}_0$	
$\psi_{it} \sim Unif(\frac{1}{2}A, A)$	$\hat{\theta}_{it}^{BLUP}$	0.1443	0.1443	0.1611
	$\theta_{it}$	0.4568	0.4568	0.4384
$\psi_{it} \sim Unif(\frac{3}{2}A, A)$	$\hat{\theta}_{it}^{BLUP}$	0.5694	0.1606	0.1922
	$\theta_{it}$	1.6288	1.1973	1.0679

ตารางที่ 3.18: ค่าเฉลี่ยของกำลังสองของความคลาดเคลื่อน (Mean squared error) ของ  $\hat{\theta}_{it}^{EBLUP}$  กรณี  $A \sim Unif(4, 6)$  ของ PML และ APML

จากตารางที่ 3.18 จะเห็นว่า ค่าเฉลี่ยของกำลังสองของความคลาดเคลื่อนของ  $\hat{\theta}_{it}^{EBLUP}$  เมื่อเทียบกับ  $\hat{\theta}_{it}^{BLUP}$  ของกรณีที่ปรับแล้ว APML มีค่ามากกว่ากรณีที่ยังไม่ปรับ PML และเมื่อเทียบกับ  $\theta_{it}$  ของกรณีที่ปรับแล้ว APML มีค่าน้อยกว่ากรณีที่ยังไม่ปรับ PML

กรณีจำลอง  $A$  จากการแจกแจงยูนิฟอร์มในช่วง  $(9,11)$  นั่นคือ  $A \sim Unif(9, 11)$

$\psi_{it}$	PML	APML
$\psi_{it} \sim Unif(\frac{1}{2}A, A)$	1	0
$\psi_{it} \sim Unif(\frac{3}{2}A, 2A)$	68	0

ตารางที่ 3.19: จำนวนตัวประมาณความแปรปรวนจากการถดถอย  $\hat{A} \leq 0$  จากการจำลอง 1000 ครั้ง

จากตารางที่ 3.19 จะเห็นว่าเมื่อความแปรปรวนจากการสุ่มตัวอย่างมีค่ามากขึ้น จำนวนตัวประมาณความแปรปรวนจากการถดถอย  $\hat{A} \leq 0$  ของกรณีที่ยังไม่ได้ปรับ PML จะมีจำนวนมากขึ้น แต่กรณีที่ปรับแล้ว APML ไม่พบตัวประมาณความแปรปรวนจากการถดถอย  $\hat{A} \leq 0$

ความแปรปรวนจากการสุ่มตัวอย่าง	PML	APML
$\psi_{it} \sim Unif(\frac{1}{2}A, A)$	31.0112	41.1391
$\psi_{it} \sim Unif(\frac{3}{2}A, 2A)$	47.7899	43.3065

ตารางที่ 3.20: ค่าเฉลี่ยของกำลังสองของความคลาดเคลื่อน (Mean squared error) ของตัวประมาณความแปรปรวนจากการถดถอย  $\hat{A}$

จากตารางที่ 3.20 จะเห็นว่า เมื่อความแปรปรวนจากการสุ่มตัวอย่างมีค่ามาก ค่าเฉลี่ยของกำลังสองของความคลาดเคลื่อนของตัวประมาณความแปรปรวนจากการถดถอย  $\hat{A}$  ของกรณีที่ปรับแล้ว APML มีค่าน้อยกว่า ค่าเฉลี่ยของกำลังสองของความคลาดเคลื่อนของกรณีที่ยังไม่ได้ปรับ PML



ความแปรปรวนจากการ สุ่มตัวอย่าง	ค่าจริง	PML		APML
		$\hat{A}$	$\hat{A}_0$	
$\psi_{it} \sim Unif(\frac{1}{2}A, A)$	$\hat{\theta}_{it}^{BLUP}$	0.0570	0.0568	0.0599
	$\theta_{it}$	0.9205	0.9201	0.8584
$\psi_{it} \sim Unif(\frac{3}{2}A, A)$	$\hat{\theta}_{it}^{BLUP}$	$2 \times 10^5$	0.1555	0.0919
	$\theta_{it}$	$2 \times 10^5$	2.1670	1.8054

ตารางที่ 3.21: ค่าเฉลี่ยของกำลังสองของความคลาดเคลื่อน (Mean squared error) ของ  $\hat{\theta}_{it}^{EBLUP}$  กรณี  $A \sim Unif(9, 11)$  ของ PML และ APML

จากตารางที่ 3.21 จะเห็นว่า ค่าเฉลี่ยของกำลังสองของความคลาดเคลื่อนของ  $\hat{\theta}_{it}^{EBLUP}$  เทียบกับ  $\hat{\theta}_{it}^{BLUP}$  และเทียบกับ  $\theta_{it}$  ของกรณีที่ปรับแล้ว APML ส่วนใหญ่มีค่าน้อยกว่ากรณีที่ ยังไม่ปรับ PML

### 3.3 กรณีความแปรปรวนจากการถดถอยขึ้นกับเวลา

ในหัวข้อนี้เราจะพูดถึงการจำลองข้อมูล ในกรณีที่ความแปรปรวนจากการถดถอยขึ้นกับเวลา ซึ่งจะพิจารณาใน 2 กรณี คือ

1. กรณีปัจจัยเกิดจากพื้นที่
2. กรณีปัจจัยเกิดจากเวลา

#### 3.3.1 กรณีปัจจัยเกิดจากพื้นที่

ในหัวข้อนี้เราจะพูดถึงการจำลองข้อมูล ในกรณีที่ความแปรปรวนจากการถดถอยขึ้นกับเวลา เมื่อปัจจัยเกิดจากพื้นที่ จะทำให้ได้แบบจำลองดังนี้

1. แบบจำลองเชิงมโยง:  $\theta_{it} = x'_{it}\beta + v_i + \epsilon_{it}$   
โดยที่  $\epsilon_{it} \sim N(0, A_t)$  เมื่อ  $A_t$  คือความแปรปรวนที่เกิดจากการถดถอยที่ขึ้นกับเวลา  $t$
2. แบบจำลองการสุ่มตัวอย่าง:  $y_{it} = \theta_{it} + e_{it}$   
โดยที่  $e_{it} \sim N(0, \psi_{it})$  เมื่อ  $\psi_{it}$  คือความแปรปรวนที่เกิดจากการสุ่มตัวอย่าง

โดยแบบจำลองเชื่อมโยงจะเขียนกระจายได้ดังสมการ (3.1) และทำการแปลงเมทริกซ์เพื่อจัดให้อยู่ในรูปของการถดถอยเชิงเส้นได้ดังสมการ (3.2) และแบบจำลองการสุ่มตัวอย่างจะได้ดังสมการ (3.4) ทำให้ได้แบบจำลองเฟ-เฮรอตแบบพานาแลกรณีสัมประสิทธิ์ไม่ขึ้นกับเวลาดังสมการ (3.6) จะได้ว่า  $\mathbf{y} \sim N(\mathbf{Z}\boldsymbol{\gamma}, \boldsymbol{\Sigma})$  เมื่อ  $\boldsymbol{\Sigma} = \mathbf{A} + \boldsymbol{\psi}$  โดยที่  $\mathbf{A}$  เป็นบล็อกเมทริกซ์ของ  $\text{diag}(A_1, A_2, \dots, A_T)$  จำนวน  $n$  บล็อก และ  $\boldsymbol{\psi} = \text{diag}(\psi_{11}, \psi_{12}, \dots, \psi_{nT})$  นั่นคือ

$$\boldsymbol{\Sigma} = \begin{bmatrix} A_1 & 0 & 0 & \dots & 0 & 0 & 0 \\ 0 & \ddots & 0 & \dots & 0 & 0 & 0 \\ 0 & 0 & A_T & \dots & 0 & 0 & 0 \\ \vdots & \vdots & \vdots & \ddots & \vdots & \vdots & \vdots \\ 0 & 0 & 0 & 0 & A_1 & 0 & 0 \\ 0 & 0 & 0 & 0 & 0 & \ddots & 0 \\ 0 & 0 & 0 & 0 & 0 & 0 & A_T \end{bmatrix} + \begin{bmatrix} \psi_{11} & 0 & 0 & \dots & 0 & 0 & 0 \\ 0 & \ddots & 0 & \dots & 0 & 0 & 0 \\ 0 & 0 & \psi_{1T} & \dots & 0 & 0 & 0 \\ \vdots & \vdots & \vdots & \ddots & \vdots & \vdots & \vdots \\ 0 & 0 & 0 & 0 & \psi_{n1} & 0 & 0 \\ 0 & 0 & 0 & 0 & 0 & \ddots & 0 \\ 0 & 0 & 0 & 0 & 0 & 0 & \psi_{nT} \end{bmatrix}$$

$$= \begin{bmatrix} A_1 + \psi_{11} & 0 & 0 & \dots & 0 & 0 & 0 \\ 0 & \ddots & 0 & \dots & 0 & 0 & 0 \\ 0 & 0 & A_T + \psi_{1T} & \dots & 0 & 0 & 0 \\ \vdots & \vdots & \vdots & \ddots & \vdots & \vdots & \vdots \\ 0 & 0 & 0 & 0 & A_1 + \psi_{n1} & 0 & 0 \\ 0 & 0 & 0 & 0 & 0 & \ddots & 0 \\ 0 & 0 & 0 & 0 & 0 & 0 & A_T + \psi_{nT} \end{bmatrix} \quad (3.19)$$

### 3.3.2 กรณีปัจจัยเกิดจากเวลา

ในหัวข้อนี้เราจะพูดถึงการจำลองข้อมูล ในกรณีที่ความแปรปรวนจากการถดถอยขึ้นกับเวลา เมื่อปัจจัยเกิดจากเวลา จะทำให้ได้แบบจำลองดังนี้

1. แบบจำลองเชื่อมโยง:  $\theta_{it} = x'_{it}\beta + v_t + \epsilon_{it}$

โดยที่  $\epsilon_{it} \sim N(0, A_t)$  เมื่อ  $A_t$  คือความแปรปรวนที่เกิดจากการถดถอยที่ขึ้นกับเวลา  $t$

2. แบบจำลองการสุ่มตัวอย่าง:  $y_{it} = \theta_{it} + e_{it}$

โดยที่  $e_{it} \sim N(0, \psi_{it})$  เมื่อ  $\psi_{it}$  คือความแปรปรวนที่เกิดจากการสุ่มตัวอย่าง

จากแบบจำลองเชื่อมโยงจะทำให้เขียนเป็นเมทริกซ์ได้ดังนี้  
 สำหรับ  $i = 1, 2, \dots, n$  และ  $t = 1, 2, \dots, T$

$$\begin{bmatrix} \theta_{11} \\ \theta_{12} \\ \vdots \\ \theta_{1T} \\ \theta_{21} \\ \theta_{22} \\ \vdots \\ \theta_{2T} \\ \vdots \\ \theta_{n1} \\ \theta_{n2} \\ \vdots \\ \theta_{nT} \end{bmatrix} = \begin{bmatrix} x_{111} & x_{112} & \dots & x_{11p} \\ x_{121} & x_{122} & \dots & x_{12p} \\ \vdots & \vdots & \vdots & \vdots \\ x_{1T1} & x_{1T2} & \dots & x_{1Tp} \\ x_{211} & x_{212} & \dots & x_{21p} \\ x_{221} & x_{222} & \dots & x_{22p} \\ \vdots & \vdots & \vdots & \vdots \\ x_{2T1} & x_{2T2} & \dots & x_{2Tp} \\ \vdots & \vdots & \vdots & \vdots \\ x_{n11} & x_{n12} & \dots & x_{n1p} \\ x_{n21} & x_{n22} & \dots & x_{n2p} \\ \vdots & \vdots & \vdots & \vdots \\ x_{nT1} & x_{nT2} & \dots & x_{nTp} \end{bmatrix} \begin{bmatrix} \beta_1 \\ \beta_2 \\ \vdots \\ \beta_p \end{bmatrix} + \begin{bmatrix} v_1 \\ v_2 \\ \vdots \\ v_T \\ v_1 \\ v_2 \\ \vdots \\ v_T \\ \vdots \\ v_1 \\ v_2 \\ \vdots \\ v_T \end{bmatrix} + \begin{bmatrix} \epsilon_{11} \\ \epsilon_{12} \\ \vdots \\ \epsilon_{1T} \\ \epsilon_{21} \\ \epsilon_{22} \\ \vdots \\ \epsilon_{2T} \\ \vdots \\ \epsilon_{n1} \\ \epsilon_{n2} \\ \vdots \\ \epsilon_{nT} \end{bmatrix}$$

และเราจะทำการแปลงเมทริกซ์เพื่อจัดให้อยู่ในรูปของการถดถอยเชิงเส้น เป็นดังนี้

$$\begin{bmatrix} \theta_{11} \\ \theta_{12} \\ \vdots \\ \theta_{1T} \\ \theta_{21} \\ \theta_{22} \\ \vdots \\ \theta_{2T} \\ \vdots \\ \theta_{n1} \\ \theta_{n2} \\ \vdots \\ \theta_{nT} \end{bmatrix} = \begin{bmatrix} x_{111} & x_{112} & \dots & x_{11p} & 1 & 0 & \dots & 0 \\ x_{121} & x_{122} & \dots & x_{12p} & 0 & 1 & \dots & 0 \\ \vdots & \vdots & \vdots & \vdots & \vdots & \vdots & \ddots & \vdots \\ x_{1T1} & x_{1T2} & \dots & x_{1Tp} & 0 & 0 & \dots & 1 \\ x_{211} & x_{212} & \dots & x_{21p} & 1 & 0 & \dots & 0 \\ x_{221} & x_{222} & \dots & x_{22p} & 0 & 1 & \dots & 0 \\ \vdots & \vdots & \vdots & \vdots & \vdots & \vdots & \ddots & \vdots \\ x_{2T1} & x_{2T2} & \dots & x_{2Tp} & 0 & 0 & \dots & 1 \\ \vdots & \vdots & \vdots & \vdots & \vdots & \vdots & \vdots & \vdots \\ x_{n11} & x_{n12} & \dots & x_{n1p} & 1 & 0 & \dots & 0 \\ x_{n21} & x_{n22} & \dots & x_{n2p} & 0 & 1 & \dots & 0 \\ \vdots & \vdots & \vdots & \vdots & \vdots & \vdots & \ddots & \vdots \\ x_{nT1} & x_{nT2} & \dots & x_{nTp} & 0 & 0 & \dots & 1 \end{bmatrix} \begin{bmatrix} \beta_1 \\ \beta_2 \\ \vdots \\ \beta_p \\ v_1 \\ v_2 \\ \vdots \\ v_T \end{bmatrix} + \begin{bmatrix} \epsilon_{11} \\ \epsilon_{12} \\ \vdots \\ \epsilon_{1T} \\ \epsilon_{21} \\ \epsilon_{22} \\ \vdots \\ \epsilon_{2T} \\ \vdots \\ \epsilon_{n1} \\ \epsilon_{n2} \\ \vdots \\ \epsilon_{nT} \end{bmatrix}$$

ซึ่งสามารถเขียนในรูปเมทริกซ์ได้เป็น

$$\boldsymbol{\theta} = \mathbf{Z}\boldsymbol{\gamma} + \boldsymbol{\epsilon} \quad (3.20)$$

โดยที่  $\mathbf{Z} = [\mathbf{X}, \mathbf{B}]$  เมื่อ

$$\mathbf{B} = \begin{bmatrix} 1 & 0 & \dots & 0 \\ 0 & 1 & \dots & 0 \\ \vdots & \vdots & \ddots & \vdots \\ 0 & 0 & \dots & 1 \\ 1 & 0 & \dots & 0 \\ 0 & 1 & \dots & 0 \\ \vdots & \vdots & \ddots & \vdots \\ 0 & 0 & \dots & 1 \\ 1 & 0 & \dots & 0 \\ 0 & 1 & \dots & 0 \\ \vdots & \vdots & \ddots & \vdots \\ 0 & 0 & \dots & 1 \end{bmatrix}$$

เป็นเมทริกซ์ขนาด  $nT \times T$  และ  $\boldsymbol{\gamma} = (\beta_1, \beta_2, \dots, \beta_p, v_1, v_2, \dots, v_T)'$  เป็นเมทริกซ์ขนาด  $(p + T) \times 1$

เมื่อพิจารณาแบบจำลองการสุ่มตัวอย่าง จะสามารถเขียนในรูปเมทริกซ์ได้ดังสมการ (3.5) ดังนั้น แบบจำลองเฟ-เฮรอตแบบพานอลสามารถเขียนผ่าน (3.20) และ (3.5) ได้ดังนี้

$$\mathbf{y} = \mathbf{Z}\boldsymbol{\gamma} + \boldsymbol{\epsilon} + \mathbf{e} \quad (3.21)$$

จาก (3.21) จะได้ว่า  $\mathbf{y} \sim N(\mathbf{Z}\boldsymbol{\gamma}, \boldsymbol{\Sigma})$  เมื่อ  $\boldsymbol{\Sigma} = \mathbf{A} + \boldsymbol{\psi}$  ซึ่ง  $\boldsymbol{\Sigma}$  เป็นดังสมการ (3.19)

ในทางปฏิบัติ เราไม่รู้ค่า  $A_t$  แต่เราสามารถประมาณค่า  $A_t$  ได้ โดยค่าความแปรปรวน  $A_t$  จะถูกประมาณโดยตัวประมาณแบบภาวะความน่าจะเป็นสูงสุดโปรไฟล์ (Profile maximum likelihood estimator: PML) ตัวประมาณแบบภาวะความน่าจะเป็นสูงสุดโปรไฟล์แบบปรับ (Adjusted profile maximum likelihood estimator: APML) ตัวประมาณแบบภาวะความน่าจะเป็นสูงสุดเรสซิดัล (Residual maximum likelihood estimator: REML) และตัวประมาณแบบภาวะความน่าจะเป็นสูงสุดเรสซิดัลแบบปรับ (Adjusted residual maximum likelihood estimator: AREML) โดยการใช้คำสั่ง `optim` ในโปรแกรม R

ซึ่งการหาฟังก์ชันลึอกภาวะความน่าจะเป็นโปรไฟล์และฟังก์ชันลึอกภาวะความน่าจะเป็นเรสซิดัล จะเป็นดังสมการ (3.10) และ (3.11)

พิจารณากรณีความแปรปรวนจากการถดถอยขึ้นกับเวลา เมื่อปัจจัยเกิดจากพื้นที่ และปัจจัยเกิดจากเวลา สำหรับตัวประมาณพารามิเตอร์ที่ใช้คือ ตัวประมาณที่ไม่เอนเอียงเชิงเส้นที่ดีที่สุด (best linear unbiased predictor: BLUP)

$$\theta_{it}^{BLUP} = \frac{A_t}{A_t + \psi_{it}} y_{it} + \frac{\psi_{it}}{A_t + \psi_{it}} z'_{it} \gamma$$

เมื่อ  $A_t$  คือ ค่าความแปรปรวนที่ขึ้นกับเวลา  $t$

หลังจากได้ตัวประมาณภาวะความน่าจะเป็นสูงสุด  $\hat{A}_t$  แล้ว จะได้ว่าตัวประมาณที่ไม่เอนเอียงเชิงเส้นเชิงประจักษ์ที่ดีที่สุด ( $\hat{\theta}_{it}^{EBLUP}$ ) ดังนี้

$$\hat{\theta}_{it}^{EBLUP} = \frac{\hat{A}_t}{\hat{A}_t + \psi_{it}} y_{it} + \frac{\psi_{it}}{\hat{A}_t + \psi_{it}} z'_{it} \gamma$$

จะเห็นว่ากรณีที่  $\hat{A}_t < 0$  จะขัดแย้งกับความหมายของ  $A_t$  ซึ่งเป็นความแปรปรวน ซึ่งจะทำให้ตัวประมาณค่า  $\hat{\theta}_{it}^{EBLUP}$  คลาดเคลื่อนจากค่าจริง และกรณีที่  $\hat{A}_t = 0$  จะทำให้  $\hat{\theta}_{it}^{EBLUP} = z'_{it} \gamma$  ซึ่งเป็นปัญหาเดียวกันกับที่เกิดขึ้นจากงานของ Li และ Lahiri เมื่อ ปีค.ศ. 2010 ดังนั้น ต่อไปเราจะขยายหลักการปรับตัวประมาณ  $\hat{A}_t$  มาใช้กับแบบจำลองพานเนลเพื่อที่จะทำให้  $\hat{A}_t > 0$  ดังนี้

$$L_{adj}(\mathbf{A}|\mathbf{y}) = \prod_{t=1}^T A_t \times L(\mathbf{A}|\mathbf{y})$$

เมื่อ  $L(\mathbf{A}|\mathbf{y})$  คือฟังก์ชันภาวะความน่าจะเป็นแบบโปรไฟล์และแบบเรสซิดัล ตามลำดับ

ต่อไปจะเป็นการแสดงว่า หลักการปรับตัวประมาณ  $\hat{A}_t$  จะให้ค่า  $\hat{A}_t > 0$

พิจารณาตัวประมาณแบบภาวะความน่าจะเป็นสูงสุดโปรไฟล์แบบปรับ (APML estimator) จะได้ฟังก์ชันภาวะความน่าจะเป็นโปรไฟล์แบบปรับ ดังนี้

$$\begin{aligned} L_{APML}(\mathbf{A}|\mathbf{y}) &= \prod_{t=1}^T A_t \times L_{PML}(\mathbf{A}|\mathbf{y}) \\ &= \frac{\prod_{t=1}^T A_t}{[(2\pi)^{nT} |\boldsymbol{\Sigma}|]^{1/2}} \exp\left\{-\frac{1}{2} \mathbf{y}' \mathbf{P} \mathbf{y}\right\} \end{aligned}$$

เมื่อ  $\mathbf{P} = \boldsymbol{\Sigma}^{-1} - \boldsymbol{\Sigma}^{-1} \mathbf{Z} (\mathbf{Z}' \boldsymbol{\Sigma}^{-1} \mathbf{Z})^{-1} \mathbf{Z}' \boldsymbol{\Sigma}^{-1}$

เห็นได้ชัดว่าฟังก์ชันภาวะความน่าจะเป็นโปรไฟล์  $L_{PML}(\mathbf{A}|\mathbf{y})$  เป็นฟังก์ชันของ  $\mathbf{A}$  ซึ่งเป็นฟังก์ชันต่อเนื่องและให้ค่าบวก จะได้ว่า สำหรับ  $t = 1, \dots, T$

$$\begin{aligned} & \lim_{A_t \rightarrow \infty} \frac{\prod_{t=1}^T A_t}{[(2\pi)^{nT} |\boldsymbol{\Sigma}|]^{1/2}} \exp\{-\frac{1}{2} \mathbf{y}' \mathbf{P} \mathbf{y}\} \\ &= \frac{1}{(2\pi)^{nT}} \lim_{A_t \rightarrow \infty} \frac{\prod_{t=1}^T A_t}{\prod_{d=1}^D |\boldsymbol{\Sigma}_d|^{1/2}} \exp\{-\frac{1}{2} \mathbf{y}' \mathbf{P} \mathbf{y}\} \\ &= \frac{1}{(2\pi)^{nT}} \lim_{A_t \rightarrow \infty} \frac{\prod_{t=1}^T A_t}{\prod_{d=1}^D |\mathbf{A} + \boldsymbol{\psi}_d|^{1/2}} \exp\{-\frac{1}{2} \mathbf{y}' \mathbf{P} \mathbf{y}\} \\ &= \frac{1}{(2\pi)^{nT}} \left( \lim_{A_t \rightarrow \infty} \frac{\prod_{t=1}^T A_t}{\prod_{d=1}^D |\mathbf{A} + \boldsymbol{\psi}_d|^{1/2}} \right) \left( \lim_{A_t \rightarrow \infty} \exp\{-\frac{1}{2} \mathbf{y}' \mathbf{P} \mathbf{y}\} \right) \\ &= 0 \end{aligned}$$

เมื่อ  $\lim_{A_t \rightarrow \infty} \exp\{-\frac{1}{2} \mathbf{y}' \mathbf{P} \mathbf{y}\} = 1$

โดยบทตั้ง 3.3.1 จะได้ว่ามี  $\mathbf{A} = (A_{01}, \dots, A_{0T})'$  ซึ่ง  $L_{APML}(\mathbf{A}|\mathbf{y})$  ให้ค่าสูงสุด โดยที่  $A_{0t} > 0$  สำหรับ  $t = 1, \dots, T$  และในทำนองเดียวกัน ตัวประมาณแบบภาวะความน่าจะเป็นสูงสุดเรสซิดัลแบบปรับ (AREML estimator) มีค่ามากกว่าศูนย์

จากหลักการปรับ จะทำให้ได้ล็อกฟังก์ชันภาวะความน่าจะเป็น (log-likelihood function) ดังนี้

1. ตัวประมาณแบบภาวะความน่าจะเป็นสูงสุดโปรไฟล์แบบปรับ (Adjusted profile maximum likelihood estimator: APML)

$$l_{APML}(\mathbf{A}|\mathbf{y}) = \sum_{t=1}^T \log(A_t) - \frac{1}{2} (nT \log(2\pi) + \log |\boldsymbol{\Sigma}| + \mathbf{y}' \mathbf{P} \mathbf{y})$$

เมื่อ  $\mathbf{P} = \boldsymbol{\Sigma}^{-1} - \boldsymbol{\Sigma}^{-1} \mathbf{Z} (\mathbf{Z}' \boldsymbol{\Sigma}^{-1} \mathbf{Z})^{-1} \mathbf{Z}' \boldsymbol{\Sigma}^{-1}$

2. ตัวประมาณแบบภาวะความน่าจะเป็นสูงสุดเรสซิดัลแบบปรับ (Adjusted residual maximum likelihood estimator: AREML)

$$l_{AREML}(\mathbf{A}|\mathbf{y}) = \sum_{t=1}^T \log(A_t) - \frac{1}{2} (nT \log(2\pi) + \log |\boldsymbol{\Sigma}| + \log(|\mathbf{Z}' \boldsymbol{\Sigma}^{-1} \mathbf{Z}|) + \mathbf{y}' \mathbf{P} \mathbf{y})$$

เมื่อ  $\mathbf{P} = \boldsymbol{\Sigma}^{-1} - \boldsymbol{\Sigma}^{-1} \mathbf{Z} (\mathbf{Z}' \boldsymbol{\Sigma}^{-1} \mathbf{Z})^{-1} \mathbf{Z}' \boldsymbol{\Sigma}^{-1}$

### 3.3.3 การศึกษาเชิงตัวเลข

ในหัวข้อนี้ เราจะทำการศึกษาประสิทธิภาพของตัวประมาณค่าแบบภาวะความน่าจะเป็นสูงสุดแบบปรับ ทั้งในแบบภาวะความน่าจะเป็นสูงสุดโพรไฟล์ และแบบภาวะความน่าจะเป็นสูงสุดเรสซิดัล โดยพิจารณาแบบจำลองเฟ-เฮรอตแบบพานเนล สำหรับพื้นที่ย่อย 10 พื้นที่ และเวลา 5 ปี ดังนี้

กรณีปัจจัยเกิดจากพื้นที่

สำหรับ  $i = 1, 2, \dots, 10$  และ  $t = 1, 2, \dots, 5$

พิจารณา แบบจำลองเชื่อมโยง

$$\theta_{it} = x'_{it}\beta + v_i + \epsilon_{it} \quad (3.22)$$

เมื่อ  $\text{var}(\epsilon_{it}) = A_t$

และแบบจำลองการสุ่มตัวอย่าง

$$y_{it} = \theta_{it} + e_{it} \quad (3.23)$$

เมื่อ  $\text{var}(e_{it}) = \psi_{it}$

โดยมีการตั้งค่าการจำลองข้อมูลดังนี้

1. จำนวนครั้งที่ทำการจำลอง:  $K = 1000$  ครั้ง
2. จำนวนตัวอย่าง: 10
3. สัมประสิทธิ์การถดถอย  $\beta_1 = \beta_2 = 1$
4. จำลอง  $\theta_{it} (i = 1, 2, \dots, n)$  และ  $(t = 1, 2, \dots, T)$  จากสมการ (3.22) โดยจำลอง  $x_{it1}, x_{it2}$  จากการแจกแจงปกติ ที่มีค่าเฉลี่ย 10 ความแปรปรวน 1 และ  $1.05^2$  นั่นคือ  $x_{it1} \sim N(10, 1)$  และ  $x_{it2} \sim N(10, 1.05^2)$  ตามลำดับ
5. จำลอง  $v_i$  จากการแจกแจงปกติที่มีค่าเฉลี่ย 10 ความแปรปรวน  $1.65^2$  นั่นคือ  $v_i \sim N(10, 1.65^2)$
6. จำลอง  $A_t$  ใน 3 กรณี คือ
  - (a) จำลอง  $A_t$  จากการแจกแจงยูนิฟอร์มในช่วง (1,2) นั่นคือ  $A_t \sim Unif(1, 2)$

- (b) จำลอง  $A_t$  จากการแจกแจงยูนิฟอร์มในช่วง (4,6) นั่นคือ  $A_t \sim Unif(4, 6)$
- (c) จำลอง  $A_t$  จากการแจกแจงยูนิฟอร์มในช่วง (9,11) นั่นคือ  $A_t \sim Unif(9, 11)$
7. จำลอง  $\epsilon_{it}$  จากการแจกแจงปกติที่มีค่าเฉลี่ย 0 ความแปรปรวน  $A_t$   
 นั่นคือ  $\epsilon_{it} \sim N(0, A_t)$
8. จำลอง  $\psi_{it}$  ใน 2 กรณี คือ
- (a) จำลอง  $\psi_{it}$  จากการแจกแจงยูนิฟอร์มในช่วง  $(\frac{1}{2}A_t, A_t)$  นั่นคือ  $\psi_{it} \sim Unif(\frac{1}{2}A_t, A_t)$
- (b) จำลอง  $\psi_{it}$  จากการแจกแจงยูนิฟอร์มในช่วง  $(\frac{3}{2}A_t, 2A_t)$  นั่นคือ  $\psi_{it} \sim Unif(\frac{3}{2}A_t, 2A_t)$
9. จำลอง  $e_{it}$  จากการแจกแจงปกติที่มีค่าเฉลี่ย 0 ความแปรปรวน  $\psi_{it}$   
 นั่นคือ  $e_{it} \sim N(0, \psi_{it})$
10. จำลอง  $y_{it}$  จาก  $y_{it} = \theta_{it} + e_{it}$

### กรณีปัจจัยเกิดจากเวลา

สำหรับ  $i = 1, 2, \dots, 10$  และ  $t = 1, 2, \dots, 5$

พิจารณา แบบจำลองเชื่อมโยง

$$\theta_{it} = x'_{it}\beta + v_i + \epsilon_{it} \quad (3.24)$$

เมื่อ  $\text{var}(\epsilon_{it}) = A_t$

และแบบจำลองการสุ่มตัวอย่าง

$$y_{it} = \theta_{it} + e_{it} \quad (3.25)$$

เมื่อ  $\text{var}(e_{it}) = \psi_{it}$

โดยมีการตั้งค่าการจำลองข้อมูลดังนี้

1. จำนวนครั้งที่ทำการจำลอง:  $K = 1000$  ครั้ง
2. จำนวนตัวอย่าง: 10
3. สัมประสิทธิ์การถดถอย  $\beta_1 = \beta_2 = 1$



4. จำลอง  $\theta_{it}(i = 1, 2, \dots, n)(t = 1, 2, \dots, T)$  จากสมการ (3.24)  
โดยจำลอง  $x_{it1}, x_{it2}$  จากการแจกแจงปกติ ที่มีค่าเฉลี่ย 10 ความแปรปรวน 1 และ  $1.05^2$  นั่นคือ  $x_{it1} \sim N(10, 1)$  และ  $x_{it2} \sim N(10, 1.05^2)$  ตามลำดับ
5. จำลอง  $v_i$  จากการแจกแจงปกติที่มีค่าเฉลี่ย 10 ความแปรปรวน  $1.65^2$   
นั่นคือ  $v_t \sim N(10, 1.65^2)$
6. จำลอง  $A_t$  ใน 3 กรณี คือ
  - (a) จำลอง  $A_t$  จากการแจกแจงยูนิฟอร์มในช่วง (1,2) นั่นคือ  $A_t \sim Unif(1, 2)$
  - (b) จำลอง  $A_t$  จากการแจกแจงยูนิฟอร์มในช่วง (4,6) นั่นคือ  $A_t \sim Unif(4, 6)$
  - (c) จำลอง  $A_t$  จากการแจกแจงยูนิฟอร์มในช่วง (9,11) นั่นคือ  $A_t \sim Unif(9, 11)$
7. จำลอง  $\epsilon_{it}$  จากการแจกแจงปกติที่มีค่าเฉลี่ย 0 ความแปรปรวน  $A_t$   
นั่นคือ  $\epsilon_{it} \sim N(0, A_t)$
8. จำลอง  $\psi_{it}$  ใน 2 กรณี คือ
  - (a) จำลอง  $\psi_{it}$  จากการแจกแจงยูนิฟอร์มในช่วง  $(\frac{1}{2}A_t, A_t)$  นั่นคือ  $\psi_{it} \sim Unif(\frac{1}{2}A_t, A_t)$
  - (b) จำลอง  $\psi_{it}$  จากการแจกแจงยูนิฟอร์มในช่วง  $(\frac{3}{2}A_t, 2A_t)$  นั่นคือ  $\psi_{it} \sim Unif(\frac{3}{2}A_t, 2A_t)$
9. จำลอง  $e_{it}$  จากการแจกแจงปกติที่มีค่าเฉลี่ย 0 ความแปรปรวน  $\psi_{it}$   
นั่นคือ  $e_{it} \sim N(0, \psi_{it})$
10. จำลอง  $y_{it}$  จาก  $y_{it} = \theta_{it} + e_{it}$

หลังจากได้ข้อมูลจากการจำลองข้อมูลแล้ว จึงนำข้อมูลดังกล่าวมาหาจำนวนของการเกิด  $\hat{A}_t \leq 0$  และ หาค่าเฉลี่ยของกำลังสองของความคลาดเคลื่อน (Mean squared error) ของตัวประมาณค่าความแปรปรวนจากการถดถอย  $\hat{A}_t$  และหาค่าเฉลี่ยของกำลังสองของความคลาดเคลื่อน (Mean squared error) ของ  $\hat{\theta}_{it}^{EBLUP}$

การคำนวณค่าเฉลี่ยของกำลังสองของความคลาดเคลื่อน (Mean squared error: MSE) ของตัวประมาณค่า

1. การคำนวณ MSE ของตัวประมาณค่าความแปรปรวนจากการถดถอย  $\hat{A}_t$

$$MSE = \frac{1}{K} \sum_{k=1}^K (\hat{A}_{tk} - A_{tk})^2 \quad (3.26)$$

เป็นสูตรคำนวณค่าเฉลี่ยของกำลังสองของความคลาดเคลื่อน (Mean squared error: MSE) ระหว่างตัวประมาณค่าความแปรปรวนจากการถดถอย  $\hat{A}_t$  และ ค่าจริงของความแปรปรวนจากการถดถอย  $A_t$

2. การคำนวณ MSE ของตัวประมาณค่าที่ไม่เอนเอียงเชิงเส้นเชิงประจักษ์ที่ดีที่สุด  $\hat{\theta}_{it}^{EBLUP}$

2.1 การคำนวณ MSE แบบที่ 1

$$MSE = \frac{1}{nTK} \sum_{i=1}^n \sum_{t=1}^T \sum_{k=1}^K (\hat{\theta}_{itk}^{EBLUP} - \theta_{itk}^{BLUP})^2 \quad (3.27)$$

เป็นสูตรคำนวณค่าเฉลี่ยของกำลังสองของความคลาดเคลื่อน (Mean squared error: MSE) ระหว่างตัวประมาณค่า  $\hat{\theta}_{it}^{EBLUP}$  ที่ได้จากค่าประมาณความแปรปรวนจากการถดถอย  $\hat{A}_t$  และตัวประมาณค่า  $\theta_{it}^{BLUP}$  ที่ได้จากค่าจริงของความแปรปรวนจากการถดถอย  $A_t$

2.2 การคำนวณ MSE แบบที่ 2

$$MSE = \frac{1}{nTK} \sum_{i=1}^n \sum_{t=1}^T \sum_{k=1}^K (\hat{\theta}_{itk}^{EBLUP} - \theta_{itk})^2 \quad (3.28)$$

เป็นสูตรคำนวณค่าเฉลี่ยของกำลังสองของความคลาดเคลื่อน (Mean squared error: MSE) ระหว่างตัวประมาณค่า  $\hat{\theta}_{it}^{EBLUP}$  ที่ได้จากค่าประมาณความแปรปรวนจากการถดถอย  $\hat{A}_t$  และค่าจริง  $\theta_{it}$

## ผลการวิเคราะห์การจำลองข้อมูล กรณีปัจจัยเกิดจากพื้นที่

ผลการวิเคราะห์การจำลองข้อมูลแสดงดังตารางที่ 3.22 ถึง ตารางที่ 3.33 โดยตารางที่ 3.22 ตารางที่ 3.26 และตารางที่ 3.30 แสดงจำนวนที่ตัวประมาณความแปรปรวนจากการถดถอย  $\hat{A}_t$  มีค่าน้อยกว่าหรือเท่ากับ 0 จากแบบที่ไม่ปรับ และแบบที่ปรับแล้ว จากการจำลอง  $A_t$  จากการแจกแจงยูนิฟอร์มในช่วง (1,2), (4,6) และ (9,11) ตามลำดับ

ตารางที่ 3.23 ตารางที่ 3.27 และตารางที่ 3.31 แสดงค่าเฉลี่ยของกำลังสองของความคลาดเคลื่อนของตัวประมาณความแปรปรวนจากการถดถอย ( $\hat{A}_t$ ) เมื่อเทียบกับความแปรปรวนจากการถดถอยที่เราจำลองไว้ตั้งแต่ต้น จากแบบที่ไม่ปรับ และแบบที่ปรับแล้ว จากการจำลอง  $A_t$  จากการแจกแจงยูนิฟอร์มในช่วง (1,2), (4,6) และ (9,11) ตามลำดับ

ตารางที่ 3.24 ตารางที่ 3.28 และตารางที่ 3.32 แสดงค่าเฉลี่ยของกำลังสองของความคลาดเคลื่อนของตัวประมาณที่ไม่เอนเอียงเชิงเส้นเชิงประจักษ์ที่ดีที่สุดที่คำนวณจากตัวประมาณความแปรปรวนจากการถดถอย  $\hat{A}_t$  ( $\hat{\theta}_{it}^{EBLUP}$ ) เทียบกับที่คำนวณจากการถดถอยที่เราจำลองไว้ตั้งแต่ต้น ( $\hat{\theta}_{it}^{BLUP}$ ) และเทียบกับค่าจริงที่เราจำลองได้ตั้งแต่ต้น ( $\theta_{it}$ ) กรณี PML และ APML จากการจำลอง  $A_t$  จากการแจกแจงยูนิฟอร์มในช่วง (1,2), (4,6) และ (9,11) ตามลำดับ โดยที่  $\hat{A}_{0t}$  ที่แสดงในตารางคือ  $\hat{A}_{0t} = \max(0, \hat{A}_t)$

ตารางที่ 3.25 ตารางที่ 3.29 และตารางที่ 3.33 แสดงค่าเฉลี่ยของกำลังสองของความคลาดเคลื่อนของตัวประมาณที่ไม่เอนเอียงเชิงเส้นเชิงประจักษ์ที่ดีที่สุดที่คำนวณจากตัวประมาณความแปรปรวนจากการถดถอย  $\hat{A}_t$  ( $\hat{\theta}_{it}^{EBLUP}$ ) เทียบกับที่คำนวณจากการถดถอยที่เราจำลองไว้ตั้งแต่ต้น ( $\hat{\theta}_{it}^{BLUP}$ ) และเทียบกับค่าจริงที่เราจำลองได้ตั้งแต่ต้น ( $\theta_{it}$ ) กรณี REML และ AREML จากการจำลอง  $A_t$  จากการแจกแจงยูนิฟอร์มในช่วง (1,2), (4,6) และ (9,11) ตามลำดับ โดยที่  $\hat{A}_{0t}$  ที่แสดงในตารางคือ  $\hat{A}_{0t} = \max(0, \hat{A}_t)$

กรณีจำลอง  $A_t$  จากการแจกแจงยูนิฟอร์มในช่วง (1,2) นั่นคือ  $A_t \sim Unif(1, 2)$

$A_t \sim Unif(1, 2)$	$\hat{A}_t$	PML	APML	REML	AREML
$\psi_{it} \sim Unif(\frac{1}{2}A_t, A_t)$	$\hat{A}_1$	247	0	179	0
	$\hat{A}_2$	264	0	181	0
	$\hat{A}_3$	240	0	169	0
	$\hat{A}_4$	244	0	160	0
	$\hat{A}_5$	242	0	179	0
$\psi_{it} \sim Unif(\frac{3}{2}A_t, A_t)$	$\hat{A}_1$	330	0	269	0
	$\hat{A}_2$	307	0	203	0
	$\hat{A}_3$	299	0	312	0
	$\hat{A}_4$	313	0	333	0
	$\hat{A}_5$	292	0	316	0

ตารางที่ 3.22: จำนวนตัวประมาณความแปรปรวนจากการถดถอย  $\hat{A}_t \leq 0$  กรณี  $A_t \sim Unif(1, 2)$

จากตารางที่ 3.22 จะเห็นว่า ตัวประมาณความแปรปรวนจากการถดถอย  $\hat{A}_t \leq 0$  จะมีจำนวนเพิ่มมากขึ้นเมื่อ ความแปรปรวนจากการสุ่มตัวอย่างมีค่าเพิ่มขึ้น เช่น ในกรณีที่ ความแปรปรวนจากการสุ่มตัวอย่างมีค่าน้อย นั่นคือกรณี  $\psi_{it} \sim Unif(\frac{1}{2}A_t, A_t)$  จะเห็นว่า จำนวน  $\hat{A}_1 \leq 0$  มีจำนวน 247 ซึ่งน้อยกว่า จำนวน  $\hat{A}_1 \leq 0$  ที่มีจำนวน 330 ในกรณีที่ ความแปรปรวนจากการสุ่มตัวอย่างมีค่ามาก ในขณะที่ไม่พบตัวประมาณความแปรปรวนจากการถดถอย  $\hat{A}_t \leq 0$  ของกรณีที่ปรับแล้ว

$A_t \sim Unif(1, 2)$	ค่าจริง	PML	APML	REML	AREML
$\psi_{it} \sim Unif(\frac{1}{2}A_t, A_t)$	$A_1$	1.9455	1.9422	2.7796	2.3141
	$A_2$	2.0796	1.8444	2.8324	2.3241
	$A_3$	1.9657	2.0369	2.5758	2.1508
	$A_4$	1.9496	1.8709	3.2600	2.4629
	$A_5$	2.0150	1.9725	2.3722	1.8293
$\psi_{it} \sim Unif(\frac{3}{2}A_t, A_t)$	$A_1$	5.0485	2.7919	6.5389	4.3944
	$A_2$	4.9876	3.2724	5.9326	3.9831
	$A_3$	4.5512	2.8444	5.7819	3.5478
	$A_4$	4.8958	3.1114	6.3099	4.1014
	$A_5$	4.6262	3.1723	6.2847	4.5999

ตารางที่ 3.23: ค่าเฉลี่ยของกำลังสองของความคลาดเคลื่อน (Mean squared error) ของ  $\hat{A}_t$  กรณี  $A_t \sim Unif(1, 2)$

จากตารางที่ 3.23 จะเห็นว่า ค่าเฉลี่ยของกำลังสองของความคลาดเคลื่อนของตัวประมาณความแปรปรวนจากการถดถอย  $\hat{A}_t$  ของกรณีที่ปรับแล้วส่วนใหญ่ มีค่าน้อยกว่า ค่าเฉลี่ยของกำลังสองของความคลาดเคลื่อนของตัวประมาณความแปรปรวนจากการถดถอย  $\hat{A}_t$  ของกรณีที่ยังไม่ปรับ เช่น เมื่อความแปรปรวนจากการสุ่มตัวอย่างมีค่าระหว่าง  $\frac{1}{2}A_t$  ถึง  $A_t$  และเมื่อพิจารณา  $A_2$  จะเห็นว่า ค่าเฉลี่ยของกำลังสองของความคลาดเคลื่อนของตัวประมาณความแปรปรวนจากการถดถอยของกรณีที่ปรับแล้ว มีค่า 1.8444 และ ค่าเฉลี่ยของกำลังสองของความคลาดเคลื่อนของตัวประมาณความแปรปรวนจากการถดถอยของกรณีที่ยังไม่ปรับ มีค่า 2.0796 และเมื่อพิจารณา  $A_3$  จะเห็นว่า เมื่อความแปรปรวนจากการสุ่มตัวอย่างมีค่ามากขึ้น ทำให้ค่าเฉลี่ยของกำลังสองของความคลาดเคลื่อนของตัวประมาณความแปรปรวนจากการถดถอย  $\hat{A}_t$  ของกรณีที่ปรับแล้ว มีค่าน้อยกว่า ค่าเฉลี่ยของกำลังสองของความคลาดเคลื่อนของตัวประมาณความแปรปรวนจากการถดถอย  $\hat{A}_t$  ของกรณีที่ยังไม่ปรับ

ความแปรปรวนจากการ สุ่มตัวอย่าง	ค่าจริง	PML		APML
		$\hat{A}_t$	$\hat{A}_{0t}$	
$\psi_{it} \sim Unif(\frac{1}{2}A_t, A_t)$	$\hat{\theta}_{it}^{BLUP}$	$8 \times 10^{25}$	0.0196	0.0233
	$\theta_{it}$	$8 \times 10^{25}$	0.0833	0.0871
$\psi_{it} \sim Unif(\frac{3}{2}A_t, A_t)$	$\hat{\theta}_{it}^{BLUP}$	$3 \times 10^{26}$	0.0221	0.0245
	$\theta_{it}$	$3 \times 10^{26}$	0.1164	0.1185

ตารางที่ 3.24: ค่าเฉลี่ยของกำลังสองของความคลาดเคลื่อน (Mean squared error) ของ  $\hat{\theta}_{it}^{EBLUP}$  กรณี  $A_t \sim Unif(1, 2)$  ของ PML และ APML

จากตารางที่ 3.24 จะเห็นว่า ค่าเฉลี่ยของกำลังสองของความคลาดเคลื่อนของ  $\hat{\theta}_{it}^{EBLUP}$  ของกรณีที่ปรับแล้ว มีค่ามากกว่า กรณีที่ยังไม่ปรับ

ความแปรปรวนจากการ สุ่มตัวอย่าง	ค่าจริง	REML		AREML
		$\hat{A}_t$	$\hat{A}_{0t}$	
$\psi_{it} \sim Unif(\frac{1}{2}A_t, A_t)$	$\hat{\theta}_{it}^{BLUP}$	$2 \times 10^{12}$	0.0183	0.0130
	$\theta_{it}$	$2 \times 10^{12}$	0.0823	0.0769
$\psi_{it} \sim Unif(\frac{3}{2}A_t, A_t)$	$\hat{\theta}_{it}^{BLUP}$	$4 \times 10^6$	0.0280	0.0208
	$\theta_{it}$	$4 \times 10^6$	0.1216	0.1145

ตารางที่ 3.25: ค่าเฉลี่ยของกำลังสองของความคลาดเคลื่อน (Mean squared error) ของ  $\hat{\theta}_{it}^{EBLUP}$  กรณี  $A_t \sim Unif(1, 2)$  ของ REML และ AREML

จากตารางที่ 3.25 จะเห็นว่า ค่าเฉลี่ยของกำลังสองของความคลาดเคลื่อนของ  $\hat{\theta}_{it}^{EBLUP}$  ของกรณีที่ปรับแล้ว มีค่าน้อยกว่า กรณีที่ยังไม่ปรับ

กรณีจำลอง  $A_t$  จากการแจกแจงยูนิฟอร์มในช่วง (4,6) นั่นคือ  $A_t \sim Unif(4, 6)$

$A_t \sim Unif(4, 6)$	$\hat{A}_t$	PML	APML	REML	AREML
$\psi_{it} \sim Unif(\frac{1}{2}A_t, A_t)$	$\hat{A}_1$	282	0	175	0
	$\hat{A}_2$	285	0	185	0
	$\hat{A}_3$	310	0	187	0
	$\hat{A}_4$	289	0	184	0
	$\hat{A}_5$	253	0	166	0
$\psi_{it} \sim Unif(\frac{3}{2}A_t, A_t)$	$\hat{A}_1$	365	0	303	0
	$\hat{A}_2$	348	0	355	0
	$\hat{A}_3$	350	0	319	0
	$\hat{A}_4$	344	0	312	0
	$\hat{A}_5$	353	0	285	0

ตารางที่ 3.26: จำนวนตัวประมาณความแปรปรวนจากการถดถอย  $\hat{A}_t \leq 0$  กรณี  $A_t \sim Unif(4, 6)$

จากตารางที่ 3.26 จะเห็นว่า ตัวประมาณความแปรปรวนจากการถดถอย  $\hat{A}_t \leq 0$  จะมีจำนวนเพิ่มมากขึ้นเมื่อ ความแปรปรวนจากการสุ่มตัวอย่างมีค่าเพิ่มขึ้น เช่น ในกรณีที่ ความแปรปรวนจากการสุ่มตัวอย่างมีค่าน้อย จำนวน  $\hat{A}_1 \leq 0$  มีจำนวน 282 ซึ่งน้อยกว่า จำนวน  $\hat{A}_1 \leq 0$  ที่มีจำนวน 365 ในกรณีที่ ความแปรปรวนจากการสุ่มตัวอย่างมีค่ามาก ในขณะที่ไม่พบตัวประมาณความแปรปรวนจากการถดถอย  $\hat{A}_t \leq 0$  ของกรณีที่ปรับแล้ว

$A_t \sim Unif(4, 6)$	ค่าจริง	PML	APML	REML	AREML
$\psi_{it} \sim Unif(\frac{1}{2}A_t, A_t)$	$A_1$	24.5909	23.3996	28.9144	24.0294
	$A_2$	24.7967	22.7299	25.0939	20.6999
	$A_3$	25.9264	23.0417	27.9466	22.1794
	$A_4$	26.6893	20.8597	25.8634	22.2384
	$A_5$	23.4931	17.3090	25.4975	21.0219
$\psi_{it} \sim Unif(\frac{3}{2}A_t, A_t)$	$A_1$	61.8607	32.3360	63.1715	44.3585
	$A_2$	52.1603	30.7480	68.0060	42.0421
	$A_3$	58.8310	32.9307	65.8627	44.5358
	$A_4$	49.6401	26.5054	62.0695	43.4865
	$A_5$	50.9668	34.8136	66.0254	54.3486

ตารางที่ 3.27: ค่าเฉลี่ยของกำลังสองของความคลาดเคลื่อน (Mean squared error) ของ  $\hat{A}_t$  กรณี  $A_t \sim Unif(4, 6)$

จากตารางที่ 3.27 จะเห็นว่า ค่าเฉลี่ยของกำลังสองของความคลาดเคลื่อนของ  $\hat{A}_t$  ของกรณีที่ปรับแล้ว ทั้ง APML และ AREML มีค่าน้อยกว่า ค่าเฉลี่ยของกำลังสองของความคลาดเคลื่อนของ  $\hat{A}_t$  ของกรณีที่ยังไม่ปรับ ทั้ง PML และ REML

ความแปรปรวนจากการสุ่มตัวอย่าง	ค่าจริง	PML		APML
		$\hat{A}_t$	$\hat{A}_{0t}$	
$\psi_{it} \sim Unif(\frac{1}{2}A_t, A_t)$	$\hat{\theta}_{it}^{BLUP}$	$4 \times 10^{26}$	0.0886	0.0607
	$\theta_{it}$	$4 \times 10^{26}$	0.3028	0.2738
$\psi_{it} \sim Unif(\frac{3}{2}A_t, A_t)$	$\hat{\theta}_{it}^{BLUP}$	$8 \times 10^{26}$	0.0759	0.0779
	$\theta_{it}$	$8 \times 10^{26}$	0.3921	0.3948

ตารางที่ 3.28: ค่าเฉลี่ยของกำลังสองของความคลาดเคลื่อน (Mean squared error) ของ  $\hat{\theta}_{it}^{EBLUP}$  กรณี  $A_t \sim Unif(4, 6)$  ของ PML และ APML

จากตารางที่ 3.28 จะเห็นว่า เมื่อความแปรปรวนจากการสุ่มตัวอย่างมีค่าระหว่าง  $\frac{1}{2}A_t$  ถึง  $A_t$  ค่าเฉลี่ยของกำลังสองของความคลาดเคลื่อนของ  $\hat{\theta}_{it}^{EBLUP}$  ของกรณีที่ปรับแล้ว มีค่าน้อยกว่า กรณีที่ยังไม่ปรับ แต่เมื่อความแปรปรวนจากการสุ่มตัวอย่างมีค่าระหว่าง  $\frac{3}{2}A_t$  ถึง  $2A_t$



ค่าเฉลี่ยของกำลังสองของความคลาดเคลื่อนของ  $\hat{\theta}_{it}^{EBLUP}$  ของกรณีที่ปรับแล้ว มีค่ามากกว่ากรณีที่ยังไม่ปรับ

ความแปรปรวนจากการ สุ่มตัวอย่าง	ค่าจริง	REML		AREML
		$\hat{A}_t$	$\hat{A}_{0t}$	
$\psi_{it} \sim Unif(\frac{1}{2}A_t, A_t)$	$\hat{\theta}_{it}^{BLUP}$	$1 \times 10^5$	0.0611	0.0428
	$\theta_{it}$	$1 \times 10^5$	0.2726	0.2546
$\psi_{it} \sim Unif(\frac{3}{2}A_t, A_t)$	$\hat{\theta}_{it}^{BLUP}$	$7 \times 10^4$	0.0896	0.0670
	$\theta_{it}$	$7 \times 10^4$	0.4068	0.3834

ตารางที่ 3.29: ค่าเฉลี่ยของกำลังสองของความคลาดเคลื่อน (Mean squared error) ของ  $\hat{\theta}_{it}^{EBLUP}$  กรณี  $A_t \sim Unif(4, 6)$  ของ REML และ AREML

จากตารางที่ 3.29 จะเห็นว่า ค่าเฉลี่ยของกำลังสองของความคลาดเคลื่อนของ  $\hat{\theta}_{it}^{EBLUP}$  ของกรณีที่ปรับแล้ว มีค่าน้อยกว่า กรณีที่ยังไม่ปรับ

กรณีจำลอง  $A_t$  จากการแจกแจงยูนิฟอิร์มในช่วง (9,11) นั่นคือ  $A_t \sim Unif(9, 11)$

$A_t \sim Unif(9, 11)$	$\hat{A}_t$	PML	APML	REML	AREML
$\psi_{it} \sim Unif(\frac{1}{2}A_t, A_t)$	$\hat{A}_1$	287	0	153	0
	$\hat{A}_2$	283	0	171	0
	$\hat{A}_3$	324	0	196	0
	$\hat{A}_4$	321	0	193	0
	$\hat{A}_5$	314	0	174	0
$\psi_{it} \sim Unif(\frac{3}{2}A_t, A_t)$	$\hat{A}_1$	411	0	281	0
	$\hat{A}_2$	384	0	279	0
	$\hat{A}_3$	363	0	315	0
	$\hat{A}_4$	422	0	338	0
	$\hat{A}_5$	381	0	325	0

ตารางที่ 3.30: จำนวนตัวประมาณความแปรปรวนจากการถดถอย  $\hat{A}_t \leq 0$  กรณี  $A_t \sim Unif(9, 11)$

จากตารางที่ 3.30 จะเห็นว่า ตัวประมาณความแปรปรวนจากการถดถอย  $\hat{A}_t \leq 0$  จะมีจำนวนเพิ่มมากขึ้นเมื่อ ความแปรปรวนจากการสุ่มตัวอย่างมีค่าเพิ่มขึ้น เช่น ในกรณีที่ ความแปรปรวนจากการสุ่มตัวอย่างมีค่าน้อย จำนวน  $\hat{A}_1 \leq 0$  มีจำนวน 287 ซึ่งน้อยกว่า จำนวน  $\hat{A}_1 \leq 0$  ที่มีจำนวน 411 ในกรณีที่ ความแปรปรวนจากการสุ่มตัวอย่างมีค่ามาก ในขณะที่ไม่พบตัวประมาณความแปรปรวนจากการถดถอย  $\hat{A}_t \leq 0$  ของกรณีที่ปรับแล้ว

$A_t \sim Unif(9, 11)$	ค่าจริง	PML	APML	REML	AREML
$\psi_{it} \sim Unif(\frac{1}{2}A_t, A_t)$	$A_1$	109.2809	89.2547	113.7953	101.0025
	$A_2$	134.9824	85.1508	116.8868	98.6747
	$A_3$	116.0919	79.7114	108.2653	80.4849
	$A_4$	130.5931	88.8596	123.2960	95.0106
	$A_5$	110.0836	83.0024	113.4962	82.4317
$\psi_{it} \sim Unif(\frac{3}{2}A_t, A_t)$	$A_1$	236.8013	115.5385	226.4998	115.8273
	$A_2$	233.1346	115.9303	230.4199	148.6772
	$A_3$	234.2293	122.5360	234.3832	127.1627
	$A_4$	247.3689	119.7243	233.5138	115.4806
	$A_5$	220.0148	139.9154	224.2997	134.6341

ตารางที่ 3.31: ค่าเฉลี่ยของกำลังสองของความคลาดเคลื่อน (Mean squared error) ของ  $\hat{A}_t$  กรณี  $A_t \sim Unif(9, 11)$

จากตารางที่ 3.31 จะเห็นว่า ค่าเฉลี่ยของกำลังสองของความคลาดเคลื่อนของ  $\hat{A}_t$  ของกรณีที่ปรับแล้ว ทั้ง APML และ AREML มีค่าน้อยกว่า ค่าเฉลี่ยของกำลังสองของความคลาดเคลื่อนของ  $\hat{A}_t$  ของกรณีที่ยังไม่ปรับ ทั้ง PML และ REML

ความแปรปรวนจากการ สุ่มตัวอย่าง	ค่าจริง	PML		APML
		$\hat{A}_t$	$\hat{A}_{0t}$	
$\psi_{it} \sim Unif(\frac{1}{2}A_t, A_t)$	$\hat{\theta}_{it}^{BLUP}$	$3 \times 10^{27}$	0.2013	0.1588
	$\theta_{it}$	$3 \times 10^{27}$	0.6227	0.5837
$\psi_{it} \sim Unif(\frac{3}{2}A_t, A_t)$	$\hat{\theta}_{it}^{BLUP}$	$1 \times 10^{28}$	0.1592	0.1544
	$\theta_{it}$	$1 \times 10^{28}$	0.7964	0.7877

ตารางที่ 3.32: ค่าเฉลี่ยของกำลังสองของความคลาดเคลื่อน (Mean squared error) ของ  $\hat{\theta}_{it}^{EBLUP}$  กรณี  $A_t \sim Unif(9, 11)$  ของ PML และ APML

จากตารางที่ 3.32 จะเห็นว่า ค่าเฉลี่ยของกำลังสองของความคลาดเคลื่อนของ  $\hat{\theta}_{it}^{EBLUP}$  ของกรณีที่ปรับแล้ว มีค่าน้อยกว่า กรณีที่ยังไม่ปรับ

ความแปรปรวนจากการ สุ่มตัวอย่าง	ค่าจริง	PML		APML
		$\hat{A}_t$	$\hat{A}_{0t}$	
$\psi_{it} \sim Unif(\frac{1}{2}A_t, A_t)$	$\hat{\theta}_{it}^{BLUP}$	$6 \times 10^7$	0.1266	0.0930
	$\theta_{it}$	$6 \times 10^7$	0.5528	0.5184
$\psi_{it} \sim Unif(\frac{1}{2}A_t, A_t)$	$\hat{\theta}_{it}^{BLUP}$	$8 \times 10^4$	0.1849	0.1425
	$\theta_{it}$	$8 \times 10^4$	0.8156	0.7744

ตารางที่ 3.33: ค่าเฉลี่ยของกำลังสองของความคลาดเคลื่อน (Mean squared error) ของ  $\hat{\theta}_{it}^{EBLUP}$  กรณี  $A_t \sim Unif(9, 11)$  ของ REML และ AREML

จากตารางที่ 3.33 จะเห็นว่า ค่าเฉลี่ยของกำลังสองของความคลาดเคลื่อนของ  $\hat{\theta}_{it}^{EBLUP}$  ของกรณีที่ปรับแล้ว มีค่าน้อยกว่า กรณีที่ยังไม่ปรับ

จากตารางที่ 3.23, ตารางที่ 3.24, ตารางที่ 3.27, ตารางที่ 3.28, ตารางที่ 3.31 และ ตารางที่ 3.32 ทำให้สรุปได้ว่า เมื่อความแปรปรวนจากการสุ่มตัวอย่งมีค่าน้อย การประมาณแบบภาวะความน่าจะเป็นสูงสุดโปรไฟล์ (PML) จะยังคงให้ผลที่ดีกว่า แต่เมื่อความแปรปรวนจากการสุ่มตัวอย่างมีค่าเพิ่มขึ้น การประมาณแบบภาวะสูงสุดโปรไฟล์แบบปรับ (APML) จะให้ผลที่ดีกว่า

## ผลการวิเคราะห์การจำลองข้อมูล กรณีปัจจัยเกิดจากเวลา

ผลการวิเคราะห์การจำลองข้อมูลแสดงดังตารางที่ 3.34 ถึง ตารางที่ 3.45 โดยตารางที่ 3.34 ตารางที่ 3.38 และตารางที่ 3.42 แสดงจำนวนที่ตัวประมาณความแปรปรวนจากการถดถอย  $\hat{A}_t$  มีค่าน้อยกว่าหรือเท่ากับ 0 จากแบบที่ไม่ปรับ และแบบที่ปรับแล้ว จากการจำลอง  $A_t$  จากการแจกแจงยูนิฟอร์มในช่วง (1,2), (4,6) และ (9,11) ตามลำดับ

ตารางที่ 3.35 ตารางที่ 3.39 และตารางที่ 3.43 แสดงค่าเฉลี่ยของกำลังสองของความคลาดเคลื่อนของตัวประมาณความแปรปรวนจากการถดถอย ( $\hat{A}_t$ ) เมื่อเทียบกับความแปรปรวนจากการถดถอยที่เราจำลองไว้ตั้งแต่ต้น จากแบบที่ไม่ปรับ และแบบที่ปรับแล้ว จากการจำลอง  $A_t$  จากการแจกแจงยูนิฟอร์มในช่วง (1,2), (4,6) และ (9,11) ตามลำดับ

ตารางที่ 3.36 ตารางที่ 3.40 และตารางที่ 3.44 แสดงค่าเฉลี่ยของกำลังสองของความคลาดเคลื่อนของตัวประมาณที่ไม่เอนเอียงเชิงเส้นเชิงประจักษ์ที่ดีที่สุดที่คำนวณจากตัวประมาณความแปรปรวนจากการถดถอย  $\hat{A}_t$  ( $\hat{\theta}_{it}^{EBLUP}$ ) เทียบกับที่คำนวณจากการถดถอยที่เราจำลองไว้ตั้งแต่ต้น ( $\hat{\theta}_{it}^{BLUP}$ ) และเทียบกับค่าจริงที่เราจำลองได้ตั้งแต่ต้น ( $\theta_{it}$ ) กรณี PML และ APML จากการจำลอง  $A_t$  จากการแจกแจงยูนิฟอร์มในช่วง (1,2), (4,6) และ (9,11) ตามลำดับ โดยที่  $\hat{A}_{0t}$  ที่แสดงในตารางคือ  $\hat{A}_{0t} = \max(0, \hat{A}_t)$

ตารางที่ 3.37 ตารางที่ 3.41 และตารางที่ 3.45 แสดงค่าเฉลี่ยของกำลังสองของความคลาดเคลื่อนของตัวประมาณที่ไม่เอนเอียงเชิงเส้นเชิงประจักษ์ที่ดีที่สุดที่คำนวณจากตัวประมาณความแปรปรวนจากการถดถอย  $\hat{A}_t$  ( $\hat{\theta}_{it}^{EBLUP}$ ) เทียบกับที่คำนวณจากการถดถอยที่เราจำลองไว้ตั้งแต่ต้น ( $\hat{\theta}_{it}^{BLUP}$ ) และเทียบกับค่าจริงที่เราจำลองได้ตั้งแต่ต้น ( $\theta_{it}$ ) กรณี REML และ AREML จากการจำลอง  $A_t$  จากการแจกแจงยูนิฟอร์มในช่วง (1,2), (4,6) และ (9,11) ตามลำดับ โดยที่  $\hat{A}_{0t}$  ที่แสดงในตารางคือ  $\hat{A}_{0t} = \max(0, \hat{A}_t)$

กรณีจำลอง  $A_t$  จากการแจกแจงยูนิฟอร์มในช่วง (1,2) นั่นคือ  $A_t \sim Unif(1, 2)$

$A_t \sim Unif(1, 2)$	$\hat{A}_t$	PML	APML	REML	AREML
$\psi_{it} \sim Unif(\frac{1}{2}A_t, A_t)$	$\hat{A}_1$	19	0	12	0
	$\hat{A}_2$	20	0	9	0
	$\hat{A}_3$	19	0	10	0
	$\hat{A}_4$	29	0	12	0
	$\hat{A}_5$	18	0	12	0
$\psi_{it} \sim Unif(\frac{3}{2}A_t, A_t)$	$\hat{A}_1$	107	0	56	0
	$\hat{A}_2$	120	0	57	0
	$\hat{A}_3$	114	0	66	0
	$\hat{A}_4$	104	0	66	0
	$\hat{A}_5$	90	0	67	0

ตารางที่ 3.34: จำนวนตัวประมาณความแปรปรวนจากการถดถอย  $\hat{A}_t \leq 0$  กรณี  $A_t \sim Unif(1, 2)$

จากตารางที่ 3.34 จะเห็นว่า ตัวประมาณความแปรปรวนจากการถดถอย  $\hat{A}_t \leq 0$  จะมีจำนวนเพิ่มมากขึ้นเมื่อ ความแปรปรวนจากการสุ่มตัวอย่างมีค่าเพิ่มขึ้น เช่น ในกรณีที่ ความแปรปรวนจากการสุ่มตัวอย่างมีค่าน้อย จำนวน  $\hat{A}_1 \leq 0$  มีจำนวน 19 ซึ่งน้อยกว่า จำนวน  $\hat{A}_1 \leq 0$  ที่มีจำนวน 107 ในกรณีที่ ความแปรปรวนจากการสุ่มตัวอย่างมีค่ามาก ในขณะที่ไม่พบตัวประมาณความแปรปรวนจากการถดถอย  $\hat{A}_t \leq 0$  ของกรณีที่ปรับแล้ว

$A_t \sim Unif(1, 2)$	ค่าจริง	PML	APML	REML	AREML
$\psi_{it} \sim Unif(\frac{1}{2}A_t, A_t)$	$A_1$	11.0106	13.6703	15.3377	22.1219
	$A_2$	8.9050	9.6967	23.9830	16.7186
	$A_3$	9.1474	9.7036	14.0996	15.5775
	$A_4$	10.2475	11.1299	14.2955	15.3420
	$A_5$	10.5954	11.6838	14.8174	16.1737
$\psi_{it} \sim Unif(\frac{3}{2}A_t, A_t)$	$A_1$	12.8316	13.8853	20.3155	22.7061
	$A_2$	11.4861	12.2535	18.5847	21.6193
	$A_3$	12.611	13.1104	17.0474	20.1178
	$A_4$	12.4471	13.3787	17.9514	20.9566
	$A_5$	12.9847	14.708	18.8384	19.7291

ตารางที่ 3.35: ค่าเฉลี่ยของกำลังสองของความคลาดเคลื่อน (Mean squared error) ของ  $\hat{A}_t$  กรณี  $A_t \sim Unif(1, 2)$

จากตารางที่ 3.35 จะเห็นว่า ค่าเฉลี่ยของกำลังสองของความคลาดเคลื่อนของตัวประมาณความแปรปรวนจากการถดถอย  $\hat{A}_t$  ของกรณีที่ปรับแล้วส่วนใหญ่ มีค่ามากกว่า ค่าเฉลี่ยของกำลังสองของความคลาดเคลื่อนของตัวประมาณความแปรปรวนจากการถดถอย  $\hat{A}_t$  ของกรณีที่ยังไม่ปรับ

ความแปรปรวนจากการ สุ่มตัวอย่าง	ค่าจริง	PML		APML
		$\hat{A}_t$	$\hat{A}_{0t}$	
$\psi_{it} \sim Unif(\frac{1}{2}A_t, A_t)$	$\hat{\theta}_{it}^{BLUP}$	$6 \times 10^{27}$	0.0141	0.0138
	$\theta_{it}$	$6 \times 10^{27}$	0.0770	0.0767
$\psi_{it} \sim Unif(\frac{3}{2}A_t, A_t)$	$\hat{\theta}_{it}^{BLUP}$	$8 \times 10^{22}$	0.0314	0.0303
	$\theta_{it}$	$8 \times 10^{22}$	0.1257	0.1246

ตารางที่ 3.36: ค่าเฉลี่ยของกำลังสองของความคลาดเคลื่อน (Mean squared error) ของ  $\hat{\theta}_{it}^{EBLUP}$  กรณี  $A_t \sim Unif(1, 2)$  ของ PML และ APML

จากตารางที่ 3.36 จะเห็นว่า ค่าเฉลี่ยของกำลังสองของความคลาดเคลื่อนของ  $\hat{\theta}_{it}^{EBLUP}$  ของกรณีที่ปรับแล้ว มีค่าน้อยกว่า กรณีที่ยังไม่ปรับ

ความแปรปรวนจากการ สุ่มตัวอย่าง	ค่าจริง	REML		AREML
		$\hat{A}_t$	$\hat{A}_{0t}$	
$\psi_{it} \sim Unif(\frac{1}{2}A_t, A_t)$	$\hat{\theta}_{it}^{BLUP}$	$8 \times 10^2$	0.0157	0.0157
	$\theta_{it}$	$8 \times 10^2$	0.0767	0.0767
$\psi_{it} \sim Unif(\frac{3}{2}A_t, A_t)$	$\hat{\theta}_{it}^{BLUP}$	$6 \times 10^3$	0.0375	0.0378
	$\theta_{it}$	$6 \times 10^3$	0.1321	0.1322

ตารางที่ 3.37: ค่าเฉลี่ยของกำลังสองของความคลาดเคลื่อน (Mean squared error) ของ  $\hat{\theta}_{it}^{EBLUP}$  กรณี  $A_t \sim Unif(1, 2)$  ของ REML และ AREML

จากตารางที่ 3.37 จะเห็นว่า ค่าเฉลี่ยของกำลังสองของความคลาดเคลื่อนของ  $\hat{\theta}_{it}^{EBLUP}$  ของกรณีที่ปรับแล้ว มีค่ามากกว่า กรณีที่ยังไม่ปรับ

กรณีจำลอง  $A_t$  จากการแจกแจงยูนิฟอร์มในช่วง (4,6) นั่นคือ  $A_t \sim Unif(4, 6)$

$A_t \sim Unif(4, 6)$	$\hat{A}_t$	PML	APML	REML	AREML
$\psi_{it} \sim Unif(\frac{1}{2}A_t, A_t)$	$\hat{A}_1$	82	0	44	0
	$\hat{A}_2$	79	0	41	0
	$\hat{A}_3$	74	0	42	0
	$\hat{A}_4$	69	0	47	0
	$\hat{A}_5$	75	0	37	0
$\psi_{it} \sim Unif(\frac{3}{2}A_t, A_t)$	$\hat{A}_1$	232	0	172	0
	$\hat{A}_2$	242	0	162	0
	$\hat{A}_3$	275	0	170	0
	$\hat{A}_4$	234	0	171	0
	$\hat{A}_5$	243	0	158	0

ตารางที่ 3.38: จำนวนตัวประมาณความแปรปรวนจากการถดถอย  $\hat{A}_t \leq 0$  กรณี  $A_t \sim Unif(4, 6)$

จากตารางที่ 3.38 จะเห็นว่า ตัวประมาณความแปรปรวนจากการถดถอย  $\hat{A}_t \leq 0$  จะมีจำนวนเพิ่มมากขึ้นเมื่อ ความแปรปรวนจากการสุ่มตัวอย่างมีค่าเพิ่มขึ้น เช่น ในกรณีที่ ความ

แปรปรวนจากการสุ่มตัวอย่างมีค่าน้อย จำนวน  $\hat{A}_1 \leq 0$  มีจำนวน 82 ซึ่งน้อยกว่า จำนวน  $\hat{A}_1 \leq 0$  ที่มีจำนวน 232 ในกรณีที่ ความแปรปรวนจากการสุ่มตัวอย่างมีค่ามาก ในขณะที่ไม่พบตัวประมาณความแปรปรวนจากการถดถอย  $\hat{A}_t \leq 0$  ของกรณีที่ปรับแล้ว

$A_t \sim Unif(4, 6)$	ค่าจริง	PML	APML	REML	AREML
$\psi_{it} \sim Unif(\frac{1}{2}A_t, A_t)$	$A_1$	25.7889	25.5525	38.7732	39.5964
	$A_2$	28.2141	27.4266	40.6827	43.2698
	$A_3$	27.2895	28.1973	46.7276	45.8648
	$A_4$	25.9643	28.5671	39.6351	42.1338
	$A_5$	26.7632	25.6961	41.5184	42.8258
$\psi_{it} \sim Unif(\frac{3}{2}A_t, A_t)$	$A_1$	52.5534	45.4596	82.4417	70.9135
	$A_2$	56.7883	53.7488	74.1973	74.8909
	$A_3$	53.7038	45.518	74.6331	77.3044
	$A_4$	50.6123	35.8663	73.5495	76.1944
	$A_5$	49.5669	45.7854	67.4801	67.5902

ตารางที่ 3.39: ค่าเฉลี่ยของกำลังสองของความคลาดเคลื่อน (Mean squared error) ของ  $\hat{A}_t$  กรณี  $A_t \sim Unif(4, 6)$

จากตารางที่ 3.39 จะเห็นว่า วิธีโปรไฟล์ ค่าเฉลี่ยของกำลังสองของความคลาดเคลื่อนของ  $\hat{A}_t$  ของกรณีที่ปรับแล้วส่วนใหญ่ มีค่าน้อยกว่า ค่าเฉลี่ยของกำลังสองของความคลาดเคลื่อนของ  $\hat{A}_t$  ของกรณีที่ยังไม่ปรับ และ วิธีเรสซิดัล ค่าเฉลี่ยของกำลังสองของความคลาดเคลื่อนของ  $\hat{A}_t$  ของกรณีที่ปรับแล้วส่วนใหญ่ มีค่ามากกว่า ค่าเฉลี่ยของกำลังสองของความคลาดเคลื่อนของ  $\hat{A}_t$  ของกรณีที่ยังไม่ปรับ



ความแปรปรวนจากการ สุ่มตัวอย่าง	ค่าจริง	PML		APML
		$\hat{A}_t$	$\hat{A}_{0t}$	
$\psi_{it} \sim Unif(\frac{1}{2}A_t, A_t)$	$\hat{\theta}_{it}^{BLUP}$	$2 \times 10^{25}$	0.0428	0.0359
	$\theta_{it}$	$2 \times 10^{25}$	0.2535	0.2456
$\psi_{it} \sim Unif(\frac{3}{2}A_t, A_t)$	$\hat{\theta}_{it}^{BLUP}$	$9 \times 10^{24}$	0.0790	0.0670
	$\theta_{it}$	$9 \times 10^{24}$	0.3967	0.3844

ตารางที่ 3.40: ค่าเฉลี่ยของกำลังสองของความคลาดเคลื่อน (Mean squared error) ของ  $\hat{\theta}_{it}^{EBLUP}$  กรณี  $A_t \sim Unif(4, 6)$  ของ PML และ APML

จากตารางที่ 3.40 จะเห็นว่า ค่าเฉลี่ยของกำลังสองของความคลาดเคลื่อนของ  $\hat{\theta}_{it}^{EBLUP}$  ของกรณีที่ปรับแล้ว มีค่าน้อยกว่า กรณีที่ยังไม่ปรับ

ความแปรปรวนจากการ สุ่มตัวอย่าง	ค่าจริง	REML		AREML
		$\hat{A}_t$	$\hat{A}_{0t}$	
$\psi_{it} \sim Unif(\frac{1}{2}A_t, A_t)$	$\hat{\theta}_{it}^{BLUP}$	$1 \times 10^4$	0.0374	0.0354
	$\theta_{it}$	$1 \times 10^4$	0.2472	0.2445
$\psi_{it} \sim Unif(\frac{3}{2}A_t, A_t)$	$\hat{\theta}_{it}^{BLUP}$	$4 \times 10^4$	0.0862	0.0772
	$\theta_{it}$	$4 \times 10^4$	0.3997	0.3893

ตารางที่ 3.41: ค่าเฉลี่ยของกำลังสองของความคลาดเคลื่อน (Mean squared error) ของ  $\hat{\theta}_{it}^{EBLUP}$  กรณี  $A_t \sim Unif(4, 6)$  ของ REML และ AREML

จากตารางที่ 3.41 จะเห็นว่า ค่าเฉลี่ยของกำลังสองของความคลาดเคลื่อนของ  $\hat{\theta}_{it}^{EBLUP}$  ของกรณีที่ปรับแล้ว มีค่าน้อยกว่า กรณีที่ยังไม่ปรับ

กรณีจำลอง  $A_t$  จากการแจกแจงยูนิฟอร์มในช่วง (9,11) นั่นคือ  $A_t \sim Unif(9, 11)$

$A_t \sim Unif(9, 11)$	$\hat{A}_t$	PML	APML	REML	AREML
$\psi_{it} \sim Unif(\frac{1}{2}A_t, A_t)$	$\hat{A}_1$	111	0	75	0
	$\hat{A}_2$	124	0	74	0
	$\hat{A}_3$	94	0	73	0
	$\hat{A}_4$	92	0	59	0
	$\hat{A}_5$	97	0	67	0
$\psi_{it} \sim Unif(\frac{3}{2}A_t, A_t)$	$\hat{A}_1$	290	0	223	0
	$\hat{A}_2$	299	0	223	0
	$\hat{A}_3$	314	0	227	0
	$\hat{A}_4$	284	0	211	0
	$\hat{A}_5$	287	0	214	0

ตารางที่ 3.42: จำนวนตัวประมาณความแปรปรวนจากการถดถอย  $\hat{A}_t \leq 0$  กรณี  $A_t \sim Unif(9, 11)$

จากตารางที่ 3.42 จะเห็นว่า ตัวประมาณความแปรปรวนจากการถดถอย  $\hat{A}_t \leq 0$  จะมีจำนวนเพิ่มมากขึ้นเมื่อ ความแปรปรวนจากการสุ่มตัวอย่างมีค่าเพิ่มขึ้น เช่น ในกรณีที่ ความแปรปรวนจากการสุ่มตัวอย่างมีค่าน้อย จำนวน  $\hat{A}_1 \leq 0$  มีจำนวน 111 ซึ่งน้อยกว่า จำนวน  $\hat{A}_1 \leq 0$  ที่มีจำนวน 290 ในกรณีที่ ความแปรปรวนจากการสุ่มตัวอย่างมีค่ามาก ในขณะที่ไม่พบตัวประมาณความแปรปรวนจากการถดถอย  $\hat{A}_t \leq 0$  ของกรณีที่ปรับแล้ว

$A_t \sim Unif(9, 11)$	ค่าจริง	PML	APML	REML	AREML
$\psi_{it} \sim Unif(\frac{1}{2}A_t, A_t)$	$A_1$	88.9036	76.3790	150.7957	111.6700
	$A_2$	83.3698	75.3235	156.7024	119.3193
	$A_3$	92.5606	79.7479	185.0442	106.7151
	$A_4$	74.7842	70.9137	129.3061	119.1655
	$A_5$	78.7051	76.8828	141.4963	121.6293
$\psi_{it} \sim Unif(\frac{3}{2}A_t, A_t)$	$A_1$	177.0682	147.6599	227.8032	205.4590
	$A_2$	183.8852	166.2443	256.7421	244.1974
	$A_3$	177.8427	130.2867	240.4514	214.9400
	$A_4$	161.6913	108.2753	210.0101	190.3035
	$A_5$	177.9881	130.3682	193.9313	194.5964

ตารางที่ 3.43: ค่าเฉลี่ยของกำลังสองของความคลาดเคลื่อน (Mean squared error) ของ  $\hat{A}_t$  กรณี  $A_t \sim Unif(9, 11)$

จากตารางที่ 3.43 จะเห็นว่า ค่าเฉลี่ยของกำลังสองของความคลาดเคลื่อนของ  $\hat{A}_t$  ของกรณีที่ปรับแล้วส่วนใหญ่ มีค่าน้อยกว่า ค่าเฉลี่ยของกำลังสองของความคลาดเคลื่อนของ  $\hat{A}_t$  ของกรณีที่ยังไม่ปรับ

ความแปรปรวนจากการ สุ่มตัวอย่าง	ค่าจริง	PML		APML
		$\hat{A}_t$	$\hat{A}_{0t}$	
$\psi_{it} \sim Unif(\frac{1}{2}A_t, A_t)$	$\hat{\theta}_{it}^{BLUP}$	$1 \times 10^{27}$	0.0922	0.0813
	$\theta_{it}$	$1 \times 10^{27}$	0.5166	0.5065
$\psi_{it} \sim Unif(\frac{3}{2}A_t, A_t)$	$\hat{\theta}_{it}^{BLUP}$	$2 \times 10^{23}$	0.1529	0.1225
	$\theta_{it}$	$2 \times 10^{23}$	0.7898	0.7595

ตารางที่ 3.44: ค่าเฉลี่ยของกำลังสองของความคลาดเคลื่อน (Mean squared error) ของ  $\hat{\theta}_{it}^{EBLUP}$  กรณี  $A_t \sim Unif(9, 11)$  ของ PML และ APML

จากตารางที่ 3.44 จะเห็นว่า ค่าเฉลี่ยของกำลังสองของความคลาดเคลื่อนของ  $\hat{\theta}_{it}^{EBLUP}$  ของกรณีที่ปรับแล้ว มีค่าน้อยกว่า กรณีที่ยังไม่ปรับ

ความแปรปรวนจากการ สุ่มตัวอย่าง	ค่าจริง	REML		AREML
		$\hat{A}_t$	$\hat{A}_{0t}$	
$\psi_{it} \sim Unif(\frac{1}{2}A_t, A_t)$	$\hat{\theta}_{it}^{BLUP}$	$3 \times 10^9$	0.0790	0.0751
	$\theta_{it}$	$3 \times 10^9$	0.5027	0.4992
$\psi_{it} \sim Unif(\frac{3}{2}A_t, A_t)$	$\hat{\theta}_{it}^{BLUP}$	$3 \times 10^3$	0.1592	0.1392
	$\theta_{it}$	$3 \times 10^3$	0.7935	0.7738

ตารางที่ 3.45: ค่าเฉลี่ยของกำลังสองของความคลาดเคลื่อน (Mean squared error) ของ  $\hat{\theta}_{it}^{BLUP}$  กรณี  $A_t \sim Unif(9, 11)$  ของ REML และ AREML

จากตารางที่ 3.45 จะเห็นว่า ค่าเฉลี่ยของกำลังสองของความคลาดเคลื่อนของ  $\hat{\theta}_{it}^{BLUP}$  ของกรณีที่ปรับแล้ว มีค่าน้อยกว่า กรณีที่ยังไม่ปรับ

จากตารางที่ 3.35 ตารางที่ 3.39 และตารางที่ 3.43 ทำให้สรุปได้ว่า เมื่อความแปรปรวนจากการสุ่มตัวอย่างมีค่าน้อย การประมาณแบบภาวะความน่าจะเป็นสูงสุดโปรไฟล์ (PML) จะยังคงให้ผลที่ดีกว่า แต่เมื่อความแปรปรวนจากการสุ่มตัวอย่างมีค่าเพิ่มขึ้น การประมาณแบบภาวะสูงสุดโปรไฟล์แบบปรับ (APML) จะให้ผลที่ดีกว่า และจากตารางที่ 3.37 ตารางที่ 3.41 และตารางที่ 3.45 ทำให้สรุปได้ว่า เมื่อความแปรปรวนจากการสุ่มตัวอย่างมีค่าน้อย การประมาณแบบภาวะความน่าจะเป็นสูงสุดเรสซิดัล (REML) จะยังคงให้ผลที่ดีกว่า แต่เมื่อความแปรปรวนจากการสุ่มตัวอย่างมีค่าเพิ่มขึ้น การประมาณแบบภาวะสูงสุดเรสซิดัลแบบปรับ (AREML) จะให้ผลที่ดีกว่า

### 3.4 บทประยุกต์ของการปรับฟังก์ชันภาวะความน่าจะเป็น

ในหัวข้อนี้เราศึกษาบทประยุกต์ของการปรับฟังก์ชันภาวะความน่าจะเป็นกับข้อมูลรายได้เฉลี่ยต่อครัวเรือนของประเทศไทย โดยในลำดับแรก เราจะอธิบายโครงสร้างของข้อมูล ตามด้วยการคำนวณค่าประมาณจากตัวอย่างของรายได้เฉลี่ยต่อครัวเรือน และผลการศึกษาเชิงตัวเลขของวิธีการปรับฟังก์ชันภาวะความน่าจะเป็นกับข้อมูลรายได้เฉลี่ยต่อครัวเรือนของประเทศไทย

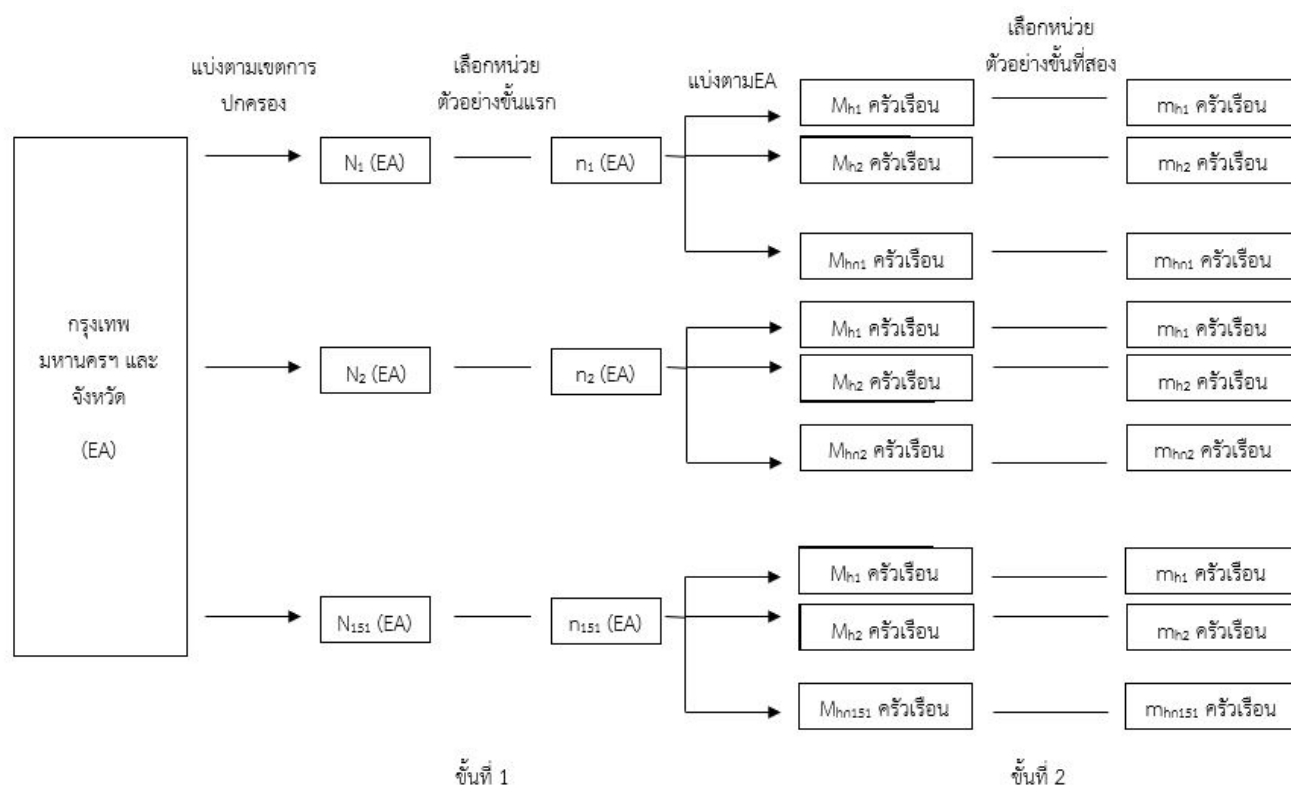
ข้อมูลที่ใช้ในการศึกษาเป็นข้อมูลรายได้เฉลี่ยของประชากรไทย ในปี พ.ศ. 2552, 2554, 2556, 2558 และ 2560 และข้อมูลสำมะโนประชากรและเคหะ พ.ศ. 2553 โดยมีที่มาของข้อมูลคือ สำนักงานสถิติแห่งชาติ

จำนวนกลุ่มตัวอย่างของข้อมูลรายได้เฉลี่ยของประชากรไทยในปี พ.ศ. 2552 คือ 43,844 คร่าวเรือน ปี พ.ศ. 2554 คือ 42,083 คร่าวเรือน ปี พ.ศ. 2556 คือ 42,738 คร่าวเรือน ปี พ.ศ. 2558 คือ 43,400 คร่าวเรือน ปี พ.ศ. 2560 คือ 43,210 คร่าวเรือน และในปี พ.ศ. 2554 ประเทศไทยได้มีการแบ่งและแยกจังหวัดหนองคาย ออกมาเป็นจังหวัดหนองคายและจังหวัดบึงกาฬ แต่เราจะใช้ข้อมูลโดยการคิดเป็นเพียงหนึ่งจังหวัด คือ จังหวัดหนองคาย ในงานวิจัยนี้จะแยกพิจารณาเป็น 5 ภาค รวมทั้งสิ้น 75 จังหวัด ไม่รวมกรุงเทพมหานคร

สำหรับแผนการสุ่มตัวอย่างที่ใช้คือ แผนการสุ่มตัวอย่างแบบแบ่งสองชั้นภูมิ (stratified two-stage sampling) โดยมีกรุงเทพมหานครฯ และจังหวัดเป็นชั้นภูมิ (76 ชั้นภูมิ) และในแต่ละชั้นภูมิ (ยกเว้นกรุงเทพมหานครฯ) ได้ทำการแบ่งออกเป็นอีก 2 ชั้นภูมีย่อย ตามเขตการปกครอง คือ ในเขต และ นอกเขต แล้วทำการเลือกตัวอย่าง ดังนี้

1. การเลือกหน่วยตัวอย่างชั้นแรก เป็นการเลือก enumeration area (EA) ตัวอย่าง จากแต่ละชั้นภูมีย่อย (แต่ละเขตการปกครอง) อย่างอิสระต่อกัน โดยให้ความน่าจะเป็นในการเลือกเป็นปฏิภาคกับจำนวนคร่าวเรือนของ EA นั้น ๆ
2. การเลือกหน่วยตัวอย่างชั้นที่สอง เป็นการเลือกคร่าวเรือนตัวอย่าง จากทุก EA ที่เลือกในชั้นแรก โดยกำหนดขนาดตัวอย่างในแต่ละเขต คือ
  - (a) ในเขต เลือก 15 คร่าวเรือนต่อตัวอย่าง EA
  - (b) นอกเขต เลือก 12 คร่าวเรือนต่อตัวอย่าง EA

การเลือกหน่วยตัวอย่างในแต่ละชั้น จะแสดงดังรูปที่ 3.1



รูปที่ 3.1: แสดงการเลือกตัวอย่างในแต่ละชั้น

สำหรับตัวแปรที่สนใจจะศึกษาเป็นดังนี้

1. ตัวแปรตาม ได้แก่ รายได้เฉลี่ยต่อครัวเรือนของประชากรไทย
2. ตัวแปรต้น ได้แก่
  - สัดส่วนของครัวเรือนที่เป็นบ้านเดี่ยว
  - สัดส่วนของครัวเรือนที่มีที่อยู่เป็นของตัวเอง
  - สัดส่วนของครัวเรือนที่เช่าที่อยู่
  - สัดส่วนของครัวเรือนที่เป็นบ้านอิฐหรือปูน
  - สัดส่วนของครัวเรือนที่มีที่ดินเป็นของตัวเอง
  - สัดส่วนของครัวเรือนที่ใช้แก๊สในการทำอาหาร
  - สัดส่วนของครัวเรือนที่ใช้ชักโครก
  - สัดส่วนของประชากรชาย
  - ค่าเฉลี่ยของประชากรต่อครัวเรือนส่วนตัว

- สัดส่วนที่ของครัวเรือนที่หัวหน้าครัวเรือนทำงาน
- สัดส่วนของประชากรที่ศึกษาจบระดับมัธยมศึกษา
- สัดส่วนของประชากรที่ศึกษาจบระดับปริญญาตรี และ ปวส.
- ตัวบ่งชี้พื้นที่เขต

ตัวแปรเหล่านี้เป็นข้อมูลที่ได้มาจากข้อมูลสำมะโนประชากรและเคหะ พ.ศ. 2553 สำนักงานสถิติแห่งชาติ

ในงานวิจัยนี้ใช้แบบจำลองเฟ-เฮรอตแบบพานเนล กับตัวแปรตามและตัวแปรต้นดังกล่าว โดยค่าความแปรปรวน  $A$  จะถูกประมาณโดยตัวประมาณแบบภาวะความน่าจะเป็นสูงสุดโปรไฟล์ (Profile maximum likelihood estimator: PML) ตัวประมาณแบบภาวะความน่าจะเป็นสูงสุดโปรไฟล์แบบปรับ (Adjusted profile maximum likelihood estimator: APML) ตัวประมาณแบบภาวะความน่าจะเป็นสูงสุดเรสซิดัล (Residual maximum likelihood estimator: REML) และตัวประมาณแบบภาวะความน่าจะเป็นสูงสุดเรสซิดัลแบบปรับ (Adjusted residual maximum likelihood estimator: AREML) โดยการใช้คำสั่ง `optim` ในโปรแกรม R โดยงานวิจัยของเราจะแบ่งพิจารณาเป็น 10 พื้นที่ขนาดเล็กตามภาคและเขตการปกครอง

### 3.4.1 การคำนวณการประมาณค่าของรายได้ประชากรไทย

ในหัวข้อนี้ เราแสดงสูตรการหาค่าประมาณจากตัวอย่างของรายได้ของประชากร และสูตรการคำนวณความแปรปรวน ตามโครงสร้างการสุ่มตัวอย่างที่ได้นำเสนอในหัวข้อ 2.1.7

ให้

$\bar{y}_h$  คือ ค่าเฉลี่ยของรายได้

$V(\bar{y}_h)$  คือ ค่าความแปรปรวนของรายได้เฉลี่ย

$y_{hij}$  คือ รายได้ของคน  $j$  ของชั้นภูมีย่อยที่  $i$  ของชั้นภูมิที่  $h$

$N_h$  คือ จำนวน EA ของชั้นภูมิที่  $h$

$n_h$  คือ จำนวนตัวอย่างที่สุ่มจาก EA ของชั้นภูมิที่  $h$

$M_{hi}$  คือ จำนวนครัวเรือนของชั้นภูมีย่อยที่  $i$  ของชั้นภูมิที่  $h$

$m_{hi}$  คือ จำนวนตัวอย่างที่สุ่มจากครัวเรือนของชั้นภูมีย่อยที่  $i$  ของชั้นภูมิที่  $h$

$\bar{M}_h$  คือ ค่าเฉลี่ยของจำนวนตัวอย่างครัวเรือน

ในการประมาณค่าเฉลี่ยรายได้นั้นจะปรับใช้ (2.16) และ (2.18) ดังนี้

ดังนั้น รายได้ทั้งหมดจะเป็น

$$\frac{N_h}{n_h} \sum_{i=1}^{n_h} \frac{M_{hi}}{m_{hi}} \sum_{j=1}^{m_{hi}} y_{hij}$$

โดยที่  $\frac{N_h}{n_h}$  และ  $\frac{M_{hi}}{m_{hi}}$  คือน้ำหนักที่เกิดจากแบ่งและสุ่มครั้งที่ 1 และครั้งที่ 2 ตามลำดับ เนื่องจากเราไม่ทราบจำนวนตัวอย่างทั้งหมดที่สุ่มมาแน่ชัด เราจึงประมาณจำนวนตัวอย่างทั้งหมดด้วย  $N_h \times \bar{M}_h$

ดังนั้น ค่าเฉลี่ยของรายได้ของชั้นภูมิที่  $h$  คือ

$$\begin{aligned} \bar{y}_h &= \frac{1}{N_h \times \bar{M}_h} \times \frac{N_h}{n_h} \sum_{i=1}^{n_h} \frac{M_{hi}}{m_{hi}} \sum_{j=1}^{m_{hi}} y_{hij} \\ &= \frac{1}{n_h \bar{M}_h} \sum_{i=1}^{n_h} \frac{M_{hi}}{m_{hi}} \sum_{j=1}^{m_{hi}} y_{hij} \end{aligned}$$

และในการประมาณค่าความแปรปรวนของค่าประมาณค่าเฉลี่ยรายได้ จะปรับใช้ (2.17) และ (2.19) ดังนี้

$$V(\bar{y}_h) = \left( \frac{1}{n_h} - \frac{1}{N_h} \right) S_h^2 + \frac{1}{n_h N_h} \sum_{i=1}^{N_h} \frac{M_{hi}^2}{\bar{M}_h^2} \left( \frac{1}{m_{hi}} - \frac{1}{M_{hi}} \right) S_{hi}^2$$

เมื่อ

$$\begin{aligned} S_h^2 &= \frac{1}{N_h - 1} \sum_{i=1}^{N_h} \left( \frac{M_{hi}}{\bar{M}_h} \bar{Y}_{hi} - \bar{Y}_h \right)^2 && \text{คือความแปรปรวนจากชั้นภูมิที่ 1} \\ S_{hi}^2 &= \frac{1}{M_{hi} - 1} \sum_{j=1}^{M_{hi}} (y_{hij} - \bar{Y}_{hi})^2 && \text{คือความแปรปรวนจากชั้นภูมิที่ 2} \end{aligned}$$

### 3.4.2 ผลการศึกษาเชิงตัวเลขของวิธีการปรับฟังก์ชันภาวะความน่าจะเป็นกับข้อมูลรายได้เฉลี่ยต่อครัวเรือนของประเทศไทย

ในหัวข้อนี้เป็นการแสดงผลของการปรับฟังก์ชันภาวะความน่าจะเป็นกับข้อมูลรายได้ของประชากรไทย ซึ่งพิจารณาใน 2 กรณีคือ

1. กรณีความแปรปรวนจากการถดถอยไม่ขึ้นกับเวลา
2. กรณีความแปรปรวนจากการถดถอยขึ้นกับเวลา



### กรณีความแปรปรวนจากการถดถอยไม่ขึ้นกับเวลา

ผลแสดงดังตารางที่ 3.46 และตารางที่ 3.47 โดยตารางที่ 3.46 แสดงค่าตัวประมาณความแปรปรวนจากการถดถอย ( $\hat{A}$ ) จากข้อมูลรายได้ของประชากรไทย เมื่อสัมประสิทธิ์การถดถอยไม่ขึ้นกับเวลา และตารางที่ 3.47 แสดงค่าตัวประมาณความแปรปรวนจากการถดถอย ( $\hat{A}$ ) จากข้อมูลรายได้ของประชากรไทย เมื่อสัมประสิทธิ์การถดถอยขึ้นกับเวลา

ภูมิภาค	จำนวน จังหวัด	PML		APML	
		ในเขต	นอกเขต	ในเขต	นอกเขต
ภาคกลาง	18	0.2310	0.1967	0.2362	0.2013
ภาคตะวันออก	7	0.3015	0.1926	0.3203	0.2043
ภาคเหนือ	17	0.0556	0.0670	0.0569	0.0687
ภาคตะวันออกเฉียงเหนือ	19	0.1463	0.0725	0.1494	0.0741
ภาคใต้	14	0.1565	0.1296	0.1612	0.1334

ตารางที่ 3.46: ค่าประมาณความแปรปรวนจากการถดถอย  $\hat{A}$  กรณีความแปรปรวนจากการถดถอยไม่ขึ้นกับเวลา

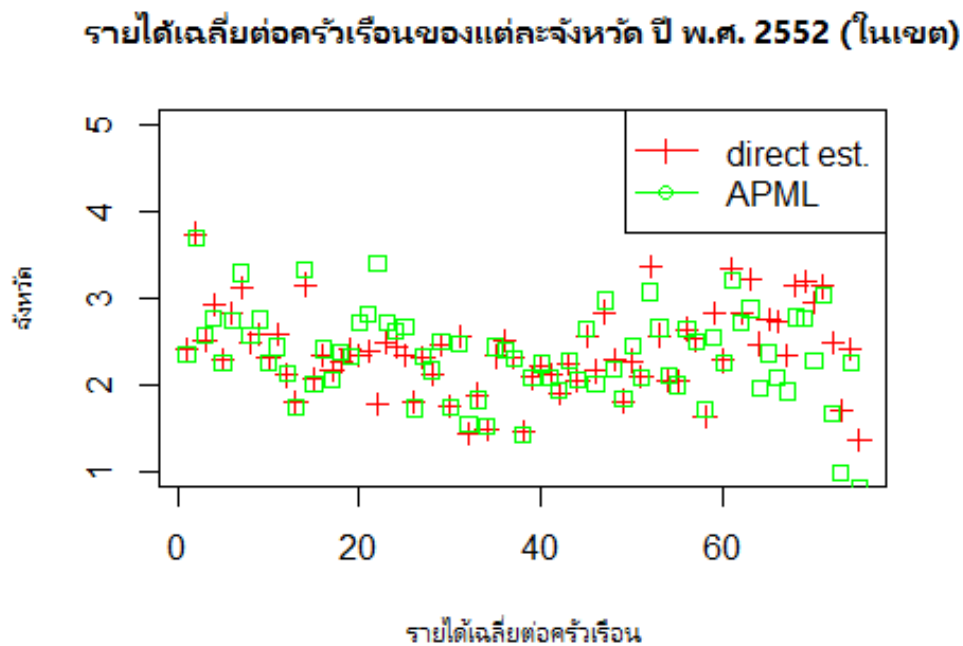
จากตารางที่ 3.46 จะเห็นว่า ไม่พบค่าตัวประมาณความแปรปรวนจากการถดถอย  $\hat{A} = 0$  เนื่องจากตัวอย่างมีขนาดใหญ่ ซึ่งจะให้ค่าตัวประมาณความแปรปรวนจากการถดถอย  $\hat{A}$  โดยวิธีตัวประมาณแบบภาวะความน่าจะเป็นสูงสุดโพรไฟล์ (PML) ที่ดี และเมื่อใช้ตัวประมาณแบบภาวะความน่าจะเป็นสูงสุดโพรไฟล์แบบปรับ (APML) จะได้ค่าประมาณความแปรปรวนจากการถดถอย  $\hat{A}$  ที่ไม่แตกต่างกันมาก

ภูมิภาค	จำนวน จังหวัด	PML		APML	
		ในเขต	นอกเขต	ในเขต	นอกเขต
ภาคกลาง	18	0.0346	0.0395	0.0356	0.0404
ภาคตะวันออก	7	$-5 \times 10^{-10}$	$-1 \times 10^{-10}$	$4 \times 10^{-8}$	$3 \times 10^{-9}$
ภาคเหนือ	17	$-1 \times 10^{-4}$	$-1 \times 10^{-4}$	0.0106	0.0057
ภาคตะวันออกเฉียงเหนือ	19	$-4 \times 10^{-4}$	$-4 \times 10^{-4}$	0.0424	0.0108
ภาคใต้	14	0.0282	0.0137	0.0290	0.0139

ตารางที่ 3.47: ค่าประมาณความแปรปรวนจากการถดถอย  $\hat{A}$  กรณีความแปรปรวนจากการถดถอยไม่ขึ้นกับเวลา

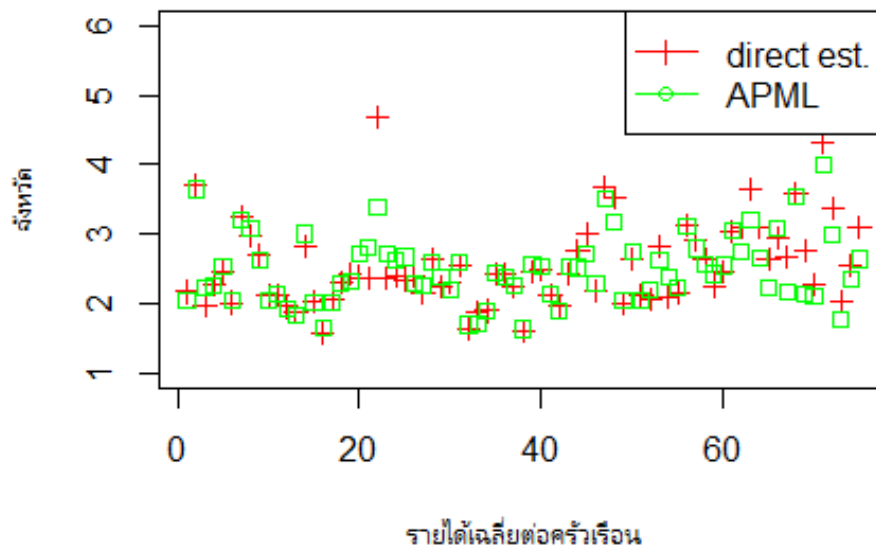
จากตารางที่ 3.47 จะเห็นว่าค่าตัวประมาณความแปรปรวนจากการถดถอย ( $\hat{A}$ ) จากแบบที่ยังไม่ปรับ บางค่าจะยังคงมีค่าน้อยกว่า 0 เช่น ตัวประมาณความแปรปรวนจากการถดถอยของภาคตะวันออก ในเขต มีค่า  $-5 \times 10^{-10}$  แต่จะไม่พบค่าตัวประมาณความแปรปรวนจากการถดถอย ( $\hat{A}$ ) จากแบบที่ปรับแล้ว

หลังจากคำนวณหาตัวประมาณความแปรปรวนจากการถดถอย ( $\hat{A}$ ) เมื่อสัมประสิทธิ์การถดถอยขึ้นกับเวลา ได้แล้ว เราจึงใช้ค่าดังกล่าวในการหาตัวประมาณที่ไม่เอนเอียงเชิงเส้นเชิงประจักษ์ที่ดีที่สุด ( $\theta_{it}^{EBLUP}$ ) กรณี APML และ เปรียบเทียบกับค่าเฉลี่ยรายได้ประชากรไทย ได้ดังรูปที่ 3.2 ถึง รูปที่ 3.11



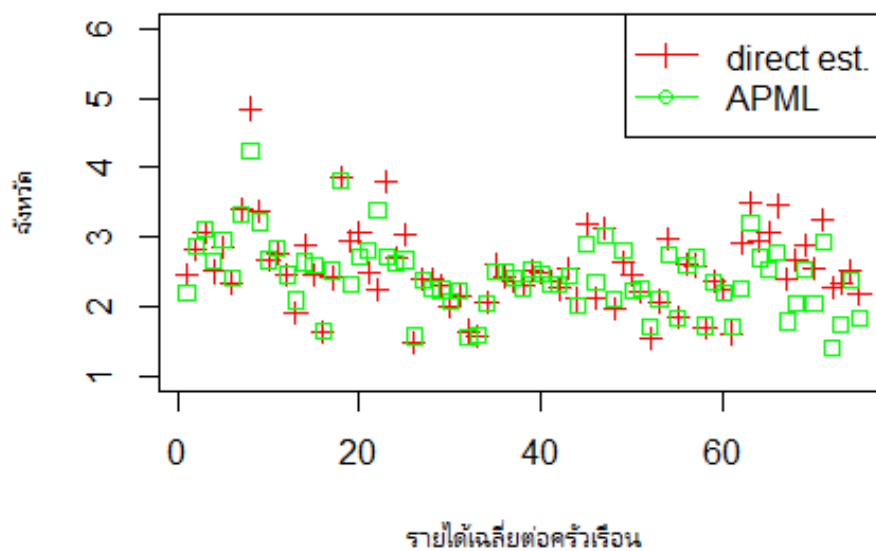
รูปที่ 3.2: รายได้เฉลี่ยต่อครัวเรือนของแต่ละจังหวัด ปี พ.ศ. 2552 (ในเขต)

รายได้เฉลี่ยต่อครัวเรือนของแต่ละจังหวัด ปี พ.ศ. 2554 (ในเขต)



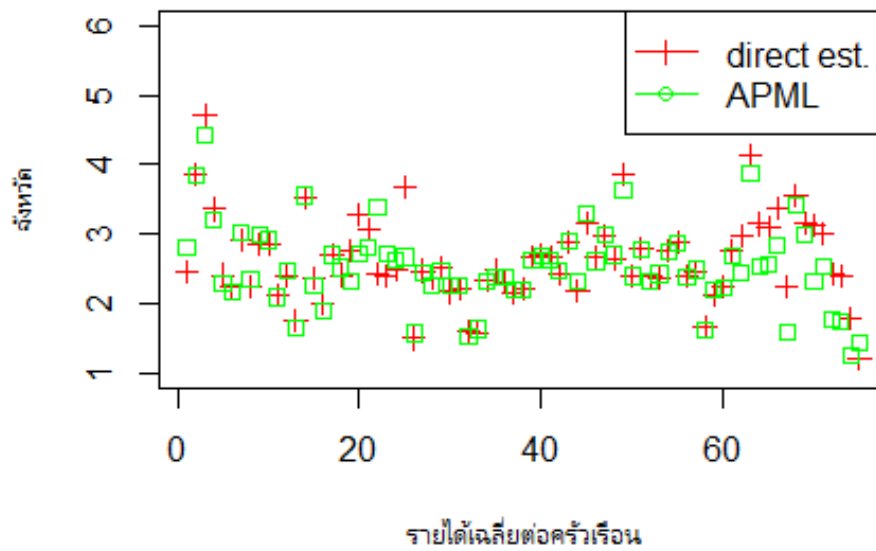
รูปที่ 3.3: รายได้เฉลี่ยต่อครัวเรือนของแต่ละจังหวัด ปี พ.ศ. 2554 (ในเขต)

รายได้เฉลี่ยต่อครัวเรือนของแต่ละจังหวัด ปี พ.ศ. 2556 (ในเขต)



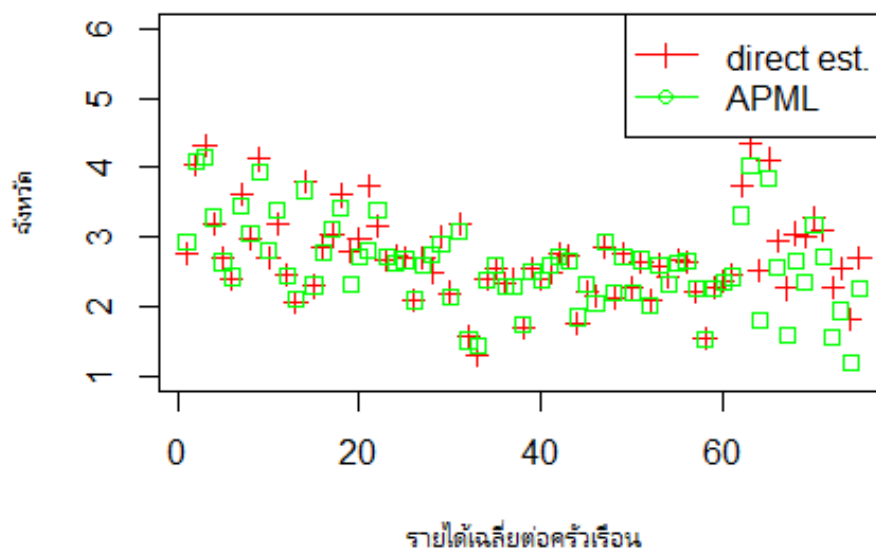
รูปที่ 3.4: รายได้เฉลี่ยต่อครัวเรือนของแต่ละจังหวัด ปี พ.ศ. 2556 (ในเขต)

รายได้เฉลี่ยต่อครัวเรือนของแต่ละจังหวัด ปี พ.ศ. 2558 (ในเขต)



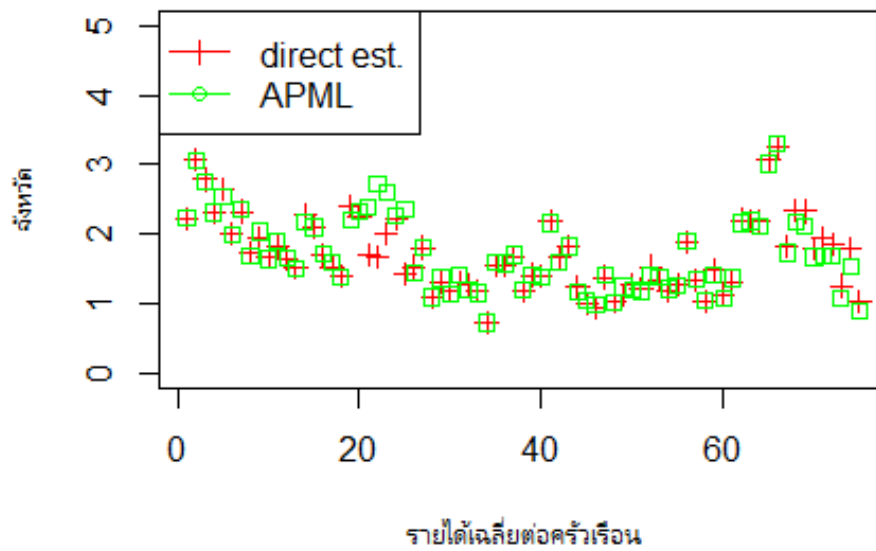
รูปที่ 3.5: รายได้เฉลี่ยต่อครัวเรือนของแต่ละจังหวัด ปี พ.ศ. 2558 (ในเขต)

รายได้เฉลี่ยต่อครัวเรือนของแต่ละจังหวัด ปี พ.ศ. 2560 (ในเขต)



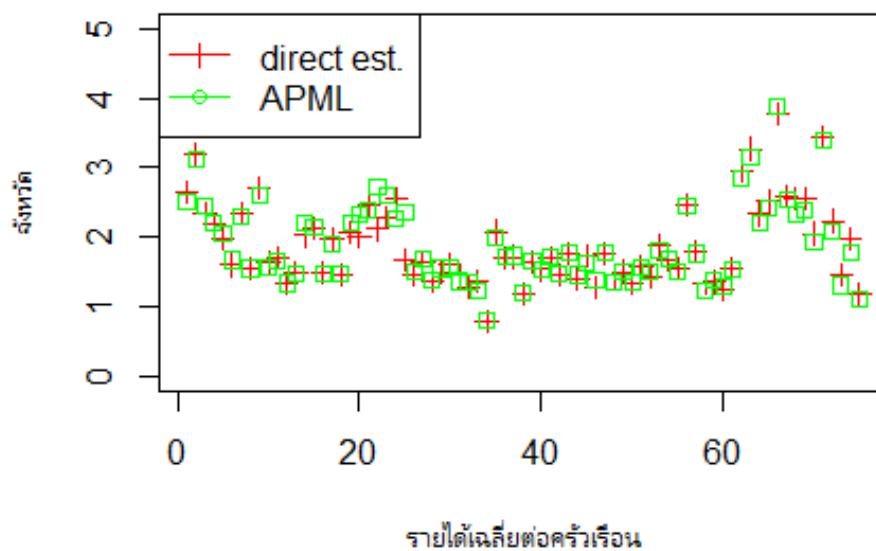
รูปที่ 3.6: รายได้เฉลี่ยต่อครัวเรือนของแต่ละจังหวัด ปี พ.ศ. 2560 (ในเขต)

รายได้เฉลี่ยต่อครัวเรือนของแต่ละจังหวัด ปี พ.ศ. 2552 (นอกเขต)



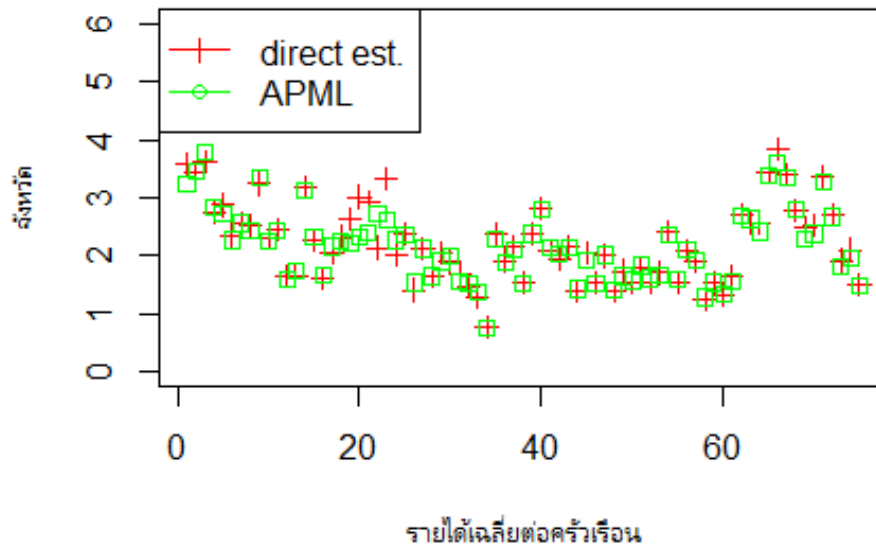
รูปที่ 3.7: รายได้เฉลี่ยต่อครัวเรือนของแต่ละจังหวัด ปี พ.ศ. 2552 (นอกเขต)

รายได้เฉลี่ยต่อครัวเรือนของแต่ละจังหวัด ปี พ.ศ. 2554 (นอกเขต)



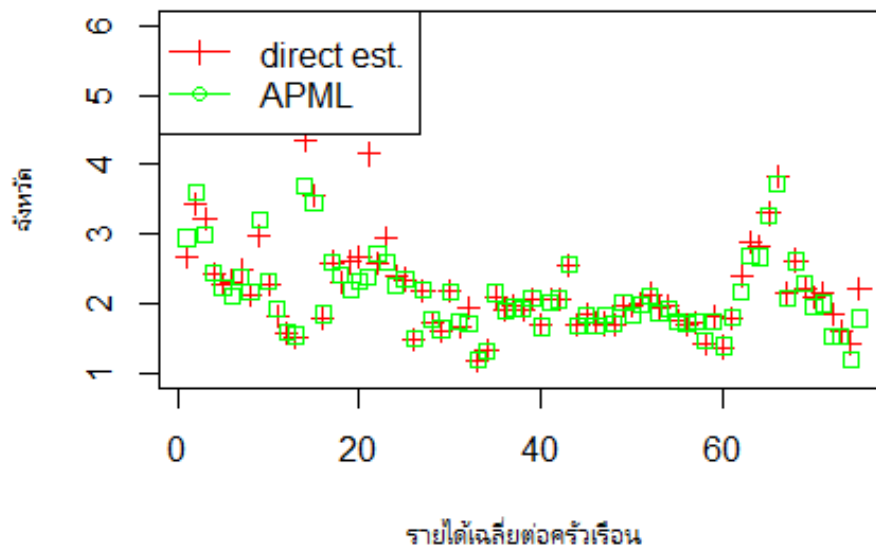
รูปที่ 3.8: รายได้เฉลี่ยต่อครัวเรือนของแต่ละจังหวัด ปี พ.ศ. 2554 (นอกเขต)

รายได้เฉลี่ยต่อครัวเรือนของแต่ละจังหวัด ปี พ.ศ. 2556 (นอกเขต)



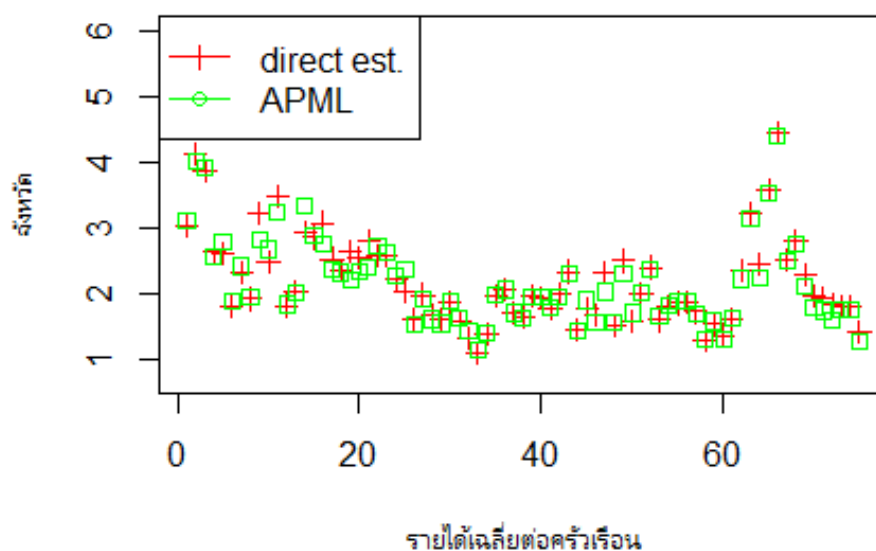
รูปที่ 3.9: รายได้เฉลี่ยต่อครัวเรือนของแต่ละจังหวัด ปี พ.ศ. 2556 (นอกเขต)

รายได้เฉลี่ยต่อครัวเรือนของแต่ละจังหวัด ปี พ.ศ. 2558 (นอกเขต)



รูปที่ 3.10: รายได้เฉลี่ยต่อครัวเรือนของแต่ละจังหวัด ปี พ.ศ. 2558 (นอกเขต)

### รายได้เฉลี่ยต่อครัวเรือนของแต่ละจังหวัด ปี พ.ศ. 2560 (นอกเขต)



รูปที่ 3.11: รายได้เฉลี่ยต่อครัวเรือนของแต่ละจังหวัด ปี พ.ศ. 2560 (นอกเขต)

จากรูปที่ 3.2 ถึง รูปที่ 3.11 จะเห็นว่า ตัวประมาณที่ไม่เอนเอียงเชิงเส้นเชิงประจักษ์ที่ดีที่สุด ( $\hat{\theta}_{it}^{EBLUP}$ ) กรณี APML มีแนวโน้มใกล้เคียงกับค่าประมาณจากการสุ่มตัวอย่าง (direct estimate) แต่ต่างกันเล็กน้อย ซึ่งเป็นผลเนื่องมาจากการปรับจากความคลาดเคลื่อนจากการสุ่มตัวอย่าง

### กรณีความแปรปรวนจากการถดถอยขึ้นกับเวลา

ผลแสดงดังตารางที่ 3.48 ถึงตารางที่ 3.53 โดยตารางที่ 3.48 แสดงค่าตัวประมาณความแปรปรวนจากการถดถอย ( $\hat{A}_t$ ) จากข้อมูลรายได้ของประชากรไทยจากแบบที่ยังไม่ปรับ เมื่อปัจจัยเกิดจากพื้นที่ ตารางที่ 3.49 แสดงค่าตัวประมาณความแปรปรวนจากการถดถอย ( $\hat{A}_t$ ) จากข้อมูลรายได้ของประชากรไทยจากแบบที่ปรับแล้ว เมื่อปัจจัยเกิดจากพื้นที่ ตารางที่ 3.50 แสดงค่าตัวประมาณความแปรปรวนจากการถดถอย ( $\hat{A}_t$ ) จากข้อมูลรายได้ของประชากรไทยจากแบบที่ยังไม่ปรับ (PML) เมื่อปัจจัยเกิดจากเวลา ตารางที่ 3.51 แสดงค่าตัวประมาณความแปรปรวนจากการถดถอย ( $\hat{A}_t$ ) จากข้อมูลรายได้ของประชากรไทยจากแบบที่ปรับแล้ว (APML) เมื่อปัจจัยเกิดจากเวลา ตารางที่ 3.52 แสดงค่าตัวประมาณความแปรปรวนจากการถดถอย ( $\hat{A}_t$ ) จากข้อมูลรายได้ของประชากรไทยจากแบบที่ยังไม่ปรับ (REML) เมื่อปัจจัยเกิดจากเวลา และตารางที่ 3.53 แสดงค่าตัวประมาณความแปรปรวนจากการถดถอย ( $\hat{A}_t$ ) จาก

ข้อมูลรายได้ของประชากรไทยจากแบบที่ปรับแล้ว (AREML) เมื่อปัจจัยเกิดจากเวลา

ภูมิภาค	เขตการปกครอง	จำนวนจังหวัด	$A_1$	$A_2$	$A_3$	$A_4$	$A_5$
ภาคกลาง	ในเขต	18	0.2078	-0.2306	1.5968	1.0490	1.0671
	นอกเขต		-0.0939	0.0399	0.7223	0.9557	1.7209
ภาคตะวันออก	ในเขต	7	0.6761	1.6681	-0.1224	0.8882	-0.4425
	นอกเขต		1.2161	0.9652	-0.1080	1.4550	-0.0802
ภาคเหนือ	ในเขต	17	-0.0972	0.8942	0.7213	0.0460	1.5667
	นอกเขต		0.0064	-0.0528	1.9105	0.8633	0.5000
ภาคตะวันออกเฉียงเหนือ	ในเขต	19	0.7828	1.6001	-0.2617	-0.2390	0.8844
	นอกเขต		1.8120	-0.1098	0.6613	1.0112	0.0545
ภาคใต้	ในเขต	14	-0.1819	0.8453	0.0960	0.8859	1.8445
	นอกเขต		-0.1100	0.8342	1.6403	0.6711	0.0441

ตารางที่ 3.48: ค่าประมาณความแปรปรวนจากการถดถอย  $\hat{A}_t$  กรณีไม่ปรับ (PML) กรณีความแปรปรวนจากการถดถอยขึ้นกับเวลา เมื่อปัจจัยเกิดจากพื้นที่

จากตารางที่ 3.48 จะเห็นว่าค่าตัวประมาณความแปรปรวนจากการถดถอย ( $\hat{A}_t$ ) จากแบบที่ยังไม่ปรับ บางค่าจะยังคงมีค่าน้อยกว่า 0 เช่น ตัวประมาณความแปรปรวนจากการถดถอยของภาคกลาง นอกเขต ( $\hat{A}_1$ ) มีค่า -0.0939



ภูมิภาค	เขตการปกครอง	จำนวนจังหวัด	$A_1$	$A_2$	$A_3$	$A_4$	$A_5$
ภาคกลาง	ในเขต	18	0.0095	0.0282	1.0741	0.0185	0.3046
	นอกเขต		0.0088	0.1371	0.3799	0.1374	1.9765
ภาคตะวันออก	ในเขต	7	0.0433	0.0617	0.2687	0.2514	2.1486
	นอกเขต		0.4366	0.0985	0.2323	2.3093	0.0272
ภาคเหนือ	ในเขต	17	0.1157	0.6899	0.0045	0.0062	1.7510
	นอกเขต		0.0896	0.0042	0.0979	1.9657	0.0264
ภาคตะวันออกเฉียงเหนือ	ในเขต	19	0.0307	2.5748	0.0208	0.0666	0.0969
	นอกเขต		2.5512	0.0124	0.0021	0.0696	0.0345
ภาคใต้	ในเขต	14	0.1123	1.4890	0.0664	0.0190	0.0533
	นอกเขต		0.0507	0.9104	1.5744	0.0098	0.0098

ตารางที่ 3.49: ค่าประมาณความแปรปรวนจากการถดถอย  $\hat{A}_t$  กรณีปรับ (APML) กรณีความแปรปรวนจากการถดถอยขึ้นกับเวลา เมื่อปัจจัยเกิดจากพื้นที่

จากตารางที่ 3.49 จะเห็นว่า ไม่พบค่าตัวประมาณความแปรปรวนจากการถดถอย  $\hat{A}_t = 0$

ภูมิภาค	เขตการปกครอง	จำนวนจังหวัด	$A_1$	$A_2$	$A_3$	$A_4$	$A_5$
ภาคกลาง	ในเขต	18	-0.3083	0.7166	1.2319	-0.0628	1.7013
	นอกเขต		0.9487	0.0354	-0.4117	0.7067	1.9299
ภาคตะวันออก	ในเขต	7	0.5949	1.6889	0.0002	0.8899	-0.3276
	นอกเขต		0.842	-0.2129	1.9340	-0.6540	0.3959
ภาคเหนือ	ในเขต	17	0.8122	-0.0388	0.0087	0.7028	1.5773
	นอกเขต		0.7916	0.6725	1.6441	-0.0773	-0.0214
ภาคตะวันออกเฉียงเหนือ	ในเขต	19	0.2246	1.5149	-0.0675	0.9936	0.4983
	นอกเขต		0.7308	0.8658	-0.0811	-0.0212	1.5978
ภาคใต้	ในเขต	14	1.0736	0.8739	-0.0083	-0.0526	1.6083
	นอกเขต		0.8040	-0.3296	1.6117	-0.1267	0.8707

ตารางที่ 3.50: ค่าประมาณความแปรปรวนจากการถดถอย  $\hat{A}_t$  กรณีไม่ปรับ (PML) กรณีความแปรปรวนจากการถดถอยขึ้นกับเวลา เมื่อปัจจัยเกิดจากเวลา

จากตารางที่ 3.50 จะเห็นว่าค่าตัวประมาณความแปรปรวนจากการถดถอย ( $\hat{A}_t$ ) จากแบบที่ยังไม่ปรับ บางค่าจะยังคงมีค่าน้อยกว่า 0 เช่น ตัวประมาณความแปรปรวนจากการถดถอยของภาคกลาง ในเขต ( $\hat{A}_1$ ) มีค่า -0.3083

ภูมิภาค	เขตการปกครอง	จำนวนจังหวัด	$A_1$	$A_2$	$A_3$	$A_4$	$A_5$
ภาคกลาง	ในเขต	18	0.0154	0.1046	1.6808	0.0647	1.3993
	นอกเขต		0.0856	0.0057	0.0699	1.7283	1.8341
ภาคตะวันออก	ในเขต	7	0.0285	0.0462	0.0787	0.1501	2.9041
	นอกเขต		0.1390	0.0384	0.4091	2.1077	0.0315
ภาคเหนือ	ในเขต	17	0.0998	0.3322	0.0054	0.0072	2.3811
	นอกเขต		0.0209	0.0036	2.6697	0.0196	0.0276
ภาคตะวันออกเฉียงเหนือ	ในเขต	19	0.0962	0.2190	2.2801	0.0659	0.0293
	นอกเขต		0.0072	1.0879	0.0359	0.0241	1.3529
ภาคใต้	ในเขต	14	0.1733	0.9539	0.0239	0.0240	0.0369
	นอกเขต		0.0072	0.0791	1.9350	0.0176	0.5889

ตารางที่ 3.51: ค่าประมาณความแปรปรวนจากการถดถอย  $\hat{A}_t$  กรณีปรับ (APML) กรณีความแปรปรวนจากการถดถอยขึ้นกับเวลา เมื่อปัจจัยเกิดจากเวลา

จากตารางที่ 3.51 จะเห็นว่า ไม่พบค่าตัวประมาณความแปรปรวนจากการถดถอย  $\hat{A}_t = 0$

ภูมิภาค	เขตการปกครอง	จำนวนจังหวัด	$A_1$	$A_2$	$A_3$	$A_4$	$A_5$
ภาคกลาง	ในเขต	18	-0.1208	0.6877	1.2877	-0.0486	1.6076
	นอกเขต		0.8909	0.2617	-0.0741	0.5944	1.8363
ภาคตะวันออก	ในเขต	7	0.1058	-0.1454	0.1787	0.4924	3.0321
	นอกเขต		0.8206	-0.2150	1.9114	-0.6476	0.4058
ภาคเหนือ	ในเขต	17	0.8064	-0.1196	0.7643	-0.0684	1.6154
	นอกเขต		0.9788	0.7346	1.4644	-0.0346	0.1859
ภาคตะวันออกเฉียงเหนือ	ในเขต	19	1.0297	0.7685	1.7790	-0.1874	0.1688
	นอกเขต		0.7034	0.8302	-0.2588	-0.0216	1.6468
ภาคใต้	ในเขต	14	1.4529	0.3880	-0.0317	-0.0780	1.4356
	นอกเขต		0.0015	0.3935	0.8657	2.0084	-0.0636

ตารางที่ 3.52: ค่าประมาณความแปรปรวนจากการถดถอย  $\hat{A}_t$  กรณีไม่ปรับ (REML) กรณีความแปรปรวนจากการถดถอยขึ้นกับเวลา เมื่อปัจจัยเกิดจากเวลา

จากตารางที่ 3.52 จะเห็นว่าค่าตัวประมาณความแปรปรวนจากการถดถอย ( $\hat{A}_t$ ) จากแบบที่ยังไม่ปรับ บางค่าจะยังคงมีค่าน้อยกว่า 0 เช่น ตัวประมาณความแปรปรวนจากการถดถอยของภาคกลาง ในเขต ( $\hat{A}_1$ ) มีค่า -0.1208

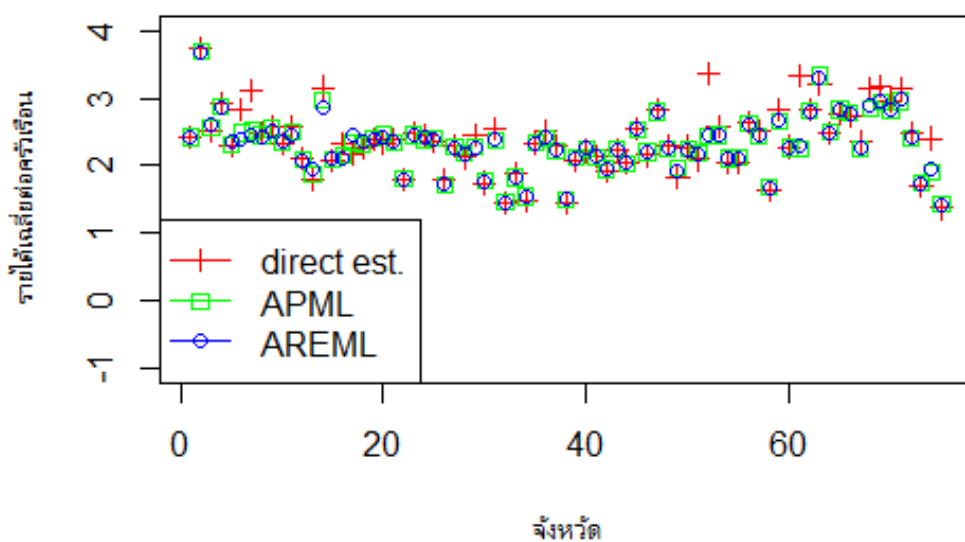
ภูมิภาค	เขตการปกครอง	จำนวนจังหวัด	$A_1$	$A_2$	$A_3$	$A_4$	$A_5$
ภาคกลาง	ในเขต	18	0.0451	0.1409	0.3362	0.0986	0.1502
	นอกเขต		0.1150	0.3520	0.1295	0.0954	0.1158
ภาคตะวันออก	ในเขต	7	0.1011	0.0462	0.0787	0.3680	2.9041
	นอกเขต		0.1498	0.0911	0.4192	2.1710	0.0948
ภาคเหนือ	ในเขต	17	0.0684	0.0455	0.1766	0.0410	0.0508
	นอกเขต		0.0371	0.0137	2.6075	0.0321	0.0376
ภาคตะวันออกเฉียงเหนือ	ในเขต	19	0.1195	0.2572	0.1688	0.0961	0.0550
	นอกเขต		0.0180	0.0131	0.0411	0.0392	2.9258
ภาคใต้	ในเขต	14	0.1694	2.9169	0.0552	0.0468	0.1245
	นอกเขต		0.0283	0.1648	0.0837	0.0454	0.0338

ตารางที่ 3.53: ค่าประมาณความแปรปรวนจากการถดถอย  $\hat{A}_t$  กรณีปรับ (AREML) กรณีความแปรปรวนจากการถดถอยขึ้นกับเวลา เมื่อปัจจัยเกิดจากเวลา

จากตารางที่ 3.53 จะเห็นว่า ไม่พบค่าตัวประมาณความแปรปรวนจากการถดถอย  $\hat{A}_t = 0$

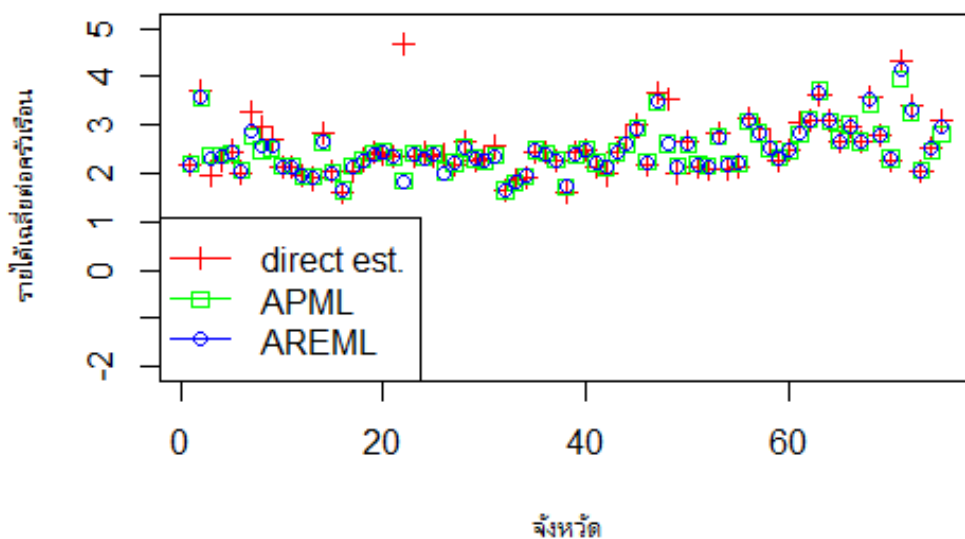
หลังจากคำนวณหาตัวประมาณความแปรปรวนจากการถดถอย ( $\hat{A}_t$ ) เมื่อปัจจัยเกิดจากเวลาได้แล้ว เราจึงใช้ค่าดังกล่าวในการหาตัวประมาณที่ไม่เอนเอียงเชิงเส้นเชิงประจักษ์ที่ดีที่สุด ( $\theta_{it}^{EBLUP}$ ) ทั้งกรณี APML และ AREML พร้อมทั้งเปรียบเทียบกับค่าเฉลี่ยรายได้ประชากรไทย ได้ดังรูปที่ 3.12 ถึง รูปที่ 3.21

รายได้เฉลี่ยต่อครัวเรือนของแต่ละจังหวัด ปี พ.ศ. 2552 (ในเขต)



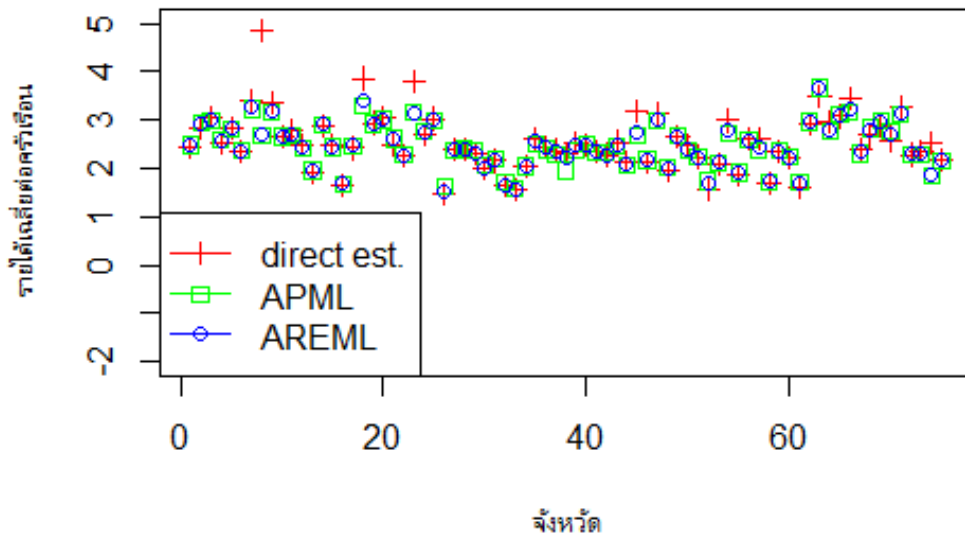
รูปที่ 3.12: รายได้เฉลี่ยต่อครัวเรือนของแต่ละจังหวัด ปี พ.ศ. 2552 (ในเขต)

รายได้เฉลี่ยต่อครัวเรือนของแต่ละจังหวัด ปี พ.ศ. 2554 (ในเขต)



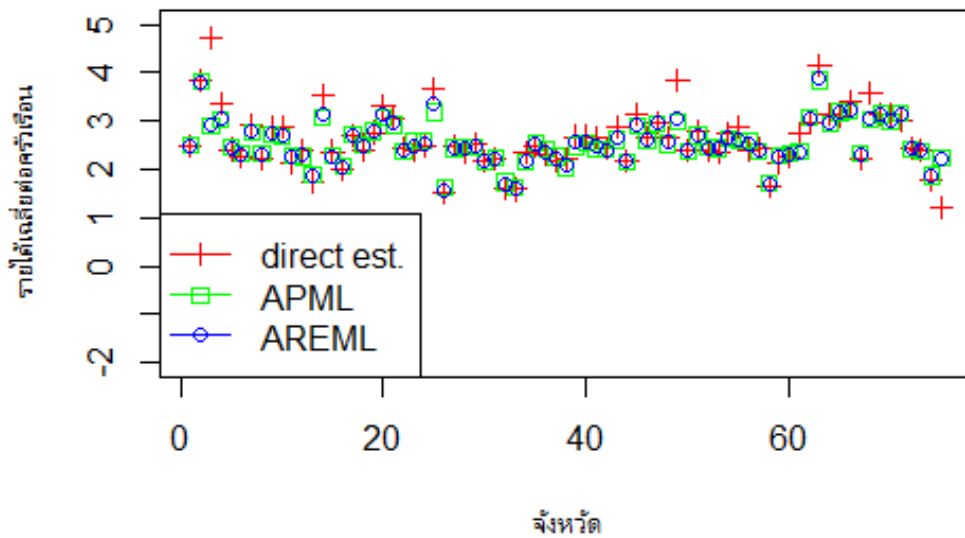
รูปที่ 3.13: รายได้เฉลี่ยต่อครัวเรือนของแต่ละจังหวัด ปี พ.ศ. 2554 (ในเขต)

รายได้เฉลี่ยต่อครัวเรือนของแต่ละจังหวัด ปี พ.ศ. 2556 (ในเขต)



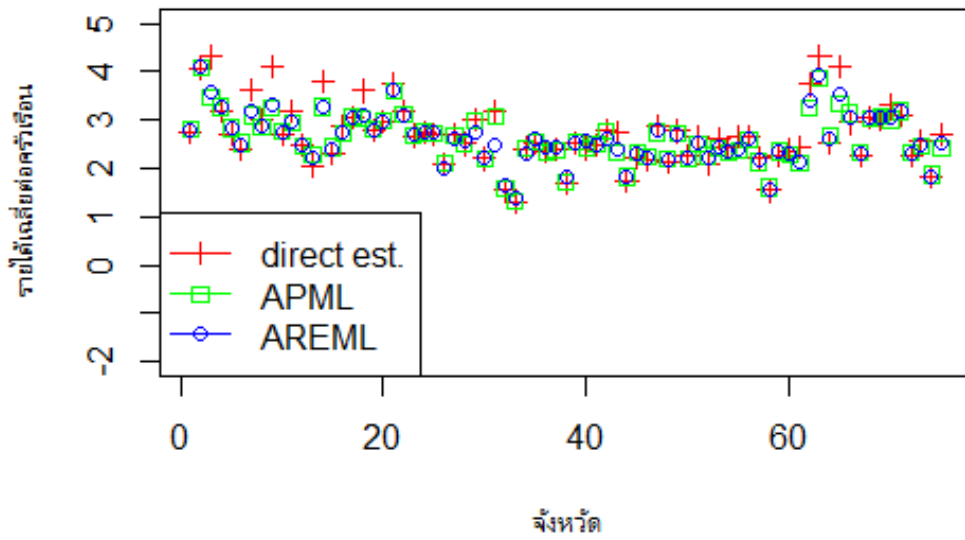
รูปที่ 3.14: รายได้เฉลี่ยต่อครัวเรือนของแต่ละจังหวัด ปี พ.ศ. 2556 (ในเขต)

รายได้เฉลี่ยต่อครัวเรือนของแต่ละจังหวัด ปี พ.ศ. 2558 (ในเขต)



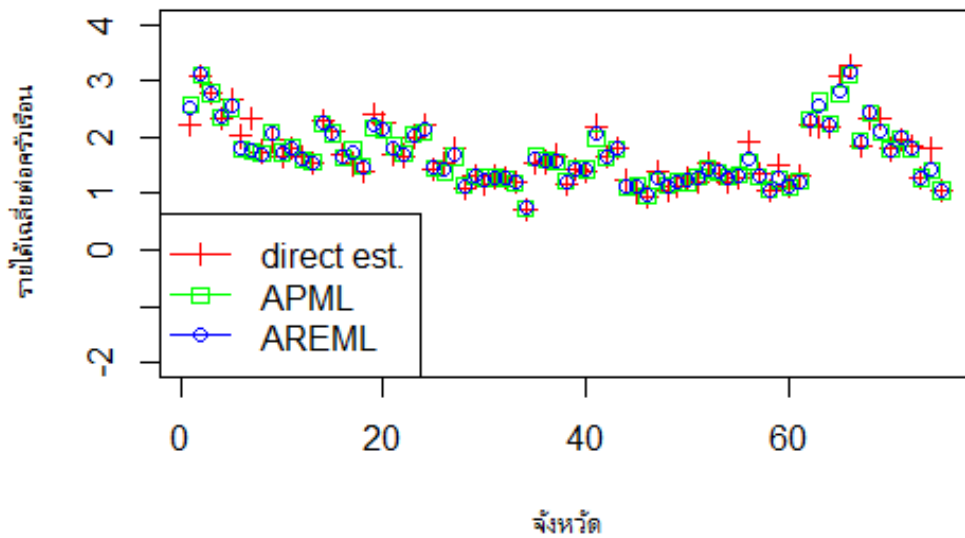
รูปที่ 3.15: รายได้เฉลี่ยต่อครัวเรือนของแต่ละจังหวัด ปี พ.ศ. 2558 (ในเขต)

รายได้เฉลี่ยต่อครัวเรือนของแต่ละจังหวัด ปี พ.ศ. 2560 (ในเขต)



รูปที่ 3.16: รายได้เฉลี่ยต่อครัวเรือนของแต่ละจังหวัด ปี พ.ศ. 2560 (ในเขต)

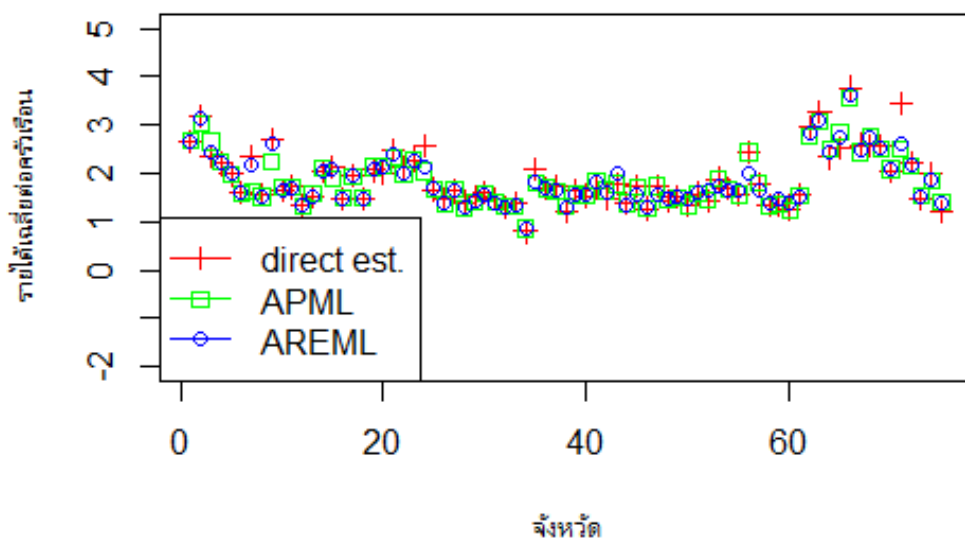
รายได้เฉลี่ยต่อครัวเรือนของแต่ละจังหวัด ปี พ.ศ. 2552 (นอกเขต)



รูปที่ 3.17: รายได้เฉลี่ยต่อครัวเรือนของแต่ละจังหวัด ปี พ.ศ. 2552 (นอกเขต)

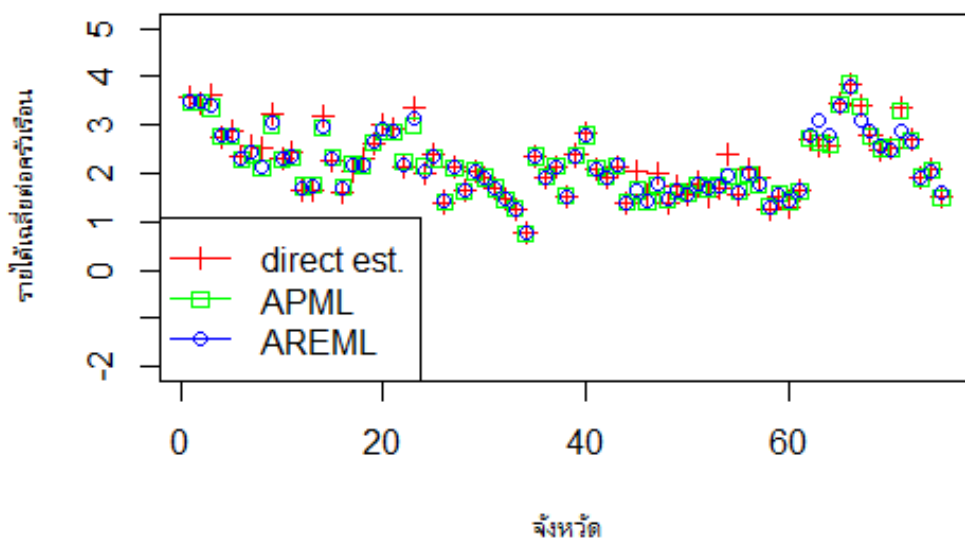


รายได้เฉลี่ยต่อครัวเรือนของแต่ละจังหวัด ปี พ.ศ. 2554 (นอกเขต)



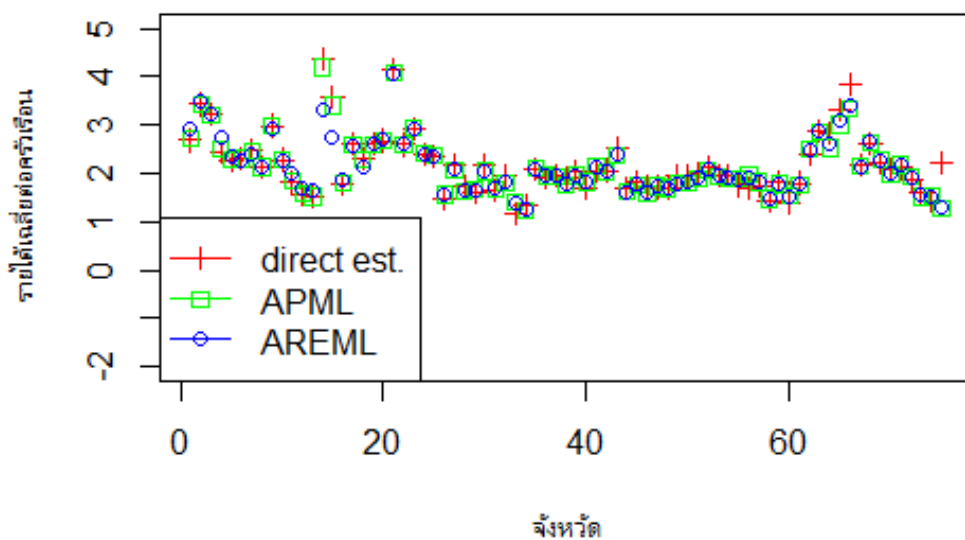
รูปที่ 3.18: รายได้เฉลี่ยต่อครัวเรือนของแต่ละจังหวัด ปี พ.ศ. 2554 (นอกเขต)

รายได้เฉลี่ยต่อครัวเรือนของแต่ละจังหวัด ปี พ.ศ. 2556 (นอกเขต)



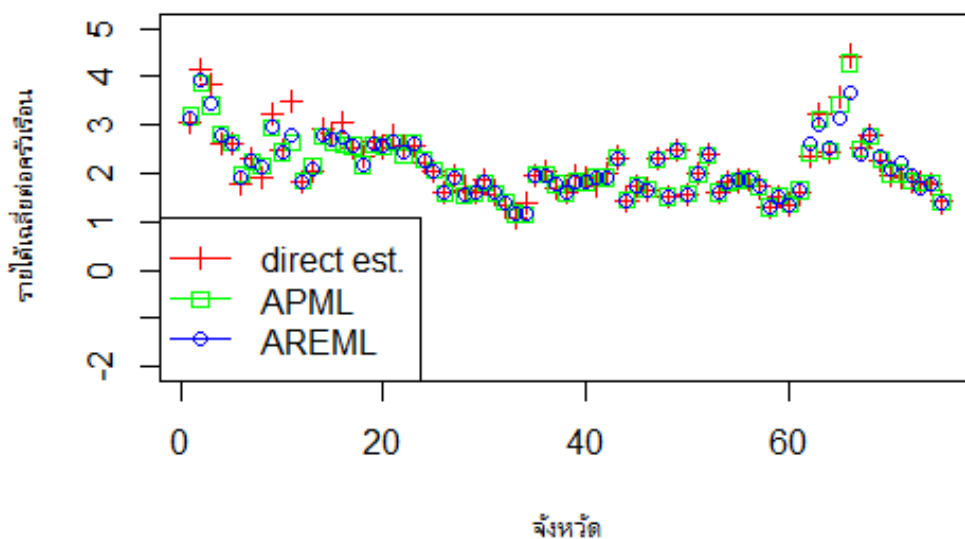
รูปที่ 3.19: รายได้เฉลี่ยต่อครัวเรือนของแต่ละจังหวัด ปี พ.ศ. 2556 (นอกเขต)

รายได้เฉลี่ยต่อครัวเรือนของแต่ละจังหวัด ปี พ.ศ. 2558 (นอกเขต)



รูปที่ 3.20: รายได้เฉลี่ยต่อครัวเรือนของแต่ละจังหวัด ปี พ.ศ. 2558 (นอกเขต)

รายได้เฉลี่ยต่อครัวเรือนของแต่ละจังหวัด ปี พ.ศ. 2560 (นอกเขต)



รูปที่ 3.21: รายได้เฉลี่ยต่อครัวเรือนของแต่ละจังหวัด ปี พ.ศ. 2560 (นอกเขต)

จากรูปที่ 3.12 ถึง รูปที่ 3.21 จะเห็นว่า ตัวประมาณที่ไม่เอนเอียงเชิงเส้นเชิงประจักษ์ที่ดีที่สุด ( $\hat{\theta}_{it}^{EBLUP}$ ) ทั้งกรณี APMLE และ AREML ให้ค่าที่ใกล้เคียงกันและมีแนวโน้มใกล้เคียงกับค่าประมาณจากการสุ่มตัวอย่าง (direct estimate) แต่ต่างกันเล็กน้อย ซึ่งเป็นผลเนื่องมาจากการปรับจากความคลาดเคลื่อนจากการสุ่มตัวอย่าง

### 3.5 สรุปผลการศึกษา

ในโครงการนี้ เราได้ศึกษาวิธีการปรับภาวะความน่าจะเป็นสูงสุด สำหรับแบบจำลองเฟ-เฮโรท และใช้แนวคิดนี้ขยายไปสู่แบบจำลองที่ทั่วไปกว่า คือแบบจำลองเฟ-เฮโรทแบบพานาลซึ่งเป็นแบบจำลองที่สามารถรองรับการวิเคราะห์ข้อมูลพร้อมกันหลายปี รวมถึงได้มีการสร้างฟังก์ชันภาวะความน่าจะเป็นแบบปรับ เพื่อศึกษาข้อมูลเชิงตัวเลข และจากผลการศึกษาชี้ว่าหลักการปรับแบบภาวะความน่าจะเป็นสูงสุดสามารถให้ค่าประมาณความแปรปรวนที่มีค่ามากกว่า 0 ได้ ซึ่งได้มีการพิสูจน์ไว้ในหัวข้อ 3.2 และหัวข้อ 3.3 อีกทั้งตัวประมาณที่ได้จากวิธีการปรับภาวะความน่าจะเป็นสูงสุดยังให้ค่าเฉลี่ยของกำลังสองของความคลาดเคลื่อนที่ต่ำกว่า ตัวประมาณที่ไม่ได้ปรับ สุดท้ายเราได้นำหลักการปรับภาวะความน่าจะเป็นสูงสุดไปประยุกต์ใช้กับข้อมูลรายได้เฉลี่ยต่อครัวเรือนของประเทศไทยในปี พ.ศ. 2552 - 2560 ซึ่งผลการศึกษาแสดงในหัวข้อ 3.4.2

## รายการอ้างอิง

- กลุ่มระเบียบวิธีสถิติ สำนักนโยบายและวิชาการสถิติ, เทคนิคการสุ่มตัวอย่างและการประมาณค่า, (กรุงเทพฯ), หน้า 35-54
- คณาจารย์ภาควิชาคณิตศาสตร์และวิทยาการคอมพิวเตอร์ คณะวิทยาศาสตร์ จุฬาลงกรณ์มหาวิทยาลัย, ความน่าจะเป็นและสถิติ, 13 (กรุงเทพฯ: โรงพิมพ์แห่งจุฬาลงกรณ์มหาวิทยาลัย, 2555), หน้า 93-94.
- Alvin C. Rencher., Linear Models in Statistics, 1 (New York: A Wiley-Interscience Publication, 2000), pp. 67-68,121-131.
- Angkunsit A. and Suntornchost J., “Bivariate Fay-Herriot Models With Application to Thai Socio-Economic Data,” Naresuan University Journal 29 (2021): 36-49
- J.N.K. Rao and Isabel Monila., SMALL AREA ESTIMATION, 2 (United States of America: A Wiley-Interscience Publication, 2015), pp. 98-99,118-120.
- J.N.K. Rao and Yu M., “Small-area estimation by combining time-series and cross-sectional data,” Canadian Journal of Statistics 22, 4 (December 1994): 511-528
- Li H., “SMALL AREA ESTIMATION: AN EMPIRICAL BEST LINEAR UNBIASED PREDICTION APPROACH”, (Doctoral dissertation, Graduate School, University of Maryland, 2007) P.19
- Li H. and Lahiri P., “An adjusted maximum likelihood method for solving small area estimation problems,” Journal of Multivariate Analysis 101 (2010): 882-892
- Yoshimori M. and Lahiri P., “A new adjusted maximum likelihood method for the Fay–Herriot small area model,” Journal of Multivariate Analysis 124 (2014): 281-294

ภาคผนวก

# แบบเสนอหัวข้อโครงการ รายวิชา 2301399 Project Proposal

## ปีการศึกษา 2563

ชื่อโครงการ (ภาษาไทย) การประมาณค่าพารามิเตอร์แบบปรับของภาวะความน่าจะเป็นสำหรับ  
แบบจำลองพหุเนล

ชื่อโครงการ (ภาษาอังกฤษ) Adjusted Likelihood Parameter Estimation for Panel Data Models

อาจารย์ที่ปรึกษา ผู้ช่วยศาสตราจารย์ ดร. จิราพรรณ สุนทรโชติ

ผู้ดำเนินการ พิชรสิริ เพ็ญฟู เลขประจำตัวนิสิต 6033529923

สาขาวิชาคณิตศาสตร์ ภาควิชาคณิตศาสตร์และวิทยาการคอมพิวเตอร์

คณะวิทยาศาสตร์ จุฬาลงกรณ์มหาวิทยาลัย

### หลักการและเหตุผล

Small area estimation (SAE) หรือ การประมาณค่าสำหรับพื้นที่ขนาดเล็ก เป็นหนึ่งในวิธีการทางสถิติที่เป็นที่รู้จัก สำหรับการประมาณพารามิเตอร์ เมื่อขนาดของตัวอย่างไม่เพียงพอที่จะให้ตัวประมาณค่าที่ได้มีความน่าเชื่อถือ หลักการของ small area estimation คือ การยืมข้อมูลจากแหล่งข้อมูลที่มีความเกี่ยวข้องกันผ่านแบบจำลอง โดยแบบจำลองที่ใช้อย่างกว้างขวางคือแบบจำลองเฟ-เฮรอต ซึ่งเป็นแบบจำลองที่นิยามผ่าน 2 แบบจำลอง คือ แบบจำลองเชื่อมโยง (Linking Model) และแบบจำลองการสุ่มตัวอย่าง (Sampling Model) ดังนี้

แบบจำลองเชื่อมโยง:  $\theta_i = x_i^T \beta + v_i$

แบบจำลองการสุ่มตัวอย่าง:  $y_i = \theta_i + e_i$

สำหรับ  $i = 1, 2, 3, \dots, n$

เมื่อ  $n$  คือ ขนาดตัวอย่าง

$\theta_i$  คือ ค่าเฉลี่ยสำหรับพื้นที่  $i$

$v_i$  คือ ความคลาดเคลื่อนจากการถดถอย ซึ่งสมมติให้มีการแจกแจงปกติ ที่มีค่าเฉลี่ยเท่ากับ 0 และความแปรปรวนเท่ากับ  $A$

$y_i$  คือ ค่าประมาณของ  $\theta_i$  ที่ได้จากการสำรวจ

$e_i$  คือ ความคลาดเคลื่อนที่เกิดจากการสุ่มตัวอย่าง ซึ่งสมมติให้มีการแจกแจงปกติ ที่มีค่าเฉลี่ยเท่ากับ 0 และความแปรปรวน  $\psi_i$

การประมาณตัวแปรเป้าหมายในแบบจำลองเฟ-เฮรอต คือ ตัวประมาณที่ไม่เอนเอียงเชิงเส้นเชิงประจักษ์ที่ดีที่สุด (Empirical best linear unbiased predictor: EBLUP)

$$\hat{\theta}_i^{EBLUP} = \frac{\hat{A}}{\hat{A} + \psi_i} y_i + \frac{\psi_i}{\hat{A} + \psi_i} x_i^T \beta$$

ซึ่งเป็นผลบวกแบบถ่วงน้ำหนักระหว่างค่าที่ได้จากการสำรวจ ( $y_i$ ) และ ค่าที่ได้จากสมการถดถอย ( $x_i^T \beta$ ) โดยน้ำหนักจะขึ้นกับความแปรปรวนจากการถดถอย ( $\hat{A}$ ) จะเห็นว่ากรณีที่มีความแปรปรวนจากการถดถอยมีค่าเท่ากับ 0 จะทำให้  $\hat{\theta}_i^{EBLUP} = x_i^T \beta$  ซึ่งเป็นตัวประมาณที่ได้มาจากการถดถอยเพียงส่วนเดียว ทำให้ตัวประมาณที่ได้ไม่น่าเชื่อถือ เราจึงต้องการหลีกเลี่ยงกรณีความแปรปรวนจากการถดถอยมีค่าเท่ากับ 0 ซึ่งในปี ค.ศ. 2010 Li และ Lahiri ได้เสนอแนวคิดการปรับภาวะความน่าจะเป็นสูงสุด (Adjusted maximum likelihood: ADM) เพื่อไม่ให้เกิดกรณีดังกล่าว นอกจากนี้ต่อมาในปี ค.ศ. 2014 Yoshimori และ Lahiri ได้มีการปรับปรุงวิธีการของ Li และ Lahiri และปี ค.ศ. 2020 Angkunsit และ Suntornchost ได้มีการนำแนวคิดของ Li และ Lahiri ไปปรับสร้างการประมาณค่าความแปรปรวนสำหรับแบบจำลองหลายตัวแปร

ข้อมูลพานาล เป็นรูปแบบข้อมูลที่มีการบันทึกข้อมูลของหน่วยสำรวจหลายหน่วย โดยบันทึกเป็นค่าตัวแปรที่สนใจจากหลายหน่วยเวลา ทำให้สามารถศึกษาแนวโน้มระหว่างเวลาได้ ซึ่งในปี ค.ศ. 1994 Rao และ Yu ได้ทำการศึกษาเกี่ยวกับข้อมูลพานาล และการประมาณ empirical best linear unbiased predictor (EBLUP) ของข้อมูลพานาลไว้แล้ว ซึ่งข้อมูลพานาลเป็นลักษณะของข้อมูลที่พบในหลายสาขา เช่น ข้อมูลทางด้านเศรษฐศาสตร์

ดังนั้น ในโครงการนี้ เราจะขยายหลักการของ Li และ Lahiri ในการปรับตัวประมาณของความแปรปรวนของแบบจำลองถดถอยให้เป็นแบบจำลองพานาล และจะใช้แบบจำลองพานาลในการศึกษาแนวโน้มระหว่างปีของรายได้ของประชากรไทยตั้งแต่ปี พ.ศ. 2552 - พ.ศ. 2560

## วัตถุประสงค์

เพื่อประยุกต์หลักการการปรับของวิธภาวะน่าจะเป็นสูงสุดสำหรับแบบจำลองพานาล กับข้อมูลรายได้ของประชากรไทย ตั้งแต่ปี พ.ศ. 2552 ถึงปี พ.ศ. 2560

## ขอบเขตของโครงการ

ข้อมูลที่ใช้ในการศึกษาในโครงการนี้ คือข้อมูลรายได้ของประชากรไทยตั้งแต่ปี พ.ศ. 2552 ถึงปี พ.ศ. 2560 โดยเป็นข้อมูลที่มีการเก็บทุก 2 ปี

## วิธีการดำเนินงาน

ก. แผนการศึกษา :

1. ศึกษาความรู้พื้นฐานเกี่ยวกับการถดถอยเชิงเส้น (Linear Regression)
2. ศึกษาความรู้เกี่ยวกับตัวแบบเฟ-เฮร็อต (Fay-Herriot models)
3. ศึกษาความรู้เกี่ยวกับตัวแบบพาเนล (Panel models)
4. ศึกษาการใช้โปรแกรม  $R$
5. ศึกษาข้อมูลรายได้ของประชากรไทยตั้งแต่ ปี พ.ศ. 2552 ถึงปี พ.ศ. 2560
6. วิเคราะห์ความสัมพันธ์ของข้อมูล
7. สรุปผลและเขียนรายงาน

ข. ระยะเวลาที่ศึกษา :

ขั้นตอนการดำเนินงาน	กันยายน พ.ศ. 2563 – เมษายน พ.ศ. 2564							
	ก.ย.	ต.ค.	พ.ย.	ธ.ค.	ม.ค.	ก.พ.	มี.ค.	เม.ย.
1. ศึกษาความรู้พื้นฐานเกี่ยวกับการถดถอยเชิงเส้น								
2. ศึกษาความรู้เกี่ยวกับตัวแบบเฟ-เฮร็อต								
3. ศึกษาความรู้เกี่ยวกับตัวแบบพาเนล								
4. ศึกษาการใช้โปรแกรม $R$								
5. ศึกษาข้อมูลรายได้ของประชากรไทยตั้งแต่ปี พ.ศ. 2552 ถึงปี พ.ศ. 2560								
6. วิเคราะห์ความสัมพันธ์ของข้อมูล								
7. สรุปผลและเขียนรายงาน								

## ประโยชน์ที่คาดว่าจะได้รับ

ก. ในด้านความรู้และประสบการณ์ต่อตัวนิสิตเอง

1. มีความรู้เกี่ยวกับการถดถอยเชิงเส้น
2. มีความรู้เกี่ยวกับแบบจำลองเฟ-เฮร็อต
3. มีความรู้เกี่ยวกับแบบจำลองพาเนล
4. มีความรู้ ความสามารถในการใช้โปรแกรม  $R$

ข. ความรู้ ความเข้าใจที่นำไปสู่การแก้ไขปัญหาของสังคมหรือสภาพแวดล้อมสามารถคาดการณ์แนวโน้มรายได้ของประชากรไทยได้



## อุปกรณ์และเครื่องมือที่ใช้

ฮาร์ดแวร์

1. เครื่องคอมพิวเตอร์
2. เครื่องพิมพ์

ซอฟต์แวร์

1. โปรแกรม Microsoft Word
2. โปรแกรม Microsoft PowerPoint
3. โปรแกรม Microsoft Excel

โปรแกรม *R*

## งบประมาณ

ค่าวัสดุ

- |                  |       |     |
|------------------|-------|-----|
| - หนังสือ        | 2,500 | บาท |
| - ฮาร์ดดิสก์พกพา | 1,500 | บาท |
| - แฟลชไดรฟ์      | 300   | บาท |
| - กระดาษ A4      | 200   | บาท |
| - แฟ้มเอกสาร     | 200   | บาท |

## เอกสารอ้างอิง

1. Angkunsit A. and Suntornchost J., “Bivariate Fay-Herriot Models With Application to Thai Socio-Economic Data,” Naresuan University Journal 29 (2021): 36-49
2. Li H. and Lahiri P., “An adjusted maximum likelihood method for solving small area estimation problems,” Journal of Multivariate Analysis 101 (2010): 882-892
3. Rao J.N.K. and Yu M., “Small-area estimation by combining time-series and cross-sectional data,” Canadian Journal of Statistics 22,4 (December 1994): 511-528
4. Yoshimori M. and Lahiri H., “A new adjusted maximum likelihood method for the Fay-Herriot small area model,” Journal of Multivariate Analysis 124 (2014): 281-294

## ประวัติผู้เขียน



ชื่อ-สกุล นางสาวพัชรสิริ เฟื่องฟู

วัน เดือน ปีเกิด 23 ตุลาคม พ.ศ. 2541

สถานที่อยู่ปัจจุบัน 126 หมู่ 20 ตำบลรอบเวียง อำเภอเมือง  
เชียงราย จังหวัดเชียงราย 57000

ประวัติการศึกษา

2560 - ปัจจุบัน : จุฬาลงกรณ์มหาวิทยาลัย คณะวิทยาศาสตร์  
ภาควิชาคณิตศาสตร์และวิทยาการคอมพิวเตอร์ สาขา  
คณิตศาสตร์

2557 - 2559 : โรงเรียนสามัคคีวิทยาคม จังหวัดเชียงราย