ใบรับรองวิทยานิพนธ์

บัณฑิตวิทยาลัย มหาวิทยาลัยเกษตรศาสตร์

วิทยาศาสตรมหาบัณฑิต (สถิติ)

ปริญญา

สถิติ	สถิติ
สาขา	**ภาควิชา**

เรื่อง การประมาณค่าพารามิเตอร์ในตัวแบบการถดถอยเชิงเส้นอย่างง่าย เมื่อวัดค่าตัวแปรอิสระโดยมีความคลาดเคลื่อน

Estimation of Parameters in Simple Linear Regression when Independent Variable is Measured with Error

นามผู้วิจัย นางสาวศุภลักษณ์ ตัสโต

ได้พิจารณาเห็นชอบโดย

ประธานกรรมการ ....

(รองศาสตราจารย์เปรมใจ ตรีสรานุวัฒนา, M.Stat.)

กรรมการ ....

(ผู้ช่วยศาสตราจารย์ลี่ลี อิงศรีสว่าง, Ph.D.)

กรรมการ ....

(ผู้ช่วยศาสตราจารย์ปนัดดา อินทร์พรหม, พบ.ม.)

หัวหน้าภาควิชา ....

(รองศาสตราจารย์สายสุดา สมชิต, M.S.)

บัณฑิตวิทยาลัย มหาวิทยาลัยเกษตรศาสตร์รับรองแล้ว

....

(รองศาสตราจารย์วินัย อาจคงหาญ, M.A.)

คณบดีบัณฑิตวิทยาลัย

วันที่ 7 **เดือน** พฤศจิกายน **พ.ศ.** 2548

วิทยานิพนธ์

เรื่อง

การประมาณค่าพารามิเตอร์ในตัวแบบการถดถอยเชิงเส้นอย่างง่าย เมื่อวัดค่าตัวแปรอิสระโดยมีความคลาดเคลื่อน

Estimation of Parameters in Simple Linear Regression when Independent Variable is Measured with Error

โดย

นางสาวศุภลักษณ์ ตัสโต

เสนอ

บัณฑิตวิทยาลัย มหาวิทยาลัยเกษตรศาสตร์
เพื่อความสมบูรณ์แห่งปริญญาวิทยาศาสตรมหาบัณฑิต (สถิติ)
พ.ศ. 2548

ISBN 974-9846-64-8

ศุภลักษณ์ ตัสโต 2548: การประมาณค่าพารามิเตอร์ในตัวแบบการถดถอยเชิงเส้นอย่างง่ายเมื่อวัดค่าตัวแปรอิสระโดยมีความคลาดเคลื่อน ปริญญาวิทยาศาสตรมหาบัณฑิต (สถิติ) สาขาสถิติ ภาควิชาสถิติ ประธานกรรมการที่ปรึกษา: รองศาสตราจารย์เปรมใจ ตรีสรานุวัฒนา, M.Stat. 80 หน้า
ISBN 974-9846-64-8

การวิจัยครั้งนี้มีวัตถุประสงค์เพื่อศึกษาเปรียบเทียบวิธีการประมาณค่าพารามิเตอร์ในตัวแบบการถดถอยเชิงเส้นอย่างง่าย เมื่อตัวแปรอิสระถูกวัดมาโดยมีความคลาดเคลื่อนด้วยวิธีการประมาณค่า 4 วิธี ได้แก่ วิธีกำลังสองน้อยที่สุด วิธีปรับตัวประมาณกำลังสองน้อยที่สุด วิธี Orthogonal Regression และวิธี Instrumental Variables แบ่งการศึกษาออกเป็น 2 ส่วน คือ ส่วนที่ 1 เปรียบเทียบประสิทธิภาพของวิธีการประมาณค่าพารามิเตอร์ 4 วิธี โดยใช้เกณฑ์จากค่าความเอนเอียงของตัวประมาณค่าพารามิเตอร์ และส่วนที่ 2 เปรียบเทียบประสิทธิภาพของวิธีการประมาณค่าพารามิเตอร์ 4 วิธี โดยใช้เกณฑ์จากค่าความแปรปรวนของค่าประมาณพารามิเตอร์ กำหนดการแจกแจงของตัวแปรอิสระ คือ การแจกแจงแบบปกติ มีค่าเฉลี่ยเท่ากับ 0 ความแปรปรวนเท่ากับ 1 และการแจกแจงของค่าคลาดเคลื่อนที่ศึกษาคือการแจกแจงแบบปกติ มีค่าเฉลี่ยเท่ากับ 0 ความแปรปรวนเท่ากับ 5 ขนาดของกลุ่มตัวอย่างที่ทำการศึกษามี 3 ระดับ คือ กลุ่มตัวอย่างขนาดเล็ก ได้แก่ ขนาดตัวอย่างเท่ากับ 10, 20 และ 30 กลุ่มตัวอย่างขนาดกลาง ได้แก่ ขนาดตัวอย่างเท่ากับ 50 และ 70 และกลุ่มตัวอย่างขนาดใหญ่ ได้แก่ ขนาดตัวอย่างเท่ากับ 100, 200 และ 300 ข้อมูลที่ใช้ในการวิจัยได้จากการจำลองด้วยเทคนิคมอนติคาร์โล และทำการทดลองซ้ำๆ กัน 1,000 ครั้ง ในสถานการณ์ที่กำหนด ผลการศึกษาพบว่า เมื่อกลุ่มตัวอย่างมีขนาดเล็ก วิธีกำลังสองน้อยที่สุดให้ค่าความเอนเอียงและค่าความแปรปรวนต่ำสุด กลุ่มตัวอย่างขนาดกลางและกลุ่มตัวอย่างขนาดใหญ่ วิธี Orthogonal Regression ให้ค่าความเอนเอียงและค่าความแปรปรวนต่ำสุดและมีค่าใกล้เคียงกับวิธีอื่น สรุปได้ว่า เมื่อตัวอย่างมีขนาดเล็ก ($n < 30$) วิธีกำลังสองน้อยที่สุดเป็นวิธีการที่ดีที่สุด แต่เมื่อตัวอย่างมีขนาดใหญ่ ($n \geq 30$) ทุกวิธีให้ค่าความเอนเอียงและค่าความแปรปรวนใกล้เคียงกัน

ศุภลักษณ์ ตัสโต

ลายมือชื่อนิสิต

27 / ก.ค. / 48

ลายมือชื่อประธานกรรมการ

Supalak Tusto 2005: Estimation of Parameters in Simple Linear Regression when Independent Variable is Measured with Error. Master of Science (Statistics), Major Field: Statistics, Department of Statistics. Thesis Advisor: Associate Professor Premjai Trisaranuwatana, M.Stat. 80 pages.
ISBN 974-9846-64-8

The purpose of this research was to compare the methods for estimated parameters in simple linear regression when independent variable (X) is measured with error. The method of ordinary least square, adjusted least square, orthogonal regression and instrumental variables were studied. Criterias for best method were based on biased and variance of the estimators. In this study, X were considered as normally distribuld with mean 0 and variance, alse random error were normal distribution with mean 0 but variance equal to 5. The sample sizes investigated were divided into 3 groups, group of small sizes (10, 20 and 30), medium sizes (50 and 70) and large sizes (100, 200 and 300). Data were generated through Monte Carlo simulation techniques with 1000 replicates under each conditions. The result showed that ordinary least square gave minimum biased and variance for small sizes, orthogonal regression gave biased and variance of the estimated parameters as low as other methods for medium and large sizes. It was concluded that ordinary least square was best for small sizes ($n < 30$) , for medium and large sizes ($n \geq 30$) all methods gave biased and variance of the estimated parameters by the same amount.

Supalak Tusto — Student's signature

Premjai Trisarant — Thesis Advisor's signature 27 / ก.ค. / 48

กิตติกรรมประกาศ

วิทยานิพนธ์ฉบับนี้สำเร็จลุล่วงได้ด้วยดี ด้วยความอนุเคราะห์จากบุคคลหลายฝ่าย ผู้วิจัยขอกราบขอบพระคุณ รองศาสตราจารย์เปรมใจ ตรีสรานุวัฒนา ประธานกรรมการที่ปรึกษา ที่ได้ช่วยเหลือในการวางแผนงานวิจัยในวิทยานิพนธ์ฉบับนี้ ตลอดจนการให้คำปรึกษาแนะนำ และตรวจสอบแก้ไขข้อบกพร่องต่าง ๆ จนวิทยานิพนธ์ฉบับนี้สำเร็จลุล่วงได้ด้วยดี

ผู้วิจัยขอกราบขอบพระคุณ ผู้ช่วยศาสตราจารย์ ดร. ลี่ลี อิงศรีสว่าง กรรมการที่ปรึกษาวิชาเอก และผู้ช่วยศาสตราจารย์ปนัดดา อินทร์พรหม กรรมการที่ปรึกษาวิชารอง ที่กรุณาให้คำปรึกษาแนะนำ และช่วยเหลือในการทำวิทยานิพนธ์ฉบับนี้ให้สมบูรณ์ยิ่งขึ้น ทั้งนี้ผู้วิจัยขอกราบขอบพระคุณคณาจารย์ภาควิชาสถิติทุกท่าน ที่ให้โอกาสทางการศึกษาและประสิทธิประสาทความรู้ให้แก่ผู้วิจัยจนกระทั่งสำเร็จการศึกษา

ท้ายนี้ผู้วิจัยขอขอบพระคุณครอบครัว ที่คอยสนับสนุนการศึกษา และเพื่อน ๆ ทุกคนที่คอยให้ความช่วยเหลือและให้กำลังใจ สนับสนุนการทำวิทยานิพนธ์จนสำเร็จลุล่วงได้ด้วยดี

ศุภลักษณ์ ตัสโต

ตุลาคม 2548

สารบัญ

สารบัญตาราง

สารบัญภาพ

สารบัญภาพ (ต่อ)

การประมาณค่าพารามิเตอร์ในตัวแบบการถดถอยเชิงเส้นอย่างง่าย เมื่อวัดค่าตัวแปรอิสระโดยมีความคลาดเคลื่อน

Estimation of Parameters in Simple Linear Regression when Independent Variable is Measured with Error

คำนำ

การวิเคราะห์การถดถอยเชิงเส้นอย่างง่ายเป็นการศึกษาความสัมพันธ์ระหว่างตัวแปร 2 ตัว โดยตัวแปรหนึ่งเป็นตัวแปรอิสระและอีกตัวแปรหนึ่งเป็นตัวแปรตาม การวิเคราะห์การถดถอยเชิงเส้นอย่างง่ายเป็นวิธีทางสถิติที่ไม่ยุ่งยากในด้านการคำนวณและวิเคราะห์ข้อมูล และมีบทบาทในศาสตร์สาขาต่างๆ มากมาย เช่น ทางวิทยาศาสตร์ ทางการเกษตร ทางการแพทย์ ทางการบริหาร ทางด้านเศรษฐศาสตร์ เป็นต้น การวิเคราะห์การถดถอยเชิงเส้นอย่างง่ายจึงเป็นที่นิยมและยอมรับอย่างแพร่หลายในเกือบทุกสาขาวิชา โดยมีตัวแบบการถดถอยเชิงเส้นอย่างง่ายเป็น

$$Y_i = \beta_0 + \beta_1 X_i + \varepsilon_i \qquad ; i = 1,2,\ldots,n$$

โดยที่

Y_i เป็นค่าของตัวแปรตาม Y จากตัวอย่างหน่วยที่ i

β_0 เป็นค่าพารามิเตอร์แทนค่าเฉลี่ยของ Y เมื่อ X = 0

β_1 เป็นค่าพารามิเตอร์แทนอัตราการเปลี่ยนแปลงของค่าเฉลี่ย Y เมื่อ X เพิ่มขึ้นหนึ่งหน่วย

X_i เป็นค่าของตัวแปรอิสระจากตัวอย่างหน่วยที่ i

ε_i เป็นค่าความคลาดเคลื่อนที่เกิดขึ้นโดยสุ่มจากตัวอย่างหน่วยที่ i

และมีข้อสมมติเบื้องต้นในการวิเคราะห์การถดถอยเชิงเส้นอย่างง่ายว่า ε_i เป็นตัวแปรสุ่มที่มีค่าเฉลี่ยศูนย์ ค่าความแปรปรวนของความคลาดเคลื่อน (ε_i) คงที่เท่ากับ σ^2 และ ε_i, ε_j เป็นอิสระกัน ทั้งนี้ค่าของตัวแปรอิสระ X จะต้องวัดมาโดยไม่มีความคลาดเคลื่อนในการประมาณค่าพารามิเตอร์ β_0 และ β_1 จะใช้วิธีกำลังสองน้อยที่สุดซึ่งเป็นวิธีที่ให้ตัวประมาณที่มีคุณสมบัติที่ดี

คือเป็นตัวประมาณเชิงเส้นที่ไม่เอนเอียงและมีความแปรปรวนต่ำสุดในบรรดาตัวประมาณเชิงเส้นที่ไม่เอนเอียงทั้งหมด (Best Linear Unbiased Estimator : BLUE)

ในบางปัญหาพบว่า ค่าสังเกตหรือค่าที่เกิดจากการวัดของตัวแปรอิสระที่เก็บรวบรวมเพื่อนำมาวิเคราะห์การถดถอยอาจไม่สามารถวัดค่าได้โดยไม่มีความคลาดเคลื่อนตามที่ตัวแบบกำหนดไว้ ตัวอย่างเช่น

1. ข้อมูลที่เก็บรวบรวมไว้ในแหล่งต่างๆ ซึ่งเป็นข้อมูลทุติยภูมิ ส่วนใหญ่เก็บรวบรวมขึ้นจากการสำรวจด้วยตัวอย่าง ในกระบวนการสุ่มตัวอย่างนั้น แม้ว่าผู้เก็บรวบรวมข้อมูลจะควบคุมงานสำรวจได้ดีเพียงใด ความผิดพลาดก็สามารถเกิดขึ้นได้เสมอในหลายๆ ขั้นตอน เริ่มตั้งแต่ขั้นการวางแผนเตรียมการ ขั้นปฏิบัติงานภาคสนาม รวมถึงขั้นการประมวลผลและการนำเสนอผล ความผิดพลาดอาจเกิดขึ้น เนื่องจากผู้ให้ข้อมูลจงใจให้ข้อมูลผิดพลาด ผู้บันทึกข้อมูลบันทึกข้อมูลผิดพลาด แบบสำรวจขาดความถี่ถ้วนและไม่เจาะปัญหาให้ลึกลงไปในทุกด้าน ผู้ประมวลผลทำงานบกพร่อง ตลอดถึงการปัดเศษและปรับค่าข้อมูลให้สอดคล้องกัน จะเห็นได้ว่าความผิดพลาดนั้นย่อมปรากฏอยู่แล้วเป็นปกติ และโดยปรัชญาของการทำงานสำรวจ เป็นที่ยอมรับกันโดยทั่วไปแล้วว่าความผิดพลาดอาจมีอยู่ในระดับหนึ่ง ด้วยเหตุนี้ทั้งตัวแปรตาม Y และตัวแปรอิสระ X จึงรวมความผิดพลาดเหล่านี้ไว้ในตัวอยู่แล้ว

2. ในบางสถานการณ์ผู้วิจัยไม่สามารถใช้ตัวแปรที่เป็นปัจจัยที่แท้จริงที่มีผลต่อความเคลื่อนไหวของ Y มาใช้ได้ เพราะไม่สามารถวัดค่าได้โดยตรง ทำให้ต้องเลือกใช้ตัวแปรทดแทน (Proxy Variable หรือ Proxy Index) มาใช้แทน เช่น ระดับการศึกษา (X) ซึ่งถือว่าเป็นตัวแปรสำคัญที่สามารถอธิบายการเปลี่ยนแปลงของรายได้ (Y) แต่เนื่องจากการศึกษาเป็นตัวแปรที่วัดค่าได้ยาก เพราะการศึกษาเป็นกระบวนการที่ต่อเนื่องและถี่ถ้วนลึกซึ้งเกินกว่าที่จะประเมินค่าเป็นตัวเลขได้ ทำให้เราต้องใช้จำนวนปีที่ได้รับการศึกษาในสถาบันการศึกษามาใช้แทน ซึ่งก็พอจะนำมาใช้แทนได้แต่ก็ไม่อาจแทนได้ทั้งหมด ด้วยเหตุนี้เมื่อนำตัวแปรจำนวนปีที่ได้รับการศึกษาในสถาบันการศึกษามาใช้แทนตัวแปรการศึกษา จึงมีความผิดพลาดเกิดขึ้น

3. การใช้ตัวแปรเชิงคุณภาพแทนตัวแปรเชิงปริมาณ มักทำให้เกิดความคลาดเคลื่อนในข้อมูลเสมอ ทั้งนี้เพราะตัวแปรที่แท้จริงนั้นมีค่าต่อเนื่อง เมื่อจำเป็นต้องใช้ตัวแปรเพื่อแบ่งกลุ่มเป็นมาตรวัดแทนค่าตัวแปรที่มีค่าต่อเนื่อง ความผิดพลาดจึงปรากฏขึ้นอย่างแน่นอนและไม่อาจ

เลี่ยงได้ เช่น ให้ X_1 แทนรสนิยม เมื่อพิจารณา X_1 จะพบว่าค่าของ X_1 ค่อนข้างละเอียดเพราะเป็นตัวแปรที่ต้องอาศัยหลักเกณฑ์เชิงอัตตนิยมมาใช้ประเมินค่า เราไม่อาจสรุปได้ตรงกันว่านาย ก. มีรสนิยมดีเลวเพียงใด ทั้งนี้ขึ้นอยู่กับผู้พิจารณาว่ามีภูมิหลังแตกต่างกันเพียงใด ถ้าผู้พิจารณามีภูมิหลังต่างกันจะมองนาย ก. ต่างกัน เป็นต้น การกำหนดให้ X_1 เป็นตัวแปรเชิงคุณภาพ แสดงว่าเรายอมให้ X_1 มีค่าหยาบกว่าความเป็นจริง ความผิดพลาดในค่าของ X_1 จึงเกิดขึ้น

4. การใช้เลขดัชนีเป็นค่าของตัวแปรแทนที่จะใช้ค่าที่แท้จริงของตัวแปร มีผลให้เกิดความผิดพลาดได้ ทั้งนี้เพราะเลขดัชนีเป็นตัวเลขที่รวมความคลาดเคลื่อนไว้ในตัวเองมากอยู่แล้ว กล่าวคือ การคำนวณเลขดัชนีต้องอาศัยการกำหนดปีฐาน ซึ่งหากกำหนดปีฐานไม่ถูกต้องเลขดัชนีก็มีความบกพร่อง ซึ่งเลขดัชนีต้องอาศัยปริมาณ (Quantity), ราคา (Price) และอื่นๆ ซึ่งต้องอาศัยผลจากการสำรวจ ความผิดพลาดในข้อมูลเหล่านี้ย่อมเกิดขึ้น และเลขดัชนีเป็นค่าประมาณที่เรานิยมปัดเศษทิ้ง ความผิดพลาดของข้อมูลเนื่องจากการปัดเศษจึงปรากฏอยู่ และเมื่อนำเลขดัชนีมาใช้เป็นค่าของตัวแปรอิสระหรือตัวแปรตามตัวแปรเหล่านี้จึงรวมเอาความผิดพลาดไว้อย่างไม่อาจเลี่ยงได้

Berkson (1950) ได้ทำการศึกษาเกี่ยวกับการวัดค่าตัวแปรอิสระ X_i โดยมีความคลาดเคลื่อนเกิดขึ้น ซึ่งสามารถนำวิธีกำลังสองน้อยที่สุดมาประยุกต์ใช้ได้ โดยการกำหนดให้ค่าสังเกตจากตัวแปรอิสระที่วัดได้เป็น W_i ถ้าค่าจริงของตัวแปรอิสระคือ X_i ดังนั้น $W_i = X_i + U_i$; U_i เป็นค่าความคลาดเคลื่อนจากการวัด สำหรับตัวอย่างสถานการณ์ข้างต้นที่พบ ได้แก่ การไหลเวียนของกระแสไฟฟ้าในวงจรไฟฟ้า ถ้าให้การไหลเวียนของกระแสไฟฟ้าเป็นตัวแปรอิสระ ซึ่งในการทดลองต้องการกำหนดค่าหรือตั้งค่าการไหลเวียนของกระแสไฟฟ้าเป็น 100 แอมป์, 125 แอมป์, 150 แอมป์ และ 175 แอมป์ แต่ในทางปฏิบัติแล้วการไหลเวียนของกระแสไฟฟ้าที่วัดได้อาจจะไม่ตรงกับค่าเป้าหมายที่ตั้งไว้ ซึ่งปัญหาประเภทนี้มักจะพบในงานด้านวิศวกรรมศาสตร์และด้านวิทยาศาสตร์ ตัวแปรต่างๆ เช่น อุณหภูมิ ความดัน หรือระดับการไหลเวียน ที่กำหนดไว้ในการทดลองอาจจะมีความคลาดเคลื่อนเกิดขึ้นในเครื่องมือทำให้วัดค่าสังเกตของตัวแปรนั้นๆ ได้แตกต่างจากค่าที่ตั้งไว้

Fuller (1987) ได้ยกตัวอย่างสถานการณ์ซึ่งค่าของตัวแปรอิสระ X ไม่สามารถบอกค่าที่แท้จริงได้ โดยพิจารณาความสัมพันธ์ระหว่างผลผลิตของข้าวโพด (Y_i) และปริมาณไนโตรเจนในดิน (X_i) ค่าสัมประสิทธิ์การถดถอย β_1 คือจำนวนผลผลิตที่เพิ่มขึ้นเมื่อปริมาณไนโตรเจน

ในดินเพิ่มขึ้น 1 หน่วย ผู้วิจัยสามารถหาค่าประมาณของปริมาณไนโตรเจนในดินได้ โดยสุ่มตัวอย่างดินเพื่อนำตัวอย่างดินไปวิเคราะห์ในห้องทดลอง ผลจากการสุ่มตัวอย่างและการวิเคราะห์ผลอาจทำให้ไม่ได้ค่าสังเกต x_i ที่ถูกต้อง แต่จะได้ค่าประมาณของค่าสังเกต x_i ถ้าให้ปริมาณไนโตรเจนในดินที่วิเคราะห์ได้เป็น w_i ดังนั้น $w_i = x_i + u_i$ โดยที่ u_i เป็นค่าความคลาดเคลื่อนที่เกิดจากการวัดค่า x_i เป็นต้น

งานวิจัยนี้ สนใจวิธีการประมาณค่าพารามิเตอร์ในตัวแบบการถดถอยเชิงเส้นอย่างง่าย เมื่อการวัดค่าตัวแปรอิสระมีความคลาดเคลื่อนเกิดขึ้น โดยศึกษาเพื่อหาวิธีการที่เหมาะสมกว่าวิธีกำลังสองน้อยที่สุด เนื่องจากถ้าใช้วิธีกำลังสองน้อยที่สุดประมาณค่าพารามิเตอร์แล้วจะให้ค่าประมาณที่มีความเอนเอียง จากการศึกษางานวิจัยที่ผ่านมาพบว่าวิธี Orthogonal Regression และวิธี Instrumental Variables เป็นวิธีการที่มีการนำมาใช้แก้ปัญหาเมื่อการวัดค่าตัวแปรอิสระมีความคลาดเคลื่อนเกิดขึ้น โดยจะให้ค่าประมาณที่ไม่มีความเอนเอียงและมีความคงตัว ซึ่งเป็นคุณสมบัติของตัวประมาณที่ดี

ดังนั้น จึงสนใจศึกษาเปรียบเทียบวิธีการประมาณค่าพารามิเตอร์ในตัวแบบการถดถอยเชิงเส้นอย่างง่าย โดยวิธีกำลังสองน้อยที่สุด วิธีปรับตัวประมาณกำลังสองน้อยที่สุด วิธี Orthogonal Regression และวิธี Instrumental Variables โดยศึกษากับตัวแบบการถดถอยเชิงเส้นอย่างง่าย เมื่อข้อมูลมีการวัดความคลาดเคลื่อนของตัวแปรอิสระ เพื่อหาวิธีการที่เหมาะสมภายใต้ขอบเขตของการศึกษาเดียวกัน โดยทำการศึกษาจากข้อมูลที่สร้างด้วยเทคนิคมอนติคาร์โล และเปรียบเทียบประสิทธิภาพของวิธีการประมาณค่าทุกวิธีจากค่าความเอนเอียง และค่าความแปรปรวนของค่าประมาณพารามิเตอร์

วัตถุประสงค์

งานวิจัยครั้งนี้มีวัตถุประสงค์เพื่อ

1. ศึกษาปัญหา สาเหตุ และผลกระทบที่อาจเกิดขึ้นในการวิเคราะห์การถดถอยเมื่อตัวแปรอิสระถูกวัดมาโดยมีความคลาดเคลื่อน

2. เพื่อศึกษาเปรียบเทียบวิธีการประมาณค่าพารามิเตอร์ในตัวแบบการถดถอยเชิงเส้นอย่างง่ายเมื่อตัวแปรอิสระถูกวัดมาโดยมีความคลาดเคลื่อน ด้วยวิธีการประมาณค่า 4 วิธี คือ

2.1 วิธีกำลังสองน้อยที่สุด

2.2 วิธีปรับตัวประมาณกำลังสองน้อยที่สุด

2.3 วิธี Orthogonal Regression

2.4 วิธี Instrumental Variables

3. หาวิธีประมาณพารามิเตอร์ที่เหมาะสม ในกรณีที่วัดค่าตัวแปรอิสระโดยมีความคลาดเคลื่อนจากการวัดเกิดขึ้น เมื่อตัวอย่างมีขนาดต่างๆ กัน

ประโยชน์ที่คาดว่าจะได้รับ

1. สามารถหาวิธีการประมาณค่าพารามิเตอร์ที่เหมาะสมเพื่อใช้ในการวิเคราะห์การถดถอยเชิงเส้นอย่างง่ายในกรณีที่วัดค่าตัวแปรอิสระโดยมีความคลาดเคลื่อนเกิดขึ้น

2. ช่วยหาค่าทำนายที่มีความแม่นยำและมีความคลาดเคลื่อนน้อย ในกรณีที่ข้อมูลมีปัญหาเกี่ยวกับความคลาดเคลื่อนจากการวัดค่าตัวแปรอิสระ ซึ่งวิธีกำลังสองน้อยที่สุดไม่อาจให้ค่าประมาณที่มีความแม่นยำและมีความคลาดเคลื่อนต่ำ

ขอบเขตการวิจัย

การวิจัยครั้งนี้จะทำภายใต้ขอบเขต ดังนี้

1. กำหนดรูปแบบความคลาดเคลื่อนในการวิเคราะห์การถดถอยเชิงเส้นอย่างง่าย เมื่อตัวแปรอิสระวัดมาโดยมีความคลาดเคลื่อนเกิดขึ้น จากตัวแบบการถดถอยเชิงเส้นอย่างง่าย

$$Y_i = \beta_0 + \beta_1 X_i + \varepsilon_i \qquad ; i = 1,2, \ldots, n$$

เมื่อ ε_i เป็นค่าความคลาดเคลื่อนที่เกิดขึ้นโดยสุ่มของตัวอย่างหน่วยที่ i

ปัญหาสำคัญของการวิเคราะห์เมื่อการวัดค่าตัวแปรอิสระมีความคลาดเคลื่อน คือ การกำหนดรูปแบบของความคลาดเคลื่อนที่เหมาะสม ซึ่งขึ้นอยู่กับสถานการณ์และข้อมูลที่นำมาใช้ในการวิเคราะห์ รูปแบบของความคลาดเคลื่อนที่นำมาศึกษา คือ Classical Error Model ซึ่งเป็นรูปแบบที่เข้าใจง่ายและไม่ยุ่งยาก เหมาะสมที่จะใช้เมื่อไม่สามารถทราบ X ที่ถูกต้องได้โดยตรงเนื่องจากมีความคลาดเคลื่อนที่เกิดจากการวัดค่า

2. กำหนดขนาดตัวอย่างที่จะทำการศึกษาเป็น 3 กลุ่ม คือ

2.1 กลุ่มตัวอย่างขนาดเล็ก ได้แก่ กลุ่มตัวอย่างขนาด 10, 20 และ 30

2.2 กลุ่มตัวอย่างขนาดกลาง ได้แก่ กลุ่มตัวอย่างขนาด 50 และ 70

2.3 กลุ่มตัวอย่างขนาดใหญ่ ได้แก่ กลุ่มตัวอย่างขนาด 100, 200 และ 300

3. กำหนด $\beta_0 = 0$ และ $\beta_1 = 1$ เนื่องจากได้ทำการทดลอง ณ ค่าต่างๆ ของ β_0 และ β_1 พบว่า ไม่ว่า β_0 และ β_1 จะเป็นค่าใดก็ตาม ผลสรุปไม่เปลี่ยนแปลง ดังนั้นจึงกำหนดค่าให้กับ β_0 และ β_1 ดังกล่าว

4. ในการตรวจสอบข้อมูลที่สร้างขึ้นว่ามีปัญหาเกี่ยวกับการวัดค่าของตัวแปรอิสระ โดยมีความคลาดเคลื่อนเกิดขึ้นหรือไม่ จะทำการทดสอบสมมติฐานโดยใช้ระดับนัยสำคัญเท่ากับ 0.05

5. กำหนดค่าสัมประสิทธิ์สหสัมพันธ์ของตัวแปรอิสระกับตัวแปร Instrumental ในกรณีของวิธี Instrumental Variables เป็น 0.20, 0.40, 0.50, 0.60, 0.80 และ 0.90

6. เกณฑ์การตัดสินใจว่า วิธีกำลังสองน้อยที่สุด วิธีปรับตัวประมาณกำลังสองน้อยที่สุด วิธี Orthogonal Regression และวิธี Instrumental Variables วิธีใดเป็นวิธีที่มีประสิทธิภาพมากที่สุด ในการประมาณค่าพารามิเตอร์ในตัวแบบการถดถอยเชิงเส้นอย่างง่ายของรูปแบบต่างๆ จะพิจารณา โดยการหาค่าความเอนเอียง (Bias) และค่าความแปรปรวน (Variance) ของค่าประมาณพารามิเตอร์ β_1 โดยวิธีที่แตกต่างกันทั้งหมด 4 วิธี และจะเลือกวิธีที่ให้ค่าความเอนเอียงและค่าความแปรปรวนต่ำที่สุด ซึ่งถือว่าเป็นตัวประมาณที่ดีที่สุด การคำนวณหาค่าต่างๆ ทำได้ดังนี้

$$\text{Bias}(\hat{\beta}_1) = \frac{\sum_{t=1}^{1000}\left|\hat{\beta}_{1t} - \beta_1\right|}{1000}$$

$$S^2(\hat{\beta}_1) = \frac{\sum_{t=1}^{1000}\left(\hat{\beta}_{1t} - \beta_1\right)^2}{1000}$$

โดยที่

β_1 คือ ค่าพารามิเตอร์แทนอัตราการเปลี่ยนแปลงในค่าคาดหวังของ Y เมื่อ X มีค่าเพิ่มขึ้นหนึ่งหน่วย

$\hat{\beta}_{1t}$ คือ ค่าประมาณ β_1 ในการทำซ้ำรอบที่ t

t คือ รอบที่ทำซ้ำ

7. การศึกษาครั้งนี้ได้ทำการจำลองชุดข้อมูลให้มีสถานการณ์ตามที่กำหนด โดยใช้เทคนิคการจำลองแบบมอนติคาร์โล (Monte Carlo Simulation Technique) และทำการจำลองชุดข้อมูลในแต่ละสถานการณ์ซ้ำจำนวน 1,000 ชุด

การตรวจเอกสาร

การตรวจเอกสารแบ่งออกเป็น 2 ส่วน ส่วนแรกจะกล่าวถึงวิธีการทางสถิติที่ใช้ในการวิจัยครั้งนี้ และส่วนที่สองจะกล่าวถึงงานวิจัยที่เกี่ยวข้อง ดังนี้

<u>วิธีการทางสถิติ</u>

1. <u>วิธีการตรวจสอบการเกิดปัญหาการวัดความคลาดเคลื่อน</u>

มนตรี (2529) ได้อธิบายว่า สำหรับปัญหาการวัดความคลาดเคลื่อนในตัวแปรอิสระนับว่าเป็นปัญหาที่ยังมีการพัฒนาวิธีการทดสอบไม่กว้างขวางเท่าปัญหาอื่นๆ วิธีหนึ่งที่ตรวจสอบได้ก็คือ การทดสอบนัยสำคัญของสหสัมพันธ์ระหว่าง $|e_i|$ กับ X ซึ่งมีขั้นตอนดังนี้

1.1 ใช้ตัวแบบการถดถอย $Y_i = \beta_0 + \beta_1 X_i + \varepsilon_i$ ประมาณพารามิเตอร์โดยวิธีกำลังสองน้อยที่สุด (Ordinary Least Squares Method : OLS Method) และคำนวณหาเศษตกค้าง (Residual) : $e_i = Y_i - \hat{Y}_i$; i = 1,2, …, n

1.2 คำนวณสหสัมพันธ์ระหว่าง $|e|$ กับ X_i ดังนี้

$$r = \frac{\sum |e_i| x_i}{\sqrt{\sum e_i^2 \sum x_i^2}}; \qquad i = 1,2, \ldots, n$$

1.3 ทดสอบสมมติฐาน

$$H_0 : \rho = 0$$
$$H_1 : \rho \neq 0$$

โดยเราจะปฏิเสธสมมติฐานหลัก ณ ระดับนัยสำคัญ α เมื่อ $|t| > t_{n-2,1-\frac{\alpha}{2}}$ เมื่อ

$$t = \frac{r\sqrt{n-2}}{\sqrt{1-r^2}}$$

หรือนัยหนึ่งถ้าพบว่า $|t| > t_{n-2,1-\frac{\alpha}{2}}$ เราย่อมสรุปได้ว่า X_i มีปัญหาการวัดความคลาดเคลื่อนเกิดขึ้น

2. คุณสมบัติของตัวประมาณ

จากสมการถดถอย $Y_i = \beta_0 + \beta_1 X_i + \varepsilon_i$ จะต้องคำนวณหาค่า β_0 และ β_1 ซึ่งจะทำให้ทราบถึงความสัมพันธ์ระหว่าง X และ Y ว่ามีความสัมพันธ์ตามกันหรือตรงข้ามกัน และความสัมพันธ์นั้นมากหรือน้อยเพียงใด โดยใช้ข้อมูลตัวอย่างในการประมาณค่า

ในการเลือกใช้สถิติเพื่อประมาณค่าพารามิเตอร์นั้น จะอาศัยการแจกแจงของตัวอย่างเป็นหลักในการพิจารณา โดยใช้คุณสมบัติของสถิติหลายอย่างประกอบกันในการเลือกตัวประมาณดังนี้

2.1 ความไม่เอนเอียง (unbiasedness)

ถ้าหากตัวประมาณ T = t(X) ใดๆ มีคุณสมบัติว่า ค่าเฉลี่ยของการแจกแจงจากตัวอย่างของ $\tau(\tilde{\theta})$ เท่ากับ $\tau(\theta)$ ก็ย่อมจะเป็นสิ่งที่ดี เพราะแสดงว่าค่าเฉลี่ยของ T อยู่เท่ากับค่าจริง ซึ่งจะเรียกคุณสมบัตินี้ว่าความไม่เอนเอียง (unbiasedness)

ให้ $X_1, X_2, ..., X_n$ เป็นตัวอย่างสุ่มขนาด n และ ตัวสถิติ T = t(X) เป็นตัวประมาณของ $\tau(\theta)$ ความเอนเอียงของ T จะถูกกำหนดให้มีค่าเท่ากับ

$$\text{bias (T)} = E\left[(T(X)) - \tau(\theta)\right]$$

ตัวสถิติ T จะเรียกว่าเป็น ตัวประมาณที่ไม่เอนเอียง ของ $\tau(\theta)$ ถ้า bias (T) มีค่าเท่ากับศูนย์

2.2 <u>ความคงตัว (consistency)</u>

โดยปกติการคำนวณหาค่าของตัวประมาณ T สำหรับประมาณ $\tau(\theta)$ ใดๆ มักจะคำนวณจากข้อมูลที่สังเกตในตัวอย่างสุ่มขนาด n ดังนั้นถ้าขนาดตัวอย่างมีขนาดใหญ่ขึ้นแล้วค่าของตัวประมาณ T ที่ได้ควรจะมีค่าเข้าใกล้ค่า $\tau(\theta)$ มากยิ่งขึ้น นั่นคือ ค่าประมาณที่ได้มีความถูกต้องมากขึ้นเมื่อเพิ่มขนาดตัวอย่างสุ่ม และถ้าเป็นไปได้ถ้าเพิ่มขนาดตัวอย่างได้มากจนเท่ากับขนาดของประชากร ตัวสถิติ T ก็จะให้ค่าเท่ากับค่าของ $\tau(\theta)$ เราเรียกคุณสมบัตินี้ว่า ความคงตัว (consistency)

ตัวประมาณ T ใดๆ จะมีคุณสมบัติที่เรียกว่าความคงเส้นคงวาของ $\tau(\theta)$ ถ้า T ลู่เข้าในเชิงความน่าจะเป็นไปสู่ค่า $\tau(\theta)$ กล่าวคือ

$$\lim_{n\to\infty} P(|T-\tau(\theta)| < \varepsilon) = 1$$

หรือ

$$\lim_{n\to\infty} P(|T-\tau(\theta)| > \varepsilon) = 0 \qquad \text{สำหรับทุกๆ ค่าของ } \theta \in \Theta$$

2.3 <u>ความพอเพียง (sufficiency)</u>

ให้ $g(x\,;\theta)$ เป็นฟังก์ชันความหนาแน่นร่วมของตัวอย่างสุ่ม X และตัวสถิติ T = t (X) ซึ่งมีฟังก์ชันความหนาแน่น $f(t(X);\theta)$ จะถูกเรียกว่าสถิติพอเพียงสำหรับพารามิเตอร์ θ ก็ต่อเมื่อฟังก์ชันความน่าจะเป็นแบบมีเงื่อนไขของ X เมื่อกำหนดให้ T = t(X); $f(X = x \mid T = t(X))$ ไม่ขึ้นอยู่กับค่าพารามิเตอร์ θ สำหรับทุกๆ ค่าของ T = t(X) นั่นคือ

$$f(X = x \mid T = t(X)) = \frac{g(x,\theta)}{f(t(x),\theta)} = h(x)$$ เมื่อ h(x) เป็นฟังก์ชันที่ไม่ขึ้นอยู่กับค่า θ

หรืออาจกล่าวอีกนัยหนึ่งว่า T = t(X) จะเป็นสถิติพอเพียงสำหรับพารามิเตอร์ θ ถ้า

$$g(x,\theta) = f(t(x),\theta).\ h(x)$$

3. <u>รูปแบบของความคลาดเคลื่อน</u>

Buzas and Stefanski (2003) ได้กล่าวถึงรูปแบบของความคลาดเคลื่อนในการวิเคราะห์การถดถอยเชิงเส้นอย่างง่ายเมื่อตัวแปรอิสระวัดมาโดยมีความคลาดเคลื่อนเกิดขึ้น 2 รูปแบบ ดังนี้

3.1 <u>Classical Error Model</u>

สำหรับตัวแบบในกรณีที่ไม่สามารถทราบ X ที่ถูกต้องได้โดยตรงเนื่องจากมีความคลาดเคลื่อนที่เกิดจากการวัดค่านั้นจะพบว่า

$$W_i = X_i + U_i$$

โดยที่

X_i เป็นค่าจริงของตัวแปรอิสระ

W_i เป็นค่าสังเกต X_i ที่วัดมาโดยมีความคลาดเคลื่อน

U_i เป็นค่าความคลาดเคลื่อนในการวัดค่าของ X_i

$U_i \sim N(0, \sigma^2_U)$

จากตัวแบบ

$$Y_i = \beta_0 + \beta_1 X_i + \varepsilon_i \qquad \text{--------------- (1)}$$

กำหนดให้ ε_i และ U_i เป็นอิสระกัน เขียนตัวแบบ (1) ได้เป็น

$$\begin{aligned} Y_i &= \beta_0 + \beta_1(W_i - U_i) + \varepsilon_i \\ &= \beta_0 + \beta_1 W_i + (\varepsilon_i - \beta_1 U_i) \end{aligned}$$

เป็นตัวแบบการถดถอยเชิงเส้นที่มีค่าคลาดเคลื่อนเป็น $\delta_i = \varepsilon_i - \beta_1 U_i$ ที่มี $\mathrm{cov}(X_i, \delta_i) = -\beta_1 \sigma_u^2$ ถ้า $\beta_1 \neq 0$, W_i และ ε_i มีสหสัมพันธ์กันแล้ว ตัวประมาณโดยวิธีกำลังสองน้อยที่สุดของตัวแบบนี้จะเป็นตัวประมาณที่มีความเอนเอียง

3.2 Berkson Error Model

Berkson (1950) กล่าวถึงรูปแบบของการถดถอยเชิงเส้นอย่างง่ายเมื่อตัวแปรอิสระ X ถูกวัดมาโดยมีความคลาดเคลื่อน ตัวแบบนี้กำหนดให้การทดลองมีการควบคุมค่าตัวแปรอิสระเป็น W_i ทำให้ W_i เป็นค่าคงที่ ดังนั้นค่าจริง X_i มีค่าเป็น $X_i = W_i - U_i$
จากตัวแบบ (1) จึงสามารถเขียนได้เช่นเดียวกับ Classical Error Model เป็น

$$Y_i = \beta_0 + \beta_1 W_i + (\varepsilon_i - \beta_1 U_i)$$

เนื่องจาก X_i เป็นตัวแปรสุ่ม W_i เป็นค่าคงที่ ดังนั้น $E(\varepsilon_i - \beta_1 U_i) = 0$ และ $\varepsilon_i - \beta_1 U_i$ ไม่มีสหสัมพันธ์กับ X_i หรือ $\mathrm{Cov}(X, \delta) = 0$ ดังนั้นตัวแบบของ Berkson จึงมีคุณสมบัติตามที่ตัวแบบการถดถอยต้องการ และตัวประมาณพารามิเตอร์โดยวิธีกำลังสองน้อยที่สุดจะมีคุณสมบัติที่เป็นตัวประมาณที่มีความเอนเอียงและมีค่าความแปรปรวนต่ำสุด

4. การประมาณค่าพารามิเตอร์ของสมการถดถอยเมื่อวัดค่าตัวแปรอิสระ X มาโดยมีความคลาดเคลื่อน

สำหรับการวิจัยครั้งนี้จะได้ศึกษาวิธีการประมาณค่าพารามิเตอร์ในกรณีที่การวัดค่าตัวแปร X ความคลาดเคลื่อนเกิดขึ้นในกรณีของ Classical Error Model ด้วยวิธีการประมาณค่าพารามิเตอร์ 4 วิธี ดังต่อไปนี้

4.1 <u>วิธีกำลังสองน้อยที่สุด</u>

Fuller (1987) กล่าวถึงการประมาณค่าสัมประสิทธิ์การถดถอยโดยวิธีกำลังสองน้อยที่สุด มีหลักการคือหาค่าประมาณสัมประสิทธิ์การถดถอยที่ทำให้ผลบวกกำลังสองของความแตกต่างระหว่างค่าจริงกับค่าประมาณหรือผลบวกกำลังสองของความคลาดเคลื่อนมีค่าน้อยที่สุด

ตัวประมาณของ β_1 คือ

$$\beta_1 = \frac{\sum_{i=1}^{N}\left(W_i - \overline{W}\right)\left(Y_i - \overline{Y}\right)}{\sum_{i=1}^{N}\left(W_i - \overline{W}\right)^2}$$

การประมาณค่าโดยใช้วิธีกำลังสองน้อยที่สุด เมื่อตัวแบบวัดค่าตัวแปรอิสระมาโดยมีความคลาดเคลื่อน จะทำให้ได้ตัวประมาณที่มีความเอนเอียง (Biased Estimators) และเป็นตัวประมาณที่ไม่คงตัว (Inconsistent Estimators)

4.2 <u>วิธีปรับตัวประมาณกำลังสองน้อยที่สุด</u>

จากตัวแบบ

$$Y_i = \beta_0 + \beta_1 W_i + (\varepsilon_i - \beta_1 U_i) \qquad \text{--------------- (2)}$$

ให้ $\hat{\gamma}_1$ เป็นตัวประมาณพารามิเตอร์โดยวิธีกำลังสองน้อยที่สุดที่มีค่าความคลาดเคลื่อนในตัวแบบเป็น $\varepsilon_i - \beta_1 U_i$ มีค่า $Cov(X,\delta)$ เป็น $-\beta_1\sigma_u^2$

ซึ่งตัวประมาณ $\hat{\gamma}_1$ เป็นตัวประมาณที่มีความเอนเอียงของ β_1 เนื่องจาก

$$E(\hat{\gamma}_1) = \beta_1 \cdot \frac{\sigma_X^2}{\sigma_X^2 + \sigma_U^2} = \beta_1 \cdot \frac{\sigma_X^2}{\sigma_W^2}$$

วิธีปรับตัวประมาณกำลังสองน้อยที่สุดนี้ จะหาตัวประมาณที่ไม่เอนเอียง โดยการปรับค่าตัวประมาณกำลังสองน้อยที่สุด ($\hat{\gamma}_1$) ด้วย reliability ratio K_{XX}

$$K_{XX} = \frac{\sigma_X^2}{\sigma_W^2}$$

ประมาณ K_{XX} โดยใช้ข้อมูลจากตัวอย่าง

Fuller (1987) แสดงว่า ตัวประมาณที่ไม่เอนเอียงของ β_1 ในตัวแบบ (2) คือ

$$\beta_1 = \frac{\hat{\gamma}}{K_{XX}}$$

ที่มีความแปรปรวน $\sigma^2\left(\hat{\beta}_1\right) = K_{XX}^{-2}\left[\sum_{i=1}^{n}\left(X_i - \overline{X}\right)^2\right]^{-1} S_\ell^2$

การประมาณค่าโดยใช้วิธีปรับตัวประมาณกำลังสองน้อยที่สุด จะทำให้ได้ตัวประมาณที่มีคุณสมบัติดังนี้

1) เป็นตัวประมาณที่ไม่เอนเอียง (Unbiased Estimators)

2) เป็นตัวประมาณที่คงตัว (Consistent Estimators)

4.3 <u>วิธี Orthogonal Regression</u>

Carroll and David (1996) ได้เสนอวิธี Orthogonal Regression ซึ่งเป็นวิธีการที่เก่าแก่วิธีการหนึ่งของการถดถอยเชิงเส้นที่เหมาะสมสำหรับปัญหาที่เกิดขึ้นเนื่องจากความคลาดเคลื่อนจากการวัด เป็นเทคนิคที่นำมาประยุกต์ใช้ไม่บ่อยนัก เนื่องจากวิธีการนี้จะนำส่วนของสมการความคลาดเคลื่อน รวมทั้งความแปรปรวนของความคลาดเคลื่อนจากการวัดมาพิจารณาด้วย จะได้ค่าประมาณของ β_1 ดังนี้

$$\hat{\beta}_1 = \frac{s_Y^2 - \eta s_W^2 + \left\{\left(s_Y^2 - \eta s_W^2\right)^2 + 4\eta s_{WY}^2\right\}^{1/2}}{2S_{WY}}$$

โดยที่ $$\eta = \frac{\sigma_\varepsilon^2}{\sigma_U^2}$$

ตัวประมาณ β_1 โดยวิธีนี้มีคุณสมบัติสำคัญคือ มีค่าอยู่ระหว่างค่าความชันของเส้นถดถอยแทนความสัมพันธ์ของ Y กับ W และค่าส่วนกลับของความชันของเส้นถดถอยระหว่าง W กับ Y

การประมาณค่าโดยใช้วิธี Orthogonal Regression จะทำให้ได้ตัวประมาณที่มีคุณสมบัติดังนี้

1) เป็นตัวประมาณที่ไม่เอนเอียง (Unbiased Estimators)

2) เป็นตัวประมาณที่คงตัว (Consistent Estimators)

4.4 <u>วิธี Instrumental Variables</u>

Carter and Fuller (1980) กล่าวว่าปัญหาในการประมาณค่าพารามิเตอร์ในกรณีที่ตัวแปรอิสระถูกวัดมาโดยมีความคลาดเคลื่อน จะทำให้ค่าประมาณของสัมประสิทธิ์การถดถอยมีความเอนเอียงและเป็นตัวประมาณที่ไม่คงตัว เนื่องจากตัวแปรอิสระมีความสัมพันธ์กับความคลาดเคลื่อน

วิธี Instrumental Variables เป็นอีกวิธีหนึ่งที่นำมาประยุกต์ใช้ เพื่อแก้ปัญหาดังกล่าว โดยการหาตัวแปรอีก 1 ตัว คือตัวแปร T ที่มีคุณสมบัติดังนี้

1. ตัวแปร T มีสหสัมพันธ์อย่างสูงกับตัวแปรอิสระ X

2. ตัวแปร T ไม่มีสหสัมพันธ์กับความคลาดเคลื่อนในการวัดค่า X และไม่มีสหสัมพันธ์กับเทอมของความคลาดเคลื่อนในตัวแบบ นั่นคือ

$$\text{Cov}(T_i, U_i) = 0$$

และ $$\text{Cov}(T_i, \varepsilon_i) = 0$$

จะได้ค่าประมาณของ β_1 ของตัวแบบการถดถอย คือ

$$\hat{\beta}_1 = \frac{\sum_{i=1}^{N}\left(Y_i - \overline{Y}\right)\left(T_i - \overline{T}\right)}{\sum_{i=1}^{N}\left(W_i - \overline{W}\right)\left(T_i - \overline{T}\right)}$$

การประมาณค่าโดยใช้วิธี Instrumental Variables จะทำให้ได้ตัวประมาณที่มีคุณสมบัติดังนี้

1) เป็นตัวประมาณที่ไม่เอนเอียง (Unbiased Estimators)

2) เป็นตัวประมาณที่คงตัว (Consistent Estimators)

5. <u>เทคนิคการจำลองแบบมอนติคาร์โล (Monte Carlo Simulation Technique)</u>

เป็นเทคนิคที่ใช้ในการสร้างข้อมูลเพื่อจำลองปัญหาที่ต้องการศึกษา จากการศึกษาข้อมูลที่จำลองขึ้นนี้ จะทำให้ผู้ศึกษาสามารถแก้ปัญหาทางคณิตศาสตร์และปัญหาในงานด้านอื่นๆ ภายใต้สถานการณ์ที่ต้องการได้ เทคนิคการจำลองแบบมีอยู่ด้วยกันหลายวิธี เทคนิคการจำลองแบบมอนติคาร์โลเป็นวิธีหนึ่งที่อาศัยหลักของตัวเลขสุ่ม (Random Number) มาช่วยหาคำตอบของปัญหาที่ต้องการศึกษา ซึ่งเป็นวิธีหนึ่งที่มีผู้นิยมใช้เพื่อการศึกษาถึงการแก้ปัญหากันอย่างแพร่หลายในปัจจุบัน งานวิจัยครั้งนี้จึงใช้เทคนิคการจำลองแบบมอนติคาร์โลมาช่วยในการสร้างข้อมูลที่มีลักษณะต่างๆ ตามที่ต้องการศึกษา

ขั้นตอนของเทคนิคการจำลองแบบมอนติคาร์โล มีดังนี้

1. การสร้างตัวเลขสุ่ม ซึ่งเป็นขั้นตอนที่สำคัญมากในเทคนิคการจำลองแบบมอนติคาร์โล โดยจะสร้างเลขสุ่มที่มีการแจกแจงแบบยูนิฟอร์ม (Uniform Distribution) ที่เป็นอิสระกัน ในช่วง 0 ถึง 1

สำหรับการสร้างเลขสุ่มนั้น จะใช้ฟังก์ชันในโปรแกรม MATLAB ในการสร้างตัวเลขแบบสุ่มที่มีการแจกแจงแบบยูนิฟอร์ม ดังนี้

Random (‘Uniform’,M,N) คือ เมตริกซ์ขนาด M×N ที่ประกอบด้วยเลขสุ่มที่มีการแจกแจงแบบยูนิฟอร์ม

2. ประยุกต์การใช้เลขสุ่มกับปัญหาที่ต้องการศึกษา ขั้นตอนนี้เป็นขั้นตอนที่ขึ้นอยู่กับลักษณะของปัญหา โดยใช้เลขสุ่มที่มีการแจกแจงแบบยูนิฟอร์ม เพื่อสร้างข้อมูลที่มีการแจกแจงแบบอื่นแล้วใช้ข้อมูลที่มีการแจกแจงดังกล่าวในการแก้ปัญหา

สำหรับการสร้างเลขสุ่มที่มีการแจกแจงแบบปกติ จะใช้ฟังก์ชันในโปรแกรม MATLAB ในการสร้างตัวเลขแบบสุ่มที่มีการแจกแจงแบบปกติ ดังนี้

Random(‘Normal’,A,B,M,N) คือ เมตริกซ์ขนาด M×N ที่ประกอบด้วยเลขสุ่มที่มีการแจกแจงแบบปกติ โดยมีค่าเฉลี่ยเท่ากับ A และความแปรปรวนเท่ากับ B

3. เมื่อประยุกต์ใช้เลขสุ่มกับปัญหาที่ต้องการศึกษาได้แล้ว ขั้นตอนต่อไปจะเป็นการทดลองกระทำซ้ำๆ โดยใช้กระบวนการของเลขสุ่ม (Random Process) มากระทำในลักษณะต่างๆ กัน เพื่อหาคำตอบของปัญหาที่ต้องการศึกษาเมื่อประยุกต์ใช้เลขสุ่มกับปัญหาที่ต้องการศึกษา

งานวิจัยที่เกี่ยวข้อง

สำหรับผลงานวิจัยที่เกี่ยวข้องกับการศึกษานี้ได้แก่

Anderson (1951) ได้ทำการศึกษาตัวประมาณโดยวิธีภาวะความน่าจะเป็นสูงสุด เมื่อตัวแปรอิสระวัดมาโดยมีความคลาดเคลื่อน โดยพิจารณากรณีประมาณค่าพารามิเตอร์ภายใต้ข้อจำกัดเชิงเส้น พบว่าสามารถหาคุณสมบัติของตัวประมาณเมื่อตัวอย่างมีขนาดใหญ่ ภายใต้ข้อกำหนดว่า ความแปรปรวนของค่าสัมประสิทธิ์การถดถอยเข้าใกล้ศูนย์

Dorff and Gurland (1961) ได้ศึกษาเกี่ยวกับความแปรปรวนของตัวประมาณในตัวแบบการถดถอยเชิงเส้นอย่างง่ายที่มีตัวแปรอิสระเพียง 1 ตัว และเมื่อพิจารณาจากค่าประมาณความแปรปรวนของตัวประมาณโดยวิธีภาวะความน่าจะเป็นสูงสุดดีกว่าตัวประมาณโดยวิธีโมเมนต์

Farrell (1978) ได้ทำการศึกษาเรื่องความคลาดเคลื่อนจากการวัดและนัยสำคัญทางสถิติของตัวแปรอิสระ พบว่าการวัดค่าตัวแปรอิสระโดยมีความคลาดเคลื่อนนั้นจะได้ตัวประมาณที่มีความเอนเอียงเมื่อใช้การประมาณค่าโดยวิธีกำลังสองน้อยที่สุด ทำให้สัมประสิทธิ์การถดถอยมีค่าเข้าใกล้ศูนย์ สำหรับการทดสอบสมมติฐานว่าสัมประสิทธิ์การถดถอยมีค่าเท่ากับศูนย์โดยใช้ตัวสถิติ t ในการทดสอบ แสดงให้เห็นว่าความคลาดเคลื่อนจากการวัดจะทำให้ความน่าจะเป็นในการปฏิเสธสมมติฐานหลักน้อยลง

Carroll and David (1996) ได้ศึกษาวิธี Orthogonal regression ซึ่งเป็นวิธีการหนึ่งของการถดถอยเชิงเส้นที่นำมาใช้เมื่อมีปัญหาความคลาดเคลื่อนจากการวัดค่าตัวแปรอิสระ พบว่าวิธีการนี้มักจะนำมาใช้ไม่ถูกต้องนัก เนื่องจากเรามักละเลยเทอมของความคลาดเคลื่อน ทำให้ค่าประมาณที่ได้มีค่ามากกว่าค่าประมาณที่แท้จริง ดังนั้น การวิเคราะห์ความคลาดเคลื่อนจากการวัดด้วยวิธี Orthogonal regression จะต้องนำเทอมของความคลาดเคลื่อนจากการวัด รวมทั้งค่าประมาณสัดส่วนความแปรปรวนของความคลาดเคลื่อนจากการวัดมาพิจารณาด้วย

Marcia (2002) ได้ศึกษาวิธีการประมาณค่าพารามิเตอร์ในตัวแบบการถดถอยเชิงเส้นอย่างง่ายเมื่อวัดค่าตัวแปรอิสระ โดยมีความคลาดเคลื่อนด้วยวิธี Instrumental Variables พบว่าปัญหาในการประมาณค่าพารามิเตอร์ คือค่าสัมประสิทธิ์การถดถอยจะมีความเอนเอียงและไม่คงตัว เนื่องจากตัวแปรอิสระมีความสัมพันธ์กับเทอมของความคลาดเคลื่อน จึงได้หาตัวแปรที่มีความสัมพันธ์ค่อนข้างสูงกับตัวแปรอิสระ X และไม่มีความสัมพันธ์กับความคลาดเคลื่อนจากการวัดหรือเทอมของความคลาดเคลื่อน นั่นคือตัวแปร T ดังนั้น จะได้ค่าสัมประสิทธิ์การถดถอยคือ

$$\hat{\beta} = \frac{\sum_{i=1}^{N}\left(Y_i - \overline{Y}\right)\left(T_i - \overline{T}\right)}{\sum_{i=1}^{N}\left(W_i - \overline{W}\right)\left(T_i - \overline{T}\right)}$$

อุปกรณ์และวิธีการ

อุปกรณ์

เครื่องไมโครคอมพิวเตอร์ ภาควิชาสถิติ คณะวิทยาศาสตร์ มหาวิทยาลัยเกษตรศาสตร์ โดยใช้ โปรแกรม MATLAB Version 6.5 และโปรแกรม SAS Version 9.0 ในการวิเคราะห์ ดังนี้

1. เพื่อทำการประมวลผลข้อมูล

2. สร้างกราฟ เพื่อเปรียบเทียบวิธีการประมาณค่าพารามิเตอร์ทั้ง 4 วิธี

วิธีการ

ในการศึกษาวิจัยครั้งนี้ ข้อมูลที่ใช้ในการวิจัยได้มาจากการจำลองด้วยเทคนิคมอนติคาร์โล และเขียนโปรแกรมคอมพิวเตอร์ MATLAB Version 6.5 และ SAS Version 9.0 โดยมีขั้นตอนในการวิจัยประกอบด้วย

1. จำลองข้อมูลที่ใช้ในการวิจัยด้วยเทคนิคมอนติคาร์โล โดยสร้างข้อมูลให้มีการแจกแจงตามที่กำหนด ดังนี้

1.1 สร้างค่าจริงของตัวแปรอิสระ (X_i) ให้มีการแจกแจงแบบปกติมาตรฐาน มีค่าเฉลี่ย 0 และความแปรปรวน 1 นั่นคือ $x_i \sim N(0,1)$ โดยมีตัวอย่างขนาด n

1.2 สร้างค่าความคลาดเคลื่อน (ε_i) ให้มีการแจกแจงแบบปกติ มีค่าเฉลี่ย 0 และความแปรปรวน 0.5 นั่นคือ $\varepsilon_i \sim N(0,5)$

1.3 กำหนดค่า $\beta_0 = 0$ และ $\beta_1 = 1$

1.4 สร้างค่าความคลาดเคลื่อนในการวัดค่าของ x_i (u_i) ให้มีการแจกแจงแบบปกติมาตรฐาน มีค่าเฉลี่ย 0 และความแปรปรวน 1 นั่นคือ $u_i \sim N(0,1)$

1.5 สร้าง $W_i = X_i + U_i$

1.6 สร้างตัวแบบการถดถอยเชิงเส้นอย่างง่ายที่มีความคลาดเคลื่อนเป็น $\gamma_i = \varepsilon_i - \beta_1 U_i$ คือ

$$Y_i = \beta_0 + \beta_1 W_i + (\varepsilon_i - \beta_1 U_i)$$

1.7 ประมาณค่าพารามิเตอร์ β_0 และ β_1

1.8 คำนวณหาเศษตกค้าง (Residual) และทดสอบปัญหาการวัดความคลาดเคลื่อน

2. สร้างข้อมูลที่ใช้ศึกษาตามลักษณะการแจกแจงและตัวแบบที่กำหนด ในแต่ละขนาดตัวอย่างจำนวน 1,000 ชุด

3. ประมาณค่าพารามิเตอร์ด้วยวิธีการประมาณ 4 วิธี ดังนี้

3.1 วิธีกำลังสองน้อยที่สุด

ค่าประมาณของ β_1 เป็นดังนี้

$$\hat{\beta}_1 = \frac{\sum_{i=1}^{N}\left(W_i - \overline{W}\right)\left(Y_i - \overline{Y}\right)}{\sum_{i=1}^{N}\left(W_i - \overline{W}\right)^2}$$

3.2 วิธีปรับตัวประมาณกำลังสองน้อยที่สุด

หาค่าประมาณกำลังสองน้อยที่สุดของค่าความชัน ($\hat{\gamma}$) และใช้ $\hat{\gamma}$ ประมาณ β_1 ดังนี้

$$\hat{\beta}_1 = \frac{\hat{\gamma}}{K_{xx}} = \frac{\sum_{i=1}^{N}\left(W_i - \overline{W}\right)\left(Y_i - \overline{Y}\right)}{\sum_{i=1}^{N}\left(W_i - \overline{W}\right)^2} \cdot \frac{\sigma_w^2}{\sigma_x^2}$$

ประมาณ σ_W^2 , σ_X^2 และ K_{XX} ด้วยข้อมูลตัวอย่าง โดยที่

$$K_{xx} = \frac{\sigma_x^2}{\sigma_w^2}$$

3.3 วิธี Orthogonal Regression

ประมาณ β_1 โดย

$$\hat{\beta}_1 = \frac{s_y^2 - \eta s_w^2 + \left\{\left(s_y^2 - \eta s_w^2\right)^2 + 4\eta s_{wy}^2\right\}^{1/2}}{2S_{wy}}$$

ประมาณ σ_w^2 , s_Y^2 , s_{WY}^2 ด้วยข้อมูลตัวอย่าง

3.4 วิธี Instrumental Variables (IV Method)

สร้างตัวแปร T ที่มีความสัมพันธ์กับค่าของตัวแปรอิสระ X เป็น 0.2, 0.4, 0.5, 0.6, 0.8 และ 0.9 โดยที่ตัวแปร T ไม่มีความสัมพันธ์กับความคลาดเคลื่อนในการวัดค่า X และไม่มีความสัมพันธ์กับเทอมของความคลาดเคลื่อน

ค่าประมาณของ β_1 เป็นดังนี้

$$\hat{\beta}_1 = \frac{\sum_{i=1}^{N}\left(Y_i - \overline{Y}\right)\left(T_i - \overline{T}\right)}{\sum_{i=1}^{N}\left(W_i - \overline{W}\right)\left(T_i - \overline{T}\right)}$$

4. เปรียบเทียบวิธีการประมาณค่าพารามิเตอร์ โดยใช้เกณฑ์จากค่าความเอนเอียงและค่าความแปรปรวนของตัวประมาณ $\hat{\beta}_1$ จากทั้ง 4 วิธี โดยมีการทำซ้ำ 1,000 ครั้ง ดังนี้

$$\text{Bias}(\hat{\beta}_1) = \frac{\sum_{t=1}^{1000} \left|\hat{\beta}_{1t} - \beta_1\right|}{1000}$$

$$s^2(\hat{\beta}_1) = \frac{\sum_{t=1}^{1000} \left(\hat{\beta}_{1t} - \beta_1\right)^2}{1000}$$

โดยที่

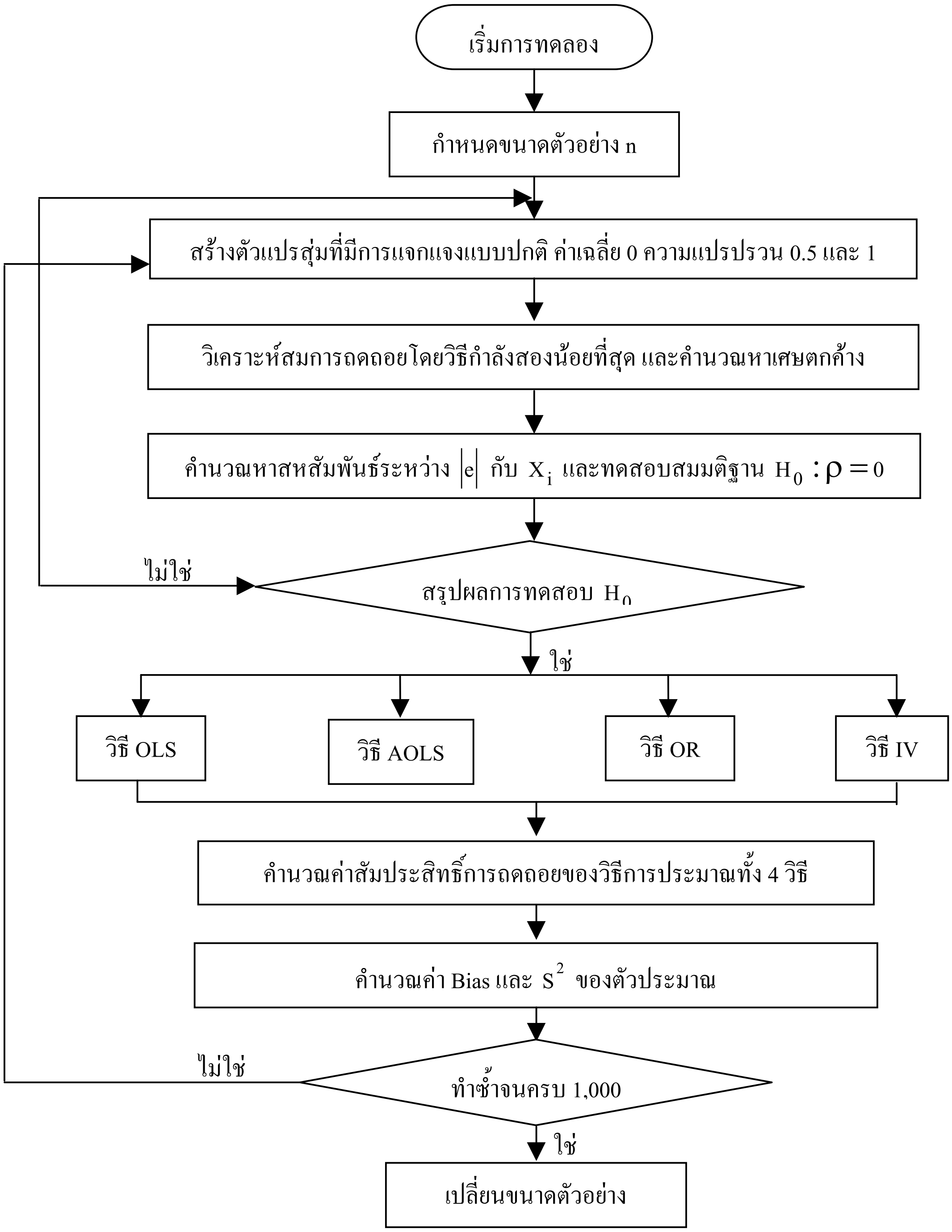
β_1 คือ ค่าพารามิเตอร์ที่กำหนดไว้ในการศึกษานี้

$\hat{\beta}_{1t}$ คือ ค่าประมาณ β_1 ในการทำซ้ำรอบที่ t

t คือ รอบที่ทำซ้ำ

5. สรุปผลการวิจัยว่าวิธีการประมาณใดเหมาะสมกับการประมาณค่าพารามิเตอร์ในกรณีที่ตัวแปรอิสระที่วัดค่ามามีความคลาดเคลื่อน เปรียบเทียบประสิทธิภาพของวิธีการประมาณค่าพารามิเตอร์ทั้ง 4 วิธี โดยพิจารณาจากค่าความเอนเอียง (Bias($\hat{\beta}_1$)) และค่าความแปรปรวน ($s^2(\hat{\beta}_1)$) เลือกวิธีที่ให้ค่าความเอนเอียงและค่าความแปรปรวนต่ำที่สุด ซึ่งถือว่าเป็นตัวประมาณที่ดีที่สุด

ขั้นตอนในการทำงานของโปรแกรม



ภาพที่ 1 แผนผังแสดงขั้นตอนการทำงานของโปรแกรม

<u>สถานที่และระยะเวลาทำการวิจัย</u>

ทำการวิจัย ณ ห้องปฏิบัติการคอมพิวเตอร์ของภาควิชาสถิติ คณะวิทยาศาสตร์ มหาวิทยาลัยเกษตรศาสตร์ โดยมีระยะเวลาในการวิจัย ตั้งแต่เดือนพฤศจิกายน 2546 ถึงเดือนกันยายน 2548

ผลและวิจารณ์

การวิจัยครั้งนี้เป็นการศึกษาเปรียบเทียบวิธีการประมาณค่าพารามิเตอร์ในตัวแบบการถดถอยเชิงเส้นอย่างง่ายเมื่อค่าตัวแปรอิสระวัดมาโดยมีความคลาดเคลื่อน 4 วิธี คือ วิธีกำลังสองน้อยที่สุด วิธีปรับตัวประมาณกำลังสองน้อยที่สุด วิธี Orthogonal Regression และวิธี Instrumental Variables โดยศึกษาค่าประมาณสัมประสิทธิ์การถดถอยที่คำนวณโดยวิธีดังกล่าว ในแต่ละสถานการณ์ที่จำลองขึ้น และจะพิจารณาว่าวิธีการประมาณใดให้ค่าความเอนเอียงและค่าความแปรปรวนของค่าประมาณพารามิเตอร์ต่ำสุด จะถือว่าวิธีประมาณดังกล่าวเป็นวิธีที่เหมาะสมที่สุดสำหรับการประมาณในสถานการณ์นั้น

นำเสนอผลการวิจัยครั้งนี้ โดยแบ่งเป็น 2 ส่วน ได้แก่

1. การเปรียบเทียบประสิทธิภาพของวิธีการประมาณค่าพารามิเตอร์ในตัวแบบการถดถอยเชิงเส้นอย่างง่ายเมื่อค่าตัวแปรอิสระวัดมาโดยมีความคลาดเคลื่อน 4 วิธี โดยใช้เกณฑ์จากค่าความเอนเอียงของค่าประมาณพารามิเตอร์ ($\text{Bias}(\hat{\beta}_1)$) ถ้าผลการเปรียบเทียบพบว่าวิธีการประมาณใดให้ค่าความเอนเอียงต่ำที่สุดจะมีประสิทธิภาพสูงสุด

2. การเปรียบเทียบประสิทธิภาพของวิธีการประมาณค่าพารามิเตอร์ในตัวแบบการถดถอยเชิงเส้นอย่างง่ายเมื่อค่าตัวแปรอิสระวัดมาโดยมีความคลาดเคลื่อน 4 วิธี โดยใช้เกณฑ์จากค่าความแปรปรวนของค่าประมาณพารามิเตอร์ ($s^2(\hat{\beta}_1)$) ถ้าผลการเปรียบเทียบพบว่าวิธีการประมาณใดให้ค่าความแปรปรวนต่ำที่สุดจะมีประสิทธิภาพสูงสุด

สำหรับการนำเสนอในรูปตาราง จะใช้สัญลักษณ์ต่อไปนี้แทนความหมายต่างๆ คือ

n หมายถึง ขนาดของกลุ่มตัวอย่าง

ρ หมายถึง ค่าสัมประสิทธิ์สหสัมพันธ์ของ X และ T

OLS หมายถึง วิธีกำลังสองน้อยที่สุด

AOLS หมายถึง วิธีปรับตัวประมาณกำลังสองน้อยที่สุด

OR หมายถึง วิธี Orthogonal Regression

IV หมายถึง วิธี Instrumental Variables

Bias($\hat{\beta}_1$) หมายถึง ค่าความเอนเอียงของตัวประมาณ $\hat{\beta}_1$

$s^2(\hat{\beta}_1)$ หมายถึง ค่าความแปรปรวนของตัวประมาณ $\hat{\beta}_1$

ส่วนการนำเสนอในรูปกราฟ จะใช้สัญลักษณ์ต่อไปนี้แทนความหมายต่างๆ คือ

…◊… หมายถึง วิธีกำลังสองน้อยที่สุด

—□— หมายถึง วิธีปรับตัวประมาณกำลังสองน้อยที่สุด

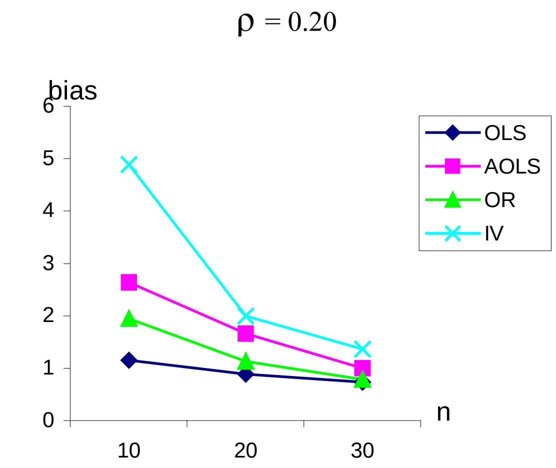
—Δ— หมายถึง วิธี Orthogonal Regression

--×-- หมายถึง วิธี Instrumental Variables

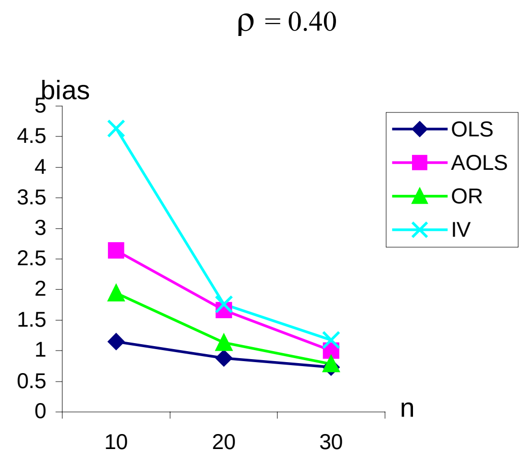
จากการทดลองทำซ้ำ 1,000 ครั้ง ในแต่ละสถานการณ์ โดยการประมาณค่าพารามิเตอร์ทั้ง 4 วิธี คำนวณค่า Bias($\hat{\beta}_1$) และ $s^2(\hat{\beta}_1)$ ได้ผลการวิเคราะห์สรุปไว้ในตารางที่ 1 สำหรับกลุ่มตัวอย่างขนาดเล็ก คือ n = 10, 20 และ 30 ตามลำดับ รายละเอียดของผลการวิเคราะห์สรุปได้ดังนี้

ตารางที่ 1 แสดงค่าความเอนเอียงและค่าความแปรปรวนของค่าประมาณพารามิเตอร์ที่คำนวณโดยวิธีการประมาณทั้ง 4 วิธี เมื่อกลุ่มตัวอย่างขนาดเล็ก

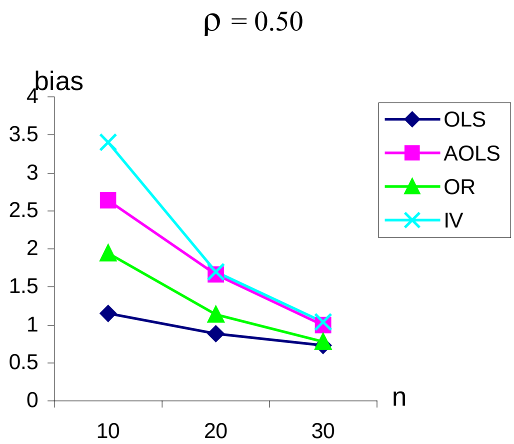
n	วิธีการประมาณค่า	ρ	Bias($\hat{\beta}_1$)	$s^2(\hat{\beta}_1)$
10	OLS	-	1.1478	2.2094
	AOLS	-	2.6417	12.2784
	OR	-	1.9431	9.5799
	IV	0.20	4.8884	88.6902
		0.40	4.6302	254.944
		0.50	3.4030	57.2363
		0.60	3.0444	24.3916
		0.80	4.9214	335.785
		0.90	8.0272	1298.27
20	OLS	-	0.8831	1.2456
	AOLS	-	1.6619	6.0705
	OR	-	1.1336	6.9405
	IV	0.20	1.9919	12.5442
		0.40	1.7555	8.6316
		0.50	1.6982	6.8675
		0.60	1.2686	4.2103
		0.80	1.2584	2.7180
		0.90	1.2482	3.0652
30	OLS	-	0.7304	0.8069
	AOLS	-	0.9974	1.8708
	OR	-	0.7821	1.5540
	IV	0.20	1.3693	4.7458
		0.40	1.1777	2.6105
		0.50	1.0408	1.8067
		0.60	0.9144	1.4197
		0.80	1.0865	2.9448
		0.90	0.8451	1.1972



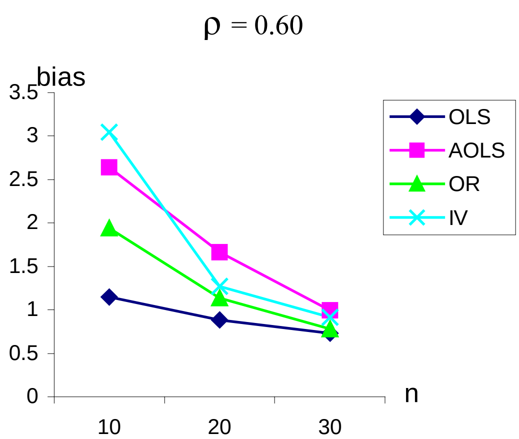
ภาพที่ 2 แสดงค่า Bias($\hat{\beta}_1$) เมื่อ n = 10, 20, 30



ภาพที่ 3 แสดงค่า Bias($\hat{\beta}_1$) เมื่อ n = 10, 20, 30

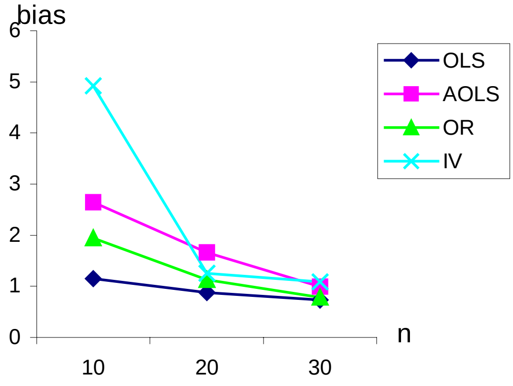


ภาพที่ 4 แสดงค่า Bias($\hat{\beta}_1$) เมื่อ n = 10, 20, 30



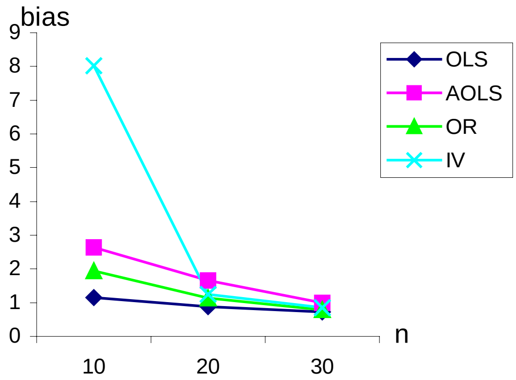
ภาพที่ 5 แสดงค่า Bias($\hat{\beta}_1$) เมื่อ n = 10, 20, 30

ρ = 0.80



ภาพที่ 6 แสดงค่า Bias($\hat{\beta}_1$) เมื่อ n = 10, 20, 30

ρ = 0.90



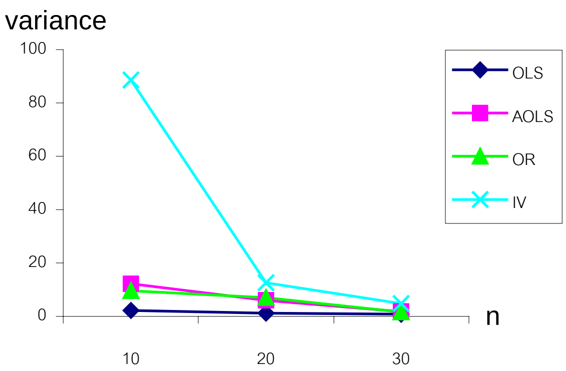
ภาพที่ 7 แสดงค่า Bias($\hat{\beta}_1$) เมื่อ n = 10, 20, 30

เกณฑ์จากค่าความเอนเอียง (กรณีตัวอย่างขนาดเล็ก n = 10, 20 และ 30)

จากตารางที่ 1 และภาพที่ 2 - 7 แสดงค่าความเอนเอียงของค่าประมาณพารามิเตอร์ที่คำนวณโดยวิธีการประมาณทั้ง 4 วิธี จากการศึกษาพบว่า เมื่อขนาดตัวอย่างเพิ่มขึ้น วิธีการประมาณทั้ง 4 วิธี ให้ค่าความเอนเอียงลดลง สำหรับวิธี Instrumental Variables กรณีขนาดตัวอย่างเท่ากับ 10 พบว่า เมื่อค่าสัมประสิทธิ์สหสัมพันธ์ของ X และ T มีค่าเพิ่มขึ้น ไม่แสดงให้เห็นว่ามีแนวโน้มเพิ่มขึ้นหรือลดลง ส่วนกรณีขนาดตัวอย่างเท่ากับ 20 และ 30 พบว่า เมื่อค่าสัมประสิทธิ์สหสัมพันธ์ของ X และ T มีค่าเพิ่มขึ้น ค่าความเอนเอียงมีแนวโน้มลดลง

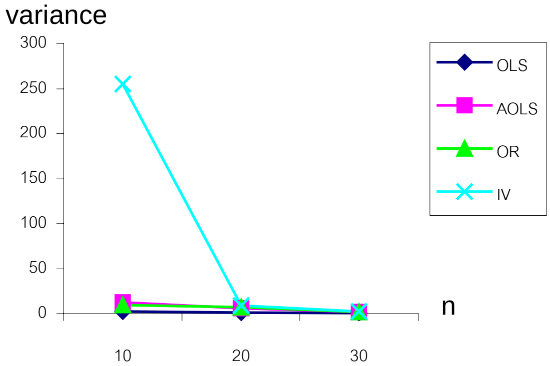
การพิจารณาโดยใช้เกณฑ์จากค่าความเอนเอียง พบว่า เมื่อขนาดตัวอย่างเท่ากับ 10 และ 20 วิธีกำลังสองน้อยที่สุดให้ค่าความเอนเอียงต่ำที่สุด รองลงมาคือ วิธี Orthogonal Regression และวิธีปรับตัวประมาณกำลังสองน้อยที่สุด ส่วนกรณีขนาดตัวอย่างมีค่าเท่ากับ 30 พบว่า ทุกวิธีให้ค่าความเอนเอียงใกล้เคียงกัน

$\rho = 0.20$

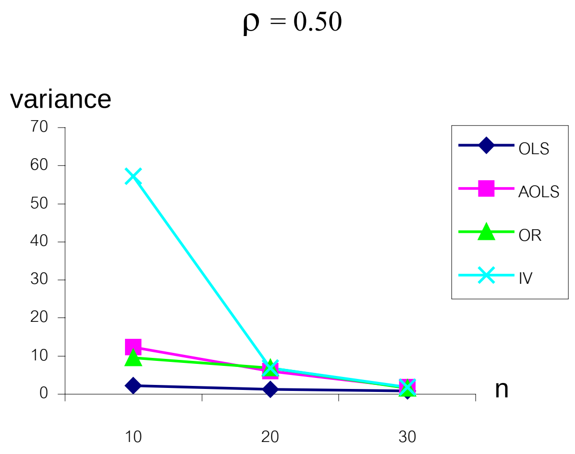


ภาพที่ 8 แสดงค่า $s^2(\hat{\beta}_1)$ เมื่อ $n = 10, 20, 30$

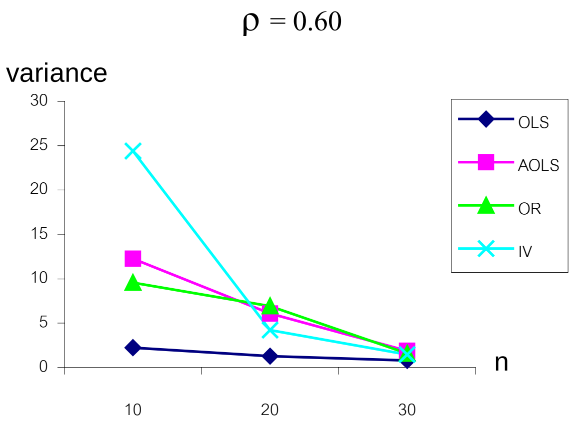
$\rho = 0.40$



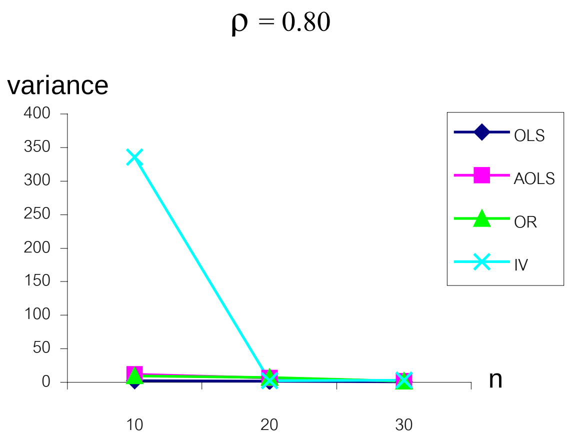
ภาพที่ 9 แสดงค่า $s^2(\hat{\beta}_1)$ เมื่อ $n = 10, 20, 30$



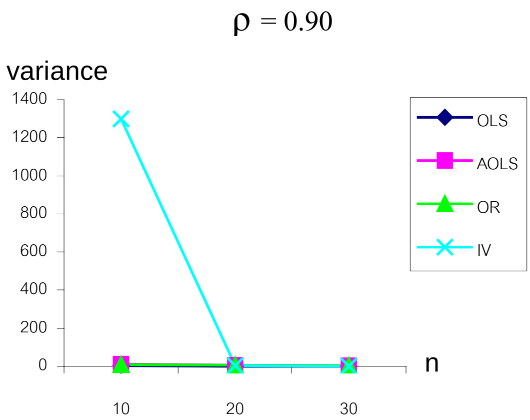
ภาพที่ 10 แสดงค่า $s^2(\hat{\beta}_1)$ เมื่อ n = 10, 20, 30



ภาพที่ 11 แสดงค่า $s^2(\hat{\beta}_1)$ เมื่อ n = 10, 20, 30



ภาพที่ 12 แสดงค่า $s^2(\hat{\beta}_1)$ เมื่อ n = 10, 20, 30



ภาพที่ 13 แสดงค่า $s^2(\hat{\beta}_1)$ เมื่อ n = 10, 20, 30

<u>เกณฑ์จากค่าความแปรปรวน</u> (กรณีตัวอย่างขนาดเล็ก n = 10, 20 และ 30)

จากตารางที่ 1 และภาพที่ 8 - 13 แสดงค่าความแปรปรวนของค่าประมาณพารามิเตอร์ที่คำนวณโดยวิธีการประมาณทั้ง 4 วิธี จากการศึกษา พบว่า เมื่อขนาดตัวอย่างเพิ่มขึ้น วิธีการประมาณทั้ง 4 วิธี ให้ค่าความแปรปรวนลดลง สำหรับวิธี Instrumental Variables กรณีขนาดตัวอย่างเท่ากับ 10 พบว่า ค่าความแปรปรวนค่อนข้างสูง และเมื่อค่าสัมประสิทธิ์สหสัมพันธ์ของ X และ T มีค่าเพิ่มขึ้น ค่าความแปรปรวนไม่แสดงให้เห็นว่ามีแนวโน้มเพิ่มขึ้นหรือลดลง ส่วนกรณีขนาดตัวอย่างเท่ากับ 20 และ 30 พบว่า เมื่อค่าสัมประสิทธิ์สหสัมพันธ์ของ X และ T มีค่าเพิ่มขึ้น ค่าความแปรปรวนมีแนวโน้มลดลง

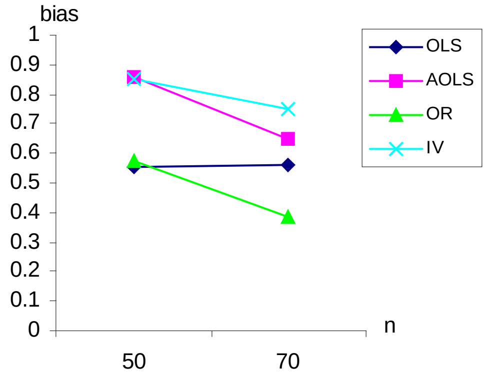
การพิจารณาโดยใช้เกณฑ์จากค่าความแปรปรวน พบว่า เมื่อขนาดตัวอย่างเท่ากับ 10 และ 20 วิธีกำลังสองน้อยที่สุดให้ค่าความแปรปรวนต่ำสุด รองลงมาคือ วิธี Orthogonal Regression และวิธีปรับตัวประมาณกำลังสองน้อยที่สุด ส่วนกรณีขนาดตัวอย่างเท่ากับ 30 ทุกวิธีให้ค่าความแปรปรวนใกล้เคียงกัน

จากการทดลองทำซ้ำ 1,000 ครั้ง ในแต่ละสถานการณ์ โดยการประมาณค่าพารามิเตอร์ทั้ง 4 วิธี คำนวณค่า Bias($\hat{\beta}_1$) และ $s^2(\hat{\beta}_1)$ ได้ผลการวิเคราะห์สรุปไว้ในตารางที่ 2 สำหรับกลุ่มตัวอย่างขนาดกลาง คือ n = 50 และ 70 ตามลำดับ รายละเอียดของผลการวิเคราะห์สรุปได้ดังนี้

ตารางที่ 2 แสดงค่าความเอนเอียงและค่าความแปรปรวนของค่าประมาณพารามิเตอร์ที่คำนวณโดยวิธีการประมาณทั้ง 4 วิธี เมื่อกลุ่มตัวอย่างขนาดกลาง

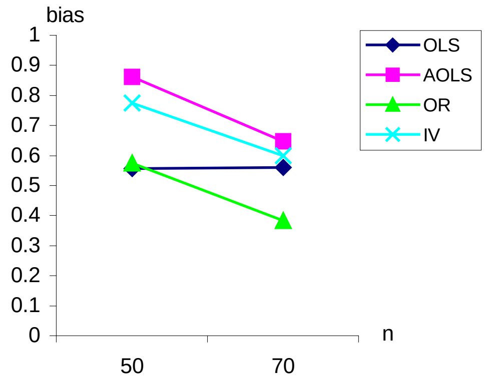
n	วิธีการประมาณค่า	ρ	Bias($\hat{\beta}_1$)	$S^2(\hat{\beta}_1)$
50	OLS	-	0.5562	0.4807
	AOLS	-	0.8614	1.2861
	OR	-	0.5743	1.0409
	IV	0.20	0.8544	1.1893
		0.40	0.7752	0.9372
		0.50	0.7332	0.8250
		0.60	0.6454	0.6828
		0.80	0.6663	0.7399
		0.90	0.6466	0.6716
70	OLS	-	0.5598	0.4633
	AOLS	-	0.6464	0.7083
	OR	-	0.3833	0.2907
	IV	0.20	0.7499	0.8870
		0.40	0.5990	0.5774
		0.50	0.6001	0.5397
		0.60	0.5628	0.56794
		0.80	0.4980	0.3976
		0.90	0.5392	0.4492

ρ = 0.20



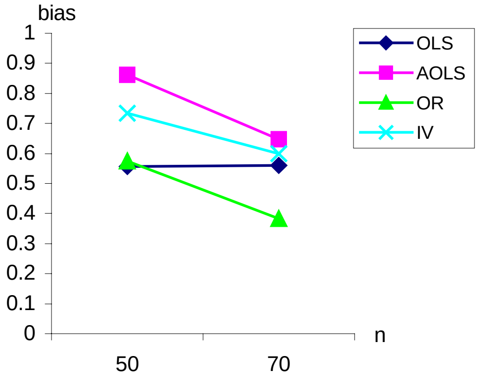
ภาพที่ 14 แสดงค่า Bias($\hat{\beta}_1$) เมื่อ n = 50, 70

ρ = 0.40



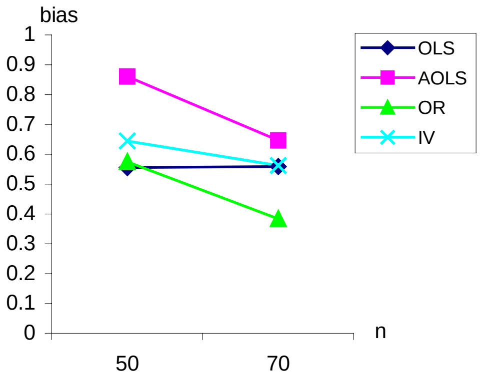
ภาพที่ 15 แสดงค่า Bias($\hat{\beta}_1$) เมื่อ n = 50, 70

ρ = 0.50

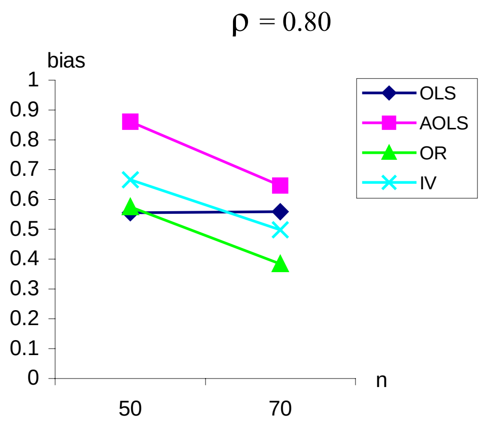


ภาพที่ 16 แสดงค่า Bias($\hat{\beta}_1$) เมื่อ n = 50, 70

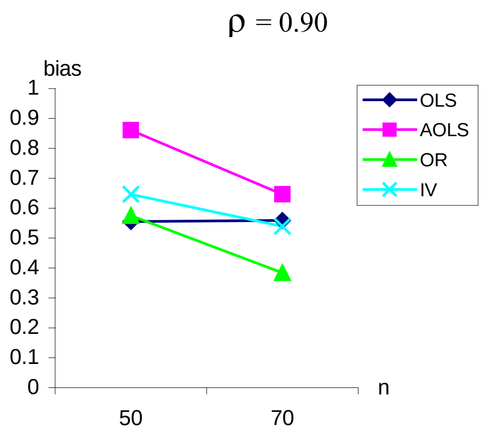
ρ = 0.60



ภาพที่ 17 แสดงค่า Bias($\hat{\beta}_1$) เมื่อ n = 50, 70



ภาพที่ 18 แสดงค่า Bias($\hat{\beta}_1$) เมื่อ n = 50, 70

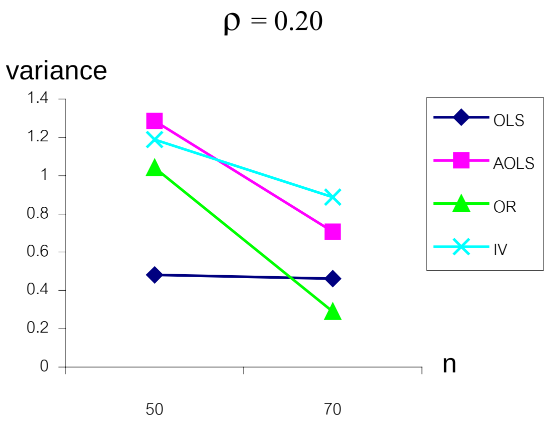


ภาพที่ 19 แสดงค่า Bias($\hat{\beta}_1$) เมื่อ n = 50, 70

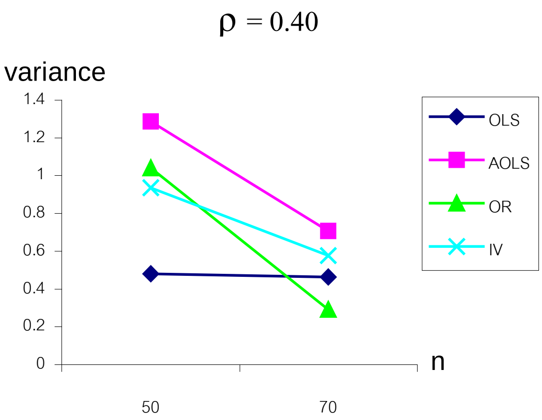
<u>เกณฑ์จากค่าความเอนเอียง</u> (กรณีตัวอย่างขนาดกลาง n = 50 และ 70)

จากตารางที่ 2 และภาพที่ 14 - 19 แสดงค่าความเอนเอียงของค่าประมาณพารามิเตอร์ที่คำนวณโดยวิธีการประมาณทั้ง 4 วิธี จากการศึกษา พบว่า เมื่อขนาดตัวอย่างเพิ่มขึ้น วิธีการประมาณทั้ง 4 วิธี ให้ค่าความเอนเอียงลดลง สำหรับวิธี Instrumental Variables พบว่าเมื่อค่าสัมประสิทธิ์สหสัมพันธ์ของ X และ T มีค่าเพิ่มขึ้น ค่าความเอนเอียงมีแนวโน้มลดลงเช่นกัน

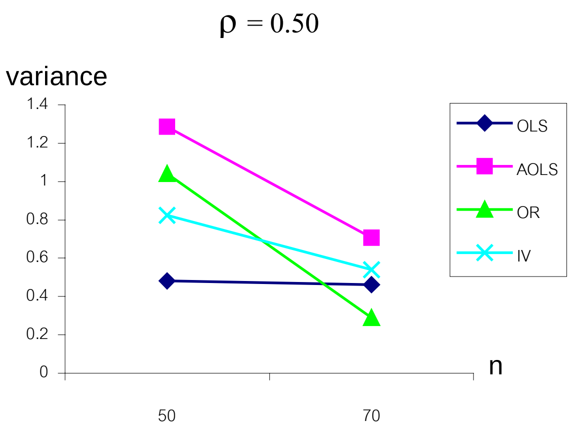
การพิจารณาโดยใช้เกณฑ์จากค่าความเอนเอียง พบว่า เมื่อขนาดตัวอย่างเท่ากับ 50 วิธีกำลังสองน้อยที่สุดให้ค่าความเอนเอียงใกล้เคียงกับวิธี Orthogonal Regression และต่ำกว่าวิธีอื่นรองลงมาคือ วิธี Instrumental Variables ทั้งนี้เมื่อขนาดตัวอย่างมีค่าเพิ่มขึ้นเป็น 70 พบว่า ทุกวิธีให้ค่าความเอนเอียงใกล้เคียงกัน



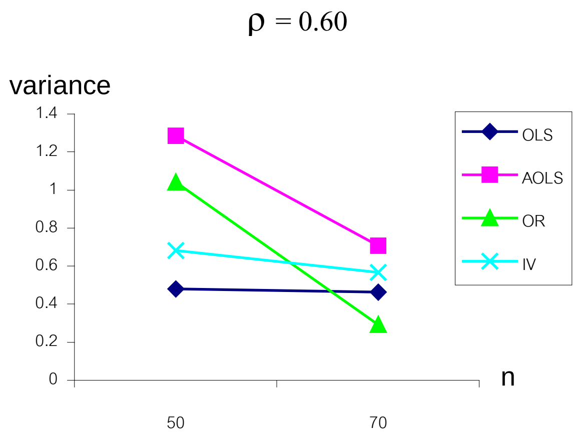
<u>ภาพที่ 20</u> แสดงค่า $s^2(\hat{\beta}_1)$ เมื่อ n = 50, 70



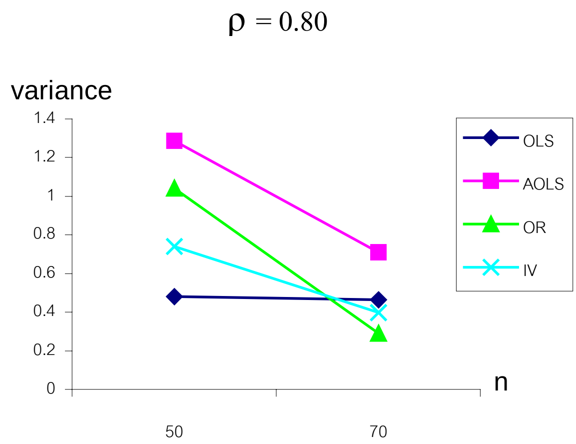
ภาพที่ 21 แสดงค่า $s^2(\hat{\beta}_1)$ เมื่อ $n = 50, 70$



ภาพที่ 22 แสดงค่า $s^2(\hat{\beta}_1)$ เมื่อ $n = 50, 70$

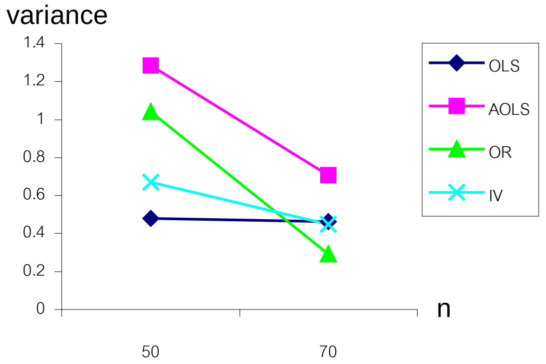


ภาพที่ 23 แสดงค่า $s^2(\hat{\beta}_1)$ เมื่อ n = 50, 70



ภาพที่ 24 แสดงค่า $s^2(\hat{\beta}_1)$ เมื่อ n = 50, 70

ρ = 0.90



ภาพที่ 25 แสดงค่า $s^2(\hat{\beta}_1)$ เมื่อ n = 50, 70

เกณฑ์จากค่าความแปรปรวน (กรณีตัวอย่างขนาดกลาง n = 50 และ 70)

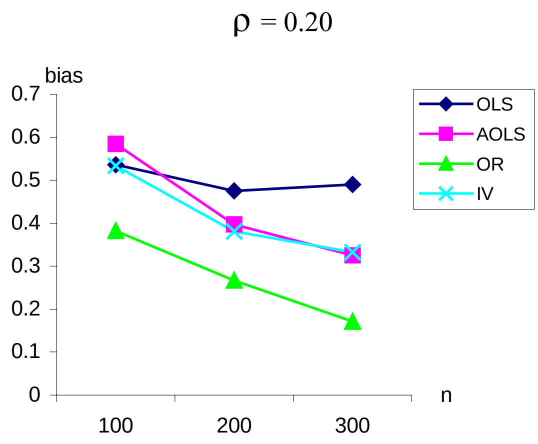
จากตารางที่ 2 และภาพที่ 20 - 25 แสดงค่าความแปรปรวนของค่าประมาณพารามิเตอร์ที่คำนวณโดยวิธีการประมาณทั้ง 4 วิธี จากการศึกษา พบว่า เมื่อขนาดตัวอย่างเพิ่มขึ้น วิธีการประมาณทั้ง 4 วิธี ให้ค่าความแปรปรวนลดลง สำหรับวิธี Instrumental Variables พบว่าเมื่อค่าสัมประสิทธิ์สหสัมพันธ์ของ X และ T มีค่าเพิ่มขึ้น ค่าความแปรปรวนมีแนวโน้มลดลงเช่นกัน

การพิจารณาโดยใช้เกณฑ์จากค่าความแปรปรวน พบว่า เมื่อขนาดตัวอย่างเท่ากับ 50 วิธีกำลังสองน้อยที่สุดให้ค่าความแปรปรวนต่ำสุด รองลงมาคือ วิธี Instrumental Variables และวิธี Orthogonal Regression ทั้งนี้เมื่อขนาดตัวอย่างมีค่าเพิ่มขึ้นเป็น 70 พบว่า ทุกวิธีให้ค่าความแปรปรวนใกล้เคียงกัน

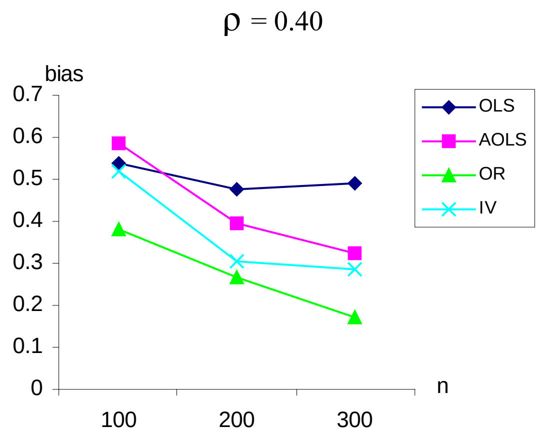
จากการทดลองทำซ้ำ 1,000 ครั้ง ในแต่ละสถานการณ์ โดยการประมาณค่าพารามิเตอร์ทั้ง 4 วิธี คำนวณค่า Bias($\hat{\beta}_1$) และ $s^2(\hat{\beta}_1)$ ได้ผลการวิเคราะห์สรุปไว้ในตารางที่ 3 สำหรับกลุ่มตัวอย่างขนาดใหญ่ คือ n = 100, 200 และ 300 ตามลำดับ รายละเอียดของผลการวิเคราะห์สรุปได้ดังนี้

ตารางที่ 3 แสดงค่าความเอนเอียงและค่าความแปรปรวนของค่าประมาณพารามิเตอร์ที่คำนวณ โดยวิธีการประมาณทั้ง 4 วิธี เมื่อกลุ่มตัวอย่างขนาดใหญ่

n	วิธีการประมาณค่า	ρ	Bias($\hat{\beta}_1$)	$s^2(\hat{\beta}_1)$
100	OLS	-	0.5362	0.3929
	AOLS	-	0.5843	0.4869
	OR	-	0.3819	0.2137
	IV	0.20	0.5333	0.4893
		0.40	0.5206	0.4308
		0.50	0.4826	0.3578
		0.60	0.4526	0.3312
		0.80	0.4717	0.3906
		0.90	0.3911	0.2349
200	OLS	-	0.4758	0.2963
	AOLS	-	0.3975	0.2529
	OR	-	0.2672	0.1245
	IV	0.20	0.3810	0.2256
		0.40	0.3032	0.1613
		0.50	0.3438	0.1746
		0.60	0.3519	0.1914
		0.80	0.3195	0.1619
		0.90	0.2883	0.1297
300	OLS	-	0.4908	0.2822
	AOLS	-	0.3252	0.1760
	OR	-	0.1712	0.0488
	IV	0.20	0.3327	0.1805
		0.40	0.2837	0.1214
		0.50	0.2763	0.1275
		0.60	0.2581	0.1091
		0.80	0.2458	0.0881
		0.90	0.2544	0.1019

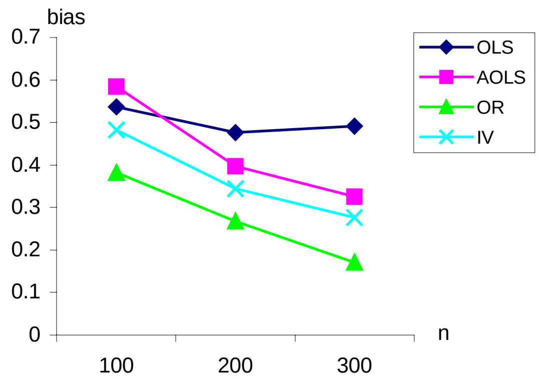


ภาพที่ 26 แสดงค่า Bias($\hat{\beta}_1$) เมื่อ n = 100, 200, 300



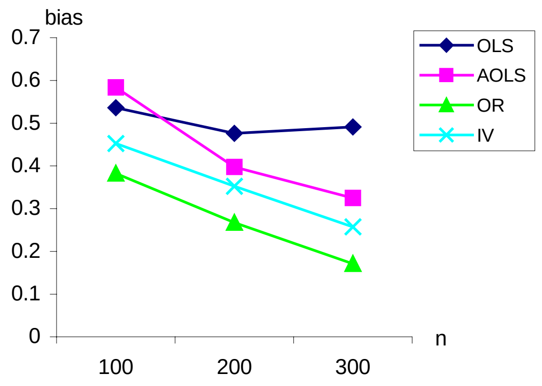
ภาพที่ 27 แสดงค่า Bias($\hat{\beta}_1$) เมื่อ n = 100, 200, 300

ρ = 0.50



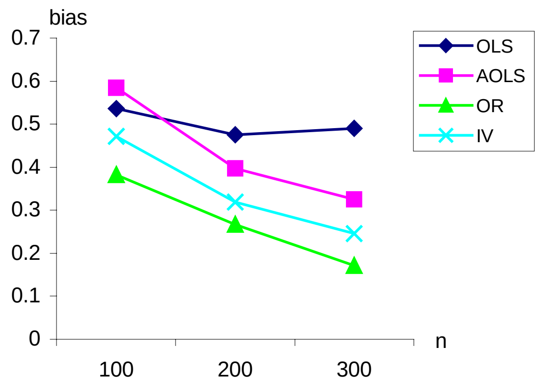
ภาพที่ 28 แสดงค่า Bias($\hat{\beta}_1$) เมื่อ n = 100, 200, 300

ρ = 0.60



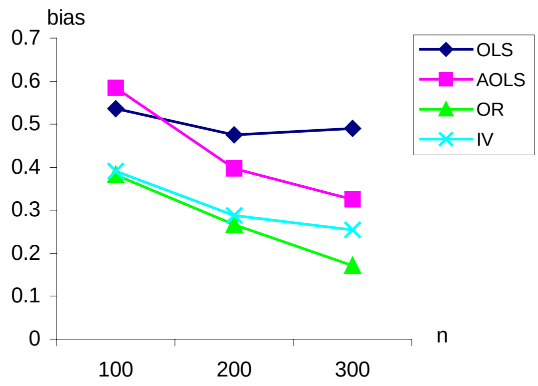
ภาพที่ 29 แสดงค่า Bias($\hat{\beta}_1$) เมื่อ n = 100, 200, 300

ρ = 0.80



ภาพที่ 30 แสดงค่า Bias($\hat{\beta}_1$) เมื่อ n = 100, 200, 300

ρ = 0.90



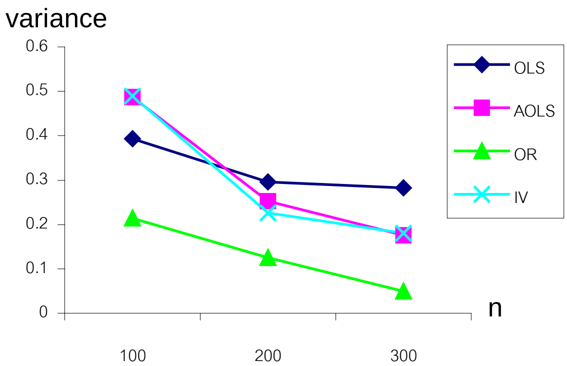
ภาพที่ 31 แสดงค่า Bias($\hat{\beta}_1$) เมื่อ n = 100, 200, 300

<u>เกณฑ์จากค่าความเอนเอียง</u> (กรณีตัวอย่างขนาดใหญ่ n = 100, 200 และ 300)

จากตารางที่ 3 และภาพที่ 26 - 31 แสดงค่าความเอนเอียงของค่าประมาณพารามิเตอร์ที่คำนวณโดยวิธีการประมาณทั้ง 4 วิธี จากการศึกษา พบว่า เมื่อขนาดตัวอย่างเพิ่มขึ้น วิธีการประมาณทั้ง 4 วิธี ให้ค่าความเอนเอียงลดลง สำหรับวิธี Instrumental Variables พบว่าเมื่อค่าสัมประสิทธิ์สหสัมพันธ์ของ X และ T มีค่าเพิ่มขึ้น ค่าความเอนเอียงมีแนวโน้มลดลงเช่นกัน

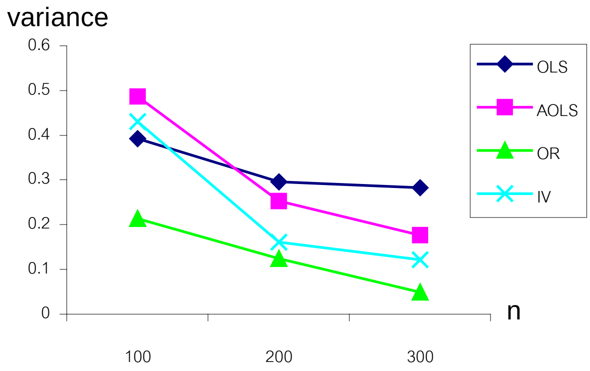
การพิจารณาโดยใช้เกณฑ์จากค่าความเอนเอียง พบว่า วิธี Orthogonal Regression ให้ค่าความเอนเอียงต่ำที่สุดและให้ค่าความเอนเอียงใกล้เคียงกับวิธีอื่นสำหรับทุกขนาดตัวอย่าง เมื่อขนาดตัวอย่างมีค่าเพิ่มขึ้น วิธีปรับตัวประมาณกำลังสองน้อยที่สุดให้ค่าความเอนเอียงต่ำกว่าวิธีกำลังสองน้อยที่สุด และเมื่อขนาดตัวอย่างมีค่าเพิ่มขึ้นทุกวิธีให้ค่าความเอนเอียงใกล้เคียงกัน

$\rho = 0.20$



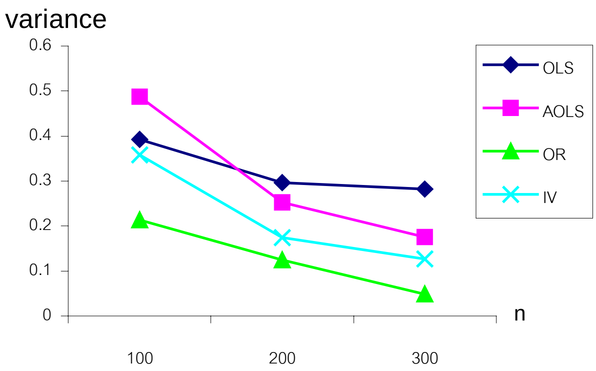
<u>ภาพที่ 32</u> แสดงค่า $s^2(\hat{\beta}_1)$ เมื่อ n = 100, 200, 300

ρ = 0.40



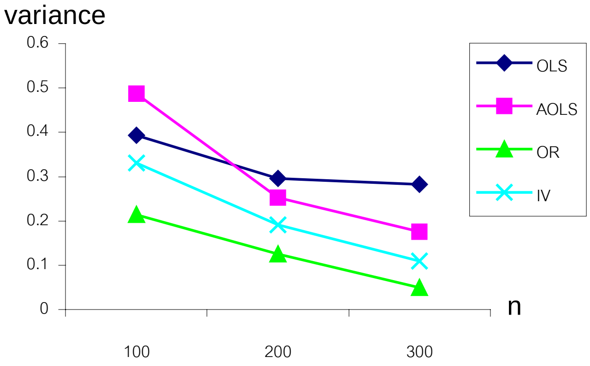
<u>ภาพที่ 33</u> แสดงค่า $s^2(\hat{\beta}_1)$ เมื่อ n = 100, 200, 300

ρ = 0.50



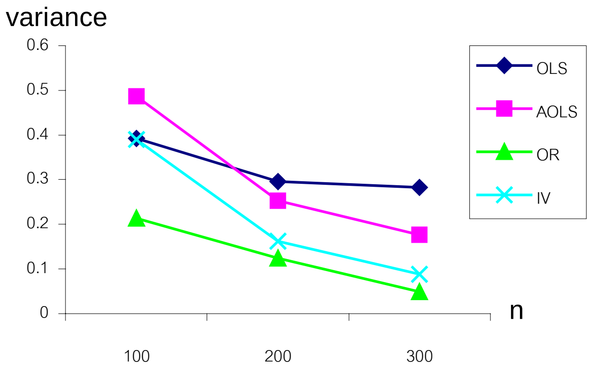
<u>ภาพที่ 34</u> แสดงค่า $s^2(\hat{\beta}_1)$ เมื่อ n = 100, 200, 300

$\rho = 0.60$

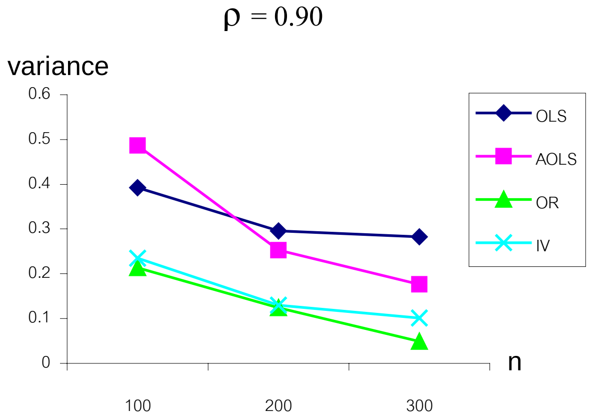


ภาพที่ 35 แสดงค่า $s^2(\hat{\beta}_1)$ เมื่อ n = 100, 200, 300

$\rho = 0.80$



ภาพที่ 36 แสดงค่า $s^2(\hat{\beta}_1)$ เมื่อ n = 100, 200, 300



ภาพที่ 37 แสดงค่า $s^2(\hat{\beta}_1)$ เมื่อ n = 100, 200, 300

เกณฑ์จากค่าความแปรปรวน (กรณีตัวอย่างขนาดใหญ่ n = 100, 200 และ 300)

จากตารางที่ 3 และภาพที่ 32 - 37 แสดงค่าความแปรปรวนของค่าประมาณพารามิเตอร์ที่คำนวณโดยวิธีการประมาณทั้ง 4 วิธี จากการศึกษา พบว่า เมื่อขนาดตัวอย่างเพิ่มขึ้น วิธีการประมาณทั้ง 4 วิธี ให้ค่าความแปรปรวนลดลง สำหรับวิธี Instrumental Variables พบว่าเมื่อค่าสัมประสิทธิ์สหสัมพันธ์ของ X และ T มีค่าเพิ่มขึ้น ค่าความแปรปรวนมีแนวโน้มลดลงเช่นกัน

การพิจารณาโดยใช้เกณฑ์จากค่าความแปรปรวน พบว่า วิธี Orthogonal Regression ให้ค่าความแปรปรวนต่ำสุดและให้ค่าความแปรปรวนใกล้เคียงกับวิธีอื่นสำหรับทุกขนาดตัวอย่าง เมื่อขนาดตัวอย่างมีค่าเพิ่มขึ้น วิธีปรับตัวประมาณกำลังสองน้อยที่สุดให้ค่าความแปรปรวนต่ำกว่าวิธีกำลังสองน้อยที่สุด และเมื่อขนาดตัวอย่างมีค่าเพิ่มขึ้นทุกวิธีให้ค่าความแปรปรวนใกล้เคียงกัน

สรุปและข้อเสนอแนะ

การวิจัยครั้งนี้ได้ทำการศึกษาปัญหา สาเหตุ และผลกระทบที่เกิดขึ้นในการวิเคราะห์การถดถอย เมื่อตัวแปรอิสระถูกวัดมาโดยมีความคลาดเคลื่อน จากการศึกษาพบว่า ค่าสังเกตหรือค่าที่เกิดจากการวัดของตัวแปรอิสระที่เก็บรวบรวมเพื่อนำมาวิเคราะห์การถดถอยอาจไม่สามารถวัดค่าได้โดยไม่มีความ คลาดเคลื่อนตามที่ตัวแบบกำหนดไว้ เนื่องจากสาเหตุต่างๆ เช่น ความผิดพลาดที่เกิดขึ้นในการเก็บรวบรวมข้อมูล ซึ่งอาจเกิดขึ้นได้ในหลายๆ ขั้นตอน ไม่ว่าจะเป็นขั้นการวางแผนเตรียมการ ขั้นปฏิบัติงานภาคสนาม ขั้นการประมวลผลและการนำเสนอผล การนำตัวแปรทดแทนมาใช้แทนตัวแปรอิสระเนื่องจากตัวแปรอิสระวัดค่าได้ยากหรือไม่สามารถวัดค่าตัวแปรอิสระได้โดยตรงและไม่สามารถประเมินค่าเป็นตัวเลขได้ เช่น ระดับการศึกษา เป็นต้น การใช้ตัวแปรเชิงคุณภาพแทนตัวแปรเชิงปริมาณมักทำให้เกิดความคลาดเคลื่อนในข้อมูลเพราะตัวแปรที่แท้จริงนั้นมีค่าต่อเนื่อง เมื่อจำเป็นต้องใช้ตัวแปรเพื่อแบ่งกลุ่มเป็นมาตรวัดแทนค่าตัวแปรที่มีค่าต่อเนื่องความผิดพลาดจึงเกิดขึ้น การใช้เลขดัชนีเป็นค่าของตัวแปรแทนที่จะใช้ค่าที่แท้จริงของตัวแปร เลขดัชนีเป็นตัวเลขที่รวมความคลาดเคลื่อนไว้ในตัวเองอยู่แล้วไม่ว่าจะเป็นการคำนวณเลขดัชนีซึ่งต้องอาศัยการกำหนดปีฐานที่ถูกต้อง การอาศัยค่าที่ได้จากการสำรวจ เช่น ปริมาณ ราคา รวมถึงค่าประมาณที่ปัดเศษทิ้ง เป็นต้น เมื่อนำตัวแปรอิสระที่ถูกวัดมาโดยมีความคลาดเคลื่อนมาใช้ในการวิเคราะห์การถดถอยทำให้ค่าประมาณที่ได้มีความเอนเอียงและมีความแปรปรวนสูง เพื่อให้ค่าประมาณที่ได้ไม่มีความเอนเอียงและมีความแปรปรวนต่ำ ผู้วิจัยได้ศึกษาวิธีการประมาณค่าพารามิเตอร์ และเปรียบเทียบวิธีการประมาณค่าพารามิเตอร์ในตัวแบบการถดถอยเชิงเส้นอย่างง่ายเมื่อวัดค่าตัวแปรอิสระโดยมีความคลาดเคลื่อน เพื่อหาวิธีประมาณพารามิเตอร์ที่เหมาะสม ในกรณีที่ตัวแปรอิสระถูกวัดมาโดยมีความคลาดเคลื่อน เมื่อตัวอย่างมีขนาดต่างๆ กัน โดยวิธีการประมาณค่าพารามิเตอร์4 วิธี ได้แก่

1. วิธีกำลังสองน้อยที่สุด

2. วิธีปรับตัวประมาณกำลังสองน้อยที่สุด

3. วิธี Orthogonal Regression

4. วิธี Instrumental Variables

ในการศึกษาวิธีการประมาณค่าพารามิเตอร์ในตัวแบบการถดถอยเชิงเส้นอย่างง่าย เมื่อตัวแปรอิสระถูกวัดมาโดยมีความคลาดเคลื่อน กำหนดขอบเขตการวิจัยดังนี้

1. กำหนดตัวแบบความคลาดเคลื่อนในการวิเคราะห์การถดถอยเชิงเส้นอย่างง่ายเมื่อวัดค่าตัวแปรอิสระ คือ Classical Error Model

2. กำหนดขนาดตัวอย่างที่ใช้ เป็น 3 กลุ่ม คือ

   2.1 กลุ่มตัวอย่างขนาดเล็ก มีค่าเป็น 10, 20 และ 30

   2.2 กลุ่มตัวอย่างขนาดกลาง มีค่าเป็น 50 และ 70

   2.3 กลุ่มตัวอย่างขนาดใหญ่ มีค่าเป็น 100, 200 และ 300

3. กำหนด $\beta_0 = 0$ และ $\beta_1 = 1$

4. กำหนด $\alpha = 0.05$

ในการเปรียบเทียบเพื่อหาข้อสรุปว่าวิธีการประมาณใดดีที่สุดในสถานการณ์ต่างๆ ที่กำหนด จะพิจารณาจากค่าความเอนเอียงและค่าความแปรปรวนในการประมาณค่าพารามิเตอร์จากการจำลอง ถ้าวิธีการประมาณใดให้จากค่าความเอนเอียงและค่าความแปรปรวนต่ำสุด จะถือว่าวิธีการประมาณนั้นเป็นวิธีการที่เหมาะสมที่สุดในสถานการณ์นั้นๆ

ในการเปรียบเทียบค่าความเอนเอียงและค่าความแปรปรวนของค่าประมาณพารามิเตอร์ในตัวแบบ Classical Error Model โดยวิธีประมาณทั้ง 4 วิธี จากสถานการณ์การทดลองที่จำลองขึ้น 1,000 ครั้ง ในแต่ละสถานการณ์ สรุปผลได้ดังนี้ คือ

1. กรณีตัวอย่างมีขนาดเล็ก (n = 10, 20 และ 30)

วิธีกำลังสองน้อยที่สุด เป็นวิธีการที่เหมาะสมที่สุด โดยให้ค่าความเอนเอียงและค่าความแปรปรวนต่ำกว่าวิธีอื่น โดยเฉพาะเมื่อขนาดตัวอย่างเท่ากับ 10 และ 20 รองลงมาคือ วิธี Orthogonal Regression และวิธีปรับตัวประมาณกำลังสองน้อยที่สุด สำหรับขนาดตัวอย่างเท่ากับ 30 พบว่า ทุกวิธีให้ค่าความเอนเอียงและค่าความแปรปรวนใกล้เคียงกัน

2. กรณีตัวอย่างมีขนาดกลาง (n = 50 และ 70)

เมื่อขนาดตัวอย่างเท่ากับ 50 วิธีกำลังสองน้อยที่สุดให้ค่าความเอนเอียงใกล้เคียงกับวิธี Orthogonal Regression โดยให้ค่าความเอนเอียงและค่าความแปรปรวนต่ำกว่าวิธีอื่น รองลงมาคือ วิธี Instrumental Variables สำหรับขนาดตัวอย่างเท่ากับ 70 พบว่า ทุกวิธีให้ค่าความเอนเอียงและค่าความแปรปรวนใกล้เคียงกัน

3. กรณีตัวอย่างมีขนาดใหญ่ (n = 100, 200 และ 300)

วิธี Orthogonal Regression เป็นวิธีการที่เหมาะสมที่สุด โดยให้ค่าความเอนเอียงและค่าความแปรปรวนต่ำกว่าวิธีอื่น รองลงมาคือ วิธี Instrumental Variables และวิธีปรับตัวประมาณกำลังสองน้อยที่สุด เมื่อขนาดตัวอย่างเพิ่มขึ้น พบว่า ทุกวิธีให้ค่าความเอนเอียงและค่าความแปรปรวนใกล้เคียงกัน

จากค่าความเอนเอียงและค่าความแปรปรวนของวิธีประมาณทั้ง 4 วิธี จำแนกตามขนาดตัวอย่าง พบว่า กรณีตัวอย่างขนาดเล็ก เมื่อขนาดตัวอย่างเท่ากับ 10 และ 20 วิธีกำลังสองน้อยที่สุดเป็นวิธีที่ดีที่สุด ส่วนกรณีขนาดตัวอย่างเท่ากับ 30 ทุกวิธีให้ค่าความเอนเอียงและค่าความแปรปรวนใกล้เคียงกัน สำหรับกรณีตัวอย่างขนาดกลางและขนาดใหญ่ ทุกวิธีให้ค่าความเอนเอียงและค่าความแปรปรวนใกล้เคียงกัน และวิธีปรับตัวประมาณกำลังสองน้อยที่สุดดีกว่าวิธีกำลังสองน้อยที่สุด เนื่องจากให้ค่าความเอนเอียงและค่าความแปรปรวนต่ำกว่า

ถึงแม้ว่าวิธีกำลังสองน้อยที่สุด จะเป็นวิธีที่ใช้กันอย่างแพร่หลายในการประมาณค่าพารามิเตอร์ในตัวแบบการถดถอยเชิงเส้นอย่างง่าย โดยเฉพาะเมื่อตัวอย่างมีขนาดเล็ก ($n < 30$) แต่เมื่อตัวแปรอิสระถูกวัดมาโดยมีความคลาดเคลื่อนและตัวอย่างมีขนาดใหญ่ ($n \geq 30$) การนำวิธีกำลังสองน้อยที่สุดมาใช้จะต้องทำให้ค่าสัมประสิทธิ์การถดถอยมีค่าเล็กลงเพื่อทำให้ค่าความเอนเอียงที่เกิดขึ้นมีค่าลดลงเข้าใกล้ศูนย์ สำหรับวิธี Instrumental Variables ผู้วิจัยจะต้องหาตัวแปรอิสระอื่นที่มีสหสัมพันธ์อย่างสูงกับตัวแปรอิสระ X และไม่มีสหสัมพันธ์กับความคลาดเคลื่อนในการวัดค่า X และความคลาดเคลื่อนในตัวแบบซึ่งวิธีการคำนวณค่อนข้างยุ่งยาก วิธี Orthogonal Regression เป็นวิธีที่คำนวณง่ายและสะดวกกว่าวิธี Instrumental Variables และสามารถนำมาใช้ได้โดยเฉพาะเมื่อตัวอย่างมีขนาดใหญ่

ข้อเสนอแนะในการทำวิจัยครั้งต่อไป

สำหรับการเปรียบเทียบวิธีการประมาณค่าพารามิเตอร์ในตัวแบบการถดถอยเชิงเส้นอย่างง่ายเมื่อวัดค่าตัวแปรอิสระโดยมีความคลาดเคลื่อนครั้งนี้ ศึกษาเฉพาะ 4 วิธีที่กล่าวมาข้างต้น นอกเหนือจาก 4 วิธีดังกล่าว พบว่ายังมีวิธีการประมาณแบบอื่นที่น่าสนใจอีก เช่น วิธีภาวะความน่าจะเป็นสูงสุด, วิธี Regression Calibration, วิธี Two-Stage Least Squares (2SLS) และวิธี Stein-Rule Least Squares (SRLS) หรืออาจจะทำการศึกษาปัญหาการวัดค่าตัวแปรอิสระโดยมีความคลาดเคลื่อนในตัวแบบอื่นๆ เช่น Multiple Regression, Nonlinear Regression และ Logistic Regression ซึ่งยังไม่ได้ทำการศึกษา จึงน่าที่จะนำมาทำการศึกษาในการวิจัยครั้งต่อไป

เอกสารและสิ่งอ้างอิง

มนตรี พิริยะกุล. 2529. **เทคนิคการวิเคราะห์การถดถอย.** สำนักพิมพ์มหาวิทยาลัยรามคำแหง, กรุงเทพฯ.

อโนทัย ตรีวานิช. 2539. **ทฤษฎีการอนุมานทางสถิติ.** หจก.โรงพิมพ์คลังนานาวิทยา, ขอนแก่น.

Anderson, T.W. 1951. Estimating Linear Restrictions on Regression Coefficient for Multivariate Normal Distributions. **Ann. Math. Statist.** 22: 327-351.

Buzas, J.S., T.D. and L.A. Stefanski. 2003. **1 Measurement Error.** Institute Of Statistics Mimeo Series No. 2544.

Carroll, R.J. and R. David. 1996. The Use and Misuse of Orthogonal Regression in Linear Error-in-Variables Model. **The American Statistician** 50 (1): 1-6.

Daniel, L.R. and S.P. Robert. 1998. **Econometric Models and Economic Forecasts.** The McGraw-Hill Companies, Inc., Singapore.

Dorff, M. and J. Gurland. 1961. Estimation of the Parameters of a Linear Functional Relation. **J. Roy. Statist. Soc. Ser. B** 23: 160-170.

Douglas, C.M. and A.P. Elizabeth. 1981**. Introduction to Linear Regression Analysis.** John Wiley and Sons, Inc., New York.

Farrell, E.B. 1978. Measurement Error and Statistical Significance of an Independent Variable. **The American Statistician** 32 (1): 26-27.

Jed, L.D. and M.L. Jeffrey. 1995. **Error in Variables and Lending Discrimination.** Federal Reserve Bank of Richmond Economic Quarterly. 81 (3): 19-32.

Jeffrey, S.B. and A.S. Leonard. 1996. Instrumental Variable Estimation in Generalized Linear Measurement Error Models. **Journal of the American Statistical Association** 91 (435): 999-1006.

Marcia, M.A. Schafgans. 2002. Instrumental Variable Estimation in Binary Measurement Error Models. **Journal of the American Statistical Association** 90: 541-550.

Neter, J. 1996. **Applied Linear Regression Models.** 3rd ed. The McGraw-Hill Companies, Inc., United States of America.

Stefanski, L.A. 2000. Measurement Error Models. **Journal of the American Statistical Association** 95 (452): 1353-1358.

Wayne, A.F. 1980. Properties of some Estimators for the Error-in-Variables Model. **The Annals of Statistics** 8 (2): 407-422.

_______. 1987. **Measurement Error Models.** John Wiley and Sons, Inc., United States of America.

_______ and R.L. Carter. 1980. Instrumental Variable Estimation of the Simple Errors-in-Variables Model. **Journal of the American Statistical Association** 75 (371): 687-692.

_______, K.W. Joan and D.W. Richard. 1974. An Error-in-Variable Analysis of Managerial Role Performance. **Journal of the American Statistical Association** 69 (348): 886-893.

ภาคผนวก

โปรแกรม MATLAB Version 6.5 ที่ใช้สำหรับประมาณค่าพารามิเตอร์

```
%%%%%%%%%%%%%%%%%%%%%%%%%%%%%%%%%%%%%%%%%%%%%%%%%%%%%%%%%%%%%%%%%%%%%%%%%%%%%%%%%%%%%%%%%%%%%%%%%%
%
 M-file Thesis.m
% Program for Thesis 'Estimation of Parameters in Simple Linear Regression when
Independent Variable is Measured with Error'
%
%%%%%%%%%%%%%%%%%%%%%%%%%%%%%%%%%%%%%%%%%%%%%%%%%%%%%%%%%%%%%%%%%%%%%%%%%%%%%%%%%%%%%%%%%%%%%%%%%%
%
% Each Sample Size have 2 Cases
% Case 1 Classical Error Model
% Case 2 Berkson Error Model
%
%%%%%%%%%%%%%%%%%%%%%%%%%%%%%%%%%%%%%%%%%%%%%%%%%%%%%%%%%%%%%%%%%%%%%%%%%%%%%%%%%%%%%%%%%%%%%%%%%%
%
fprintf('Test Model \n');
fprintf('Classical Error Model \n');
%
fprintf('Test Method \n');
fprintf('1. Ordinary Least Squares \n');
fprintf('2. Adjusted Least Squares \n');
fprintf('3. Orthogonal Regression \n');
fprintf('4. Instrumental Variable \n');
method=input('Enter Test Method [1,2,3 or 4] ==>');
%
fprintf('Choose Number of Sample Size \n');
fprintf('Sample Size [10,20,30,50,70,100,200,300 \n');
n=input('Number of Sample =');
%
Cyc=input('Enter Number of Cycle for Run =');
```

```
%
fprintf('Choose Correlation Coefficients [0.2,0.4,0.5,0.6,0.7,0.8,0.9] \n');
R=input('Enter Correlation Coefficients =');
%
%--------------------------------------------------------------------------------------------------------------
%
for C=1:Cyc
   X=randn(n,1);
   E=random('Normal',0,5,n,1);
   B0=0;
   B1=1;
   U=randn(n,1);
   W=X+U;
   for i=1:n
      Y(i)=B0+(B1*W(i))+(E(i)-(B1*U(i)));
   end;
   Y=reshape(Y,n,1);
   meanX1=mean(X);
   meanY1=mean(Y);
   meanX=repmat(meanX1,n,1);
   meanY=repmat(meanY1,n,1);
   b1=sum((X-meanX).*(Y-meanY))/sum((X-meanX).^2);
   b0=meanY-(b1.*meanX);
   Yest=b0+(b1.*X);
   e=Y-Yest;
   r=sum(abs(e).*X)/sqrt((sum(e.^2)).*(sum(X.^2)));
   t=(r*sqrt(n-2))/sqrt(1-(r^2));
   tt = tinv(0.95,n-1);
%
%------------------------------------------Measurement Error Model-------------------------------------
```

```
%
if abs(t)>tt
   if model==1
      W=X+U;
   else
      W=X+U;
   end;
%
%-----------------------------------------Ordinary Least Squares----------------------------------------
%
if method==1
   meanW=mean(W);
   varX=var(X,1);
   varW=var(W,1);
   b1OLS=(sum((W-meanW).*(Y-meanY))/sum((W-meanW).^2));
   s2=(sum(b1OLS-b1).^2)/n;
   disp(b1OLS);
   disp(se);
end;
%
%-----------------------------------------Adjusted Least Squares----------------------------------------
%
if method==2
   meanW=mean(W);
   varX=var(X,1);
   varW=var(W,1);
   b1AOLS=(sum((W-meanW).*(Y-meanY))/sum((W-meanW).^2))*(varW/varX);
   s2=(sum(b1AOLS-b1).^2)/n;
   disp(b1AOLS);
   disp(se);
end;
```

```
%
%-------------------------------------------Orthogonal Regression---------------------------------------
%
if method==3
   meanW=mean(W);
   vary=var(Y);
   varw=var(W);
   covwy=sum((W-meanW).*(Y-meanY))/(n-1);
   vare=var(e,1);
   varU=var(U,1);
   A=vare/varU;
   b1OR=(vary-(A*varw)+(((vary-(A*varw))^2+(4*A*covwy))^0.5))/(2*sqrt(covwy));
   s2=(sum(b1OR-b1).^2)/n;
   disp(b1OR);
   disp(se);
end;
%
%-------------------------------------------Instrumental Variable-----------------------------------------
%
if method==4
   T=randn(n,1);
   meanW=mean(W);
   meanT=mean(T);
%-----------------------------------------------------------------------------------------------------------
if R==0.2
      F=[0.77460 0.63246;
        0.77460 -0.63246];
      data=[X T];
      data=data';
      Z=F*data;
```

```
      Z=Z';
      X=Z(:,1);
      T=Z(:,2);
      Rxt=corrcoef(X,T);
      R1=Rxt(1,2);
      b1IV=sum((Y-meanY).*(T-meanT))/sum((W-meanW).*(T-meanT));
      s2=(sum(b1IV-b1).^2)/n;
      disp(b1IV);
      disp(se);
   end;
%-----------------------------------------------------------------------------------------------------------------
   if R==0.4
      F=[0.83666 0.54772;
        0.83666 -0.54772];
      data=[X T];
      data=data';
      Z=F*data;
      Z=Z';
      X=Z(:,1);
      T=Z(:,2);
      Rxt=corrcoef(X,T);
      R1=Rxt(1,2);
      b1IV=sum((Y-meanY).*(T-meanT))/sum((W-meanW).*(T-meanT));
      s2=(sum(b1IV-b1).^2)/n;
      disp(b1IV);
      disp(se);
   end;
%-----------------------------------------------------------------------------------------------------------------
      if R==0.5
      F=[0.86603 0.50000;
```

```
      0.86603 -0.50000];
    data=[X T];
    data=data';
    Z=F*data;
    Z=Z';
    X=Z(:,1);
    T=Z(:,2);
    Rxt=corrcoef(X,T);
    R1=Rxt(1,2);
    b1IV=sum((Y-meanY).*(T-meanT))/sum((W-meanW).*(T-meanT));
    s2=(sum(b1IV-b1).^2)/n;
    disp(b1IV);
    disp(se);
  end;
%------------------------------------------------------------------------------------------------------------
  if R==0.6
    F=[0.89443 0.44721;
       0.89443 -0.44721];
    data=[X T];
    data=data';
    Z=F*data;
    Z=Z';
    X=Z(:,1);
    T=Z(:,2);
    Rxt=corrcoef(X,T);
    R1=Rxt(1,2);
    b1IV=sum((Y-meanY).*(T-meanT))/sum((W-meanW).*(T-meanT));
    s2=(sum(b1IV-b1).^2)/n;
    disp(b1IV);
    disp(se);
  end;
%------------------------------------------------------------------------------------------------------------
```

```
if R==0.8
    F=[0.94868 0.31623;
      0.97468 -0.22361];
    data=[X T];
    data=data';
    Z=F*data;
    Z=Z';
    X=Z(:,1);
    T=Z(:,2);
    Rxt=corrcoef(X,T);
    R1=Rxt(1,2);
    b1IV=sum((Y-meanY).*(T-meanT))/sum((W-meanW).*(T-meanT));
    s2=(sum(b1IV-b1).^2)/n;
    disp(b1IV);
    disp(se);
  end;
%------------------------------------------------------------------------------------------------------------
  if R==0.9
    F=[0.97468 0.22361;
      0.97468 -0.22361];
    data=[X T];
    data=data';
    Z=F*data;
    Z=Z';
    X=Z(:,1);
    T=Z(:,2);
    Rxt=corrcoef(X,T);
    R1=Rxt(1,2);
    b1IV=sum((Y-meanY).*(T-meanT))/sum((W-meanW).*(T-meanT));
    s2=(sum(b1IV-b1).^2)/n;
```

```
        disp(b1IV);
        disp(se);
    end;
end;
end;
end;
%
%%%%%%%%%%%%%%%%%%%%%%%%%%%%%%%%%%%%%%%%%%%%%
```

โปรแกรม SAS Version 9.0 ที่ใช้สำหรับประมาณค่าพารามิเตอร์

FOR CORR = 0.20

```
DATA A (TYPE = CORR); _TYPE_ = 'CORR';
        INPUT X1 X2;
CARDS;
1.00 .
0.20 1.00
;
PROC FACTOR N = 2;
RUN;
```

The SAS System 07:15 Sunday, October 10, 2004 1

The FACTOR Procedure

Initial Factor Method: Principal Components

Prior Communality Estimates: ONE

Eigenvalues of the Correlation Matrix: Total = 2 Average = 1

	Eigenvalue	Difference	Proportion	Cumulative
1	20000000	0.40000000	0.6000	0.6000
2	0.80000000		0.4000	1.0000

2 factors will be retained by the NFACTOR criterion.

Factor Pattern

	Factor1	Factor2
X1	0.77460	0.63246
X2	0.77460	-0.63246

Variance Explained by Each Factor

Factor1	Factor2
1.2000000	0.8000000

Final Communality Estimates: Total = 2.000000

X1	X2
1.0000000	1.0000000

```
PROC IML;
F = {0.77460 0.63246,
   0.77460 -0.63246};
DATA = RANNOR(J(10,2,0));
DATA = DATA`;
Z = F*DATA;
Z = Z`;
X1 = Z[,1];
X2 = Z[,2];
```

```
Z = X1||X2;
CREATE A FROM Z [COLNAME = {X1 X2}];
APPEND FROM Z;
PROC MEANS DATA = A N MEAN STD SKEWNESS KURTOSIS;
    VAR X1 X2;
PROC CORR DATA = A NOSIMPLE;
    VAR X1 X2;
RUN;
PROC PRINT DATA = A;
RUN;
```

FOR CORR = 0.40

```
DATA A (TYPE = CORR); _TYPE_ = 'CORR';
        INPUT X1 X2;
CARDS;
1.00 .
0.40 1.00
;
PROC FACTOR N = 2;
RUN;
```

The SAS System 07:15 Sunday, October 10, 2004 3

The FACTOR Procedure

Initial Factor Method: Principal Components

Prior Communality Estimates: ONE

Eigenvalues of the Correlation Matrix: Total = 2 Average = 1

	Eigenvalue	Difference	Proportion	Cumulative
1	1.40000000	0.80000000	0.7000	0.7000
2	0.60000000		0.3000	1.0000

2 factors will be retained by the NFACTOR criterion.

Factor Pattern

	Factor1	Factor2
X1	0.83666	0.54772
X2	0.83666	-0.54772

Variance Explained by Each Factor

Factor1	Factor2
1.4000000	0.6000000

Final Communality Estimates: Total = 2.000000

X1	X2
1.0000000	1.0000000

```
PROC IML;
F = {0.83666 0.54772,
   0.83666 -0.54772};
DATA = RANNOR(J(10,2,0));
DATA = DATA`;
```

```
Z = F*DATA;
Z = Z`;
X1 = Z[,1];
X2 = Z[,2];
Z = X1||X2;
CREATE A FROM Z [COLNAME = {X1 X2}];
APPEND FROM Z;
PROC PRINT DATA = A;
RUN;
DATA RANDOM;
E = RAND('NORMAL',10,0.5,1);
OUTPUT;
PROC PRINT DATA = RANDOM;
RUN;
```

FOR CORR = 0.50

```
DATA A (TYPE = CORR); _TYPE_ = 'CORR';
        INPUT X1 X2;
CARDS;
1.00 .
0.50 1.00
;
PROC FACTOR N = 2;
RUN;
```

The SAS System 07:15 Sunday, October 10, 2004 4

The FACTOR Procedure

Initial Factor Method: Principal Components

Prior Communality Estimates: ONE

Eigenvalues of the Correlation Matrix: Total = 2 Average = 1

	Eigenvalue	Difference	Proportion	Cumulative
1	1.50000000	1.00000000	0.7500	0.7500
2	0.50000000		0.2500	1.0000

2 factors will be retained by the NFACTOR criterion.

Factor Pattern

	Factor1	Factor2
X1	0.86603	0.50000
X2	0.86603	-0.50000

Variance Explained by Each Factor

Factor1	Factor2
1.5000000	0.5000000

Final Communality Estimates: Total = 2.000000

X1	X2
1.0000000	1.0000000

```
PROC IML
F = {0.86603 0.50000,
    0.86603 -0.50000};
DATA = RANNOR(J(10,2,0));
DATA = DATA';
Z = F*DATA'
Z = Z'
X1 = Z[,1]
X2 = Z[,2]
Z = X1//X2;
CREATE A FROM Z [COLNOME = {X1 X2}];
SPPEND FROM Z;
PROC MEANS DATA = A N MEAN STDB SKEWNESS KURTEST;
     VAR X1 X2;
PROC CORRB DATA = A NOSIMPLE;
     VAR X1 X2;
RUN;
```

FOR CORR = 0.60

```
DATA A (TYPE = CORR); _TYPE_ = 'CORR';
        INPUT X1 X2;
CARDS;
1.00 .
0.60 1.00
;
PROC FACTOR N = 2;
RUN;
```

The SAS System 07:15 Sunday, October 10, 2004 5

The FACTOR Procedure

Initial Factor Method: Principal Components

Prior Communality Estimates: ONE

Eigenvalues of the Correlation Matrix: Total = 2 Average = 1

	Eigenvalue	Difference	Proportion	Cumulative
1	1.60000000	1.20000000	0.8000	0.8000
2	0.40000000		0.2000	1.0000

2 factors will be retained by the NFACTOR criterion.

Factor Pattern

	Factor1	Factor2
X1	0.89443	0.44721
X2	0.89443	-0.44721

Variance Explained by Each Factor

Factor1	Factor2
1.6000000	0.4000000

Final Communality Estimates: Total = 2.000000

X1 X2

1.0000000 1.0000000

PROC IML;

F = {**0.89443 0.44721**,

0.89443 -0.44721};

DATA = RANNOR(J(**10**,**2**,**0**));

DATA = DATA`;

Z = F*DATA;

Z = Z`;

X1 = Z[,**1**];

X2 = Z[,**2**];

Z = X1||X2;

CREATE A FROM Z [COLNAME = {X1 X2}];

APPEND FROM Z;

PROC MEANS DATA = A N MEAN STD SKEWNESS KURTOSIS;

VAR X1 X2;

PROC CORR DATA = A NOSIMPLE;

VAR X1 X2;

RUN;

PROC PRINT DATA = A;

RUN;

FOR CORR = 0.80

DATA A (TYPE = CORR); _TYPE_ = 'CORR';

INPUT X1 X2;

CARDS;

```
1.00 .
0.80 1.00
;
PROC FACTOR N = 2;
RUN;
```

The SAS System 07:15 Sunday, October 10, 2004 6

The FACTOR Procedure

Initial Factor Method: Principal Components

Prior Communality Estimates: ONE

Eigenvalues of the Correlation Matrix: Total = 2 Average = 1

	Eigenvalue	Difference	Proportion	Cumulative
1	1.80000000	1.60000000	0.9000	0.9000
2	0.20000000		0.1000	1.0000

2 factors will be retained by the NFACTOR criterion.

Factor Pattern

	Factor1	Factor2
X1	0.94868	0.31623
X2	0.94868	-0.31623

Variance Explained by Each Factor

Factor1	Factor2
1.8000000	0.2000000

Final Communality Estimates: Total = 2.000000

X1	X2
1.0000000	1.0000000

```
PROC IML;
F = {0.94868 0.31623,
   0.94868 -0.31623};
DATA = RANNOR(J(10,2,0));
DATA = DATA`;
Z = F*DATA;
Z = Z`;
X1 = Z[,1];
X2 = Z[,2];
Z = X1||X2;
CREATE A FROM Z [COLNAME = {X1 X2}];
APPEND FROM Z;
PROC MEANS DATA = A N MEAN STD SKEWNESS KURTOSIS;
     VAR X1 X2;
PROC CORR DATA = A NOSIMPLE;
     VAR X1 X2;
RUN;
```

```
PROC PRINT DATA = A;
RUN;
```

FOR CORR = 0.90

```
DATA A (TYPE = CORR); _TYPE_ = 'CORR';
	INPUT X1 X2;
CARDS;
1.00 .
0.90 1.00
;
PROC FACTOR N = 2;
RUN;
```

The SAS System 07:15 Sunday, October 10, 2004 7

The FACTOR Procedure

Initial Factor Method: Principal Components

Prior Communality Estimates: ONE

Eigenvalues of the Correlation Matrix: Total = 2 Average = 1

	Eigenvalue	Difference	Proportion	Cumulative
1	1.90000000	1.80000000	0.9500	0.9500
2	0.10000000		0.0500	1.0000

2 factors will be retained by the NFACTOR criterion.

Factor Pattern

	Factor1	Factor2
X1	0.97468	0.22361
X2	0.97468	-0.22361

Variance Explained by Each Factor

Factor1	Factor2
1.9000000	0.1000000

Final Communality Estimates: Total = 2.000000

X1	X2
1.0000000	1.0000000

```
PROC IML;
F = {0.97468 0.22361,
    0.97468 -0.22361};
DATA = RANNOR(J(10,2,0));
DATA = DATA`;
Z = F*DATA;
Z = Z`;
X1 = Z[,1];
X2 = Z[,2];
Z = X1||X2;
CREATE A FROM Z [COLNAME = {X1 X2}];
```

```
APPEND FROM Z;
PROC MEANS DATA = A N MEAN STD SKEWNESS KURTOSIS;
    VAR X1 X2;
PROC CORR DATA = A NOSIMPLE;
    VAR X1 X2;
RUN;
PROC PRINT DATA = A;
RUN;
```
