

บทความวิจัย

การวิเคราะห์ข้อมูลจำนวนครั้งการเรียกร่องลินไหมทดแทน ของการประกันภัยรถยนต์ที่มีการวัดซ้ำ ด้วยสมการประมาณ ค่าวงนัยทั่วไป และตัวแบบผสมเชิงเส้นวงนัยทั่วไป

อุไรวรรณ เจริญกิติกุล และ ลีลี อิงศรีสว่าง*

บทคัดย่อ

การวิจัยนี้มีวัตถุประสงค์เพื่อสร้างตัวแบบปัจจัยที่มีอิทธิพลต่อจำนวนครั้งการเรียกร่องลินไหมทดแทนของการประกันภัยรถยนต์ที่มีการวัดซ้ำ สำหรับภาพรวมของกลุ่มเสี่ยงภัย ด้วยวิธีสมการประมาณค่าวงนัยทั่วไป (Generalized Estimating Equation: GEE) และสำหรับหน่วยเสี่ยงภัยเฉพาะราย ด้วยวิธีตัวแบบผสมเชิงเส้นวงนัยทั่วไป (Generalized Linear Mixed Model: GLMM) เมื่อตัวแปรตาม คือ จำนวนครั้งการเรียกร่องลินไหมทดแทนมีลักษณะแจกแจงแบบปัวซอง (Poisson distribution) ซึ่งข้อมูลที่นำมาวิเคราะห์เป็นข้อมูลประกันภัยรถยนต์ชั้น 1 ที่ซื้อสัญญาของบริษัทประกันภัยแห่งหนึ่งในประเทศไทย ตั้งแต่ปีกรรรม พ.ศ. 2547 ถึงปีกรรรม พ.ศ. 2550 ประกอบด้วยตัวแปรปัจจัย คือ ลักษณะการใช้งานกลุ่มรถยนต์ ขนาดเครื่องยนต์ อายุรถ จำนวนที่นั่ง ค่าส่วนลด-เพิ่มจากประวัติการเรียกร่องลินไหมทดแทน และเบี้ยประกันภัยสุทธิ

ผลการวิเคราะห์พบว่า สำหรับภาพรวมของกลุ่มเสี่ยงภัยด้วยวิธี GEE มีเพียงปัจจัยขนาดเครื่องยนต์ และส่วนลด-เพิ่มจากประวัติการเรียกร่องลินไหมทดแทนที่มีอิทธิพลต่อจำนวนครั้งการเรียกร่องลินไหมทดแทนที่ระดับนัยสำคัญ 0.05 และสำหรับหน่วยเสี่ยงภัยเฉพาะรายในการประกันภัยรถยนต์ ด้วยวิธี GLMM พบว่ามีเพียงปัจจัยขนาดเครื่องยนต์ที่เป็นปัจจัยอิทธิพลคงที่ และลักษณะหน่วยเสี่ยงภัยเฉพาะแต่ละรายซึ่งเป็นปัจจัยอิทธิพลสุ่มมีผลต่อจำนวนครั้งการเรียกร่องลินไหมทดแทนที่ระดับนัยสำคัญ 0.05

คำสำคัญ: จำนวนครั้งการเรียกร่องลินไหมทดแทน การแจกแจงแบบปัวซอง สมการประมาณค่าวงนัยทั่วไป (GEE) ตัวแบบผสมเชิงเส้นวงนัยทั่วไป (GLMM)

Analysis of Repeated Motor Insurance Claim Count Data Using a Generalized Estimating Equation and a Generalized Linear Mixed Model

Uraiwan Jaroengeratikun and Lily Ingsrisawang*

ABSTRACT

The objective of this research was to build an impacted factor model for repeated motor insurance claim count data. A Generalized Estimating Equation (GEE) and a Generalized Linear Mixed Model (GLMM) were the methodologies used as the population average model of the exposure unit group and as the subject specific model of the exposure unit, respectively. In this study, the dependent variable with a Poisson distribution was the claim count of the comprehensive motor insurance. The claim count data was from a motor insurance contract and renewal its subsequent covering the policy years 2004 to 2007, which was provided by one of non-life insurance companies in Thailand. The studied factor variables were usage of the car, car group, engine size, age of the car, seat number, no-claims bonus value, and the premium.

The results of this study show that for GEE, the population average model of the exposure unit group, the factors significantly related to motor insurance claim count at the 0.05 level consisted of the engine size of the car and the no-claims bonus value. For GLMM, the subject specific model of the exposure unit, the engine size of the car and the specifics of each individual risk were the fixed effect factor and the random effect factor, respectively, and were significantly related to motor insurance claim count at the 0.05 level.

Keywords: claim count, Poisson distribution, Generalized Estimating Equation (GEE), Generalized Linear Mixed Model (GLMM)

บทนำ

ธุรกิจประกันภัย เป็นธุรกิจที่ให้บริการในลักษณะจัดการความเสี่ยงภัยบนความไม่แน่นอนของเหตุการณ์อนาคตที่เกิดขึ้นให้กับผู้เอาประกันภัย การจัดการความเสี่ยงภัยที่รับประกันภัยที่ดีมีประสิทธิภพย่อมส่งผลให้ธุรกิจนั้นมีฐานะการเงินที่มั่นคงเป็นที่เชื่อถือของลูกค้า ผลิตภัณฑ์หรือสัญญาประกันภัยมี 2 ประเภทตามสิ่งหรือวัตถุที่เอาประกันภัย (subject matter) ประกอบด้วย การประกันชีวิต และการประกันวินาศภัย ด้วยรูปแบบเฉพาะผลิตภัณฑ์ของการประกันวินาศภัย ปกติจะมีระยะเวลาสัญญาประกันภัยเพียงหนึ่งปี และต่ออายุสัญญาเป็นปีต่อปี ทำให้กระทบต่อการวางแผนการลงทุนของบริษัทที่นำส่วนหนึ่งของเบี้ยประกันภัยนิยมไปลงทุนกับแหล่งทุนที่มีสภาพคล่องซึ่งย่อมจะให้ผลตอบแทนค่อนข้างไม่สูงเมื่อเทียบกับแหล่งทุนบางแหล่ง นั่นก็นับว่าเป็นการสูญเสียโอกาสของการลงทุนของบริษัท เนื่องจากความไม่แน่นอนของฐานลูกค้าหรือผู้เอาประกันภัยที่ยังคงต่อสัญญาความคุ้มครองกับบริษัท และด้วยพฤติกรรมหน่วยเสี่ยงภัย (exposure unit) ต่อการเกิดมูลค่าความเสียหาย (loss or claim severity) และความถี่การเรียกร้องสินไหมทดแทนหรือจำนวนครั้งการเรียกร้องสินไหมทดแทน (claim frequency or claim count) โดยนักคณิตศาสตร์ประกันภัยในแต่ละบริษัทประกันภัยควรให้ความสำคัญศึกษาวิเคราะห์ข้อมูลการแจกแจงค่าสินไหมทดแทน (claim amount) การแจกแจงของจำนวนครั้งการเรียกร้องสินไหมทดแทน รวมถึงปัจจัยที่มีอิทธิพลต่อมูลค่าสินไหมทดแทนและจำนวนครั้งการเรียกร้องสินไหมทดแทน จากข้อมูลในอดีตที่ผ่านมาของกลุ่มผู้เอาประกันภัยที่มีเสี่ยงภัยเหมือนกัน เพื่อนำมาเป็นสารสนเทศสำหรับการดำเนินการในส่วนต่างๆ เช่น การจัดจำแนกกลุ่มเสี่ยงภัยที่ใช้ในการพิจารณารับประกันภัย (underwriting) กำหนดเบี้ยประกันภัย (premium pricing) การกำหนดเงินสำรองสินไหมทดแทน (loss reserve) ที่เป็นพันธหนี้สินของบริษัท และเงินสำรองเบี้ยประกันภัย (premium reserve) ให้มีความเหมาะสม การวางแผนการลงทุน การวางแผนการประกันภัยต่อ (reinsurance) เป็นต้น ในความสำคัญนี้ Renshaw (1994) [1] ได้ประยุกต์วิธีตัวแบบเชิงเส้นวางนัยทั่วไป (Generalized Linear Model: GLM) กับการศึกษาตัวแบบกระบวนการเรียกร้องสินไหมทดแทน (claim process) ที่ประกอบด้วยทั้งจำนวนครั้งการเรียกร้องสินไหมทดแทน และมูลค่าความเสียหายในการประกันวินาศภัยที่รวมสาระสำคัญของปัจจัยที่เป็นอัตราส่วน (rating factor) ที่ให้เป็นตัวแปรร่วม (covariate) ซึ่งลักษณะการวิเคราะห์ตัวแบบเชิงเส้นวางนัยทั่วไปเป็นวิธีการวิเคราะห์ตัวแบบความสัมพันธ์เชิงเส้นระหว่างตัวแปรตาม (dependent variable) หรือตัวแปรตอบสนอง (response variable) กับตัวแปรอิสระ (independent variable) หรือปัจจัย (factor) หรือตัวแปรทำนาย (predictor variable) หรือที่เรียกว่าตัวแปรร่วม สำหรับข้อมูลที่ไม่จำเป็นต้องเป็นการแจกแจงแบบปกติ (Non-Normal distribution) และ Nelder และ Verrall (1997) [2] และ Ohlsson และ Johansson (2004) [3] ประยุกต์เทคนิควิธี GLM ในการประมาณค่าปัจจัยความน่าเชื่อถือได้ (credibility factor) เพื่อสร้างสูตรสมการความน่าเชื่อถือได้ (credibility formula) สำหรับใช้ปรับปรุค่าเบี้ยประกันภัย ซึ่งในการวิเคราะห์ศึกษาข้อมูลส่วนใหญ่รวมทั้งที่กล่าวข้างต้นจะสมมติให้เป็นข้อมูลที่เป็นอิสระต่อกันโดยสารสนเทศจากผลการวิเคราะห์จะพิจารณาภาพรวมในกลุ่มที่มีความเสี่ยงเหมือนกัน และจะสมมติให้ลักษณะพฤติกรรมการเกิดเหตุเหมือนกัน Yau และคณะ (2003) [4] ประยุกต์วิธีตัวแบบผสมเชิงเส้นวางนัยทั่วไป (Generalized Linear Mixed Model: GLMM) ในการวิเคราะห์ข้อมูลจำนวนความถี่ของการเรียกร้องสินไหมทดแทนที่ข้อมูลมีการวัดซ้ำ (repeated claim frequency) หรือข้อมูลจำนวนครั้งการเรียกร้องสินไหมทดแทนที่มีการวัดซ้ำ ภายใต้

เงื่อนไขกำหนดเวกเตอร์ของอิทธิพลสุ่ม (random effect) ที่มีรูปแบบเชิงเส้นของตัวแปรทำนาย (linear predictor) ภายในตัวแบบที่คำนึงถึงคุณลักษณะที่แตกต่างโดยธรรมชาติของข้อมูล (inherent correlation) และแสดงข้อได้เปรียบการประยุกต์วิธี GLMM เทียบกับวิธี GLM ในข้อมูลจำนวนครั้งการเรียกร้องสินไหมทดแทนที่มีการวัดซ้ำที่ให้ความสำคัญของความสัมพันธ์ หรือความผันแปรในข้อมูลที่เก็บซ้ำของแต่ละผู้เอาประกันภัยหรือหน่วยเสี่ยงภัย พบว่าในความผันแปรของข้อมูลที่เก็บซ้ำมีค่าสูงขึ้น วิธี GLMM จะค่อนข้างเหมาะสม (fit) กับข้อมูลจำนวนครั้งการเรียกร้องสินไหมทดแทนที่มีการวัดซ้ำได้ดีกว่าวิธี GLM ดังนั้นจึงควรให้ความสำคัญของการเลือกวิธีที่นำมาประยุกต์วิเคราะห์ต้องมีความเหมาะสมกับลักษณะธรรมชาติของข้อมูล

บทความวิจัยนี้ได้นำเสนอการวิเคราะห์ข้อมูลจำนวนครั้งการเรียกร้องสินไหมทดแทนที่มีลักษณะเก็บข้อมูลซ้ำหรือที่มีการวัดซ้ำ (repeated data) เป็นรายปีกรมธรรม์ (policy year) ซึ่งได้จากการติดตามเก็บผลข้อมูลจำนวนครั้งการเรียกร้องสินไหมทดแทนของหน่วยเสี่ยงภัยในกลุ่มการประกันภัยรถยนต์ภาคสมัครใจในรูปแบบประกันภัยรถยนต์ชั้น 1 ประเภทรถยนต์นั่ง ทั่วประเทศที่ซื้อสัญญาประกันภัยของบริษัทประกันภัยแห่งหนึ่งในประเทศไทยที่เริ่มต้นสัญญาตั้งแต่ปีกรมธรรม์ พ.ศ. 2547 จนถึงปีกรมธรรม์ พ.ศ. 2550 รวมเป็นระยะเวลา 4 ปีกรมธรรม์ โดยหน่วยเสี่ยงภัยนี้ยังต่ออายุสัญญาอย่างต่อเนื่องเป็นรายปีกรมธรรม์ นับถึงปีกรมธรรม์ พ.ศ. 2551 มีทั้งหมด 137,888 ราย แต่มีจำนวนข้อมูลที่สมบูรณ์ทั้งสิ้น 331 รายการวิเคราะห์นี้ได้คำนึงถึงลักษณะความสัมพันธ์ในข้อมูลจำนวนครั้งการเรียกร้องสินไหมทดแทนที่มีการวัดซ้ำ ประกอบรวมกับการวิเคราะห์ปัจจัยที่มีอิทธิพลต่อจำนวนครั้งการเรียกร้องสินไหมทดแทน และตัวแบบประมาณค่าจำนวนครั้งการเรียกร้องสินไหมทดแทนเฉลี่ย สำหรับภาพรวมของกลุ่มเสี่ยงภัยในการประกันภัยรถยนต์ โดยใช้วิธีสมการประมาณค่าวางนัยทั่วไป (Generalized Estimating Equation: GEE) เพื่อเป็นประโยชน์ด้านสารสนเทศสำหรับปรับปรุงค่าเบี้ยประกันภัยให้พอเพียงเหมาะสม การวางแผนการประกันภัยต่อการกำหนดเงินสำรองสินไหมทดแทนได้เหมาะสมไม่สูงเกินไปจนแสดงฐานะหนี้สินเกินกว่าความเป็นจริงซึ่งก่อให้เกิดการลดความเชื่อมั่นให้กับผู้เอาประกันภัยที่มีต่อบริษัท เป็นต้น นอกจากนี้จะรวมถึงการวิเคราะห์ข้อมูลจำนวนครั้งการเรียกร้องสินไหมทดแทนที่มีการวัดซ้ำโดยใช้วิธี GLMM เพื่อทราบปัจจัยอิทธิพลคงที่ (fixed effect) ปัจจัยอิทธิพลสุ่ม (random effect) ที่แสดงลักษณะความแตกต่างเฉพาะของแต่ละหน่วยเสี่ยงภัยที่ต่างจากโดยรวมของกลุ่มเสี่ยงภัยที่กระทบต่อจำนวนครั้งการเรียกร้องสินไหมทดแทน และวิเคราะห์ตัวแบบประมาณค่าจำนวนครั้งการเรียกร้องสินไหมทดแทนเฉลี่ยของหน่วยเสี่ยงภัยเฉพาะรายในการประกันภัยรถยนต์ เพื่อจะเป็นประโยชน์ในส่วนการพิจารณารับประกันภัย การกำหนดเบี้ยประกันภัยให้ยุติธรรมและสอดคล้องกับสภาพความเสี่ยงภัยของแต่ละหน่วยเสี่ยงภัย กำหนดกลยุทธ์การตลาดเพื่อยังคงฐานลูกค้าเดิม เป็นต้น

วัตถุประสงค์

สร้างตัวแบบปัจจัยที่มีอิทธิพลต่อจำนวนครั้งการเรียกร้องสินไหมทดแทนของการประกันภัยรถยนต์ที่มีการวัดซ้ำ สำหรับภาพรวมของกลุ่มเสี่ยงภัย ด้วยวิธีสมการประมาณค่าวางนัยทั่วไป (Generalized Estimating Equation: GEE) และสำหรับหน่วยเสี่ยงภัยเฉพาะราย ด้วยวิธีตัวแบบผสมเชิงเส้นวางนัยทั่วไป (Generalized Linear Mixed Model: GLMM)

ข้อมูลการประกันภัยรถยนต์

ตารางที่ 1 แสดงข้อมูลพื้นฐานประเภทรถยนต์นั่งของหน่วยเสี่ยงภัย

ส่วนของข้อมูลที่เป็นแบบกลุ่มและอันดับ	ร้อยละของหน่วยเสี่ยงภัย	
1. ลักษณะการใช้งาน		
การใช้ส่วนบุคคล (รหัส 110)		87.3
การใช้เพื่อการพาณิชย์ (รหัส 120)		12.7
2. กลุ่มรถยนต์จำแนกตามยี่ห้อและรุ่น (brand and model)		
กลุ่ม 1		0.0
กลุ่ม 2		3.9
กลุ่ม 3		63.4
กลุ่ม 4		13.7
กลุ่ม 5		19.0
3. ขนาดเครื่องยนต์ (หน่วย C.C.)		
CC ≤ 1500		19.0
1500 < CC ≤ 2000 หรือ (1500, 2000]		29.6
2000 < CC ≤ 2500 หรือ (2000, 2500]		20.9
CC > 2500		30.5
ส่วนของข้อมูลที่เป็นค่าตัวเลข	ค่าเฉลี่ย (mean)	ส่วนเบี่ยงเบน มาตรฐาน (standard deviation: sd)
4. อายุรถยนต์ (หน่วยเป็นปี)	2.0151	1.5734
5. จำนวนที่นั่ง	6.8369	0.5483
6. ค่าส่วนลด-เพิ่มจากประวัติการเรียกร้องสินไหมทดแทน (No-Claims Bonus :NCB หน่วยเป็นร้อยละของ เบี้ยประกันภัย)		
ปีกรมธรรม์ที่ 1	15.6103	12.6767
ปีกรมธรรม์ที่ 2	22.7734	15.5717
ปีกรมธรรม์ที่ 3	22.1752	15.5538
ปีกรมธรรม์ที่ 4	26.8127	18.0288
7. เบี้ยประกันภัยสุทธิ (pure premium หน่วยเป็นพันบาท)		
ปีกรมธรรม์ที่ 1	20.7830	12.2576
ปีกรมธรรม์ที่ 2	19.5323	11.1875
ปีกรมธรรม์ที่ 3	17.1407	8.6047
ปีกรมธรรม์ที่ 4	16.0484	8.2020

จากตารางที่ 1 ในหน่วยเสี่ยงภัยทั้งสิ้น 331 ราย มีลักษณะการใช้งานส่วนบุคคลมากถึงร้อยละ 87.3 และมีเพียงร้อยละ 12.7 ที่ใช้งานเพื่อการพาณิชย์ มีกลุ่มรถยนต์ที่จำแนกตามยี่ห้อและรุ่นอยู่ในกลุ่ม 3 มากที่สุดร้อยละ 63.4 รองลงมาเป็นกลุ่ม 5 กลุ่ม 4 และกลุ่ม 2 ร้อยละ 19, 13.7 และ 3.9 ตามลำดับ และที่ไม่มีหน่วยเสี่ยงภัยใดเลยในกลุ่ม 1 มีขนาดเครื่องยนต์ขนาด $CC > 2500$ และขนาด $1500 < CC \leq 2000$ จำนวนใกล้เคียงกัน คือ ร้อยละ 30.5 และ 29.6 ตามลำดับ รองลงมาเป็นขนาด $2000 < CC \leq 2500$ และขนาด $CC \leq 1500$ ที่มีจำนวนใกล้เคียงกัน คือ ร้อยละ 20.9 และ 19 ตามลำดับ เป็นรถยนต์ที่มีอายุเฉลี่ยประมาณ 2 ปี ด้วยค่าเบี่ยงเบนมาตรฐานเท่ากับ 1.5734 มีจำนวนที่นั่งเฉลี่ยประมาณ 7 ที่นั่งด้วยค่าเบี่ยงเบนมาตรฐานเล็กน้อยเท่ากับ 0.5483 ตลอดช่วงระยะเวลา 4 ปีกรมธรรม์หน่วยเสี่ยงภัยส่วนใหญ่จะได้รับค่าส่วนลดจากประวัติการเรียกร้องสินไหมทดแทนโดยช่วงประมาณเฉลี่ยร้อยละ 15 ถึง 25 ของเบี้ยประกันภัยด้วยค่าเบี่ยงเบนมาตรฐานในช่วงประมาณ 12 ถึง 18 และค่าเบี้ยประกันภัยสุทธิในปีกรมธรรม์ที่ 1 มีค่ามากที่สุดเฉลี่ยเท่ากับ 20.7830 พันบาท ด้วยค่าเบี่ยงเบนมาตรฐาน 12.2576 พันบาท รองลงมาเป็นเบี้ยประกันภัยสุทธิในปีกรมธรรม์ที่ 2, 3 และ 4 เท่ากับ 19.5323, 17.1407 และ 16.0484 พันบาท ตามลำดับ ด้วยค่าเบี่ยงเบนมาตรฐานของเบี้ยประกันภัยสุทธิในปีกรมธรรม์ที่ 2, 3 และ 4 เท่ากับ 11.1875, 8.6047 และ 8.2020 พันบาท

ขอบเขตการศึกษา

1. ปัจจัยหรือตัวแปรทำนาย (X) ที่นำมาศึกษาอิทธิพลที่มีต่อจำนวนครั้งการเรียกร้องสินไหมทดแทนของหน่วยเสี่ยงภัย ที่มีการวัดซ้ำ ดังตารางที่ 2 [5]
2. ตัวแปรตาม Y_{ij} เป็นจำนวนครั้งการเรียกร้องสินไหมทดแทนของหน่วยเสี่ยงภัยที่ i ที่มีการวัดซ้ำในปีกรมธรรม์ที่ j โดยมีลักษณะการแจกแจงแบบปัวซอง (Poisson distribution) [4] $Y_{ij} \sim \text{Poisson}(\mu_{ij})$ เมื่อ $i = 1, 2, 3, \dots, S$; $j = 1, 2, 3, 4$; $S = 331$ และ $t = 4$

วิธีการวิเคราะห์ข้อมูล

1. การวิเคราะห์ตัวแบบปัจจัยที่มีอิทธิพลต่อจำนวนครั้งการเรียกร้องสินไหมทดแทนที่มีการวัดซ้ำ สำหรับภาพรวมของหน่วยเสี่ยงภัยด้วยวิธี GEE โดยจะเกี่ยวข้องกับ 3 ส่วนต่อไปนี้ [7]

1.1. ค่าเฉลี่ย Y_{ij} ที่มีความสัมพันธ์กับปัจจัย หรือตัวแปร X_{ij} โดยการเชื่อมโยงความสัมพันธ์ด้วยฟังก์ชันลอการิทึม (log link function)

$$\log(\mu_{ij}) = X'_{ij}\beta \quad (1)$$

เมื่อ $\mu_{ij} = E(Y_{ij})$ เป็นค่าเฉลี่ย Y_{ij} และ $X'_{ij} = [X_{ij1} \ X_{ij2} \ \dots \ X_{ijp}]$ โดยที่ X_{ij} เป็นเวกเตอร์ของปัจจัยหรือตัวแปรทำนายซึ่งมี k ตัวแปร มีขนาดเท่ากับ $(k+1) \times 1 = p \times 1$ สำหรับหน่วยเสี่ยงภัยที่ i ปีกรมธรรม์ที่ j และ β เป็นเวกเตอร์ของพารามิเตอร์ในตัวแบบที่มีขนาดเท่ากับ $p \times 1$

1.2. ความแปรปรวน Y_{ij} เป็นฟังก์ชันของค่าเฉลี่ย Y_{ij}

$$\text{Var}(Y_{ij}) = \phi v(\mu_{ij}) = \mu_{ij}$$

ตารางที่ 2 แสดงปัจจัยที่นำมาศึกษา

ปัจจัย	สเกล	ตัวแปร (X)	คำอธิบาย (description)
1. ลักษณะการใช้งาน	แบบกลุ่ม	code1 code2 (base)	การใช้ส่วนบุคคล (รหัส 110) การใช้เพื่อการพาณิชย์ (รหัส 120)
2. กลุ่มรถยนต์จำแนกตามยี่ห้อและรุ่น (brand and model) [6]	แบบกลุ่ม	car_gr1 car_gr2 car_gr3 car_gr4 car_gr5 (base)	กลุ่ม 1 กลุ่ม 2 กลุ่ม 3 กลุ่ม 4 กลุ่ม 5
3. ขนาดเครื่องยนต์	อันดับ	cc_gr1 cc_gr2 cc_gr3 cc_gr4 (base)	$CC \leq 1500$ $1500 < CC \leq 2000$ หรือ (1500, 2000] $2000 < CC \leq 2500$ หรือ (2000, 2500] $CC > 2500$
4. อายุรถยนต์	ค่าตัวเลข	car_age	หน่วยเป็นปี มีค่าตั้งแต่ 1 ปีขึ้นไป
5. จำนวนที่นั่ง	ค่าตัวเลข	seat	
6. ค่าส่วนลด-เพิ่มจากประวัติการเรียกรถยนต์ใหม่ทดแทน	ค่าตัวเลข	NCB	หน่วยเป็นร้อยละของเบี้ยประกันภัยมีค่าลด-เพิ่มที่ละ 5% ของเบี้ยประกันภัย - กรณีส่วนเพิ่มจะมีค่าเป็นจำนวนลบ - กรณีส่วนลดจะมีค่าเป็นจำนวนบวก
7. เบี้ยประกันภัยสุทธิ	ค่าตัวเลข	premium	หน่วยเป็นพันบาท

เมื่อ $\phi = 1$, ϕ เป็น scale หรือ dispersion parameter และ $v(\mu_{ij}) = \mu_{ij}$ เป็นฟังก์ชันความแปรปรวน (variance function) ของค่าเฉลี่ย Y_{ij}

1.3. กำหนด Correlation Structure (CS) ขนาดเมตริกซ์ $t \times t$ ในการวิเคราะห์ที่พิจารณาจาก 4 รูปแบบดังนี้ [7-9]

$$1.3.1. \text{ independence มีค่า } R_i = \text{Corr}(Y_{ij}, Y_{ij'}) = \begin{cases} 1 & ; j = j' \\ 0 & ; j \neq j' \end{cases}$$

$$1.3.2. \text{ exchangeable มีค่า } R_i = \text{Corr}(Y_{ij}, Y_{ij'}) = \begin{cases} 1 & ; j = j' \\ \rho & ; j \neq j' \end{cases} \quad \text{โดยที่}$$

ค่าประมาณ \hat{R}_i เรียกว่า working correlation matrix มีค่า $\hat{\rho} = \sum_{i=1}^s \sum_{j < j'} \frac{e_{ij}e_{ij'}}{\phi \left(\sum_{t=1}^s \frac{1}{2} t(t-1) - p \right)}$

เมื่อ $e_{ij} = \frac{Y_{ij} - \hat{\mu}_{ij}}{\sqrt{V(\hat{\mu}_{ij})}}$ เป็น standardized residual

1.3.3. AR(1) มีค่า $R_i = \text{Corr}(Y_{ij}, Y_{i(j+h)}) = \rho^h ; h = 0, 1, 2, \dots, t-j$ โดยที่ค่า

ประมาณ \hat{R}_i มีค่า $\hat{\rho}_{jj'} = \sum_{i=1}^s \sum_{j=1}^{t-h} \frac{e_{ij}e_{i(j+h)}}{\phi(S(t-1)-p)}$

1.3.4. unstructured มีค่า $R_i = \text{Corr}(Y_{ij}, Y_{ij'}) = \begin{cases} 1 & ; j = j' \\ \rho_{jj'} & ; j \neq j' \end{cases}$ โดยที่

ค่าประมาณ \hat{R}_i มีค่า $\hat{\rho}_{jj'} = \sum_{i=1}^s \frac{e_{ij}e_{ij'}}{\phi(S-p)}$

วิธี GEE เป็นวิธีการสำหรับวิเคราะห์ข้อมูลที่มีการวัดซ้ำ t ครั้ง ด้วยค่าที่วัดแต่ละครั้งมีความสัมพันธ์กัน หรือมีความสัมพันธ์กันภายในตัวแปรตาม Y_i ในการสร้างตัวแบบวิธี GEE ที่ต้องประมาณค่าพารามิเตอร์ β ในตัวแบบนั้นจะใช้วิธีจากการขยายสมการภาวะน่าจะเป็น (likelihood equation) ในวิธี GLM มาวิเคราะห์ร่วมกับความสัมพันธ์หรือรูปแบบ CS (R_i) ภายใน Y_i ซึ่งค่าพารามิเตอร์ β ในตัวแบบวิธี GEE จะได้จากการประมาณค่าที่เป็นตัวประมาณค่าแบบ Generalized Least Squares (GLS) ด้วยวิธีทำให้ฟังก์ชัน $\sum_{i=1}^s (Y_i - \mu_i)' V_i^{-1} (Y_i - \mu_i)$ ต่ำสุด จากการอนุพันธ์เทียบกับ β และแก้สมการหาค่า β ที่ให้ $\sum_{i=1}^s D_i' V_i^{-1} (Y_i - \mu_i) = 0$ เมื่อ $D_i = \frac{\partial \mu_i}{\partial \beta}$; $V_i = \text{Var}(Y_i) = A_i^{1/2} R_i A_i^{1/2}$ ซึ่งค่าประมาณ $\hat{\beta}$ เรียกว่า working covariance matrix; และ A_i เป็นเมตริกซ์ที่แยงมุมมีค่าสมาชิกเท่ากับ $\text{Var}(Y_{ij}) = \mu_{ij}$ โดยการแก้สมการหาค่าประมาณ β จะใช้วิธีการวนซ้ำ (iteration) จนได้ค่าเหมาะสมจากสมการ

$$\hat{\beta}_{(m+1)} = \hat{\beta}_{(m)} + \left(D'_{(m)} V_{(m)}^{-1} D_{(m)} \right)^{-1} D'_{(m)} V_{(m)}^{-1} \left(Y_{(m)} - \hat{\mu}_{(m)} \right)$$

เมื่อ m เป็นรอบที่วนซ้ำ และค่า β ที่เหมาะสม คือ ให้ค่า $|\hat{\beta}_{(m+1)} - \hat{\beta}_{(m)}| \rightarrow 0$

2. การวิเคราะห์ตัวแบบปัจจัยที่มีอิทธิพลต่อจำนวนครั้งการเรียกร่องลินไหมทดแทนที่มีการวัดซ้ำสำหรับหน่วยเสี่ยงภัยเฉพาะราย ด้วยวิธี GLMM โดยจะเกี่ยวข้องกับ 3 ส่วนต่อไปนี้ [7]

2.1. ความแปรปรวน Y_{ij} ภายใต้กำหนดปัจจัยอิทธิพลสุ่มในที่นี้ศึกษาเฉพาะปัจจัยอิทธิพลสุ่มค่าคงที่ (random intercept: b_{0i}) มีค่า

$$\text{Var}(Y_{ij} | b_{0i}) = \phi V \left(E(Y_{ij} | b_{0i}) \right) = E(Y_{ij} | b_{0i}) \text{ เมื่อ } \phi = 1$$

2.2. ค่าเฉลี่ย Y_{ij} ภายใต้กำหนดปัจจัยอิทธิพลกลุ่มค่าคงที่ จะมีตัวแปรทำนายในรูปแบบเชิงเส้น (linear predictor) ประกอบด้วยเทอมความสัมพันธ์ของปัจจัยอิทธิพลคงที่ (X_{ij}) และปัจจัยอิทธิพลกลุ่มค่าคงที่ที่มีการเชื่อมโยงความสัมพันธ์ด้วยฟังก์ชันลอการิทึม

$$\log(E(Y_{ij} | b_{0i})) = X'_{ij}\beta + b_{0i} \quad (2)$$

2.3. กำหนด $b_{0i} \sim N(0, \sigma_{b_{0i}}^2)$ และความแปรปรวนร่วม (covariance) ของ Y_i

$\text{Var}(Y_i) = \sigma_{b_{0i}}^2 + \sigma_{\varepsilon_i}^2$ เมื่อ ε_i เป็นความคลาดเคลื่อนในการวัดภายในแต่ละหน่วยเสี่ยงภัย (within exposure unit measurement error) ที่มี e_i เป็นค่าประมาณเรียกว่า เศษเหลือ (residual) ในที่นี้กำหนด $\sigma_{\varepsilon_i}^2 = \text{Var}(\varepsilon_i)$ มีลักษณะ Variance Component (VC) ดังนี้

$$\sigma_{\varepsilon_i}^2 = \text{Var}(\varepsilon_i) = \sigma^2 I_t \text{ เมื่อ } I_t \text{ เป็น identity matrix ขนาดเมตริกซ์ } t \times t$$

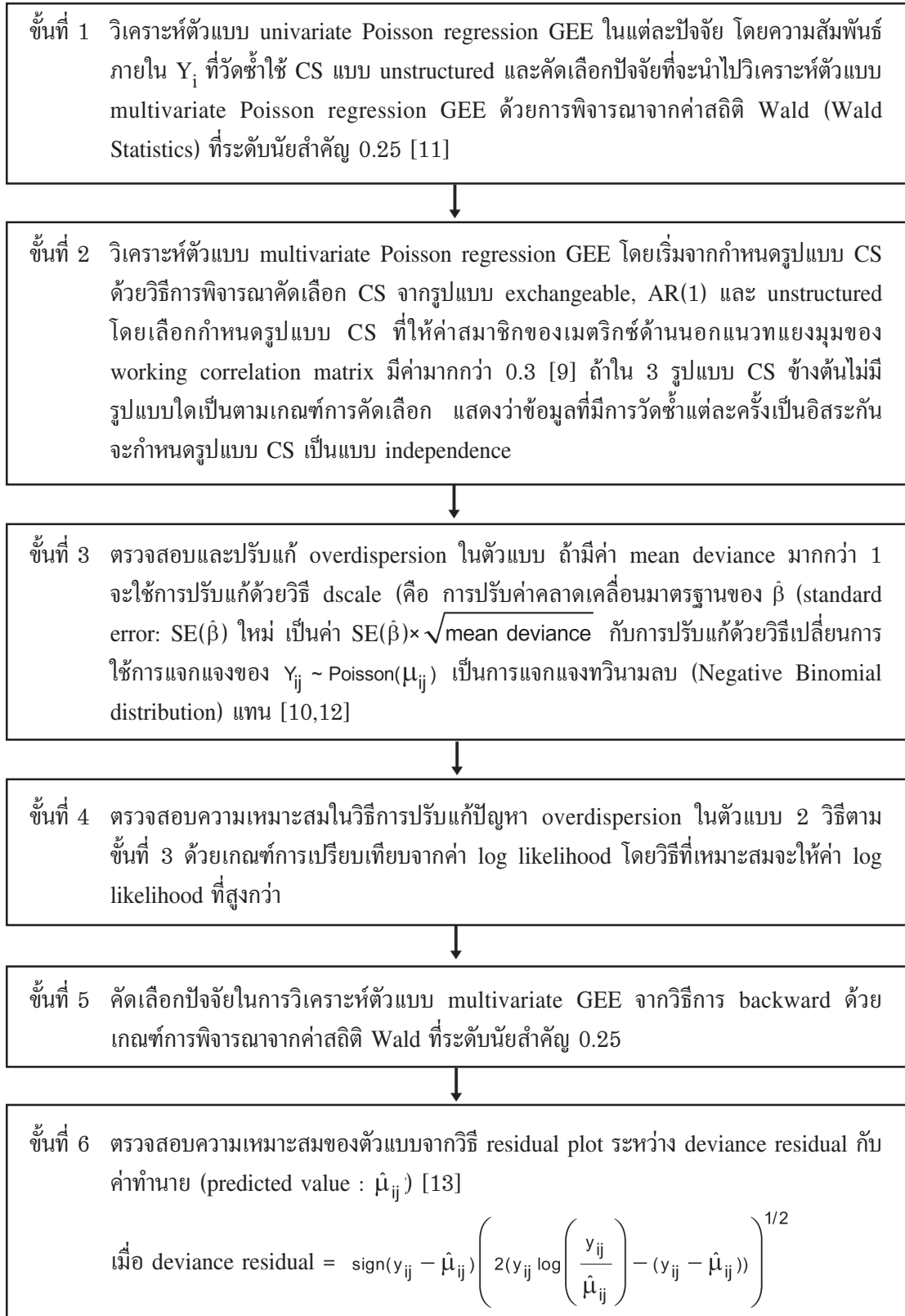
การสร้างตัวแบบวิธี GLMM จะประมาณค่าพารามิเตอร์ปัจจัยอิทธิพลคงที่ (β) [10] ด้วยวิธีภาวะน่าจะเป็นสูงสุด (Maximum Likelihood: ML) โดยการอนุพันธ์ฟังก์ชัน $\log L(\beta)$ เทียบกับ β และแก้สมการหาค่าที่ทำให้ $\frac{\partial}{\partial \beta} \log L(\beta) = 0$ เมื่อ

$$L(\beta) = \prod_{i=1}^s \prod_{j=1}^t f(y_{ij}) = \prod_{i=1}^s \int \prod_{j=1}^t f_{Y_{ij}|b_{0i}}(y_{ij} | b_{0i}) h_{b_{0i}}(b_{0i}) db_{0i}$$

เมื่อ $Y_{ij} | b_{0i} \sim \text{Poisson}(E(Y_{ij} | b_{0i}))$

ขั้นตอนการวิเคราะห์ข้อมูล

1. วิเคราะห์สร้างตัวแบบปัจจัยที่มีอิทธิพลต่อจำนวนครั้งการเรียกร่องลินไหมทดแทนที่มีการวัดซ้ำสำหรับภาพรวมของหน่วยเสี่ยงภัย ด้วยวิธี GEE จะใช้การประมวลผลจากโปรแกรม SAS[®] v.9.1 ด้วย PROC GENMOD สำหรับการวิเคราะห์ที่จำนวนครั้งการเรียกร่องลินไหมทดแทนมีการแจกแจงแบบปัวซอง



2. วิเคราะห์สร้างตัวแบบ ปัจจัยที่มีอิทธิพลคงที่ ปัจจัยที่มีอิทธิพลสุ่มต่อจำนวนครั้งการเรียกร้องสินไหมทดแทน ที่มีการวัดซ้ำของหน่วยเสี่ยงภัยเฉพาะรายด้วยวิธี GLMM จะใช้การประมวลผลจากโปรแกรม SAS® v.9.1 ด้วย PROC GLIMMIX สำหรับการวิเคราะห์ที่จำนวนครั้งการเรียกร้องสินไหมทดแทนมีการแจกแจงแบบปัวซอง

ขั้นที่ 1 วิเคราะห์ univariate Poisson regression GLMM ในแต่ละปัจจัยซึ่งเป็นปัจจัยอิทธิพลคงที่ โดยภายในตัวแบบมีปัจจัยอิทธิพลสุ่มค่าคงที่ และคัดเลือกปัจจัยอิทธิพลคงที่ที่จะนำไปวิเคราะห์ multivariate Poisson regression GLMM ด้วยเกณฑ์การพิจารณาจากค่าสถิติ t (t statistics) ที่ระดับนัยสำคัญ 0.25

ขั้นที่ 2 วิเคราะห์ multivariate Poisson regression GLMM โดยภายในตัวแบบมีปัจจัยอิทธิพลสุ่มค่าคงที่และกำหนดรูปแบบเมตริกซ์ variance-covariance ของ ϵ_i เป็น VC แล้วคัดเลือกปัจจัยอิทธิพลคงที่จากวิธีการ backward ด้วยเกณฑ์การพิจารณาจากค่าสถิติ t ที่ระดับนัยสำคัญ 0.25 และพร้อมทั้งตรวจสอบความเหมาะสมของปัจจัยอิทธิพลคงที่ที่คัดเลือกในตัวแบบด้วยการทดสอบอัตราส่วนภาวะน่าจะเป็น (Likelihood Ratio test: LR test) ที่ระดับนัยสำคัญ 0.05

ขั้นที่ 3 ตรวจสอบความเหมาะสมของปัจจัยอิทธิพลสุ่มค่าคงที่ในตัวแบบด้วยเกณฑ์การพิจารณาจากค่าสถิติ Mean Generalized Chi-Square ที่มีค่าคู่เข้าค่า 1 [14] และพิจารณาจากค่าประมาณ

$$\text{Intra-Class Correlation (ICC)} = \frac{\hat{\sigma}_{b_0}^2}{\hat{\sigma}_{b_0}^2 + \hat{\sigma}_e^2} \text{ มีค่าในช่วง } [0,1] \text{ ถ้ามีค่า ICC ต่ำ}$$

แสดงว่าลักษณะความแตกต่างกันในหน่วยเสี่ยงภัยแต่ละรายโดยรวมส่วนใหญ่ ($\hat{\sigma}_{b_0}^2$) มีอิทธิพลค่อนข้างต่ำต่อจำนวนครั้งการเรียกร้องสินไหมทดแทน

ผลการวิเคราะห์ข้อมูล

1. ผลวิเคราะห์ตัวแบบปัจจัยที่มีอิทธิพลต่อจำนวนครั้งการเรียกร้องสินไหมทดแทนที่มีการวัดซ้ำสำหรับภาพรวมของหน่วยเสี่ยงภัย ด้วยวิธี GEE

ผลการวิเคราะห์ univariate Poisson regression GEE ที่ระดับนัยสำคัญ 0.25 มีเพียงปัจจัย code, car_gr, cc_gr, NCB และ premium ที่นำไปวิเคราะห์ตัวแบบ multivariate Poisson regression GEE โดยมีผลการคัดเลือกรูปแบบ CS ในการวิเคราะห์ตัวแบบ multivariate Poisson regression GEE จากการพิจารณาค่า working correlation matrix ดังตารางที่ 3

ตารางที่ 3 แสดงค่า working correlation matrix ในแต่ละแบบของ CS

รูปแบบ CS	ปีกรมธรรม์ที่ 1	ปีกรมธรรม์ที่ 2	ปีกรมธรรม์ที่ 3	ปีกรมธรรม์ที่ 4
AR(1)				
ปีกรมธรรม์ที่ 1	1			
ปีกรมธรรม์ที่ 2	0.1354	1		
ปีกรมธรรม์ที่ 3	0.0183	0.1354	1	
ปีกรมธรรม์ที่ 4	0.0025	0.0183	0.1354	1
Exchangeable				
ปีกรมธรรม์ที่ 1	1			
ปีกรมธรรม์ที่ 2	0.1043	1		
ปีกรมธรรม์ที่ 3	0.1043	0.1043	1	
ปีกรมธรรม์ที่ 4	0.1043	0.1043	0.1043	1
Unstructured				
ปีกรมธรรม์ที่ 1	1			
ปีกรมธรรม์ที่ 2	0.1427	1		
ปีกรมธรรม์ที่ 3	0.1577	0.2205	1	
ปีกรมธรรม์ที่ 4	0.0614	0.0027	0.0696	1

ผลจากตารางที่ 3 เปรียบเทียบค่าด้านนอกแนวทแยงมุมของเมตริกซ์ในแต่ละรูปแบบ CS พบว่าทุกรูปแบบมีค่าค่อนข้างน้อย และน้อยกว่า 0.3 หมายความว่า ค่าจำนวนครั้งการเรียกร้องสินไหมทดแทนที่วัดซ้ำในแต่ละปีกรมธรรม์ สำหรับภาพรวมของหน่วยเสี่ยงภัยเป็นอิสระกัน ดังนั้นในการวิจัยนี้เลือก CS ในรูปแบบ independence สำหรับการวิเคราะห์ multivariate Poisson regression GEE โดยจะได้ตัวแบบเดียวกันกับการวิเคราะห์ตัวแบบด้วยวิธี GLM และจากผลการตรวจสอบความเหมาะสมของตัวแบบที่ในตัวแบบประกอบด้วยปัจจัย code, car_gr, cc_gr, NCB และ premium ดังตารางที่ 4 จากค่า mean deviance เท่ากับ 1.4536 มีค่าเกินกว่า 1.0 แสดงว่าเกิด overdispersion กับตัวแบบที่จะต้องปรับแก้ค่าก่อนการคัดเลือกปัจจัยในการวิเคราะห์ตัวแบบ multivariate Poisson regression GEE

ตารางที่ 4 เปรียบเทียบวิธีการปรับแก้ overdispersion

วิธี	mean deviance	log likelihood
การแจกแจงแบบปัวซองที่ปรับแก้ด้วย dscale	1.4536	-843.4758
การแจกแจงทวินามลบ	0.9483	-1157.7856

เมื่อพิจารณาผลการเปรียบเทียบวิธีการปรับแก้ overdispersion 2 วิธี ด้วยการพิจารณาค่า log likelihood พบว่าวิธีปรับแก้ที่เหมาะสม คือ การปรับแก้ด้วยวิธี dscale ในการแจกแจงแบบปัวซอง เนื่องจากมีค่า log likelihood สูงกว่า ส่วนผลการวิเคราะห์ตัวแบบ multivariate หลังการปรับแก้ค่า overdispersion จากวิธี dscale และคัดเลือกตัวแปรปัจจัยโดยวิธี backward จากเกณฑ์การพิจารณาด้วยค่าสถิติ Wald ที่ระดับนัยสำคัญ 0.25 ดังตารางที่ 5 โดยที่ระดับนัยสำคัญ 0.05 มีเพียงปัจจัย cc_gr และ NCB ที่มีอิทธิพลต่อจำนวนครั้งการเรียกร่องสินไหมทดแทน สำหรับภาพรวมของหน่วยเสี่ยงภัย ถ้าให้ค่าส่วนลดจากประวัติการเรียกร่องสินไหมทดแทนกับหน่วยเสี่ยงภัย (NCB) เพิ่มขึ้นทุกๆ 5% ของเบี้ยประกันภัยจะมีผลต่อจำนวนครั้งการเรียกร่องสินไหมทดแทนเฉลี่ยลดลงเท่ากับ $\exp(-0.0117) = 0.9884$ นั่นคือ ประมาณ 1 ครั้ง เมื่อตัวแปรปัจจัยขนาดเครื่องยนต์ (cc_gr) คงที่ และกรณีเมื่อตัวแปรปัจจัยอื่นๆ คงที่ หน่วยเสี่ยงภัยที่มีขนาดเครื่องยนต์ $CC \leq 1500$ (cc_gr1) ขนาดเครื่องยนต์ (1500, 2000] (cc_gr2) และขนาดเครื่องยนต์ (2000, 2500] (cc_gr3) จะมีจำนวนครั้งการเรียกร่องสินไหมทดแทนเฉลี่ยมากกว่าหน่วยเสี่ยงภัยที่มีขนาดเครื่องยนต์ $CC > 2500$ (cc_gr4) จำนวน 49.74%, 52.68% และ 49.14% ตามลำดับ ซึ่งมีตัวแปรปัจจัยที่มีอิทธิพลต่อจำนวนครั้งการเรียกร่องสินไหมทดแทนที่มีการวัดซ้ำ สำหรับภาพรวมของหน่วยเสี่ยงภัยในรูปแบบตัวแปรทำนายเชิงเส้น คือ

$$\begin{aligned} \log(\hat{\mu}_{ij}) = & -0.3677 + 0.4037X_{ij,cc_gr1}^* + 0.4232X_{ij,cc_gr2}^* \\ & + 0.3997X_{ij,cc_gr3}^* - 0.0117X_{ij,NCB}^* \end{aligned} \quad (3)$$

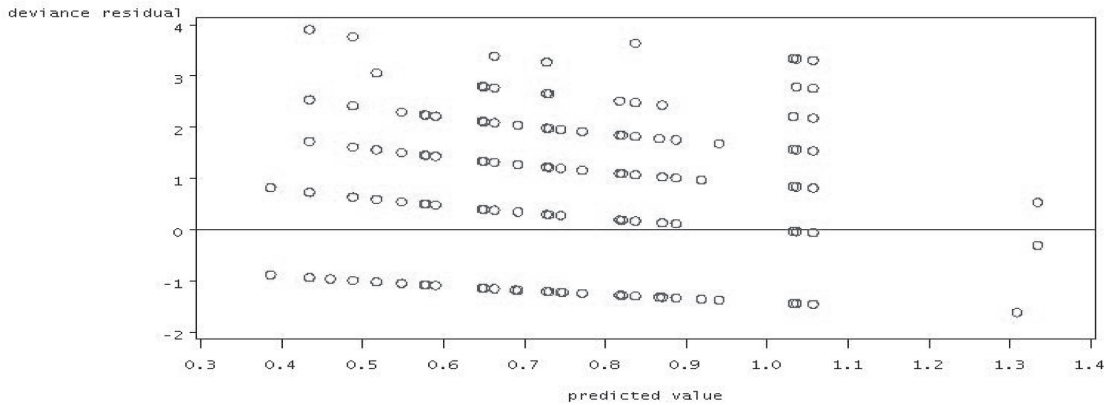
เมื่อ * ปัจจัยที่มีนัยสำคัญที่ระดับนัยสำคัญ 0.05

ตารางที่ 5 ผลวิเคราะห์ multivariate Poisson regression GEE

(ตัวแบบ : $g(\mu_{ij}) = \log(\mu_{ij}) = \eta_{ij} = X'_{ij}\beta$)

ปัจจัย	$\hat{\beta}$	SE($\hat{\beta}$)	exp($\hat{\beta}$)	ค่าสถิติ z	p-value
Intercept	-0.3677	0.1071	0.6923	-3.43	0.0006*
cc_gr1 (CC ≤ 1500)	0.4037	0.1239	1.4974	3.26	0.0011*
cc_gr2 (1500, 2000]	0.4232	0.1190	1.5268	3.55	0.0004*
cc_gr3 (2000, 2500]	0.3997	0.1377	1.4914	2.90	0.0037*
cc_gr4 (CC >2500)
NCB	-0.0117	0.0028	0.9884	-4.19	<0.0001*

หมายเหตุ: * มีนัยสำคัญที่ระดับนัยสำคัญ 0.05



รูปที่ 1 แสดง residual plot ระหว่าง deviance residual กับค่าทำนาย ($\hat{\mu}_{ij}$)

จากรูปที่ 1 ใช้ตรวจสอบความเหมาะสมของตัวแบบ พบว่าเป็นการกระจายที่ไม่มีรูปลักษณะแบบแผน (no any pattern) นั้นหมายความว่า ตัวแบบดังสมการ (3) มีความเหมาะสมกับข้อมูล

ตารางที่ 6 ผลวิเคราะห์ multivariate Poisson regression GLMM โดยที่ภายในตัวแบบมีปัจจัยอิทธิพลร่วมค่าคงที่ (ตัวแบบ : $g(E(Y_{ij} | b_{0i})) = \log(E(Y_{ij} | b_{0i})) = X'_{ij}\beta + b_{0i}$)

ปัจจัยอิทธิพลคงที่	β	SE(β)	exp(β)	ค่าสถิติ t	p-value
Intercept	-0.6932	0.2139	0.50	-3.24	0.0013*
code1 (รหัส 110)	-0.1903	0.1489	0.8267	-1.28	0.2014
code2 (รหัส 120)
cc_gr1 (CC ≤ 1500)	0.5065	0.1445	1.6595	3.51	0.0005*
cc_gr2 (1500, 2000]	0.4595	0.1280	1.5833	3.59	0.0003*
cc_gr3 (2000, 2500]	0.4055	0.1392	1.5001	2.91	0.0037*
cc_gr4 (CC > 2500)
premium	0.0075	0.0048	1.0075	1.56	0.1192
ปัจจัยอิทธิพลร่วมค่าคงที่ (b_0):	$\hat{\sigma}_{b_0}^2 = 0.3294$		SE(b_0) = 0.05693		
ค่าเศษเหลือ (e) :	$\hat{\sigma}_e^2 = 1.1143$		SE($\hat{\sigma}_e^2$) = 0.04912		
ค่า Mean Generalized Chi-Square = 1.11 ;	$ICC = \frac{\hat{\sigma}_{b_0}^2}{\hat{\sigma}_{b_0}^2 + \hat{\sigma}_e^2} = \frac{0.3294}{0.3294 + 1.1143} = 0.2282$				

หมายเหตุ: * มีนัยสำคัญที่ระดับนัยสำคัญ 0.05

2. ผลวิเคราะห์ตัวแบบปัจจัยที่มีอิทธิพลคงที่ ปัจจัยที่มีอิทธิพลสุ่ม ต่อจำนวนครั้งการเรียกร้องสินไหมทดแทน ที่มีการวัดซ้ำของหน่วยเสี่ยงภัยเฉพาะราย ด้วยวิธี GLMM ผลการวิเคราะห์ univariate Poisson regression GLMM มีปัจจัย code, car_gr, cc_gr, NCB และ premium ที่มีนัยสำคัญระดับ 0.25 เป็นปัจจัยอิทธิพลคงที่ที่คัดเลือกนำไปวิเคราะห์ใน multivariate Poisson regression GLMM

จากผลในตารางที่ 6 ปัจจัยอิทธิพลคงที่ในตัวแบบมีเพียงปัจจัยขนาดเครื่องยนต์ (cc_gr) ที่มีต่อจำนวนครั้งการเรียกร้องสินไหมทดแทนที่ระดับนัยสำคัญ 0.05 เมื่อตัวแปรปัจจัยอิทธิพลคงที่อื่นๆ คงที่ แล้วหน่วยเสี่ยงภัยที่มีขนาดเครื่องยนต์ $CC \leq 1500$ (cc_gr1), ขนาดเครื่องยนต์ (1500, 2000] (cc_gr2) และขนาดเครื่องยนต์ (2000, 2500] (cc_gr3) จะมีจำนวนครั้งการเรียกร้องสินไหมทดแทนเฉลี่ยมากกว่าหน่วยเสี่ยงภัยที่มีขนาดเครื่องยนต์ $CC > 2500$ (cc_gr4) จำนวน 65.95%, 58.33% และ 50.01% ตามลำดับ ซึ่งมีตัวแบบปัจจัยที่มีอิทธิพลคงที่ ปัจจัยที่มีอิทธิพลสุ่มค่าคงที่ ต่อจำนวนครั้งการเรียกร้องสินไหมทดแทนที่มีการวัดซ้ำของหน่วยเสี่ยงภัยเฉพาะราย ในรูปแบบตัวแปรทำนายเชิงเส้นดังนี้

$$\begin{aligned} \log(\hat{E}(Y_{ij} | b_{oi})) &= (-0.6932 + \hat{b}_{oi}) - 0.1903X_{ij,code1} \\ &\quad + 0.5065X_{ij,cc_gr1}^* + 0.4595X_{ij,cc_gr2}^* \\ &\quad + 0.4055X_{ij,cc_gr3}^* + 0.007463X_{ij,premium} \end{aligned} \quad (4)$$

หรือตัวแบบค่าประมาณจำนวนครั้งการเรียกร้องสินไหมทดแทนที่มีการวัดซ้ำของหน่วยเสี่ยงภัยเฉพาะราย คือ

$$\begin{aligned} \hat{E}(Y_{ij} | b_{oi}) &= \exp((-0.6932 + \hat{b}_{oi}) - 0.1903X_{ij,code1} \\ &\quad + 0.5065X_{ij,cc_gr1}^* + 0.4595X_{ij,cc_gr2}^* \\ &\quad + 0.4055X_{ij,cc_gr3}^* + 0.007463X_{ij,premium}) \end{aligned} \quad (5)$$

เมื่อ * ปัจจัยที่มีนัยสำคัญที่ระดับนัยสำคัญ 0.05

ผลการพิจารณาปัจจัยอิทธิพลสุ่มค่าคงที่ในหน่วยเสี่ยงภัยเฉพาะรายที่เพิ่มเข้าไปในตัวแบบพบว่ามีความเหมาะสมด้วยค่าสถิติ Mean Generalized Chi-Square มีค่าเท่ากับ 1.11 โดยที่ปัจจัยอิทธิพลสุ่มค่าคงที่จะมีผลกระทบต่อจำนวนครั้งการเรียกร้องสินไหมทดแทนของหน่วยเสี่ยงภัยเฉพาะราย แต่มีผลกระทบเฉพาะเพียงหน่วยเสี่ยงภัยบางรายเท่านั้น ซึ่งเป็นจำนวนส่วนน้อยรายจากที่มีค่า ICC เท่ากับ 0.2282 มีค่าค่อนข้างต่ำ ในข้อมูลทีวิเคราะห์นี้จำนวนหน่วยเสี่ยงภัยรวมทั้งสิ้น 331 ราย จะมีเพียงจำนวน 16 รายเท่านั้นที่ปัจจัยอิทธิพลสุ่มค่าคงที่มีผลต่อจำนวนครั้งการเรียกร้องสินไหมทดแทนของหน่วยเสี่ยงภัยในรายนั้นๆ ณ ระดับนัยสำคัญ 0.05 ดังตารางที่ 7 แสดงถึงค่าประมาณและทดสอบปัจจัยอิทธิพลสุ่มค่าคงที่ที่มีต่อจำนวนครั้งการเรียกร้องสินไหมทดแทนเฉลี่ยของหน่วยเสี่ยงภัยเฉพาะราย ในรายที่ 16 ถึงรายที่ 20 ที่ผลการทดสอบในระดับนัยสำคัญ 0.05 มีเพียงหน่วยเสี่ยงภัยเฉพาะรายที่ 16 และรายที่ 18 ที่มีปัจจัยอิทธิพลสุ่มค่าคงที่ที่มีต่อจำนวนครั้งการเรียกร้องสินไหมทดแทนเฉลี่ย

ตารางที่ 7 ค่าประมาณและทดสอบปัจจัยอิทธิพลคู่ค่าคงที่ที่มีผลต่อจำนวนครั้งการเรียกร้อยสินไหมทดแทนของหน่วยเสี่ยงภัยเฉพาะราย ในหน่วยเสี่ยงภัยรายที่ 16 ถึงรายที่ 20

หน่วยเสี่ยงภัยรายที่ (i)	(\hat{b}_{0i})	$SE(\hat{b}_{0i})$	$\exp(\hat{b}_{0i})$	ค่าสถิติ t	p-value
16	0.9290	0.3214	2.5320	2.89	0.0039*
17	-0.01512	0.4195	0.9850	-0.04	0.9713
18	0.8914	0.3218	2.4385	2.77	0.0057*
19	-0.1818	0.4339	0.8338	-0.42	0.6753
20	-0.3714	0.4908	0.6898	-0.76	0.4494

หมายเหตุ: * มีนัยสำคัญที่ระดับนัยสำคัญ 0.05

ตัวอย่างตัวแบบค่าประมาณจำนวนครั้งการเรียกร้อยสินไหมทดแทนเฉลี่ยของหน่วยเสี่ยงภัยเฉพาะรายที่ 16 คือ

$$\begin{aligned}
 \hat{E}(Y_{16} | b_{0(16)}) &= \exp((-0.6932 + \hat{b}_{0(16)}) - 0.1903X_{16,code1} \\
 &\quad + 0.5065X_{16,cc_gr1}^* + 0.4595X_{16,cc_gr2}^* \\
 &\quad + 0.4055X_{16,cc_gr3}^* + 0.007463X_{16,premium}) \\
 &= \exp((-0.6932 + 0.9290) - 0.1903X_{16,code1} \\
 &\quad + 0.5065X_{16,cc_gr1}^* + 0.4595X_{16,cc_gr2}^* \\
 &\quad + 0.4055X_{16,cc_gr3}^* + 0.007463X_{16,premium})
 \end{aligned}$$

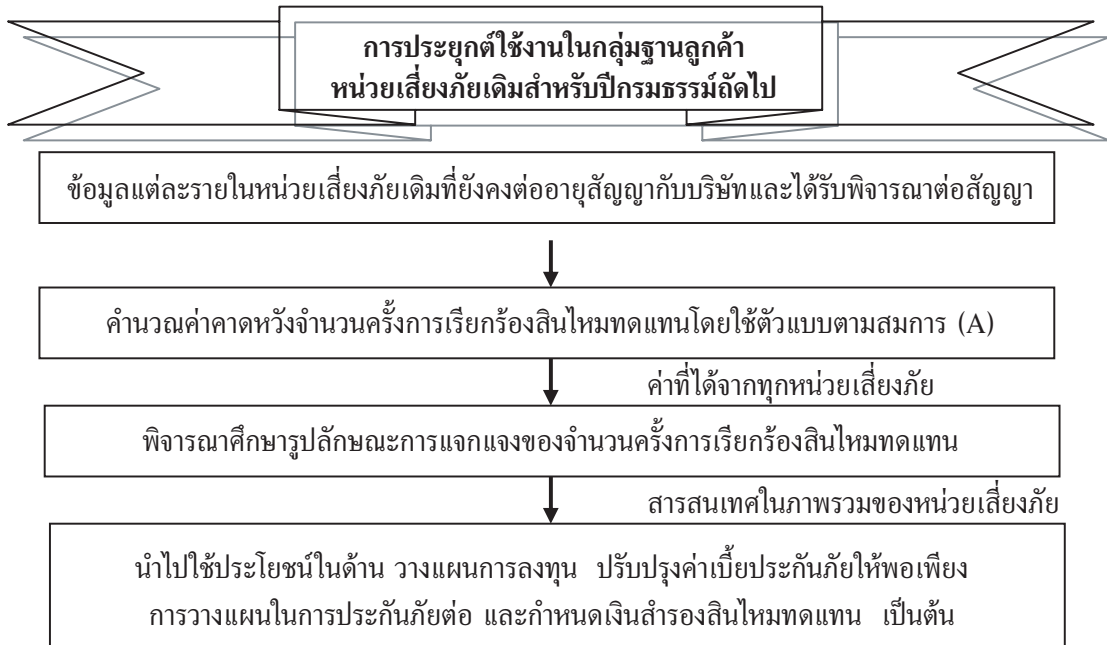
นั่นคือจากตัวอย่างหน่วยเสี่ยงภัยเฉพาะรายที่ 16 เมื่อปัจจัยอิทธิพลคงที่มีลักษณะคงที่ ด้วยคุณลักษณะเฉพาะรายในหน่วยเสี่ยงภัยรายที่ 16 จะมีจำนวนครั้งการเรียกร้อยสินไหมทดแทนเฉลี่ยเท่ากับ $\exp(-0.6932+0.9290) = 1.27$ ครั้ง หรือประมาณ 1 ครั้ง

สรุปผลและการประยุกต์ใช้งาน

1. ผลจากตัวแบบปัจจัยที่มีอิทธิพลต่อจำนวนครั้งการเรียกร่องสินไหมทดแทนที่มีการวัดซ้ำสำหรับภาพรวมของหน่วยเสี่ยงภัยในการประกันภัยรถยนต์ ด้วยวิธี GEE ในที่นี้จะเป็นตัวแบบเดียวกันกับการวิเคราะห์ข้อมูลด้วยวิธี GLM เนื่องจากผลวิเคราะห์ข้อมูลที่วัดซ้ำในหน่วยเสี่ยงภัยเดียวกันเป็นรายปีกรรมธรรม์ตลอดช่วงระยะเวลา 4 ปีกรรมธรรม์ติดต่อกัน มีความสัมพันธ์กันค่อนข้างต่ำหรือจะสรุปได้ว่าไม่มีความสัมพันธ์กันภายใน y_i โดยปัจจัยที่มีอิทธิพลต่อจำนวนครั้งการเรียกร่องสินไหมทดแทน สำหรับภาพรวมของหน่วยเสี่ยงภัย ที่ระดับนัยสำคัญ 0.05 มีเพียงปัจจัยขนาดเครื่องยนต์ (cc_gr) และส่วนลดหรือเพิ่มจากประวัติการเรียกร่องสินไหมทดแทน (NCB) ซึ่งถ้าให้ค่าส่วนลดจากประวัติการเรียกร่องสินไหมทดแทนแก่หน่วยเสี่ยงภัยเพิ่มขึ้นทุกๆ 5% ของเบี้ยประกันภัย จะมีผลต่อจำนวนครั้งการเรียกร่องสินไหมทดแทนเฉลี่ยลดลงประมาณ 1 ครั้ง เมื่อปัจจัยอื่นๆ คงที่ และเช่นกันกรณีเมื่อปัจจัยอื่นๆ คงที่ หน่วยเสี่ยงภัยที่มีขนาดเครื่องยนต์ $CC \leq 1500$ (cc_gr1) ขนาดเครื่องยนต์ (1500, 2000] (cc_gr2) และขนาดเครื่องยนต์ (2000, 2500] (cc_gr3) จะมีจำนวนครั้งการเรียกร่องสินไหมทดแทนเฉลี่ยมากกว่าหน่วยเสี่ยงภัยที่มีขนาดเครื่องยนต์ $CC > 2500$ (cc_gr4) จำนวน 49.74%, 52.68% และ 49.14% ตามลำดับ โดยมีตัวแบบปัจจัยที่มีอิทธิพลต่อจำนวนครั้งการเรียกร่องสินไหมทดแทนที่มีวัดซ้ำ สำหรับภาพรวมของหน่วยเสี่ยงภัย ในรูปแบบค่าทำนาย ($\hat{\mu}_i$) คือ

$$\begin{aligned} \hat{\mu}_i = \exp(-0.3677 + 0.4037X_{i,cc_gr1}^* + 0.4232X_{i,cc_gr2}^* \\ + 0.3997X_{i,cc_gr3}^* - 0.0117X_{i,NCB}^*) \end{aligned} \quad (A)$$

เมื่อ * ปัจจัยที่มีนัยสำคัญที่ระดับนัยสำคัญ 0.05



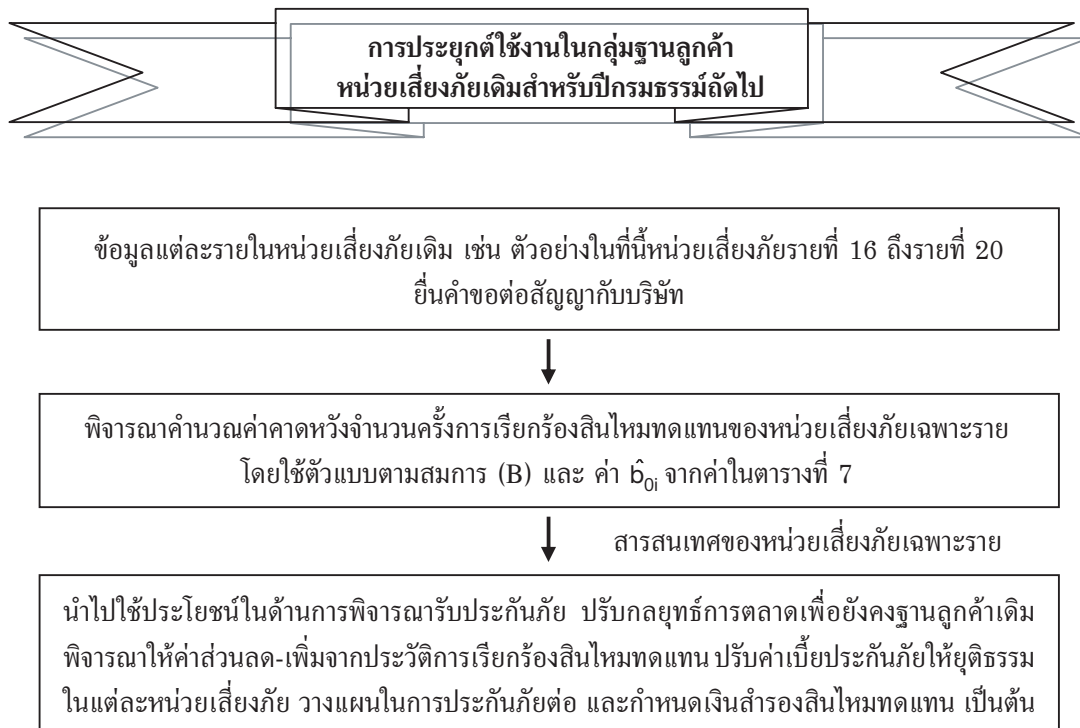
รูปที่ 2 การประยุกต์ใช้งานจากตัวแบบตามสมการ (A)

2. ผลจากตัวแบบปัจจัยที่มีอิทธิพลคงที่ ปัจจัยที่มีอิทธิพลสุ่ม ต่อจำนวนครั้งการเรียกใช้งานใหม่ทดแทนที่มีการวัดซ้ำสำหรับหน่วยเสี่ยงภัยเฉพาะรายในการประกันภัยรถยนต์ ด้วยวิธี GLMM เมื่อมีปัจจัยอิทธิพลสุ่มค่าคงที่ในตัวแบบที่วิเคราะห์ด้วยวิธี GLMM จะให้ความผันแปรของ $\hat{\beta}$ (SE($\hat{\beta}$)) แตกต่างจากวิธี GEE มีผลต่อการคัดเลือกปัจจัยอิทธิพลคงที่ จึงทำให้ได้ผลของตัวแปรปัจจัยอิทธิพลคงที่ในตัวแบบจำนวนครั้งการเรียกใช้งานใหม่ทดแทนของหน่วยเสี่ยงภัยเฉพาะรายที่เหมาะสมจากวิธี GLMM ไม่เหมือนกับในตัวแบบจำนวนครั้งการเรียกใช้งานใหม่ทดแทนสำหรับภาพรวมของหน่วยเสี่ยงภัยที่เหมาะสมจากวิธี GEE โดยปัจจัยที่มีอิทธิพลคงที่มีเพียงปัจจัยขนาดเครื่องยนต์ (cc_gr) เท่านั้นที่มีอิทธิพลต่อจำนวนครั้งการเรียกใช้งานใหม่ทดแทน ที่ระดับนัยสำคัญ 0.05 เมื่อตัวแปรปัจจัยอิทธิพลคงที่อื่นๆ คงที่ แล้วหน่วยเสี่ยงภัยที่มีขนาดเครื่องยนต์ $CC \leq 1500$ (cc_gr1) ขนาดเครื่องยนต์ (1500, 2000] (cc_gr2) และขนาดเครื่องยนต์ (2000, 2500] (cc_gr3) จะมีจำนวนครั้งการเรียกใช้งานใหม่ทดแทนเฉลี่ยมากกว่าหน่วยเสี่ยงภัยที่มีขนาดเครื่องยนต์ $CC > 2500$ (cc_gr4) จำนวน 65.95%, 58.33% และ 50.01% ตามลำดับ

ในส่วนของปัจจัยที่มีอิทธิพลสุ่มในหน่วยเสี่ยงภัยเฉพาะราย (b_{0i}) สรุปได้ว่าปัจจัยอิทธิพลสุ่มค่าคงที่ที่มีอิทธิพลต่อจำนวนครั้งการเรียกใช้งานใหม่ทดแทนของหน่วยเสี่ยงภัยเฉพาะรายในจำนวนไม่มากคือมีเพียง 16 ราย จากหน่วยเสี่ยงภัยทั้งสิ้น 331 ราย ที่มีนัยสำคัญ ณ ระดับนัยสำคัญ 0.05 โดยมีตัวแบบปัจจัยที่มีอิทธิพลคงที่ ปัจจัยที่มีอิทธิพลสุ่ม ต่อจำนวนครั้งการเรียกใช้งานใหม่ทดแทนที่มีการวัดซ้ำของหน่วยเสี่ยงภัยเฉพาะรายในรูปแบบค่าทำนาย ($\hat{E}(Y_i | b_{0i})$) คือ

$$\begin{aligned} \hat{E}(Y_i | b_{0i}) = & \exp((-0.6932 + \hat{b}_{0i}) - 0.1903X_{i,\text{code1}} \\ & + 0.5065X_{i,\text{cc_gr1}}^* + 0.4595X_{i,\text{cc_gr2}}^* \\ & + 0.4055X_{i,\text{cc_gr3}}^* + 0.007463X_{i,\text{premium}}) \end{aligned} \quad (B)$$

เมื่อ * ปัจจัยที่มีนัยสำคัญที่ระดับนัยสำคัญ 0.05



รูปที่ 3 การประยุกต์ใช้งานจากตัวแบบตามสมการ (B)

เอกสารอ้างอิง

1. Renshaw, A. E. 1994. Modelling the Claims Process in the Presence of Covariates. *ASTIN Bulletin* 24(2): 265-285.
2. Nelder, J. A., and Verrall, R. J. 1997. Credibility Theory and Generalized Linear Models. *ASTIN Bulletin* 27(1): 71-82.
3. Ohlsson, E., and Johansson, B. 2004. Combining Credibility and GLM for Rating of Multi-Level Factors. Available from URL: <http://www.casact.org/pubs/dpp/dpp04/04dpp315.pdf>. 5 March 2004.

4. Yau, K. K. W., Yip, K. C. H., and Yuen, H. K. 2003. Modelling Repeated Insurance Claim Frequency Data Using the Generalized Linear Mixed Model. *Journal of Applied Statistics* 30(8): 857-865.
5. สำนักงานคณะกรรมการกำกับและส่งเสริมการประกอบธุรกิจประกันภัย. พิกัดอัตราเบี้ยประกันภัยรถยนต์: ข้อบังคับทั่วไป. ได้จาก: <http://www.oic.or.th/download/premium/Recomment.pdf>. 29 ธันวาคม 2552.
6. สำนักงานคณะกรรมการกำกับและส่งเสริมการประกอบธุรกิจประกันภัย. พิกัดอัตราเบี้ยประกันภัยรถยนต์. ได้จาก: <http://www.oic.or.th/upload/motorinsurrance/download/730-5741.pdf>. 29 ธันวาคม 2552.
7. Fitzmaurice, G. M., Laird N. M., and Ware, J. H. 2004. *Applied Longitudinal Analysis*. New Jersey. John Wiley and Sons, Inc.
8. Lipsitz, S. R., Kim, K., and Zhao, L. 1994. Analysis of Repeated Categorical Data Using Generalized Estimating Equations. *Statistics in Medicine* 13: 1149-1163.
9. Myers, R. H., Montgomery, D. C., and Vining, G. G. 2002. *Generalized Linear Models with Applications in Engineering and the Sciences*. New York. John Wiley and Sons, Inc.
10. Jong, P. D., and Heller, G. Z. 2008. *Generalized Linear Models for Insurance Data*. New York. Cambridge University Press.
11. Hosmer, D. W., and Leweshow, S. 2000. *Applied Logistic Regression*. 2nd Edition. New York. John Wiley and Sons, Inc.
12. Shengwang, M., Wei, Y., and Whitmore, G. A. 1999. Accounting for Individual Over-Dispersion in a Bonus-Malus Automobile Insurance System. *ASTIN Bulletin* 29(2): 327-337.
13. Zuur, A. F., Ieno, E. N., Walker, N. J., Saveliev, A. A., and Smith, G. M. 2009. *Mixed Effects Models and Extensions in Ecology with R*. New York. Springer Science+Business Media.
14. Schabenberger, O. 2005. Introducing the GLIMMIX Procedure for Generalized Linear Mixed Models-SUGI30. Available from URL: <http://www2.SAS.com/proceedings/sugi30/196-30.pdf>. 10 April 2005.

ได้รับบทความวันที่ 20 กรกฎาคม 2553
ยอมรับตีพิมพ์วันที่ 30 กันยายน 2553