

ศิริัญญา วีระอนันต์ชัย 2552: ตัวแบบเชิงเส้นวางนัยทั่วไปสำหรับการศึกษาคิดตามระยะยาวของ  
 จำนวนการเรียกค่าสินไหมทดแทนการประกันภัยรถยนต์ในกรุงเทพมหานคร ประเทศไทย ประโยชน์  
 วิทยาศาสตร์มหาบัณฑิต (สถิติ) สาขาสถิติ ภาควิชาสถิติ อาจารย์ที่ปรึกษาวิทยานิพนธ์หลัก:  
 ผู้ช่วยศาสตราจารย์ลี อิงศรีสว่าง, Ph.D. 116 หน้า

งานวิจัยนี้มีวัตถุประสงค์เพื่อสร้างตัวแบบจำนวนการเรียกค่าสินไหมทดแทนการประกันภัยรถยนต์ใน  
 กรุงเทพมหานครที่มีการเก็บเข้าในระยะเวลาติดตามตั้งแต่ปี พ.ศ. 2544-2548 โดยใช้ข้อมูลที่มีจำนวนการเรียก  
 ค่าสินไหมทดแทนอย่างน้อยหนึ่งครั้งในระยะเวลาติดตาม 5 ปี จำนวน 3,635 กรมธรรม์ ที่ได้จากการ  
 ประกันภัย ด้วยวิธี Generalized Estimating Equations (GEE) เมื่อกำหนดโครงสร้างความสัมพันธ์ของข้อมูลเป็น  
 แบบ First-order Autoregressive (AR(1)) และ Compound Symmetry (CS) ตามลำดับ และวิธีตัวแบบผสมเชิง  
 เส้นวางนัยทั่วไป (Generalized Linear Mixed Models, GLMMs) เมื่อกำหนดโครงสร้างความแปรปรวนร่วมของ  
 ข้อมูลเป็นแบบ AR(1) และ CS ตามลำดับ พร้อมทั้งหาตัวแบบที่เหมาะสม โดยตัวแปรตาม คือ จำนวนการเรียก  
 ค่าสินไหมทดแทนซึ่งมีการแจกแจงแบบปัวซอง และตัวแปรอิสระ คือ ค่าสินไหมทดแทน อายุรถยนต์ เพศ และ  
 กลุ่มอายุผู้ทำประกันภัย และทำการคัดเลือกตัวแปรด้วยวิธีขั้นตอนก้าวหน้า (Forward Stepwise Method)

ผลการศึกษาพบว่า ตัวแบบวิธี GEE เมื่อโครงสร้างความสัมพันธ์เป็นแบบ AR(1) มีความเหมาะสม  
 สำหรับข้อมูลมากกว่าการกำหนดโครงสร้างความสัมพันธ์เป็นแบบ CS ด้วยค่าสถิติ Pearson Chi-square of  
 residual/DF ที่มีค่าเข้าใกล้ 1 มากกว่าคือมีค่าเท่ากับ 0.64 และ 0.63 ตามลำดับ รวมทั้งค่าร้อยละของการทำนาย  
 ถูกต้อง ( $R_p^2$ ) ที่มีค่าสูงกว่าคือมีค่าเท่ากับ 52.32 และ 52.05 ตามลำดับ ซึ่งปัจจัยที่มีผลต่อจำนวนการเรียกค่า  
 สินไหมทดแทนที่ระดับนัยสำคัญทางสถิติ 0.05 คือ ค่าสินไหมทดแทน อายุรถยนต์ กลุ่มอายุผู้ทำประกันภัย  
 อาชีพ และเพศ ส่วนตัวแบบวิธี GLMMs เมื่อกำหนดโครงสร้างความแปรปรวนร่วมเป็นแบบ AR(1) มีความ  
 เหมาะสมสำหรับข้อมูลมากกว่าการกำหนดโครงสร้างความแปรปรวนร่วมเป็นแบบ CS ด้วยค่าสถิติ  
 Generalized Chi-Square/DF เท่ากับ 0.18 และ 0.15 ตามลำดับ โดยมีค่า  $R_p^2$  ใกล้เคียงกันคือ 53.34 และ 53.38  
 ตามลำดับ รวมทั้งค่า Pseudo-AIC และค่า Pseudo-BIC ของโครงสร้างความแปรปรวนร่วมแบบ AR(1) ที่มีค่าต่ำ  
 กว่าโครงสร้างความแปรปรวนร่วมแบบ CS เท่ากับ 36630.43 และ 36649.03 ตามลำดับ และปัจจัยที่มีผลต่อ  
 จำนวนการเรียกค่าสินไหมทดแทนที่ระดับนัยสำคัญทางสถิติ 0.05 คือ ค่าสินไหมทดแทน อายุรถยนต์ กลุ่มอายุ  
 ผู้ทำประกันภัย อาชีพ และเพศ เมื่อเปรียบเทียบตัวแบบวิธี GEE และตัวแบบวิธี GLMMs ในภาพรวมของ  
 ค่าเฉลี่ยประชากรจะมีค่า  $R_p^2$  ใกล้เคียงกันประมาณร้อยละ 53 แต่เมื่อตัวแบบวิธี GLMMs ที่มีการพิจารณาเทอม  
 อิทธิพลค่าคงที่ผู้พบพบว่า ตัวแบบ GLMMs ที่มีโครงสร้างความแปรปรวนร่วมเป็นแบบ AR(1) ให้ประสิทธิภาพ  
 ในการทำนายจำนวนการเรียกค่าสินไหมทดแทนมีความถูกต้องสูงขึ้นเป็นร้อยละ 76.28

Sirinya Teeraananchai 2009: Generalized Linear Models for Longitudinal Study of Car Insurance Claims in Bangkok, Thailand. Master of Science (Statistics), Major Field: Statistics, Department of Statistics. Thesis Advisor: Assistant Professor Lily Ingsrisawang, Ph.D. 116 pages.

The goal of this study was to model a number of car insurance claims for car policyholders in Bangkok during the five-year insurance period of 2001-2005. The data used in this study came from the Department of Insurance in Thailand which consisted of 3,635 observations with at least one claim count in the 5-year period. The methodologies of GEE and GLMMs approaches were applied by taking account of correlation and covariance structures of data such as First-order Autoregressive (AR(1)) and Compound Symmetry (CS), respectively. The appropriate model is suggested. The dependent variable was the claim counts with Poisson distribution while independent variables were indemnity, car-age, age, gender, occupation, car-group and engine size. These independent variables were chosen to model the relationship between the claim counts and the potential predictors, using a forward stepwise procedure for variable selection.

The results showed that the GEE model with AR(1) correlation structure was more appropriate than the model with CS structure, as indicated by the values of Pearson Chi-square of residual/DF 0.64 and 0.63, respectively; and the percentages of correct prediction ( $R_p^2$ ) with 52.32 and 52.05, respectively. The statistically significant factors at a 0.05 level consisted of indemnity, car-age, age-group, occupation and gender. On the other hand, the GLMMs model with AR(1) covariance structure was also more appropriate than the model with CS structure, as indicated by the values of Generalized Chi-Square/DF 0.18 and 0.15, respectively; and  $R_p^2$  with 53.34 and 53.38, respectively. Moreover, the lower Pseudo-AIC and Pseudo-BIC of the GLMMs model with AR(1) structure was used to indicate the more effective covariance structure. The statistically significant factors at the 0.05 level consisted of indemnity, car-age, age-group, occupation and gender. The GLMMs and GEE for population averaged models had the same performance for estimating the claim counts with  $R_p^2$  about 53 %, but the GLMMs model with AR(1) structure and random intercept effect showed more efficiency in predicting the number of claim counts with 76.28 %.