วิทยานิพนธ์ฉบับนี้มีวัตถุประสงค์เพื่อเปรียบเทียบประสิทธิภาพของการพยากรณ์ระหว่างตัวแบบถดถอยโลจิสติคแบบสถิตย์ (Static model) และตัวแบบพลวัตร (Dynamic model) โดยวิธีการวิเคราะห์แบบสถิตย์จะพิจารณาช่วงเวลาเป็นเพียงหนึ่งช่วงเวลาเท่านั้น แต่วิธีการวิเคราะห์แบบพลวัตรจะทำการแบ่งช่วงเวลาออกเป็นหลายช่วงเวลาย่อยและพิจารณาข้อมูลของแต่ละหน่วยตัวอย่างในแต่ละช่วงเวลานั้น ซึ่งเป็นลักษณะของ Hazard model ในกรณีที่เวลาเป็นแบบไม่ต่อเนื่อง และทำการศึกษากับข้อมูล 2 ประเภท ได้แก่ ข้อมูลจำลองและข้อมูลจริง โดยข้อมูลจำลองอยู่ภายใต้เงื่อนไขว่า จำนวนตัวแปรอิสระที่ใช้ในแต่ละตัวแบบเท่ากับ 3 ตัวแปร (X_1, X_2, X_3) เมื่อ X_1 มีการแจกแจงแบบปกติด้วย $\mu = 0$ และ $\sigma^2 = 1$ X_2 มีการแจกแจงแบบเบอร์นูลีด้วย $p = 1/2$ และ X_3 เป็นตัวแปรที่มีค่าเปลี่ยนแปลงไปตามเวลาและมาจากการแจกแจงแบบเอกซโพเนนเชียลด้วยพารามิเตอร์ $\lambda = 1/36$ แบ่งช่วงเวลาออกเป็น 24 ช่วง ขนาดตัวอย่างที่ใช้เท่ากับ 10,000 และกระทำซ้ำจำนวน 1,000 ครั้ง และข้อมูลจริงเป็นข้อมูลทุติยภูมิเกี่ยวกับการเช่าซื้อสินค้าชนิดหนึ่ง โดยเหตุการณ์ที่สนใจคือการเกิดหนี้ NPL (Non Performing Loan) ใช้ตัวแปรอิสระจำนวน 3 ตัวแปร พิจารณาข้อมูลของลูกค้าในช่วงเวลา 1 ปี โดยแบ่งออกเป็น 4 ช่วงๆ ละ 3 เดือน และใช้ขนาดตัวอย่างเท่ากับ 10,000 เกณฑ์ที่ใช้ในการเปรียบเทียบคือ วัดประสิทธิภาพของการพยากรณ์ระหว่างตัวแบบทั้งสองด้วยโค้ง ROC ผลการศึกษาสรุปดังนี้

สำหรับข้อมูลจำลอง พบว่า ในทุกกรณีของการทดลองตัวแบบถดถอยโลจิสติคแบบพลวัตรมีพื้นที่ใต้โค้ง ROC มากกว่าตัวแบบสถิตย์อย่างมีนัยสำคัญทางสถิติที่ระดับนัยสำคัญ 0.05 สำหรับข้อมูลจริง พบว่า ทั้งสองตัวแบบมีพื้นที่ใต้โค้ง ROC ใกล้เคียงกัน ซึ่งตัวแบบพลวัตรมีพื้นที่ใต้โค้งมากกว่าเล็กน้อย แต่เราไม่สามารถสรุปได้ว่าตัวแบบพลวัตรมีพื้นที่ใต้โค้ง ROC มากกว่าตัวแบบสถิตย์ที่ระดับนัยสำคัญ 0.05

The objective of this research is to compare the prediction power between two logistic regression models: static model and dynamic model. A static model considers only one period of explanatory variable values for each sample unit. A dynamic model, or a hazard model, divides time into several periods and consider data of explanatory variables in several periods for each sample unit. We consider a discrete time hazard model for the dynamic model. We perform the comparisons on two types of data: simulated data and real data .The simulated data are generated by R program using Monte Carlo Simulation techniques. The model assumes 3 independent variables, X_1, X_2, X_3, where X_1 has normal distribution with mean 0 and variance 1, X_2 has bernoulli distribution with probability of success $1/2$ and X_3 has exponential distribution with parameter $1/36$. The simulation is divided into 24 periods with sample size equal to 10,000 and the simulation is repeated for 1,000 times. The real data is a secondary data from a credit activity. The interested event of dependent variable is non performing loan (NPL) of customers. The number of independent variables is equal to 3, and the sample size equals 10,000. The data considered for each customer is within one-year period divided into four 3-months long subperiods. The criteria employed for the comparison is the area under ROC curve for a prediction purpose. The conclusions of this study are as follows.

For simulated data, in most simulation runs, the area under ROC curve from the dynamic model is greater than the area under ROC curve of the static model at significance level of 0.05. Therefore the dynamic model performs better than the static model. For the real data set, the area under ROC curve from the dynamic model is only slightly more than that of the static model, but not significant at significance level of 0.05. Therefore, we cannot conclude that the dynamic model performs better than the static model.