استخدام الكوبيولا في النماذج الخطية المعممة للبيانات الثنائية

إعداد

عبير هادي السروجي

رسالة مقدمة كمتطلب للحصول على درجة الماجستير في الإحصاء

إشراف

د. مرفت خليفه

المساعد أستاذ الإحصاء الرياضي

)مشرف(

د. سلافه بن حمد

المساعد أستاذ الإحصاء التطبيقي

)مشرف مساعد(

قسم الإحصاء

كلية العلوم

جامعة الملك عبد العزيز

جدة – المملكة العربية السعودية

ربيع اول 1439 هـ – نوفمبر 2017 م

استخدام الكوبيولا في النماذج الخطية المعممة للبيانات الثنائية

عبير هادي عبدالله السروجي

المستخلص

تستخدم تقنيات التحليل التنبؤي على نطاق واسع في مجال التطبيق، وأكثرها شيوعا هو ربط المتغيرات التابعة (regressors) بمتغير الاستجابة (response) من خلال الدالة Y = f (X) للتنبؤ بقيمة الاستجابة Y. هذا النهج ينطوي على العديد من التقنيات المختلفة، بما في ذلك الانحدار الخطي، وهو ما يعرف باسم النموذج الخطي العادي (ordinary linear model) للتنبؤ الرقمي - ونماذج بروبيت ولوجيت الأحادية للتنبؤ الوصفي الثنائي. ووفقا لدراسة استقصائية سنوية عن البيانات، لا تزال نماذج الانحدار من بين أكثر أدوات التحليل شيوعا التي يستخدمها الممارسون اليوم. الحد من النماذج الخطية العادية هو أن الاستجابة لا يمكن أن تكون ثنائية. ومع ذلك، في كثير من الدراسات متغيرات الاستجابة هي عادة ثنائية، على سبيل المثال: إذا كانت الدراسة لتحديد ما إذا كان المريض سوف يتعافى من مرض أو يموت، أو لتحديد ما إذا كان الطالب سوف ينجح أو يفشل في الفحص، وغيرها. جميع الدراسات السابقة حالات متغيرات الاستجابة فيها ثنائية وبالتالي فإنه من غير المناسب التنبؤ بمتغير الاستجابة باستخدام النموذج الخطي العادي للعثور على العلاقة بين متغير الاستجابة والمتغيرات التابعة. وقد تم تطوير مجموعة من التقنيات لمعالجة هذه الحالات، ولكن التقنيات المعتمدة عادة هي نماذج بروبيت ولوجيت الأحادية. هذه النماذج هي أعضاء في أسرة النماذج الخطية المعممة، وتستخدم على نطاق واسع في التطبيقات الاجتماعية والاقتصادية والطبية. عندما يكون متغير الاستجابة ثنائي، يمكن استخدام كلا النموذجين لتقدير العلاقة بين متغير الاستجابة والمتغيرات التابعة.

عندما نهتم بتقدير تأثير متغير العلاج الثنائي الباطني (binary endogenous treatment) على متغير الاستجابة الثنائي، عندها، نماذج بروبيت ولوجيت الأحادية غير مناسبة بسبب مشكلة endogeneity. ويرجع السبب في ذلك إلى أن مشاكل الارتباط بين معادلة المعالجة الثنائية ومعادلة الاستجابة الثنائية التي قد لا نأخذها في الحسبان قد تؤدي إلى نتائج متحيزة إلى حد كبير. إن endogeneity تعني أن المتغيرين (المعالجة الثنائية والاستجابة الثنائية) قد يعتمدان على المتغيرات غير المرصودة المترابطة. يمكن السيطرة على endogeneity باستخدام نموذج بروبيت المتكرر ثنائي الاتجاه. هذا النموذج يتحكم في endogeneity باستخدام معادلتين، حيث تحدد المعادلة الأولى ما إذا كان أو لم يتم تلقي العلاج، في حين أن المعادلة الثانية تصف متغير الاستجابة الثنائية كدالة في العلاج الثنائي. ويكتمل هذا النموذج بافتراض أن الارتباط بين المعادلتين يتبع توزيع طبيعي قياسي، مما يوحي بأن التجانس موجود، وبالتالي هناك حاجة إلى تقدير مشترك من كلا المعادلتين. الحد من هذا النموذج هو عدم قدرته على التعامل مع الارتباطات الغير طبيعية بين معادلتي العلاج والاستجابة. ومن أجل معالجة هذه المسألة، فإن النهج القائم على أساس كوبيولا بروبيت ثنائي الاتجاه (copula bivariate probit model) يسمح بالارتباط بين معادلتي العلاج والاستجابة دون فرض الارتباطات الطبيعية. الكوبيولا هي أداة مناسبة لتوليد توزيعات مرنة، غير طبيعية للمعادلتين. في هذه الأطروحة، سوف نركز على الفرق بين نماذج بروبيت ولوجيت في الحالة الاحادية، في حالة ثنائية المتغير، تم تطبيق محاكاة لمقارنة غوسيان، FGM والفرانك كوبيولاس المتماثلة والتي تستخدم على نطاق واسع والمعروفة بين الباحثين في الكوبيولا.

Generalized Linear Using Copula in

Model for Binary Data

By

Abeer Hadi AL-Sorouji

A thesis submitted for the requirements of the degree of Master of Statistical Science

Supervised By

Dr. Mervat Khalifa
Assistant Professor of
Mathematical Statistics
(Advisor)

Dr. Sulafah Binhimd
Assistant Professor of
Applied Statistics
(Co-Advisor)

STATISTCS DEPARTMENT

FACULTY OF SCIENCES

KING ABDULAZIZ UNIVERSITY

JEDDAH - SAUDI ARABIA

Rabea I 1439 H- November 2017 G

Using Copula in Generalized Linear Model for Binary Data

By Abeer Hadi AL-Sorouji

Abstract

Univariate probit and logit models are members of the family of generalized linear models (GLM), and analyze the relationship between regressors and binary response variable. The question of which model performs better is important, and the answer was achieved using a Monte Carlo simulation to compare both the univariate probit and logit models under different conditions.

A bivariate probit model is frequently used in health economics, as its recursive form is practical when estimating the effect of binary endogenous (treatment) variable in a binary response model that is not suitable for univariate probit and logit models. Endogeneity can be controlled using a recursive bivariate probit model, which is completed by assuming that the structure errors of the model follow a standard bivariate Gaussian (Normal) distribution with correlation θ. The limitation of this model is its inability to deal with non-Gaussian dependencies between the treatment and response equations, and this is addressed by using a copula bivariate probit (CBP) model.

This thesis will focus on the difference between probit and logit models in univariate case. In bivariate case, simulation has been applied to compare Gaussian, FGM and Frank copulas with symmetric dependency that are widely-used and well-known among copula researchers.