科技部補助專題研究計畫成果報告
期末報告

計數迴歸模型之穩健診斷

計 畫 類 別 ： 個別型計畫
計 畫 編 號 ： NSC 102-2118-M-004-004-
執 行 期 間 ： 102 年 08 月 01 日至 103 年 07 月 31 日
執 行 單 位 ： 國立政治大學統計學系

計 畫 主 持 人 ： 鄭宗記

計畫參與人員 ： 碩士班研究生-兼任助理人員：賴柏華
碩士班研究生-兼任助理人員：顏伯魁

報 告 附 件 ： 移地研究心得報告

處 理 方 式 ：
1.公開資訊：本計畫涉及專利或其他智慧財產權，2 年後可公開查詢
2.「本研究」是否已有嚴重損及公共利益之發現：否
3.「本報告」是否建議提供政府單位施政參考：否

中 華 民 國 103 年 10 月 12 日

中 文 摘 要 ： 在計量經濟、社會及醫學科學等領域中，經常會對計數（count）的資料進行模型化的工作；卜瓦松（Poisson）及負二項（negative binomial）迴歸模型則是最常用來分析此類資料的兩種分析方法。然而，離群值（outliers）會影響到此二計數迴歸模型之係數的最大概似估計結果。本計畫利用最大削減概似估計（maximum trimming likelihood estimation）來處理具離群值的計數迴歸模型之穩健估計結果；並利用快速演算法（fast algorithm）以獲致此估計值。藉由模擬研究與實際資料分析來檢驗所提出的估計方法。

中文關鍵詞： 計數迴歸模型；快速演算法；最大削減概似估計；離群值。

英 文 摘 要 ：

英文關鍵詞：

Robust diagnostics for count regression models

Tsung-Chi Cheng*

1 Introduction

Poisson regression model and its various modifications have been extensively applied in many context in social sciences studies. Poisson regression is the standard method used to model count response data. However, the Poisson distribution assumes the equality of its mean and variance – a property that is rarely found in real data. Data that have greater variance than the mean are termed Poisson overdispersed, but are more commonly designated as simply overdispersed.

Negative binomial regression is a standard method used to model overdispersed Poisson data. The negative binomial is traditionally derived from a Poisson-Gamma mixture model. However, the negative binomial may also be thought of as a member of the single parameter exponential family of distributions. This family of distributions admits a characterization known as the generalized linear model (GLM), which summarizes each member of the family. Most importantly, the characterization is applicable to the negative binomial. Such interpretation allows statisticians to apply to the negative binomial model the various goodness-of-fit tests and residual analyses that have been developed for GLM.

When the negative binomial is used to model overdispersed Poisson count data, the distribution can be thought of as an extension to the Poisson model. Certainly, when the negative binomial is derived as a Poisson-Gamma mixture, thinking of it in this way makes perfect sense. The original derivation of the negative binomial

*Department of Statistics, National Chengchi University, 64 ZhihNan Road, Section 2, Taipei 11623, Taiwan; E-mail: chengt@nccu.edu.tw; TEL:+886 2 29393091#81132; FAX:+886 2 29398024

regression model stems from this manner of understanding it, and has continued to characterize the model to the present time.

2 Count regression model

Modeling count variables is a common task in microeconometrics, social and political sciences. Poisson regression is appropriate when the dependent variable is a count of events. It can be used to model the number of occurrences or the rate of occurrence of an event of interest as a function of some independent variables. An event count refers to the number of times an event occurs, for example the number of attending performing arts in this study. The events must be independent in the sense that the participation of one respondent will not make another more or less likely to participate, but the probability per participation is related to covariates.

In Poisson regression it is assumed that the response variable Y, number of occurrences of an event, has a Poisson distribution given k explanatory variables. The Poisson regression model is then defined as:

$$f(y_i|\boldsymbol{x}_i) = \exp[-\mu_i + y_i \log \mu_i - \log y_i!], \quad i = 1, 2, \cdots, n, \tag{1}$$

where

$$\mu_i = \exp(\boldsymbol{x}_i'\boldsymbol{\beta}). \tag{2}$$

The log of the mean (2) is assumed to be a linear function of the independent variables. The link function g relates the linear predictor to the expected value μ_i of y_i. That is, for $i = 1, 2, \cdots, n$,

$$g(\mu_i) = \log(\mu_i) = \boldsymbol{x}_i'\boldsymbol{\beta},$$

which is also called as the log-linear model in the context of the generalized linear model (McCullagh and Nelder, 1989).

For an overview of the count data models in econometrics, one can refer to Cameron and Trivedi (1998) as well as Long and Freese (2006). The latter also provides an introduction to Stata, which is a statistical computing package. The maximum likelihood estimation is the most popular method for estimating the coefficients, $\boldsymbol{\beta}$ and $\boldsymbol{\gamma}$, of all models discussed in this section. For other related issues, such as model selection, one can refer to McCullagh and Nelder (1989) and Cameron and Trivedi (2005).

2.1 Negative regression model

The classical Poisson regression model for count data is often of limited use in some disciplines because the empirical count data typically exhibit overdispersion and/or an excess number of zeros. The negative binomial (NB) regression model is a generalized linear model that accommodates a solution to the overdispersion problem and may function better in the case of excess zeros. Instead of assuming that the distribution of Y is Poisson, Y is assumed to follow a negative binomial distribution in the NB regression model. The negative binomial distribution is defined as:

$$f(y_i|\boldsymbol{x}_i) = \frac{\Gamma(y_i + \alpha^{-1})}{y_i!\Gamma(\alpha^{-1})} \left(\frac{\alpha^{-1}}{\alpha^{-1} + \mu_i}\right)^{\alpha^{-1}} \left(\frac{\mu_i}{\alpha^{-1} + \mu_i}\right)^{y_i}, \quad i = 1, 2, \cdots, n,$$

where $\Gamma(\cdot)$ is the gamma function and μ_i is corresponding to (2). This relaxes the assumption of equality between mean and variance (a property of the Poisson distribution) since the variance of the negative binomial distribution is equal to $\mu + \alpha\mu^2$, where $\alpha \geq 0$ is a dispersion parameter. If $\alpha = 0$, then the negative binomial distribution reduces to Poisson. Lawless (1987) discusses the statistical properties of Negative regression model.

Consider the log-likelihood for the ith observation

$$\begin{aligned} \ell_i &= \ell(\boldsymbol{\beta}, \alpha; y_i) = \log(\Gamma(\alpha^{-1} + y_i) - \log(\Gamma(y_i + 1)) - \\ &\quad \log(\Gamma(\alpha^{-1})) + y_i \log(\alpha\mu_i/(1 + \alpha\mu_i)) - \alpha^{-1} \log(1 + \alpha\mu_i) \end{aligned} \tag{3}$$

Hilbe (2008, p. 90) uses the deviance as the basis for the convergence criterion and the log-likelihood function could have been used as well. The estimation of α is rather intractable. The estimation of $\boldsymbol{\beta}$ and α are well-discussed in Hilbe (2008, chapter 5). Once the ML estimates of $\boldsymbol{\beta}$ and α are obtained, denoted by $\hat{\boldsymbol{\beta}}$ and $\hat{\alpha}$, the estimated value of the model is $\log(\hat{\mu}_i) = \boldsymbol{x}_i^T\hat{\boldsymbol{\beta}}$. The signed square-root deviance residual for the ith observation is then

$$d_i^* = d(y_i;\hat{\mu}_i,\hat{\alpha}) = \text{sign}(y_i-\hat{\mu}_i)\sqrt{2}\left\{\frac{1}{\hat{\alpha}}\log\frac{1+\hat{\alpha}\hat{\mu}_i}{1+\hat{\alpha}y_i} + y_i\log\frac{y_i(1+\hat{\alpha}\hat{\mu}_i)}{\hat{\mu}_i(1+\hat{\alpha}y_i)}\right\}^{1/2} \tag{4}$$

The function glm.nb in R is a modification of the system function glm to include estimation of the additional parameter, θ $(=1/\alpha)$, for a Negative Binomial generalized linear model. An alternating iteration process is used. For given θ the GLM is fitted using the same process as used by glm. For fixed means the θ parameter is estimated using score and information iterations. The two are alternated until convergence of both.

2.2 The maximum trimmed likelihood estimator

The trimmed likelihood principle is based on trimming the likelihood function rather than directly trimming the data, which was introduced independently by Hadi and Luceño (1997) and Vandev and Neykov (1998). Instead of summing up all values of the log likelihood function for each observation according to ML estimation, the trimmed likelihood approach considers to maximize the following objective function:

$$\sum_{i=a}^{b} w_{(i),\nu}\, l(\theta; x_{(i),\nu}), \tag{5}$$

where $a \le b,\ \nu = (a,b) \in \{1,2,\cdots,n\}$, and

$$l(\theta;x_{(1),\nu}) \ge l(\theta;x_{(2),\nu}) \ge \cdots \ge l(\theta;x_{(n),\nu}), \tag{6}$$

for any given value of θ. Here $l(\theta;x_i) = \ln f(x_i;\theta)$ is the contribution of the ith observation to the log likelihood function. The weights w_i's (≥ 0) are discussed in Vandev and Neykov (1998). The estimator $\theta(a,b,w)$ is obtained by maximizing (5). The resulting method is called as the *maximum trimmed likelihood* (MTL) method and $\hat{\theta}(a,b,w)$ is the maximum trimmed likelihood estimator (MTLE).

Neykov *et al.* (2007) give the combinatorial representation of MTLE (5) evaluating at $a=1$ and $b=q$ as follows:

$$\max_\theta \sum_{i=1}^{q} w_{(i),\mathcal{Q}}\, l(\theta;x_{(i),\mathcal{Q}}) = \max_\theta \max_{\mathcal{Q}\in Q} \sum_{i\in\mathcal{Q}} w_i l(\theta;x_i) = \max_{\mathcal{Q}\in Q} \max_\theta \sum_{i\in\mathcal{Q}} w_i l(\theta;x_i),$$

where Q is the set of all q-subsets of the set $\{1, \cdots, n\}$. Therefore, it follows that all possible $\binom{n}{q}$ partitions of the data have to be fitted by the MLE, and the MTLE is given by the partition with the maximum log-likelihood.

In order to study the breakdown properties of general estimators such as LMS and LTS, Vandev (1993) develops a d-fullness technique. The d-fullness technique allows the statistician to choose the tuning parameter q according to the expected percent of outliers in data. For computational aspect, Neykov and Müller (2003) propose a fast computing algorithm for MTLE, which is analogous to the C-step for LTS and MCD of Rousseeuw and van Driessen (1999, 2006).

3 Robust diagnostics for NB regression model

In this section, we first apply the fast algorithm to obtain the MTLE for NB regression model. Then simulated data are used to illustrate the approach. We conduct a simulation study to compare MTLE with MLE and then a real data is analyzed to find a new conclusion.

3.1 Computing algorithms

To obtain the RTML estimate of $\boldsymbol{\theta}_q$, this subsection employs both the fast algorithm for MTLE of Neykov and Müller (2003) and the forward search algorithm of Atkinson (1994). The fast algorithm is used to obtain an outlier-free subset and then the observations of the subset are incremented in such a way that outliers are unlikely to be included. As explained in Neykov *et al.* (2007), computing the MTLE is infeasible for large data sets because of its combinatorial nature. To get an approximative MTLE solution, an algorithm called FAST-TLE was developed in Neykov and Müller (2003).

The basic idea behind the FAST algorithm consists of carrying out many two-step procedures: a trial step followed by a refinement step (so-called the Concentration step). It reduces to the FAST-LTS and FAST-MCD algorithms proposed by Rousseeuw and Van Driessen (1999 and 2006) in the regression and multivariate cases, respectively. Here the subsample size of the trial step can be any values between p and

q, for example the so-called elemental sets are used in Rousseeuw and Van Driessen's papers.

The details of the proposed procedure is as follows.

- In the trial step a subsample of size s is selected randomly from the data and then the model is fitted to that subsample to get a trial ML estimate.
- The refinement step is designed by the so-called concentration procedure:
  - (a) the observations with the q largest log-likelihood based on the current estimate are found, starting with the trial MLE as initial estimator;
  - (b) fitting the model to these q observations yields an improved fit.
  - Repeating (a) and (b) leads to an iterative procedure.

The convergence is always guaranteed after a finite number of steps since there are only finitely many q-subsets out of $\binom{n}{q}$ (Neykov and Müller 2003). The one with the largest value of the sum of q largest log-likelihood is then an approximate to the solution of MTLE. The resulting estimates are denoted by $\hat{\boldsymbol{\beta}}_q$ and $\hat{\alpha}_q$.

Once the MTLE estimates of $\boldsymbol{\beta}$ is obtained, the estimated value of the model is $\log(\hat{\mu}_{i,q}) = \boldsymbol{x}_i^T \hat{\boldsymbol{\beta}}_q$. The corresponding log-likelihood (3) for the ith observation is defined as:

$$\begin{aligned} \ell_{i,q} &= \ell(\boldsymbol{\beta}_q, \hat{\alpha}_q; y_i) = \log(\Gamma(\hat{\alpha}_q^{-1} + y_i) - \log(\Gamma(y_i + 1)) - \\ &\quad \log(\Gamma(\hat{\alpha}_q^{-1})) + y_i \log(\alpha_q \hat{\mu}_{i,q}/(1 + \hat{\alpha}_q \hat{\mu}_{i,q})) - \hat{\alpha}_q^{-1} \log(1 + \hat{\alpha}_q \hat{\mu}_{i,q}) \end{aligned} \tag{7}$$

The resulting signed square-root deviance residual (4) for the ith observation is then

$$\begin{aligned} d_{i,q} &= d(y_i; \hat{\mu}_{i,q}, \hat{\alpha}_q) = \text{sign}(y_i - \hat{\mu}_{i,q})\sqrt{2} \\ &\quad \times \left\{ \frac{1}{\hat{\alpha}_q} \log \frac{1 + \hat{\alpha}_q \hat{\mu}_{i,q}}{1 + \hat{\alpha}_q y_i} + y_i \log \frac{y_i(1 + \hat{\alpha}_q \hat{\mu}_{i,q})}{\hat{\mu}_{i,q}(1 + \hat{\alpha}_q y_i)} \right\}^{1/2} \end{aligned}$$

which can be used to order the observations as well as to flag out the possible outliers.

3.2 Simulation data

In this subsections, we use a simulated data set to illustrate the procedure. Firstly, all the covariates in the simulated data, X_1, X_2, and X_3, are generated from $N(0,1)$. Then, let $\mu_i = x_{1i} + x_{2i} + x_{3i}$, where $i = 1, 2, \cdots, 300$, and the response value y_i is generated from a negative binomial distribution with mean $\exp(\mu_i)$ and dispersion parameter $\alpha = 3.33$. Finally, the first 30 observations (10% of the data) are shifted to be outliers by adding 50 to those values of y_i's ($i = 1, 2, \cdots, 30$).

Parts (a) and (b) of Figure 1 show the deviance residual plots based on the classical method and the proposed approach for the simulated data, respectively. There is no any outlier being revealed by MLE, while MTLE successfully identifies most of the first 30 cases as outliers.

3.3 Simulation study

we conduct a simulation study to evaluate the performance of the proposed approach. Consider the similar simulation design for the covariates and response variable as the previous subsection. There are 10% of the data assigned to be outlying cases Figures 2 and 3 present the simulation results for the sample sizes (n) 200 and 400, respectively. Each panel shows the boxplot for the resulting estimates of 300 replicates based on MLE, MTLE, and LTD (least trimmed deviance estimator). The vertical line in each panel indicates the true value. It is clear to see that MLE is spoiled by outliers. There is a clear departure between the center of the boxplot based on MLE and the true vertical line. Although LTD provides a better result than MLE, it seems more spreading than MTLE. MTLE outperforms both MLE and LTD in terms of both the width of the boxplot and its corresponding center being closer to the true value.

3.4 Ischemic heart disease

In this subsection a real data example is used to illustrate the MTLE approach. Theses data were collected by a health insurance plan and provide information concerning 788 subscribers who had made claims resulting from ischemic (coronary) heart disease (Kutner, Nachtsheim, and Neter 2008, pages 683-4). The description of the variables

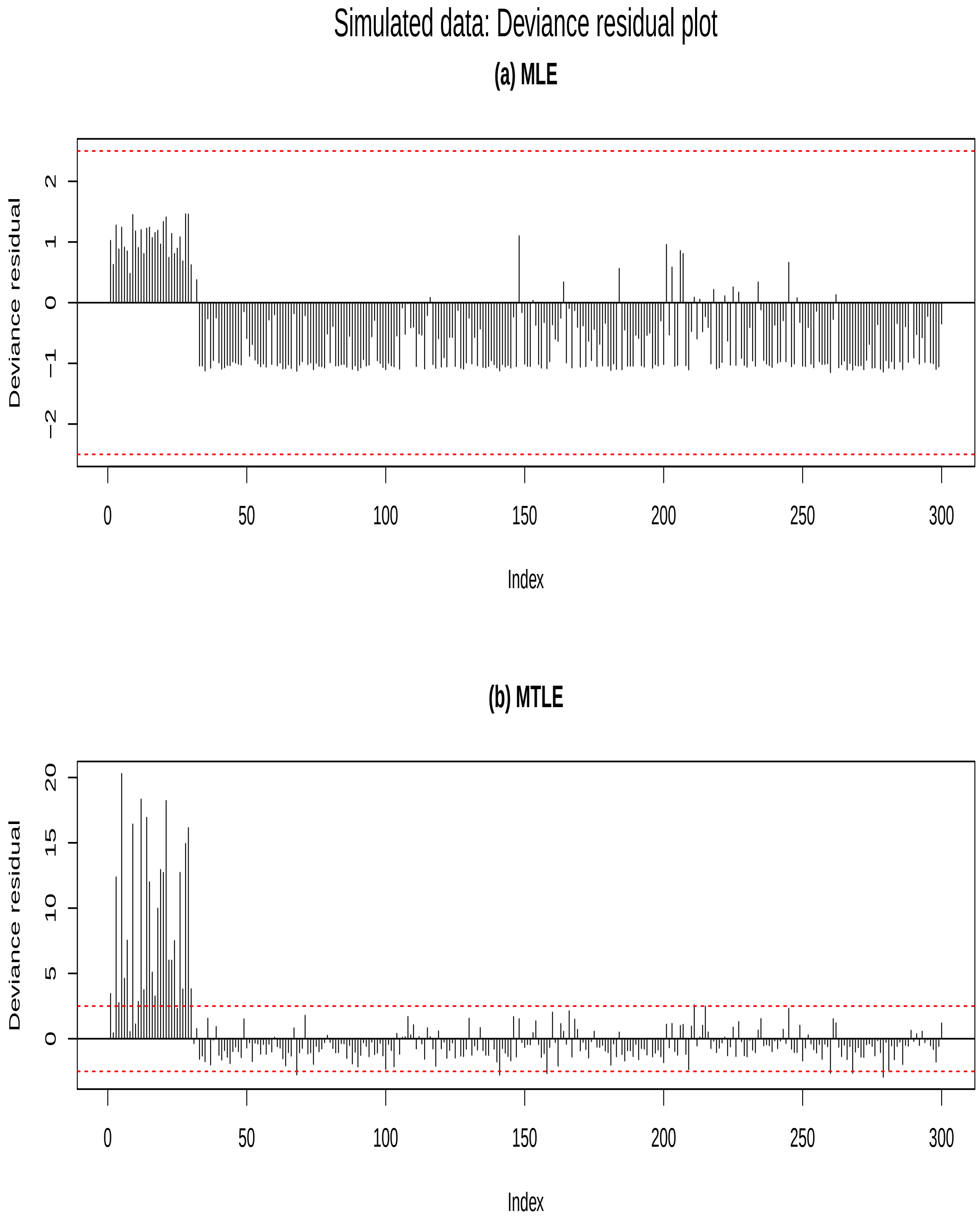

Figure 1: Simulated data: deviance residual plots based on (a) MLE and (b) MTLE

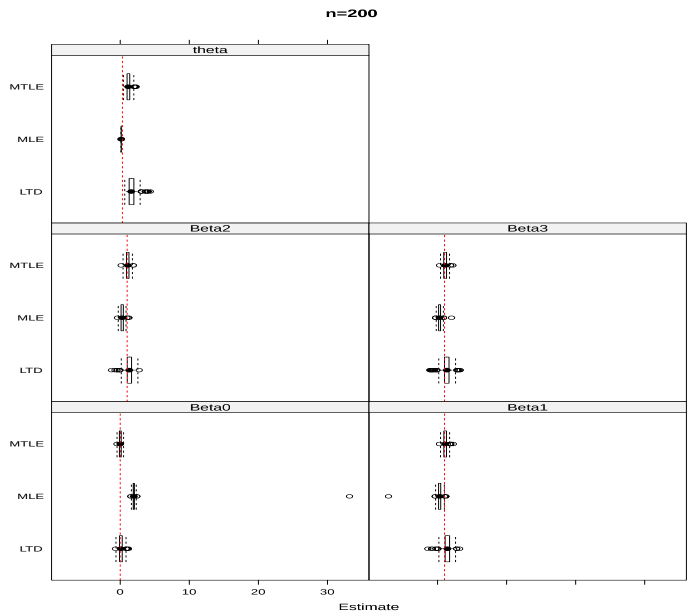

Figure 2: Simulation results: $n = 200$

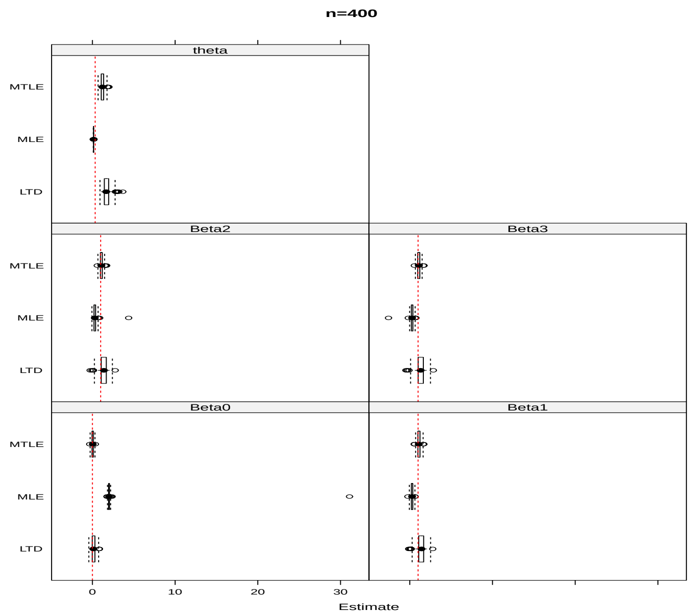

Figure 3: Simulated results: $n = 400$

is shown in Table 1. Figure 4 presents the frequency distribution of the response variable *Duration*. There is a peak at the value of zero in this plot, which indicates the possible zero-inflation situation. We here consider the following model

$$\begin{aligned} Duration &\sim Age + Gender + Intervention + Drug \\ &\quad + Emergency + Complication + Comorbidity \end{aligned}$$

to illustrate the proposed approach.

Figure 5 shows the resulting deviance plots based on MLE and MTLE. There are two functions, glm.nb and ml.nb2, for computing MLE for NB regression model in R. Both glm.nb and ml.nb2 produce quite similar (or the same) results in almost all examples. However, there exist quite different conclusions when use both functions to analyze these data. glm.nb yields a quite weird result in Figure 5 (a), which unreasonably appear too many outliers. Both ml.nb2 and MLTE lead to a quite similar pattern in deviance residuals as shown in parts (b) and (c) of Figure 5. MTLE identifies more outliers than ml.nb2 does.

To confirm the difference, Table 3 presents the estimation results. The estimate for α obtained by glm.nb is zero, which may be the main reason why we obtain the different deviance residual plot in Figure 5 (a). Both ml.nb2 and MLTE yield quite different estimates in terms of magnitude and sign for some covariates.

Variable	Description
Total cost	Total cost claims by subscriber (dollars)
Age	Age of subscriber (years)
Gender	Gender of subscriber: 1 if male; 0 female
Intervention	Total number of interventions or procedures carried out
Drug	Number of tracked drugs prescribed
Emergency	Number of emergency room visits
Complication	Number of other complications arose during heart disease treatment
Comorbidity	Number of other diseases that the subscriber had during period
Duration	Number of days of duration of treatment condition

Table 1. Ischemic heart disease: variable description

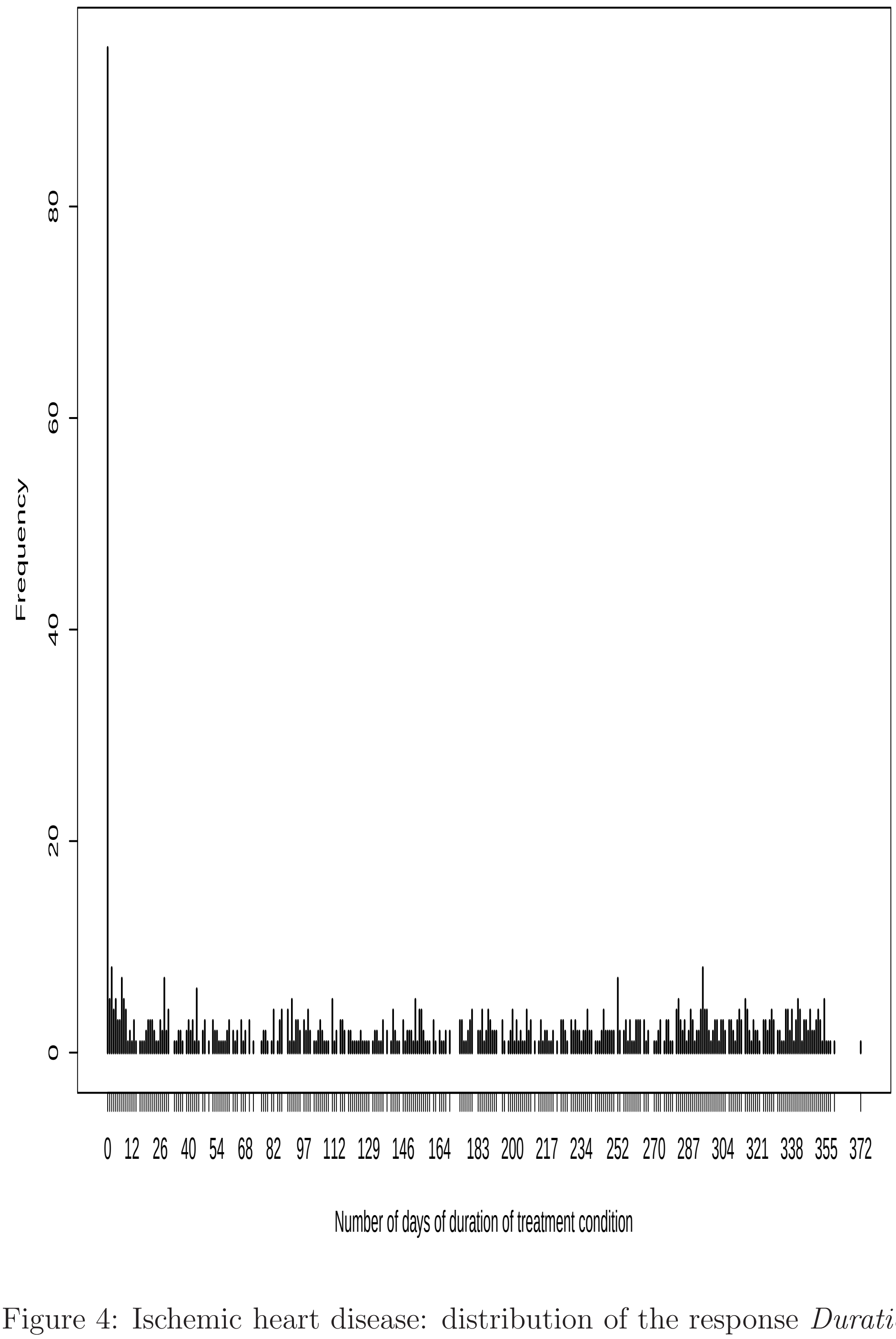

Figure 4: Ischemic heart disease: distribution of the response *Duration*

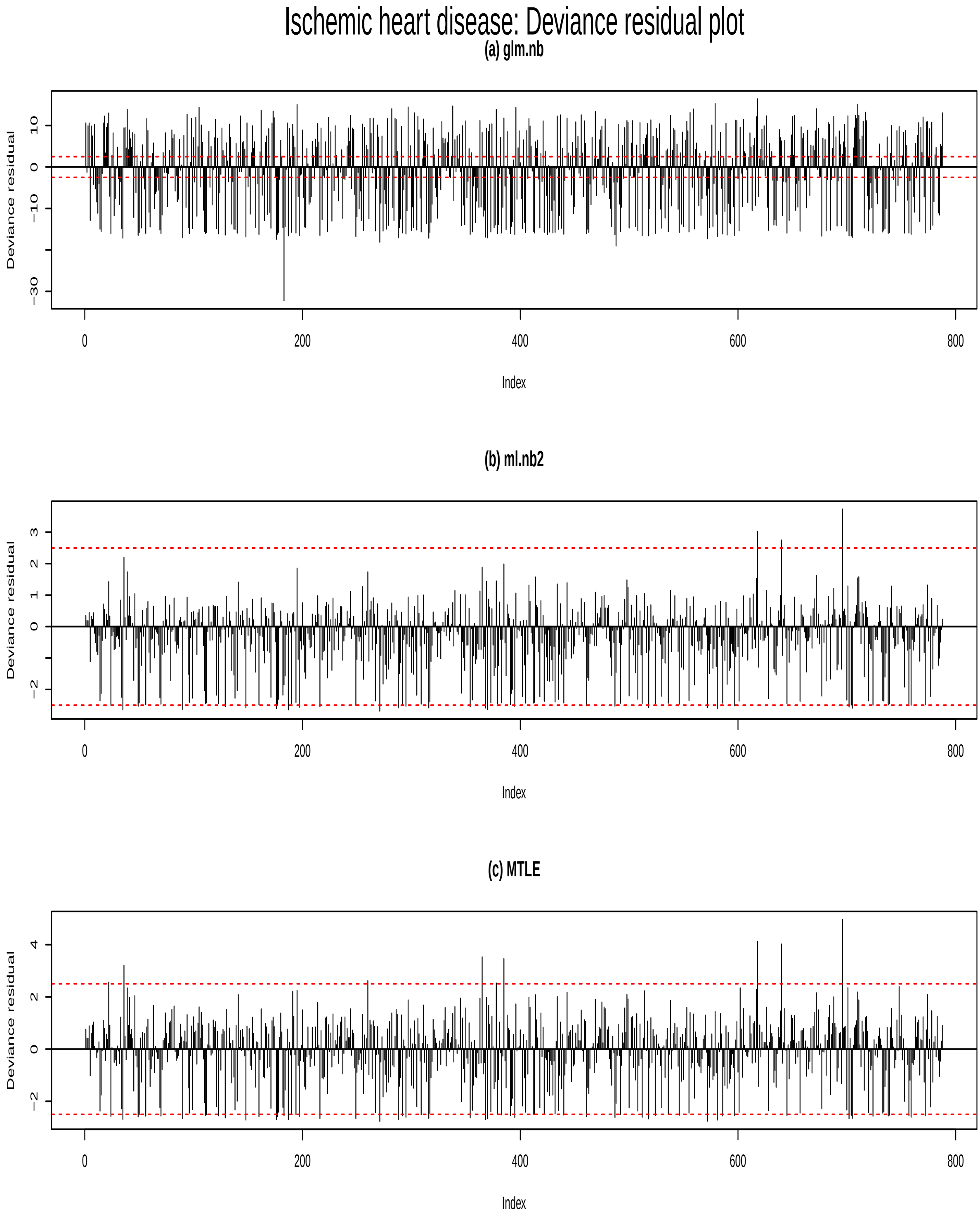

Figure 5: Ischemic heart disease: deviance residual plot

	glm.nb			ml.nb2			MTLE (75%)		
Variable	β	StdErr	t	β	StdErr	t	β	StdErr	t
Intercept	4.095	0.026	155.84	0.151	0.336	0.45	-0.324	0.431	-0.75
Age	0.011	0.000	26.08	0.073	0.006	12.42	0.072	0.007	9.90
Gender	0.053	0.007	7.94	0.262	0.118	2.22	-0.031	0.122	-0.25
Intervention	0.016	0.000	33.10	0.035	0.010	3.38	0.037	0.010	3.75
Drug	0.002	0.003	0.55	0.230	0.068	3.36	0.098	0.053	1.84
Emergency	0.017	0.001	12.91	0.003	0.023	0.15	0.059	0.023	2.59
Complication	0.081	0.010	7.77	0.888	0.317	2.80	-0.273	0.251	-1.09
Comorbiditiy	0.037	0.000	118.84	0.073	0.010	7.10	0.109	0.012	9.38
α	0.000			1.728	0.092	18.77	1.421	0.086	16.57

Table 2. Ischemic heart disease: estimation results

References

[1] Cameron, A. C., & Trivedi, P. K. (1998). *Regression analysis of count data*. Cambridge, MA: Cambridge University Press.

[2] Cameron, A. C., & Trivedi, P. K. (2005). *Microeconometrics: Methods and applications*, Cambridge, MA: Cambridge University Press.

[3] Gurmu, S., & Trivedi, P. K. (1996). Excess zeros in count models for recreational trips. *Journal of Business & Economic Statistics, 14*(4): 139-146.

[4] Hall, D. B. and Shen, J. (2009) Robust Estimation for Zero-Inflated Poisson Regression. *Scandinavian Journal of Statistics*.

[5] Hilbe, J. M. (2008) *Negative Binomial Regression*, Cambridge: Cambridge University Press.

[6] Kutner, M. H, Nachtsheim, C. J. and Neter, J. (2004) *Applied Linear Regression Models*, 4th ed. McGraw-Hill/Irwin.

[7] Lambert, D. (1992). Zero-inflated Poisson regression with an application to defects in manufacturing. *Technometrics, 34*(1), 1-14.

[8] Lawless, J. L. (1987). Negative binomial and mixed Poisson regression. *The Canadian Journal of Statistics*, 15, 209-225.

[9] Long, J. S., & Freese, J. (2006). *Regression models for categorical dependent variables using Stata*, 2nd ed. College Station, TX: StataCorp LP.

[10] McCullagh, P., & Nelder, J. A. (1989). *Generalized linear models*, 2nd ed. London: Chapman and Hall.

[11] Pregibon, D. (1982). Resistant Fits for Some Commonly Used Logistic Models with Medical Applications, *Biometrics*, 38, 485-498.

[12] Müller, C.H., Neykov, N.M., 2003. Breakdown points of the trimmed likelihood and related estimators in generalized linear models. Journal of Statistical and Planning Inference 116, 503-519.

[13] Neykov, N.M., Filzmoser, P., Dimova, R., Neytchev, P.N., 2007. Robust fitting of mixtures using the trimmed likelihood estimator. Computational Statistics and Data Analysis 52, 299-308.

[14] Rousseeuw, P.J., van Driessen, K., 1999. A fast algorithm for the minimum covariance determinant estimator. Technometrics 41, 212-223.

[15] Rousseeuw, P.J., van Driessen, K., 2006. Computing LTS regression for large data sets. Data Mining and Knowledge Discovery 12, 29-45.

[16] Tan F., Rayner,G. J., Wang X, and Peng, H. (2010). A full likelihood procedure of exchangeable negative binomials for modelling correlated and overdispersed count data. *Journal of Statistical Planning and Inference*, 140, 2849?859.

[17] Xie, F.C. and Wei, B.-C. (2008). Influence analysis for Poisson inverse Gaussian regression models based on the EM algorithm. *Metrika*, 67, 49-62.

科技部補助專題研究計畫執行國際合作與移地研究

心得報告

日期：___年___月___日

計畫編號	MOST 102-2118-M-004 -004 -		
計畫名稱	計數迴歸模型之穩健診斷		
出國人員姓名	鄭宗記	服務機構及職稱	國立政治大學統計系 教授
出國時間	103 年 7 月 8 日至 103 年 8 月 3 日	出國地點	柏林、倫敦
出國研究目的	□實驗 □田野調查 □採集樣本 ■國際合作研究 □使用國外研究設施		

一、 執行國際合作與移地研究過程

原計畫申請內容為預計至英國倫敦政經學院（London School of Economics）統計系與個人的博士論文指導教授 Anthony C. Atkinson 進行一週的學術研究交流。行前恰巧又接獲德國柏林洪堡大學 Wolfgang Karl Härdle 教授邀請，參加其於柏林近郊 Motzen 所主持之學術會議。因此整個執行國際合作與移地研究過程與原先規劃有極大的差異，時間亦較長。

完整行程如下表：

日期	地點	活動說明
7/8-7/9	飛機	
7/10-7/12	德國 Motzen	學術會議
7/13-7/18	德國柏林洪堡大學	學術交流
7/18-7/30	英國倫敦政經學院	學術交流
7/30-8/2	荷蘭阿姆斯特丹	順道參訪
8/2-8/3	飛機	

行程中因轉機之緣故，除了在荷蘭阿姆斯特丹有數日之個人參訪活動之外，其餘時間大多進行學術研究活動。此行除了在研究上的討論，亦與倫敦與柏林兩處的學者有相當互動，算是另一收穫。

二、 研究成果

計畫主要目的為進行計數迴歸模型的離群值偵測與穩健估計的問題。由於負二項迴歸模型（Negative binominal regression model）的參數估計有其棘手之處，尤其在處理穩健估計的過程，需對整體資料重複隨機抽取較小的樣本數進行最大概似估計（maximum likelihood estimation），更增加估計過程中的常發生數值問題困難。當利用現有的 R 軟體中的套件時，常無法獲得合理的收斂解。此次合作過程，藉由 Atkinson 教授之建議，自行撰寫開發負二項迴歸模型的估計的 R 程式，以解決倚賴現有程式無法避開的數值問題。此為此行在研究上進展最重要的突破之一。

另外，在倫敦期間與個人之指導教授有相當多討論的時間，也涉及一些未來研究主題的想法。

在德國 Motzen 的學術會議，個人之論文發表題目即為此專題計畫之成果。

三、 建議

此行為個人首次利用科技部計畫進行移地研究,深覺這樣的形式對於學術交流及個人的學術發展有極大助益。因過去出國皆以申請學術會議補助,而會議過程除自己的論文發表及聆聽其他學者的發表之外,無相當的時間得以進行較實質的學術交流與合作。移地研究的方式讓個人得以有較充裕與合作者進行討論；因此，此種形式實值得未來科技部計畫審查補助的鼓勵方式之一。

四、本次出國若屬國際合作研究，雙方合作性質係屬：(可複選)

□分工收集研究資料
□交換分析實驗或調查結果
■共同執行理論建立模式並驗証
□共同執行歸納與比較分析
□元件或產品分工研發
□其他 (請填寫) _______

五、其他

無。

LSE Department of Statistics

LSE THE LONDON SCHOOL OF ECONOMICS AND POLITICAL SCIENCE

Houghton Street
London WC2A 2AE
United Kingdom

tel: +44 (0)20 7955 7731
fax: +44 (0)20 7955 7416
email: statistics@lse.ac.uk

www2.lse.ac.uk/statistics/

May 29, 2014

Prof. T.C. Cheng
Department of Statistics
National Chengchi University
Taipei, Taiwan ROC

Dear George

Visit to the LSE, July 2014

It is a pleasure, on behalf of the Department of Statistics at the LSE, to invite you as a visitor from July 18th to 26th, 2014. We will be able to provide office space, computing services and library access, including electronic journals. You will also be a guest of the Senior Common Room.

I look forwards to your visit and to continuing our work on robust diagnostics.

Yours sincerely,

Anthony Atkinson

Anthony C. Atkinson
Professor of Statistics

CRC 649 "Economic Risk" Conference in Motzen
10.07 - 12.07.2014

Time	Thursday, 10. July 2014			
	Room A (Brandenburg)	Room B (Potsdam)	Room C (Berlin)	Room D (Back Office)
13:15 - 14:00	*Check-In*			
14:00 - 15:00	*Start of the conference in the plenum* **Wolfgang Härdle / Michael Burda** Welcome by CRC Coordinators **CRC 649 (Alona Zharova) / RDC (Rainer Voß, Lukas Borke)** News from CRC 649 Office and RDC **Steffen Ahrens (C10)** "Berlin Macro Network"			
15:00 - 16:00	*Plenum Lectures* **Wolfgang Härdle, Michael Burda** "TEDRIS - Tail Event Driven RIsk Structures" Discussant: **Nikolaus Wolf (B3)** **Daniel Neuhoff (C7)** "Bayesian Estimation of Autoregressive Moving-Average Processes as Exogenous Shock Processes in DSGE Models" Discussant: **Steffen Ahrens (C10)**			
16:00 - 16:30	*Coffee break (Foyer)*			
16:30 - 17:30	**Ostap Okhrin (B10)** "Efficient Iterative Maximum Likelihood Estimation of High-Parameterized Time Series Models" Discussant: **Dieter Nautz (C14)**	**Markus Bibinger & Randolf Altmeyer (C12)** "Common price and volatility jumps in noisy high-frequency data" Discussant: **Mengmeng Guo (Gast)**	**Alex Stomper (A15)** "Training the Doubtful and Timid" Discussant: **Hien Pham-Thu (Z)**	
17:30 - 19:30	*Sports activities* Swimming, football, etc. **(Andrija Mihoci)**			
from 20:00	*Dinner (Buffet)*			

Workshop 45 minutes presentation, 15 minutes discussion, Q&A

CRC 649 "Economic Risk" Conference in Motzen
10.07 - 12.07.2014

SFB 649 ÖKONOMISCHES RISIKO

Time	Friday, 11. July 2014			
	Room A (Brandenburg)	Room B (Potsdam)	Room C (Berlin)	Room D (Back Office)
from 6:30	*Breakfast (Restaurant)*			
10:00 - 11:00	**Lukas Borke (RDC)** "Quantnet basics: visualization, similarity, textmining" Discussant: **Michael Burda (C7)**	**Mayya Zhilova (B5)** "Non-asymptotic confidence bounds via multiplier bootstrap" Discussant: **Daniel Neuhoff (C7)**	**Felix Jung (B1)** "Predicting default — a dynamic Nelson-Siegel approach to forward intensities" Discussant: **Randolf Altmeyer (C12)**	
11:00 - 11:30	*Coffee break (Foyer)*			
11:30 - 12:30	**Cathy Chen (Gast)** "Surprises, sentiments, and the expectations hypothesis of the term structure of interest rates" Discussant: **George Milunovich (Gast)**	**Raphael Schöttler (B3)** "Reunification and the Real Estate Market" Discussant: **Martin Odening (C11)**	**Nooshin Nejati (Philipp Pfeiffer, Steffen Ahrens) (C10)** "Firing Taxes, Unemployment Insurance and Aggregate Fluctuations: The Role of Monetary Policy" Discussant: **Lars Winkelmann (C14)**	
12:30 - 14:00	*Lunch (Restaurant)*			
14:00 - 15:00	**Dietmar Fehr (A6)** "Talking about others: Gossip as a means to increase trust and cooperation" Discussant: **Markus Reiß (C12)**	**Frank Heinemann (C10)** "Central Bank Reputation, Transparency and Cheap Talk as Substitutes for Commitment: Experimental Evidence" Discussant: **Dorothea Kübler (A6)**	**Alexander Meyer-Gohde (C7)** "Risk Adjusted Linear Approximations: Long Run Risk In Risk-Sensitive Real Business Cycle Models" Discussant: **Jens Kolbe (B3)**	
15:00 - 15:30	*Coffee break (Foyer)*			
15:30 - 16:30	**Jin-Lung Henry Lin (Gast)** "Comparisons of Forecasting Methods with Many Predictors" Discussant: **Lei Fang (IRTG)**	**Li Ma (A13)** "Mutual funds' credit default swap strategies" Discussant: **Vladimir Spokoiny (B5)**	**Petra Burdejova (B1)** "DYTEC - DYnamic Tail Event Curves" Discussant: **Matthias Ritter (C11)**	
16:30 - 17:30	**Aleksei Netsunajev (C15)** "Is There a Technology Shock? Confronting Sign Restrictions with the Properties of the Data" Discussant: **Axel Werwatz (B3)**	**Franziska Schulz (C11)** "The impact of renewable energy production on tail events of electricity spot price indices" Discussant: **Cathy Chen (Gast)**	**Piotr Majer (Wolfgang Härdle, Hauke Heekeren) (B1)** "Portfolio Decisions and Brain Reactions via the CEAD method" Discussant: **Helmut Lütkepohl (C15)**	
17:30 - 19:30	*Sports activities* Swimming, football, etc. **(Andrija Mihoci)**			
from 19:30	*Dinner (Barbecue buffet on the terrace)*			

Workshop 45 minutes presentation, 15 minutes discussion, Q&A

CRC 649 "Economic Risk" Conference in Motzen
10.07 - 12.07.2014

Time	Saturday, 12. July 2014			
	Room A (Brandenburg)	**Room B (Potsdam)**	**Room C (Berlin)**	**Room D (Back Office)**
from 6:30	*Breakfast (Restaurant)*			
10:00 - 11:00	**Tsung-Chi Cheng (George) (Gast)** "Robust diagnostics for count regression models" Discussant: **Philipp Pfeiffer (C10)**	**Weining Wang (Z)** "Discontinuous Dynamic Semiparametric Factor models" Discussant: **Andrija Mihoci (B1)**	**Simon Voigts (C7)** "The design of the funding scheme of social security systems and its role in macroeconomic stabilization" Discussant: **Alexander Ristig (B10)**	
11:00 - 11:45	**Mentoring session** **"**Further development in science" *(including coffee)* Post-Docs: **Alexander Meyer-Gohde (C7), Markus Bibinger (C12), Andrija Mihoci (B1), Weining Wang (Z)**		**Meeting of CRC project leaders** *(including coffee)* **Project leaders**	
12:00	**Gruppenfoto**			
	Abreise			

Workshop 45 minutes presentation, 15 minutes discussion, Q&A

科技部補助計畫衍生研發成果推廣資料表

日期:2014/10/12

<table>
<tr><td rowspan="3">科技部補助計畫</td><td>計畫名稱: 計數迴歸模型之穩健診斷</td></tr>
<tr><td>計畫主持人: 鄭宗記</td></tr>
<tr><td>計畫編號: 102-2118-M-004-004-　　　學門領域: 統計方法</td></tr>
<tr><td colspan="2">無研發成果推廣資料</td></tr>
</table>

102 年度專題研究計畫研究成果彙整表

計畫主持人：鄭宗記			計畫編號：102-2118-M-004-004-				
計畫名稱：計數迴歸模型之穩健診斷							
成果項目			量化			單位	備註（質化說明：如數個計畫共同成果、成果列為該期刊之封面故事...等）
			實際已達成數（被接受或已發表）	預期總達成數(含實際已達成數)	本計畫實際貢獻百分比		
國內	論文著作	期刊論文	0	0	100%	篇	
		研究報告/技術報告	0	0	100%		
		研討會論文	0	0	100%		
		專書	0	0	100%		
	專利	申請中件數	0	0	100%	件	
		已獲得件數	0	0	100%		
	技術移轉	件數	0	0	100%	件	
		權利金	0	0	100%	千元	
	參與計畫人力（本國籍）	碩士生	2	2	100%	人次	
		博士生	0	0	100%		
		博士後研究員	0	0	100%		
		專任助理	0	0	100%		
國外	論文著作	期刊論文	0	1	100%	篇	
		研究報告/技術報告	0	0	100%		
		研討會論文	1	0	100%		
		專書	0	0	100%	章/本	
	專利	申請中件數	0	0	100%	件	
		已獲得件數	0	0	100%		
	技術移轉	件數	0	0	100%	件	
		權利金	0	0	100%	千元	
	參與計畫人力（外國籍）	碩士生	0	0	100%	人次	
		博士生	0	0	100%		
		博士後研究員	0	0	100%		
		專任助理	0	0	100%		

其他成果 (無法以量化表達之成果如辦理學術活動、獲得獎項、重要國際合作、研究成果國際影響力及其他協助產業技術發展之具體效益事項等，請以文字敘述填列。)	無

	成果項目	量化	名稱或內容性質簡述
科教處計畫加填項目	測驗工具(含質性與量性)	0	
	課程/模組	0	
	電腦及網路系統或工具	0	
	教材	0	
	舉辦之活動/競賽	0	
	研討會/工作坊	0	
	電子報、網站	0	
	計畫成果推廣之參與（閱聽）人數	0	

科技部補助專題研究計畫成果報告自評表

請就研究內容與原計畫相符程度、達成預期目標情況、研究成果之學術或應用價值（簡要敘述成果所代表之意義、價值、影響或進一步發展之可能性）、是否適合在學術期刊發表或申請專利、主要發現或其他有關價值等，作一綜合評估。

1. 請就研究內容與原計畫相符程度、達成預期目標情況作一綜合評估

■達成目標

□未達成目標（請說明，以 100 字為限）

□實驗失敗

□因故實驗中斷

□其他原因

說明：

2. 研究成果在學術期刊發表或申請專利等情形：

論文：□已發表 □未發表之文稿 ■撰寫中 □無

專利：□已獲得 □申請中 ■無

技轉：□已技轉 □洽談中 ■無

其他：（以 100 字為限）

在德國 Motzen 的學術會議，個人之論文發表題目即為此專題計畫之成果。

3. 請依學術成就、技術創新、社會影響等方面，評估研究成果之學術或應用價值（簡要敘述成果所代表之意義、價值、影響或進一步發展之可能性）（以 500 字為限）

負二項（negative binomial）迴歸模型是近年來應用相當廣泛的模型，但關於其穩健估計及離群值的偵測等相關問題的討論仍為有限。此計畫的內容與可能的成果，極具其學術上的意義，且有其應用的價值。