Comparison among Akaike Information Criterion, Bayesian Information Criterion and Vuong's test in Model Selection: A Case Study of Violated Speed Regulation in Taiwan

Kim-Hung PHO[1,*], *Sel LY*[1], *Sal LY*[1], *T. Martin LUKUSA*[2]

[1]Faculty of Mathematics and Statistics, Ton Duc Thang University, Ho Chi Minh City, Vietnam
[2]Institute of Statistical Science, Academia Sinica, Taiwan, R.O.C., Taiwan

*Corresponding Author: Kim-Hung PHO (Email: phokimhung@tdtu.edu.vn)
(Received: 4-Dec-2018; accepted: 22-Feb-2019; published: 31-Mar-2019)
DOI: http://dx.doi.org/10.25073/jaec.201931.220

Abstract. *When doing research scientific issues, it is very significant if our research issues are closely connected to real applications. In reality, when analyzing data in practice, there are frequently several models that can appropriate to the survey data. Hence, it is necessary to have a standard criteria to choose the most efficient model. In this article, our primary interest is to compare and discuss about the criteria for selecting model and its applications. The authors provide approaches and procedures of these methods and apply to the traffic violation data where we look for the most appropriate model among Poisson regression, Zero-inflated Poisson regression and Negative binomial regression to capture between number of violated speed regulations and some factors including distance covered, motorcycle engine and age of respondents by using AIC, BIC and Vuong's test. Based on results on the training, validation and test data set, we find that the criteria AIC and BIC are more consistent and robust performance in model selection than the Vuong's test. In the present paper, the authors also discuss about advantages and disadvantages of these methods and provide some of suggestions with potential directions in the future research.*

Keywords

Akaike Information Criteria (AIC), Bayesian Information Criterion (BIC), Vuong's test, Poisson regression, Zero-inflated Poisson regression, Negative binomial regression.

1. Introduction

The model selection criteria is a very crucial field in statistics, economics and several other areas and it has numerous practical applications. This issue is currently researched theoretically and practically by several statisticians and has gained many attentions in the last two decades, especially in regression and econometric models. There are three most commonly used model selection criteria including Akaike information criterion (AIC), Bayesian information criterion (BIC) and Vuong's test, which are compared and discussed in this paper. AIC is first proposed by Akaike [1] as a method to compare different models on a given outcome. Meanwhile, BIC is proposed by Schwarz [20], is a criterion for model selection among a finite set of models. Vuong's test has been proposed by Vuong [24] in the literature aiming at selecting a single model

regardless of its intended use. All three criteria are the most widespread criteria for choosing model.

Until today, these problems have been studied and utilized in numerous areas. AIC has been researched and applied extensively in literature such as: Snipes et al. [19] employ AIC and present about an example from wine ratings and prices, Taylor et al. [21] introduce indicators of hotel profitability: Model selection using AIC, Charkhi et al. [4] research about asymptotic post-selection inference for the AIC, Chang et al. [3] present about Akaike Information Criterion-based conjunctive belief rule base learning for complex system modeling, etc.

In addition, BIC is also utilized extensively in literature for example: Neath et al. [16] introduce about regression and time series model selection using variants of the Schwarz information criterion. Cavanaugh et al. [2] present about generalizing the derivation of the BIC. Weakliem [27] introduce about a critique of the Bayesian information criterion for model selection. Neath et al. [15] present about a Bayesian approach to the multiple comparisons problem. Neath et al. [17] present about the BIC: background, derivation, and applications. Nguefack-Tsague et al. [23] focus on introduce about Bayesian information criterion, etc.

Similarly to AIC and BIC, Vuong's test [24] is also used largely in literature for instance: Clarke [5] employ Vuong's test to introduce a simple distribution-free test for non-nested model selection, Theobald [22] utilize Vuong's test to present a formal test of the theory of universal common ancestry, Lukusa et al. [13] use Vuong's test to evaluate whether the zero-inflated Poisson (ZIP) regression model is consistent with the real data, Dale et al. [6] perform model comparison using Vuong's test to estimate of nested and zero-inflated ordered probit models, Schneider et al. [18] present about model selection of nested and non-nested item response models using Vuong's test, etc.

Our main objective in this paper is to provide researchers an overview of the criteria in model selection for the traffic violation data. The rest of the paper is organized as follows. In Section 2, we present approaches and procedures of the criteria for choosing model including Akaike information criterion (AIC), Bayesian information criterion (BIC) and Vuong's test. In Section 3, these methods are applied to a real data which could help readers to easily assess them. Some of suggestions and some potential directions for the further research are devoted in Section 4. Finally, some conclusions and remarks are given in Section 5.

2. Some of Criteria for Model Selection

In this section, we present approaches and procedures of ubiquitous methods to choose the most efficient model consisting of Akaike Information Criteria (AIC), Bayesian Information Criterion (BIC) and Vuong's test.

2.1. Akaike Information Criteria (AIC)

AIC is first proposed by Akaike [1] as a method to compare different models on a given outcome. The AIC for candidate model is defined as follows:

$$\mathrm{AIC} := -2\ell(\hat{\theta}|y) + 2K, \tag{1}$$

where K is the number of estimated parameters in the model including the intercept and $\ell(\hat{\theta}|y)$ is a log-likelihood at its maximum point of the estimated model. The rule of choice: the smaller the value of AIC is, the better the model is.

2.2. Bayesian Information Criterion (BIC)

BIC is first introduced by Schwarz [20], one sometimes calls the Bayesian information criterion (BIC) or Schwarz criterion (also SBC, SBIC) which is a criterion for model selection among a finite set of models. The BIC for candidate model is defined as follows:

$$\mathrm{BIC} := -2\ell(\hat{\theta}|y) + K\ln(n), \tag{2}$$

where n is a sample size; K is the number of estimated parameters in the model including the

intercept and $\ell(\hat{\theta}|y)$ is the log-likelihood at its maximum point of the estimated model. The rule of selection: the smaller the value of BIC is, the better the model is. The procedure for applying AIC and BIC are given as follows:

Step 1: Selecting candidate models which can be fitted to the data set.

Step 2: Estimating unknown parameters of models.

Step 3: Finding values of AIC and BIC by using the formulas (1) and (2), respectively.

Step 4: Basing on the rule of choice, one can decide the most suitable model.

2.3. Vuong's Test

Vuong's test [24] is one of the ubiquitous criteria for choosing model and it is often used to the data set with no missing values. Let $f_1(Y|X,Z,W;\boldsymbol{\alpha}_1)$ and $f_2(Y|X,Z,W;\boldsymbol{\alpha}_2)$ be two non-nested probability models. Let $\widehat{\boldsymbol{\alpha}}_1$ and $\widehat{\boldsymbol{\alpha}}_2$ be a consistent estimator of $\boldsymbol{\alpha}_1$ and $\boldsymbol{\alpha}_2$ under the model f_1 and f_2, respectively. Letting hypotheses

- H_0: The two models are equally closed to the true data.
- H_1: Model 1 is closer than model 2.

The Vuong's test statistics is provided as follows; (see Mouatassim and Ezzahid [14]):

$$V = V\left(\widehat{\boldsymbol{\alpha}}_1, \widehat{\boldsymbol{\alpha}}_2\right) = \frac{\sqrt{n}\left[\frac{1}{n}\sum_{i=1}^{n} m_i\left(\widehat{\boldsymbol{\alpha}}_1, \widehat{\boldsymbol{\alpha}}_2\right)\right]}{h\left(\widehat{\boldsymbol{\alpha}}_1, \widehat{\boldsymbol{\alpha}}_2\right)}, \tag{3}$$

where

$$h^2\left(\widehat{\boldsymbol{\alpha}}_1, \widehat{\boldsymbol{\alpha}}_2\right) = \frac{1}{n}\sum_{i=1}^{n} m_i^2\left(\widehat{\boldsymbol{\alpha}}_1, \widehat{\boldsymbol{\alpha}}_2\right) - \left[\frac{1}{n}\sum_{i=1}^{n} m_i\left(\widehat{\boldsymbol{\alpha}}_1, \widehat{\boldsymbol{\alpha}}_2\right)\right]^2$$

The detailed calculation of V is provided in Appendix. Note that:

- $m_i\left(\widehat{\boldsymbol{\alpha}}_1, \widehat{\boldsymbol{\alpha}}_2\right) = \ln\left(\frac{f_1\left(Y_i|X_i, Z_i, \widehat{\boldsymbol{\alpha}}_1\right)}{f_2\left(Y_i|X_i, Z_i, \widehat{\boldsymbol{\alpha}}_2\right)}\right)$, where $f_j\left(Y_i|X_i, Z_i, \widehat{\boldsymbol{\alpha}}_j\right)$, is the predicted probability of an observed count for case i from the model j, $j = 1, 2$, respectively.
- Moreover for the complete case, V can be easily obtained from the package *pscl* in R language, (Zeileis at el. [28]).

At the significant level α, the decision rule is given as follows:

- If $V > Q_{\alpha/2}$, choose model 1.
- If $V < -Q_{\alpha/2}$, choose model 2.
- If $|V| < Q_{\alpha/2}$, both models are equivalent.

where $Q_{\alpha/2}$ is an upper quantile of standard normal distribution at the level $\alpha/2$. Similar to algorithms for AIC and BIC, to perform Vuong's test, we need to do through following steps:

Step 1: Choosing candidate models which can be fitted to the data set.

Step 2: Estimating unknown coefficients of models.

Step 3: Calculating V by using (3)

Step 4: Basing on the rule of choice, one can select the most compatible model.

Note that: Step 1 is a very important step in practice, basing on characteristics of the data set, one can choose some reasonable models to fit. For example, if the data set is a binary, then candidate models are considered such as logistic regression model, probit model and so on. If the data set is class of count data, one can utilize some of models such as: Poisson regression model, binomial regression model, negative binomial regression model and so on. If the data set is a zero-inflated or imbalance data, zero-inflated Poisson (ZIP) regression model, zero-inflated binomial (ZIB) regression model, and zero-inflated negative binomial (ZINB) regression model could be more plausible candidates.

3. Models for Violated Speed Regulation

The data set utilized in this analysis is from a motorcycle survey study regarding road traffic regulations conducted in Taiwan by the Ministry of Transportation and Communication in 2007. This data set has been used in the paper "Semi-parametric estimation of a zero-inflated Poisson (ZIP) regression model with missing covariates" by Lukusa et al. [13]. This study consists of 7,386 respondents involving 1122 missing values. Before applying the criteria to select optimal models, one may require the data having no missing values. Hence, we need to remove all of missing values and displayed in the Tab. 1. The bar graph of the outcome variable Y is exhibited in Fig. 1 (Appendix). As can be observed from the Tab. 1 and the Fig. 1 that the number of people violating of speed regulations in Taiwan 2007 is very small. The data set contains most of zeros in Y which is usually called zero-inflated count data. With this type of data set, some of zero-inflated models may be more appropriate than other models. In this section, we investigate three following models: Zero-inflated Poisson (ZIP) regression model denoted by M_1, Poisson regression model called M_2 and M_3 stands for Negative binomial (NB) regression model. The forms of these models are briefly given in the Appendix. Our aim is to evaluate which model is more appropriate for modeling between the number of violated speed regulation (Y) with some factors such as Distance-covered (X), Motorcycle-engine (Z) and the Age of respondents (W). Firstly the data is randomly split into three data sets, namely, training, validation and test with respect to the percentage of $60\% - 20\% - 20\%$. This means 60% of the whole data is used to train the three models M_i, $i = 1, 2, 3$, with results as shown in the Tabs. 2, 3 and 4, respectively. Next, the validation data which is also randomly extracted by 20% of the full data is then used for selecting the most appropriate model while the remaining test data is to check accuracy when we do a performance of forecast with those models. The criteria AIC, BIC, Vuong's test, mean square error (MSE) and accuracy are respectively computed to each data set and each model for comparisons.

Descriptions	Variables	Re
Distance-covered (km a year)	X	6262
1. Under 1,000	$X = 1$	1752
2. 1,000-2,999	$X = 2$	1711
3. 3,000-9,999	$X = 3$	1856
4. Over 1,000	$X = 4$	943
Number-Violation (in a year)	Y	6262
1. Never violation	$Y = 0$	5637
2. One violation	$Y = 1$	380
3. Two violations	$Y = 2$	169
4. Three violations	$Y = 3$	59
5. Four violations	$Y = 4$	11
6. Five violations	$Y = 5$	2
7. Six violations	$Y = 6$	3
8. Seven violations	$Y = 7$	1
Motorcycle-engine (cubic centimeters (cc))	Z	6262
1. Under 50	$Z = 1$	1303
2. 50-249	$Z = 2$	4153
3. 250-549	$Z = 3$	272
4. Over 550	$Z = 4$	534
Respondent's age (years old)	W	6262
1. Under 18	$W = 1$	3
2. 18-19	$W = 2$	142
3. 20-29	$W = 3$	1395
4. 30-39	$W = 4$	1607
5. 40-49	$W = 5$	1508
6. Over 50	$W = 6$	1607

Tab. 1: Frequency of respondents (Re) in data set after deleting missing values.

The ZIP model (M_1) is composed of two parts separately, where the former is called count model with coefficients denoted by β and the latter is the so-called inflation model with coefficients denoted by γ, see Equation (5.). As can be seen from the Tab. 2, all estimated coefficients of zero-inflated part are statistically significant at the level 5% thanks to all P-values are less than 0.05. In contrast, in the count model, the Distance-covered (X) and Motorcycle-engine (Z) are not significant, except the Age (W). The factor Age affects the number of traffic violations for both parts in the sense that if W is increasing and other factors are assumed to be unchanged, then the ex-

pected number of violation is definitely reduced and the probability of not violating is clearly increasing since we have $\widehat{\beta}_3 = -0.23536 < 0$ and $\widehat{\gamma}_3 = 0.19547 > 0$, respectively.

For the Poisson regression model (M_2) and the Negative binomial regression model (M_3), we also see the statistical significance of estimated coefficients based on P-values are very small (≈ 0). The two factors X and Z with positive coefficients imply that they increase the incidence rate (see μ in (11) and (12)) of number of traffic violations while W makes it to be decreasing as in the case of ZIP model, see Tab. 3 and 4.

We now turn to discuss which model is better. Based on results represented in the Tab. 5 and 6, the smallest value AIC and BIC on validation data are respectively 1013.404 and 1033.937 and both are produced by the model M_1. One can also see this confirmation on the training and test data sets. Hence, the model M_1 (ZIP) is the most plausible model in comparison to the models M_3 and M_2. However, by Vuong's test results on the validation set, see Tab. 8, it suggests that the model M_1 is more preferable than the model M_2, but it is equivalent to the model M_3 ($P\text{-}value = 0.1 > 0.05$). This equivalence is also confirmed by the same mean square error $MSE = 0.3488$ and the same accuracy 90.42% on the validation data, see Tabs. 10 and 11. When checking on the test set, the model M_1 has a slightly better performance with the smallest MSE 0.2811, the greatest accuracy 90.60% and similarly result if using Vuong's test. Our result is consistent to Lukusa et al. It also shows that the information criteria AIC and BIC are more robust than the Vuong's test in model selection. [13].

4. Discussion and some potential directions for further research

It can be seen that, to consider the compatibility of two models, we can use some criteria such as: Vuong's test, Akaike Information Criteria (AIC) and Bayesian Information Criterion (BIC). These formulas have the same characteristics that can be derived from model's likelihood functions and results of maximum likelihood estimates (MLE). Nevertheless, if AIC or BIC is used to consider the appropriateness of models, one needs to calculate separately each formula and compare values together with the decision rule: the smaller the value of AIC or BIC is, the better the model is, but the shortcoming is sometimes one may not know how to determine whether differences between two values AIC (resp. BIC) is statistically significant or not. In case of using Vuong's test, we only need to compute the statistic given in (3) and follow the rule of choice or find the P-value which can help us differentiate two models significantly. However, the Vuong's test is not more robust than AIC and BIC in model selection as shown in the Section 3. .

For AIC and BIC, AIC is very ubiquitous in econometrics, while BIC is more commonly utilized in sociology, see Weakliem [27]. It can be seen that, BIC becomes to AIC if $K = \ln(n)$. To see the relationship between formula (1), (2), and Vuong's test, the problem is given as follows: Let D is an observed data (a real data). A number of possible models M_k for D are considered, with each model having a likelihood function $L(D|\theta_k; M_k)$ and θ_k are unknown parameters need to be estimated with p_k parameters. For simplicity's sake, let $\ell(\theta_k) = \ln[L(D|\theta_k; M_k)]$ and $\widehat{\theta}_k$ be an estimator of θ_k by using the maximum likelihood estimate (MLE). Assessment of the candidate models can be carried out as a sequence of comparisons between pairs of models. It is more convenient to consider model M_1 and M_2. The difference of two values AIC (resp. BIC) obtained from two certain models can be expressed as follows:

$$\Delta AIC := -2[\ell(\hat{\theta}_2) - \ell(\hat{\theta}_1)] + 2(p_2 - p_1) \quad (4)$$

$$\Delta BIC := -2[\ell(\hat{\theta}_2) - \ell(\hat{\theta}_1)] + (p_2 - p_1)\ln(n), \quad (5)$$

and the Vuong's test can be rewritten as:

$$V := \frac{\ell(\hat{\theta}_1) - \ell(\hat{\theta}_2)}{\sqrt{n}h(\hat{\theta}_1, \hat{\theta}_2)}, \quad (6)$$

where $h^2((\hat{\theta}_1, \hat{\theta}_2))$ denotes sample variance of the difference of log-likelihood $\ell(\hat{\theta}_1) - \ell(\hat{\theta}_2)$.

From this point of view, one may prefer the first model M_1 than the second model M_2 if $\Delta AIC, \Delta BIC$ and V are positive values.

AIC is a very widespread formula, thus there are several scholars have researched and improved it by some adjustments. List of modified AIC statistics are given as follows:

- First denoted by AICc is the corrected AIC for sample size

$$AICc := AIC + \frac{2K(K+1)}{n-K-1}. \quad (7)$$

- Next is the AIC weight of the model M_k defined by

$$AICw(k) := \frac{\exp\left(-\frac{1}{2}AICc(k)\right)}{\sum_{k=1}^{R}\exp\left(-\frac{1}{2}AICc(i)\right)}, \quad (8)$$

  where R is number of possible candidate models. The $AICw(k)$ is the weight of the evidence of the model M_k with respect to other candidate models, i.e. the model has the highest $AICw$ is considered as the strongest model.

- Evidence ratio of the model M_k is determined by

$$ER(k) := \frac{AICw_{\text{best}}}{AICw(k)}, \quad (9)$$

  where $AICw_{\text{best}}$ is the AIC weight of the best (true) model. This ratio measures how decisive the evidence in the sense that the model with the smallest ER is the most appropriate model with respect to other candidate models.

Regarding applicability, Vuong's test, Akaike Information Criteria (AIC) and Bayesian Information Criterion (BIC) are only applicable for complete data i.e. no missing values. In several practical applications, some elements in the given data set are usually missing. Hence, these traditional criteria may be no longer suitable for selecting models and if we remove all missing elements, it could lead to the biasness in inferences. Therefore, it is necessary to improve the above formulas with the possibility of dealing with missing data. To the best of our knowledge, no scholar has studied this problem yet. These are potential research directions in the next time. Some of methods to solve this issue are very ubiquitous and prevalent. Little [12] reviewed six methods to solve the missing data problem that are complete-case (CC) analysis, available-case (AC) methods, least squares (LS) on imputed data, maximum likelihood (ML), Bayesian methods and multiple imputation (MI). Zhao and Lipsitz [29] proposed the inverse probability weighting (IPW) method. Wang et al. [26] developed a regression calibration (RC) method. Wang et al. [25] introduced the joint conditional likelihood (JCL) method. In addition, we can combine methods to provide a robust tool to solve this problem. For instance: Han [8] presented multiply robust estimation in regression analysis with missing data where the IPW and MI method are combined together.

About the expansion of above issues, it is similar to the study of regression models, the traditional regression models such as logistic regression model, zero-inflated binomial (ZIB) regression model, zero-inflated Poisson (ZIP) regression model, etc, coefficients cannot be directly estimated if some covariates having missing values. Hence, one needs to have some new approaches to estimate parameters in this situation. For instance, Wang et al. [25] employed the joint conditional likelihood (JCL) estimator in logistic regression with missing covariates data. Hsieh et al. [9] extended method of Wang et al. (2002) to introduce a semiparametric analysis of randomized response data with missing covariates in logistic regression. Lee et al. [11] also extended method in Wang et al. (2002) to present a semiparametric estimation of logistic regression model with missing covariates and outcome. Pho et al. [30] discussed about three ubiquitous approaches to handle the issues having missing data. Diallo et al. [7] introduced an IPW estimator of the parameters of a ZIB regression model with missing-at-random covariates. Lukuasa et al. [13] presented a semiparametric estimation of a zero-inflated Poisson (ZIP) regression model with missing covariates, etc.

5. Conclusion

We reviewed widespread methods for selecting the most efficient model: Vuong's test, Akaike Information Criteria (AIC) and Bayesian Information Criterion (BIC). The approach and procedure of these methods and application to traffic violation data are provided step by step. Based on results on the training, validation and test data set, we find that the criteria AIC and BIC have a more consistent and robust performance in model selection than the Vuong's test in this case. Besides, some advantages and disadvantages of these methods have been discussed and compared in the paper. Furthermore, the authors also suggest some potential research directions in the next time.

References

[1] Akaike, H. (1974). A new look at the statistical model identification. In Selected Papers of Hirotugu Akaike (pp. 215-222). Springer, New York, NY.

[2] Cavanaugh, J. E., & Neath, A. A. (1999). Generalizing the derivation of the Schwarz information criterion. Communications in Statistics-Theory and Methods, 28(1), 49-66.

[3] Chang, L., Zhou, Z., Chen, Y., Xu, X., Sun, J., Liao, T., & Tan, X. (2018). Akaike Information Criterion-based conjunctive belief rule base learning for complex system modeling. Knowledge-Based Systems, 161, 47-64.

[4] Charkhi, A., & Claeskens, G. (2018). Asymptotic post-selection inference for the Akaike information criterion. Biometrika, 105(3), 645-664.

[5] Clarke, K. A. (2007). A simple distribution-free test for nonnested model selection. Political Analysis, 15(3), 347-363.

[6] Dale, D., & Sirchenko, A. (2018). Estimation of Nested and Zero-Inflated Ordered Probit Models. Higher School of Economics Research Paper No. WP BRP, 193.

[7] Diop, A., & Dupuy, J. F. (2017). Estimation in zero-inflated binomial regression with missing covariates.

[8] Han, P. (2014). Multiply robust estimation in regression analysis with missing data. Journal of the American Statistical Association, 109(507), 1159-1173.

[9] Hsieh, S. H., Lee, S. M., & Shen, P. (2009). Semiparametric analysis of randomized response data with missing covariates in logistic regression. Computational Statistics & Data Analysis, 53(7), 2673-2692.

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

[11] Lee, S. M., Li, C. S., Hsieh, S. H., & Huang, L. H. (2012). Semiparametric estimation of logistic regression model with missing covariates and outcome. Metrika, 75(5), 621-653.

[12] Little, R. J. (1992). Regression with missing X's: a review. Journal of the American Statistical Association, 87(420), 1227-1237.

[13] Lukusa, T. M., Lee, S. M., & Li, C. S. (2016). Semiparametric estimation of a zero-inflated Poisson regression model with missing covariates. Metrika, 79(4), 457-483.

[14] Mouatassim, Y., & Ezzahid, E. H. (2012). Poisson regression and zero-inflated Poisson regression: application to private health insurance data. European actuarial journal, 2(2), 187-204.

[15] Neath, A. A., & Cavanaugh, J. E. (2006). A Bayesian approach to the multiple comparisons problem. Journal of Data Science, 4(2), 131-146.

[16] Neath, A. A., & Cavanaugh, J. E. (1997). Regression and time series model selection using variants of the Schwarz information criterion. Communications in Statistics-Theory and Methods, 26(3), 559-580.

[17] Neath, A. A., & Cavanaugh, J. E. (2012). The Bayesian information criterion: background, derivation, and applications. Wiley

Interdisciplinary Reviews: Computational Statistics, 4(2), 199-203.

[18] Schneider, L., Chalmers, R. P., Debelak, R., & Merkle, E. C. (2018). Model Selection of Nested and Non-Nested Item Response Models using Vuong Tests. arXiv preprint arXiv:1810.04734.

[19] Snipes, M., & Taylor, D. C. (2014). Model selection and Akaike Information Criteria: An example from wine ratings and prices. Wine Economics and Policy, 3(1), 3-9.

[20] Schwarz, G. (1978). Estimating the dimension of a model. The annals of statistics, 6(2), 461-464.

[21] Taylor, D. C., Snipes, M., & Barber, N. A. (2018). Indicators of hotel profitability: Model selection using Akaike information criteria. Tourism and Hospitality Research, 18(1), 61-71.

[22] Theobald, D. L. (2010). A formal test of the theory of universal common ancestry. Nature, 465(7295), 219.

[23] Nguefack-Tsague, G., Bulla, I. (2014). A focused Bayesian information criterion. Advances in Statistics, 2014.

[24] Vuong, Q. H. (1989). Likelihood ratio tests for model selection and non-nested hypotheses. Econometrica: Journal of the Econometric Society, 307-333.

[25] Wang, C. Y., Chen, J. C., Lee, S. M., & Ou, S. T. (2002). Joint conditional likelihood estimator in logistic regression with missing covariate data. Statistica Sinica, 555-574.

[26] Wang, C. Y., Wang, S., Zhao, L. P., & Ou, S. T. (1997). Weighted semiparametric estimation in regression analysis with missing covariate data. Journal of the American Statistical Association, 92(438), 512-525.

[27] Weakliem, D. L. (1999). A critique of the Bayesian information criterion for model selection. Sociological Methods & Research, 27(3), 359-397.

[28] Zeileis, A., Kleiber, C., & Jackman, S. (2008). Regression models for count data in R. Journal of statistical software, 27(8), 1-25.

[29] Zhao, L. P., & Lipsitz, S. (1992). Designs and analysis of two-stage studies. Statistics in medicine, 11(6), 769-782.

[30] Pho, K. H., & Nguyen, V. T. (2018). Comparison of Newton-Raphson Algorithm and Maxlik Function. Journal of Advanced Engineering and Computation, 2(4), 281-292.

About Authors

Kim-Hung PHO is a Ph.D. Student in Applied Statistics at Feng Chia University, Taiwan. In 2014, he became a lecturer of Faculty of Mathematics and Statistics in Ton Duc Thang University, Ho Chi Minh City, Vietnam. His currently research interests include Regression models with missing data, Randomized Response Technique, Copula, Mathematics education models and Financial Mathematics.

Sel LY has worked as a lecturer at Faculty of Mathematics and Statistics, Ton Duc Thang University since 2014. He earned a Bachelor degree in Maths-Informatics Teacher Education in 2011 and a Master degrees in Probability Theory and Mathematical Statistics in 2013. Currently, he is a Ph.D. Student in Mathematical Sciences at Nanyang Technological University, Singapore. His research interests are Data mining, Copula, Stochastic Process and Financial Mathematics.

Sal LY hold Bachelor and Master degrees in Probability Theory and Mathematical Statistics in 2014 and 2016, respectively. His currently research interests include Copula Theory and Financial Mathematics.

T. Martin LUKUSA is working in Institute of Statistical Science, Academia Sinica, Taiwan, R.O.C., Taiwan. His currently research interests include Regression models with missing data, Randomized Response Technique, and Financial Mathematics.

Appendix

The detailed calculation of V

$$
\begin{aligned}
V &= \frac{\sqrt{n}\left[\frac{1}{n}\sum_{i=1}^{n} m_i\left(\widehat{\boldsymbol{\alpha}}_1, \widehat{\boldsymbol{\alpha}}_2\right)\right]}{\left\{\frac{1}{n}\sum_{i=1}^{n}\left[m_i\left(\widehat{\boldsymbol{\alpha}}_1, \widehat{\boldsymbol{\alpha}}_2\right) - \overline{m}\right]^2\right\}^{\frac{1}{2}}} \\
&= \frac{\sqrt{n}\left[\frac{1}{n}\sum_{i=1}^{n} m_i\left(\widehat{\boldsymbol{\alpha}}_1, \widehat{\boldsymbol{\alpha}}_2\right)\right]}{\{\frac{1}{n}\sum_{i=1}^{n} m_i^2\left(\widehat{\boldsymbol{\alpha}}_1, \widehat{\boldsymbol{\alpha}}_2\right) - \frac{2\overline{m}}{n}\sum_{i=1}^{n} m_i\left(\widehat{\boldsymbol{\alpha}}_1, \widehat{\boldsymbol{\alpha}}_2\right) + \overline{m}^2\}^{\frac{1}{2}}} \\
&= \frac{\sqrt{n}[\frac{1}{n}\sum_{i=1}^{n} m_i\left(\widehat{\boldsymbol{\alpha}}_1, \widehat{\boldsymbol{\alpha}}_2\right)]}{\left\{\frac{1}{n}\sum_{i=1}^{n} m_i^2\left(\widehat{\boldsymbol{\alpha}}_1, \widehat{\boldsymbol{\alpha}}_2\right) - 2\overline{m}^2 + \overline{m}^2\right\}^{\frac{1}{2}}} \\
&= \frac{\sqrt{n}\left[\frac{1}{n}\sum_{i=1}^{n} m_i\left(\widehat{\boldsymbol{\alpha}}_1, \widehat{\boldsymbol{\alpha}}_2\right)\right]}{\left\{\frac{1}{n}\sum_{i=1}^{n} m_i^2\left(\widehat{\boldsymbol{\alpha}}_1, \widehat{\boldsymbol{\alpha}}_2\right) - \overline{m}^2\right\}^{\frac{1}{2}}} \\
&= \frac{\sqrt{n}\left[\frac{1}{n}\sum_{i=1}^{n} m_i\left(\widehat{\boldsymbol{\alpha}}_1, \widehat{\boldsymbol{\alpha}}_2\right)\right]}{\left\{\frac{1}{n}\sum_{i=1}^{n} m_i^2\left(\widehat{\boldsymbol{\alpha}}_1, \widehat{\boldsymbol{\alpha}}_2\right) - \left[\frac{1}{n}\sum_{i=1}^{n} m_i\left(\widehat{\boldsymbol{\alpha}}_1, \widehat{\boldsymbol{\alpha}}_2\right)\right]^2\right\}^{\frac{1}{2}}} \\
&= \frac{\sqrt{n}\left[\frac{1}{n}\sum_{i=1}^{n} m_i\left(\widehat{\boldsymbol{\alpha}}_1, \widehat{\boldsymbol{\alpha}}_2\right)\right]}{h\left(\widehat{\boldsymbol{\alpha}}_1, \widehat{\boldsymbol{\alpha}}_2\right)}
\end{aligned}
$$

where $\overline{m} = \frac{1}{n}\sum_{i=1}^{n} m_i\left(\widehat{\boldsymbol{\alpha}}_1, \widehat{\boldsymbol{\alpha}}_2\right)$, and

$$
\begin{aligned}
h^2\left(\widehat{\boldsymbol{\alpha}}_1, \widehat{\boldsymbol{\alpha}}_2\right) &= \frac{1}{n}\sum_{i=1}^{n} m_i^2\left(\widehat{\boldsymbol{\alpha}}_1, \widehat{\boldsymbol{\alpha}}_2\right) \\
&\quad - \left[\frac{1}{n}\sum_{i=1}^{n} m_i\left(\widehat{\boldsymbol{\alpha}}_1, \widehat{\boldsymbol{\alpha}}_2\right)\right]^2.
\end{aligned}
$$

Zero-inflated Poisson (ZIP) regression model

Lambert [10] propose the parametric ZIP regression model in which the non-susceptible probability (mixing weight) p is linked to $\mathcal{X}$ via a logit-linear predictor, $p = H(\boldsymbol{\gamma}^T\mathcal{X})$ for $H(u) = [1 + \exp(-u)]^{-1}$, and the Poisson mean λ is linked to $\mathcal{X}$ via a log-linear predictor, $\lambda = \exp(\boldsymbol{\beta}^T\mathcal{X})$. where $\boldsymbol{\gamma}$ and $\boldsymbol{\beta}$ are unknown parameters need to be estimated. In the present paper, $\mathcal{X} = (X, Z, W)^T$ and so the ZIP model can be expressed as follows:

$$
\begin{aligned}
P(Y = y|X, Z, W) &= H(\boldsymbol{\gamma}^T\mathcal{X})I(y = 0) + \\
&+ [1 - H(\boldsymbol{\gamma}^T\mathcal{X})]\frac{\exp[-\exp(\boldsymbol{\beta}^T\mathcal{X})][\exp(\boldsymbol{\beta}^T\mathcal{X})]^y}{y!}
\end{aligned}
\tag{10}
$$

for $y = 0, 1, 2, \ldots$, where $\gamma = (\gamma_0, \ldots, \gamma_3)^T$ is called coefficients of zero-inflation model while $\beta = (\beta_0, \ldots, \beta_3)^T$ is called coefficients of count model, see more details in Lambert [10] and Lukusa et al. [13].

Poisson regression model

The Poisson incidence rate μ is determined by a set of p regressor variables (the $\mathbf{X}$'s). The expression relating these quantities is

$$
\mu = \exp\left(\beta_0 + \beta_1 X_1 + \cdots + \beta_n X_p\right). \tag{11}
$$

An ubiquitous Poisson regression model for an observation i is written as follows

$$
P\left(Y_i = y_i|\mu_i\right) = \frac{e^{-\mu_i}\left(\mu_i\right)^{y_i}}{y_i!},
$$

where $\mu_i = \exp\left(\beta_0 + \beta_1 X_{1i} + \cdots + \beta_p X_{pi}\right)$, and $\beta_0, \beta_1, \ldots, \beta_n$ are regression coefficients need to be estimated.

Negative binomial regression model

The mean of y is determined by the exposure time t and a set of p regressor variables (the $\mathbf{X}$'s). The expression relating these quantities is

$$
\mu_i = \exp\left(\ln(t_i) + \beta_0 + \beta_1 X_{1i} + \cdots + \beta_p X_{pi}\right). \tag{12}
$$

The widespread negative binomial regression model for an observation i is given by

$$
\begin{aligned}
P\left(Y_i = y_i|\mu_i, \alpha\right) &= \frac{\Gamma\left(y_i + \alpha^{-1}\right)}{\Gamma\left(\alpha^{-1}\right)\Gamma\left(y_i + 1\right)} \\
&\quad \times \left(\frac{1}{1 + \alpha\mu_i}\right)^{\alpha^{-1}} \left(\frac{\alpha\mu_i}{1 + \alpha\mu_i}\right)^{y_i}
\end{aligned}
$$

where $\beta_0, \beta_1, \ldots, \beta_p$ are unknown coefficients need to be estimated. In this paper, $p = 3$ and the parameter α is taken to 1 which is automatically estimated by the package "pscl" in R .

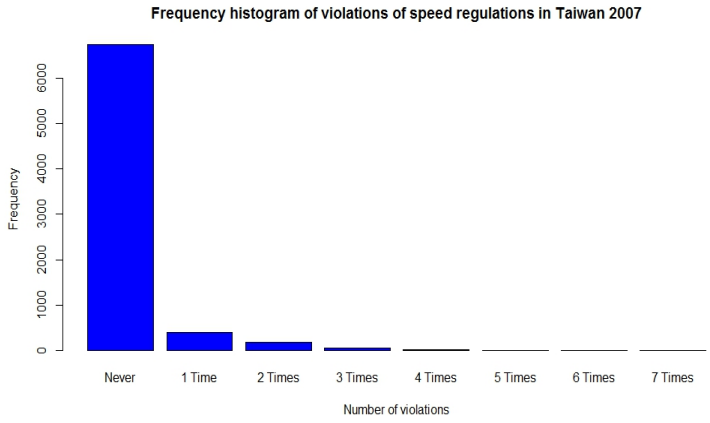

Fig. 1. Frequency of violations of speed regulations in Taiwan 2007.

Count Coeff.	Estimate	Std. Error	z value	$\Pr(>\|z\|)$
Intercept	0.31028	0.37714	0.823	0.41067
X	0.07804	0.07061	1.105	0.26904
Z	0.11969	0.08031	1.490	0.13614
W	-0.23536	0.07102	-3.314	0.00092
Zero-inflated	Estimate	Std. Error	z value	$\Pr(>\|z\|)$
Intercept	4.16321	0.50823	8.192	2.58e-16
X	-0.31510	0.09374	-3.362	0.000775
Z	-1.25280	0.14906	-8.405	$<$ 2e-16
W	0.19547	0.08847	2.209	0.027152

Tab. 2: Estimates of the model M_1 (ZIP model).

	Estimate	Std. Error	z value	$\Pr(>\|z\|)$
Intercept	-3.07651	0.22958	-13.401	$<$2e-16
X	0.29910	0.04154	7.200	6e-13
Z	0.88207	0.04166	21.174	$<$2e -16
W	-0.36958	0.03918	-9.433	$<$2e-16

Tab. 3: Estimates of the model M_2 (Poisson model).

	Estimate	Std. Error	z value	$\Pr(>\|z\|)$
Intercept	-3.23072	0.29809	-10.838	$<$ 2e-16
X	0.34635	0.05465	6.337	2.34e-10
Z	0.98898	0.06208	15.931	$<$ 2e-16
W	-0.42228	0.05024	-8.405	$<$ 2e-16

Tab. 4: Estimates of the model M_3 (NB model).

Data Set	M_1	M_2	M_3
Training	2869.835	3181.406	2943.556
Validation	1013.404	1155.584	1028.383
Test	918.2253	1025.232	945.857

Tab. 5: Results of AIC.

Data Set	M_1	M_2	M_3
Training	2894.741	3206.312	2968.462
Validation	1033.937	1176.118	1048.916
Test	938.8124	1045.819	966.4441

Tab. 6: Results of BIC.

Model	**V**	P-value	Preference
M_1 vs M_2	6.03	8.44e-10	M_1
M_1 vs M_3	4.11	1.99e-05	M_1
M_2 vs M_3	-5.14	1.00	M_3

Tab. 7: Results of Vuong's test on training data.

Model	**V**	P-value	Preference
M_1 vs M_2	5.37	5.42e-06	M_1
M_1 vs M_3	1.28	0.10	$M_1 \approx M_3$
M_2 vs M_3	-5.01	1.00	M_3

Tab. 8: Results of Vuong's test on validation data.

Model	**V**	P-value	Preference
M_1 vs M_2	4.40	4.03e-08	M_1
M_1 vs M_3	2.83	0.023	M_1
M_2 vs M_3	-3.79	1.00	M_3

Tab. 9: Results of Vuong's test on test data.

Data Set	M_1	M_2	M_3
Training	0.3145	0.3100	0.3145
Validation	0.3488	0.3416	0.3488
Test	0.2811	0.2827	0.2827

Tab. 10: Results of MSE.

Data Set	M_1	M_2	M_3
Training	0.8968	0.8933	0.8968
Validation	0.9042	0.8986	0.9042
Test	0.9063	0.9047	0.9047

Tab. 11: Results of accuracy.