Pol. J. Environ. Stud. Vol. 21, No. 3 (2012), 565-568

Original Research

Zero-Inflated Regression Models for Modeling the Effect of air Pollutants on Hospital Admissions

Mehmet Ali Cengiz*

Department of Statistics, University of Ondokuz Mayıs,
Istatıstık Bölümü, 55139 Atakum-Samsun, Turkey

Received: 13 May 2011
Accepted: 31 October 2011

Abstract

Count regression methods are the fundamental tool used for modeling the association between environmental pollution and hospital admissions. Data with many zeros are often encountered in count regression models. Failure to account for the extra zeros may result in biased parameter estimates and misleading inferences. Zero-inflated Poisson and zero-inflated negative binomial regression models have been proposed for situations where the data generating process results in too many zeros.

Keywords: count regression, zero-inflated models, air pollution

Introduction

Many environmental studies often involve the analysis of count data, such as the number of hospitalizations caused by air pollution, where Poisson Regression (PR) is the standard basic technique. However, overdispersion is widely seen in this regression model. Overdispersion in this model occurs when the response variance is greater than the mean. This may cause standard errors of the estimates to be deflated or underestimated. The Negative Binomial Regression model is a generalization of the Poisson regression model that allows for overdispersion by introducing an unobserved heterogeneity term.

The Generalized Poisson Regression (GPR) model developed by [1] is used to model dispersed count data to handle the overdispersion problem. Refer to [2-4] for a good overview of the base generalized Poisson model and its derivation. Generalized Poisson is similar to the negative binomial in that it incorporates an extra heterogeneity or dispersion parameter.

Another difficulty occurs when there are excess zeros in the data. Zero-inflated count models were first introduced by [5] to provide a method of accounting for excessive zero counts. A popular approach to the analysis of such data is to use zero-inflated Poisson (ZIP) and zero-inflated Negative Binomial (ZINB) regression models. Recent developments have discussed extending the Poisson or negative binomial distributions into models that account for extra zeros [6-8]. Some studies have used extensive ecological datasets to compare the performance of these distributions for a variety of environmental conditions [9-15].

Many of the existing studies have focused on the association between environmental pollution and hospital admissions for Chronic obstructive pulmonary disease (COPD), a group of diseases characterized by airflow obstruction that can be associated with breathing-related symptoms (e.g., cough, exertional dyspnea, expectoration, and wheeze). There is increasing interest in the use of hospital admission data in studies of short-term exposure effects attributed to air pollutants. Numerous studies have investigated the relationship between air pollution and hospital admissions for COPD [16-22].

While adverse effects of exposure to air pollutants and hospital admissions for COPD are well studied, little is known about the effect of air pollutants on COPD symptoms. This study focuses on modeling air pollution and both

*e-mail: macengiz@omu.edu.tr

hospital admissions and symptoms (assuming consecutive outcomes are independent) for COPD using different count regression models, and compares these models using Akaike Information Criteria (AIC) and Bayesian Information Criteria (BIC). The AIC introduced by [23] is a measure of the relative goodness of fit of a statistical model. The BIC developed by [24] is a criterion for model selection among a finite set of models. It is based, in part, on the likelihood function, and it is closely related to the AIC.

Methods

The most used count regression model is Poisson Regression (PR), the model may be expressed as follows:

$$f(y_i \mid x_i) = \frac{e^{-\mu_i}\mu_i^{y_{ii}}}{y_i!} \quad (1)$$

...with $\mu_i = \exp(x_i'\beta)$, where x_i is a covariate vector and β is a vector of unknown regression coefficients.

Negative Binomial Regression (NBR) is given as:

$$f(y_i \mid x_i) = \begin{pmatrix} y_i + \frac{1}{\alpha} - 1 \\ \frac{1}{\alpha} - 1 \end{pmatrix} \left(\frac{1}{1+\alpha\mu_i}\right)^{1/\alpha} \left(\frac{\alpha\mu_i}{1+\alpha\mu_i}\right)^{y_{ii}} \quad (2)$$

...where α is an ancillary parameter indicating the degree of overdispersion.

There are several different models that are referred to as Generalized Poisson Regression (GPR) models. GPR model is given as in [25, 26].

$$f(y_i \mid x_i) = \left(\frac{\mu_i}{1+\alpha\mu_i}\right)^{y_{ii}} \frac{(1+\alpha y_i)^{y_{ii}-1}}{y_i!} \exp\left[\frac{-\mu_i(1+\alpha y_i)}{1+\alpha\mu_i}\right] \quad (3)$$

To account for an extra amount of zeros, the zero-inflated Poisson Regression (ZIPR) model is given:

$$f(y_i \mid x_i) = \begin{cases} \pi_i + (1-\pi_i)\exp(-\mu_i) & \text{for } y_i = 0 \\ (1-\pi_i)\dfrac{\exp(-\mu_i)\mu_i^{y_i}}{y_i} & \text{for } y_i \geq 1 \end{cases} \quad (4)$$

...where π is the probability of being an extra zero.

Another zero-inflated count regression model is the zero-inflated Binomial Regression (ZIBR) model:

$$f(y_i \mid x_i) = \begin{cases} \pi_i + (1-\pi_i)\left(\dfrac{1}{1+\alpha\mu_i}\right)^{\frac{1}{\alpha}} & \text{for } y_i\text{=0} \\ (1-\pi_i)\dfrac{\Gamma\left(\frac{1+\alpha\mu_i}{\alpha}\right)}{\Gamma\left(\frac{1}{\alpha}\right) y_i!}\left(\dfrac{1}{1+\alpha\mu_i}\right)^{\frac{1}{\alpha}}\left(\dfrac{\alpha\mu_i}{1+\alpha\mu_i}\right)^{y_i} & \text{for } y_i\text{≥0} \end{cases} \quad (5)$$

...where α is the overdispersion parameter.

Application

This study presents the relations between the numbers of admissions with respiratory disease who applied to Afyon Respiratory Disease Hospital and the measures of air pollution at the city center. This study was performed by retrospective evaluation of the patients' records between 1 October 2006 and 30 September 2010. SO_2 (sulfur dioxide) – PM_{10} (particulate matter) values related with

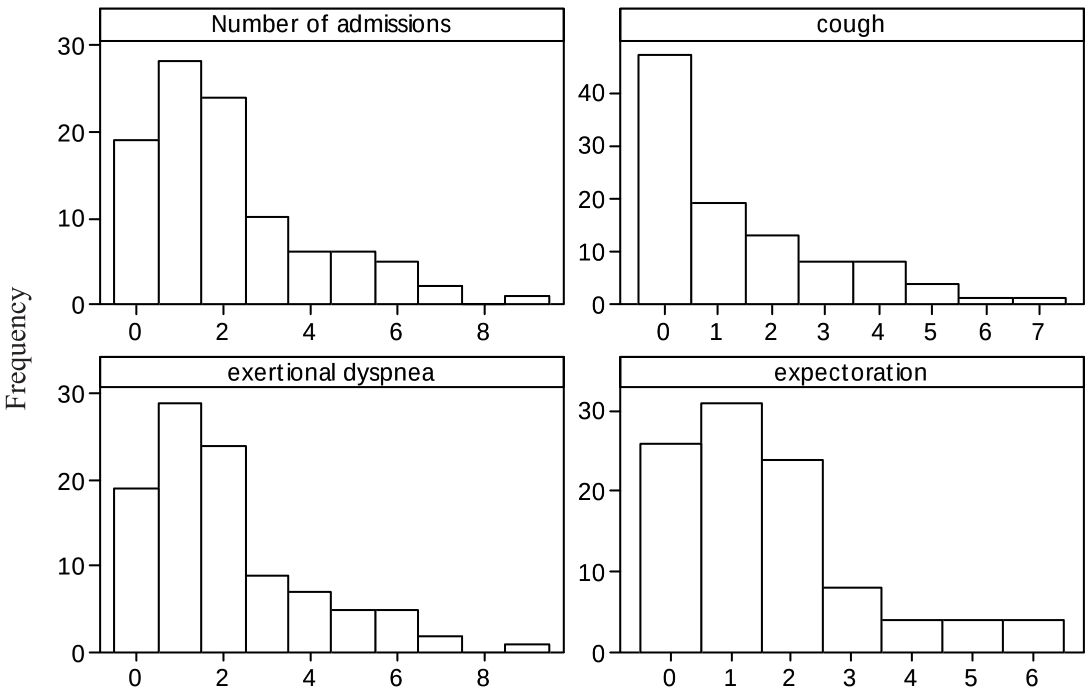

Fig. 1. Distributions of the numbers of the overall admissions, patients with cough, patients with exertional dyspnea, and patients with expectoration for COPD.

Table 1. Results of the analysis of parameter estimates using different count regression models.

Model	Parameter	PR		NBR		GPR		ZIPR		ZINBR	
		Est.	P	Est.	P	Est.	P	Est.	P	Est.	P
Number of overall admissions	Intercept	0.151	0.404	0.129	0.578	0.154	0.412	0.459	0.011	0.454	0.018
	SO_2	0.002	0.281	0.002	0.380	0.003	0.215	0.002	0.025	0.002	0.150
	PM_{10}	0.004	0.001	0.003	0.011	0.003	0.017	0.003	0.029	0.003	0.035
Number of patients with cough	Intercept	-0.111	0.589	-0.130	0.595	-0.114	0.567	0.438	0.037	0.423	0.041
	SO_2	0.001	0.315	0.001	0.396	0.002	0.322	0.018	0.032	0.018	0.025
	PM_{10}	0.004	0.004	0.005	0.016	0.005	0.003	0.018	0.08	0.019	0.018
Number of patients with exertional dyspnea	Intercept	0.121	0.589	0.143	0.395	0.134	0.567	0.354	0.039	0.387	0.045
	SO_2	0.006	0.215	0.001	0.396	0.004	0.056	0.028	0.062	0.048	0.068
	PM_{10}	0.009	0.012	0.005	0.016	0.005	0.003	0.013	0.024	0.019	0.021
Number of patients with expectoration	Intercept	0.134	0.489	0.122	0.268	0.145	0.452	0.281	0.023	0.299	0.039
	SO_2	0.007	0.243	0.008	0.314	0.006	0.061	0.032	0.064	0.035	0.06
	PM_{10}	0.014	0.009	0.005	0.021	0.004	0.007	0.021	0.017	0.026	0.017

Table 2. Result for model comparisons.

Model	PR		NBR		GPR		ZIPR		ZINBR	
	AIC	BIC	AIC	BIC	AIC	BIC	AIC	BIC	AIC	BIC
Number of overall admissions	390.76	396.61	389.21	396.43	391.26	394.25	393.26	397.25	393.123	394.12
Number of patients with cough	387.31	391.16	392.23	351.23	386.14	390.09	241.11	254.23	240.32	253.16
Number of patients with exertional dyspnea	389.24	393.37	387.12	396.19	388.78	392.15	317.56	323.12	315.21	321.80
Number of patients with expectoration	389.41	390.26	388.27	390.12	388.32	391.05	329.44	335.35	328.121	333.96

the same period were extracted from the archives of the Afyon Environmental Department Air Pollution Unit. Only PM_{10} and SO_2 values were used as air pollutants since only those values were monitored by the station in Afyon Province at that period. Weekly records of hospital admissions for COPD were obtained at Afyon State Hospital for the period from October 2006 to September 2010. Meanwhile, the number of patients with cough, exertional dyspnea, and expectoration for COPD were obtained from the patients' records of the same period.

Results

Distributions of the numbers of overall admissions, of patients with cough, of patients with exertional dyspnea, and patients with expectoration for COPD are given in Fig. 1.

Weekly counts of hospital admissions and the numbers of patients with cough, with exertional dyspnea, and with expectoration for COPD were considered as the dependent variable, and SO_2 – PM_{10} values were considered as independent variables in different count regression models. Results of the analysis of parameter estimates were given in Table 1.

For model comparison, the AIC, BIC for all count regression models used were calculated for all models. The results are shown in Table 2.

Conclusions

In this study, all models used for the numbers of admissions, of patients with exertional dyspnea, and of patients with expectoration give similar results in terms of parameter estimations. Looking at Table 1, it is easy to say that the effect of PM_{10} on hospital admissions, number of patients with exertional dyspnea, and number of patients with expectoration is highly significant ($p<0.05$), whereas the effect of SO_2 on hospital admissions, number of patients with exertional dyspnea, and number of patients with expectoration is insignificant at the 5% level.

However, the models used for the number of patients with cough give different results from the others in terms of parameter estimations. The effects of both PM_{10} and SO_2 on the number of patients with cough is highly significant ($p<0.05$).

The results in Table 2 show that the AIC and BIC values of PR, NBR, and GPR models for the number of patients with cough is dramatically larger than the AIC and BIC values of the zero inflated models (ZINBR and ZIPR). Since the number of patients with cough contains more zeros than the other datasets as shown in Fig. 1, the zero inflated models (ZINBR and ZIPR) give better results than the other models in terms of parameter estimations and model comparisons.

In this study, only two air pollutant (PM_{10} and SO_2) measurements were used in the absence of the other measurements. As a future work, the power of the study can be increased by the addition of other measures.

References

1. CONSUL P., JAIN G. A Generalization of the Poisson distribution, Technometrics. **15**, 791, **1973**.
2. CONSUL P., FAMOYE F. Generalized Poisson regression model, Commun. Stat. Theory. **21**, 89, **1992**.
3. FAMOYE F., SINGH K.P. On inflated generalized Poisson regression model. Advances and Applications in Statistics, **3**, 135, **2003**.
4. BAE S., FAMOYE F., WULU J.T., BATOLUCCİ A.A., SINGH K.P., A rich family of generalized Poisson regression models with applications. Mathematical and Computers in Simulation. **69**, (1-2), 4, **2005**.
5. LAMBERT D. Zero-inflated Poisson regression with an application to defects in manufacturing, Technometrics, **34**, 1, **1992**.
6. MARTIN T.G., WINTLE B.A., RHODES J.R., KUHNERT P.M., FIELD S.A., LOW-CHOY S.J., TYRE A.J., POSSINGHAM H.P. Zero tolerance ecology: improving ecological inference by modelling the source of zero observations. Ecol. Lett. **8**, 1235, **2005**.
7. WENGER S.J., FREEMAN M.C. Estimating species occurrence, abundance, and detection probability using zero-inflated distributions. Ecology, **89**, 2953, **2008**.
8. HILBE J.M. Negative Binomial Regressions, Second Edition, Cambridge University Press, **2011**.
9. AGARWAL D.K., GELFAND A.E., POUSTY S.C. Zero inflated models with application to spatial count data, Environ. Ecol. Stat., **9**, 342, **2002**.
10. WARTON D.I. Many zeros does not mean zero inflation: comparing the goodness-of-fit of parametric models to multivariate abundance data. Environmetrics, **16**, 275, **2005**.
11. POTTS J.M., ELITH J. Comparing species abundance models. Ecol. Model. **199**, 153, **2006**.
12. LEWIN W., FREYHOF J., HUCKSTORF V., MEHNER T., WOLTER C. When no catches matter: coping with zeros in environmental assessments. Ecological Indicators **10**, 572, **2009**.
13. EL-SHAARAWI A.H., ZHU R., JOE H. Modelling species abundance using the Poisson-Tweedie family. Environmetrics. **22**, 152, **2011**.
14. MELLES S.J., FORTIN M.J., LINDSAY K., BADZINSKI D., MELLES S.J. Expanding northward: influence of climate change, forest connectivity, and population processes on a threatened species' range shift. Glob. Change Biol., **17**, 17, **2010**.
15. VAUDOR L., LAMOUROUX N., OLIVIER J.M. Comparing distribution models for small samples of overdispersed counts of freshwater fish. Acta Oecol., **37**, 170, **2011**.
16. ANDERSON H.R., SPIX C., MEDINA S., SCHOUTEN J.P., CASTELLSAGUE J., ROSSI G., ZMIROU D., TOULOUMI G., WOJTYNIAK B., PONKA A., BACHAROVA L., SCHWARTZ J., KATSOUYANNI K. Air pollution and daily admissions for chronic obstructive pulmonary disease in 6 European cities: results from the aphea project. Eur. Respir. J. **10**, 1064, **1997**.
17. BURNETT R.T., DALES R., KREWSKI D., VINCENT R., DANN T., BROOK J.R. Associations between ambient particulate sulphate and admissions to Ontario Hospitals for cardiac and respiratory diseases. Am. J. Epidemiol. **142**, 15, **1995**.
18. CHEN Y., YANG W., JENNISON B.L., OMAYE S.T. Air particulate pollution and hospital admissions for chronic obstructive pulmonary disease in reno, Nevada. Inhal. Toxicol. **12**, 281, **2000**.
19. CHEN Y., YANG Q., KREWSKI D., SHI Y., BURNETT R.T., MCGRAIL K. Influence of relatively low level of particulate air pollution on hospitalization for coped in elderly people. Inhal. Toxicol. **16**, 21, **2004**.
20. DOMINICI F., PENG R.D., BELL M.L. Fine particulate air pollution and hospital admission for cardiovascular and respiratory diseases. JAMA. **295**, 1127, **2006**.
21. YANG C., CHEN C. Air pollution and hospital admissions for chronic obstructive pulmonary disease in a subtropical city: Taipei, Taiwan. J Toxicol. Env. Heal A, **70**, 1214, **2007**.
22. TERZI Y., CENGIZ M.A. Using of generalized additive model for model selection in multiple Poisson regression for air pollution data. Scientific Research and Essay. **4**, (9), 867, **2009**.
23. AKAIKE H., A new look at the statistical model identification, IEEE T. Automat. Contr., **19**, (6), 716, **1974**.
24. SCHWARZ G.E., Estimating the dimension of a model, Ann. Stat. **6**, (2), 461, **1978**.
25. FAMOYE F., SINGH K.P. Zero-truncated generalized Poisson regression model with an application to domestic violence. Journal of Data Science. **4**, 117, **2006**.
26. WINKELMANN R. Econometric Analysis of Count Data, Fifth Edition, New York: Springer, **2008**.