Australian Journal of Basic and Applied Sciences, 5(7): 325-331, 2011
ISSN 1991-8178

A Missing Inflated Power Series Model for regression analysis of the British Household Panel Survey (BHPS) data

E. Bahrami Samani

Department of Statistics, Faculty of Mathematical Science, Shahid Beheshti University, Tehran, 1983963113, Iran.

Abstract: A new class of Missing Inflated Power Series Model (MIPS) is proposed. The Missing Inflated Power Series Distribution (MIPSD) contains two parameters. The first parameter indicates inflation of missing data and the other parameter is that of the Power Series distribution. A full likelihood-based approach is used to obtain maximum likelihood estimates of the model parameters. To illustrate the application of such modelling the real data is analyzed.

Key words: Missing inflated model; Power serise model; Maximum Likelihood; EM algorithm; Missing mechanism.

INTRODUCTION

In applications, count variables with many missing data are very important in various scientific fields including. The problem of a high proportion of missing data have been an interest in data analysis and modeling. There Models having significantly more observed numbers of missing data are known as missing inflated models.

Rubin (1976), Little and Rubin (2002), Diggle and Kenward (1994) made important distinctions between the various types of missing mechanisms for each of the above mentioned patterns. They define the missing mechanism as missing completely at random (MCAR) if missingness is dependent neither on the observed responses nor on the missing responses, and missing at random (MAR) if, given the observed responses, it is not dependent on the missing responses. Missingness is defined as non-random if it depends on the unobserved responses. From a likelihood point of view MCAR and MAR are ignorable but not missing at random (NMAR) is non-ignorable.

The aim of this paper is to use an approach similar to that of Lambert (1992). That he described zero-inflated Poisson (ZIP) regression, a class of models for count data with excess zeros. In a ZIP model, a count response variable is assumed to be distributed as a mixture of a Poisson distribution and a distribution with point mass of one at zero, with mixing probability π. Gupta *et al*., (1996) have studied zero inflated modified power series distribution along with applications of the same for simulated data. Vandenbrock (1995) discussed a score test for testing a Poisson distribution against ZIP distribution. Yip (1988) has described an inflated Poisson distribution dealing with the number of insects per leaf. Gupta *et al*., (2004) applied a test based on asymptotic likelihood theory. Kale (1998), and Kale and Muralidharan (2000) have reported results on optimal estimating equations for discrete data with higher frequencies at a point and for mixture distributions accommodating instantaneous or early failures, respectively. Xie *et al*., (2001) have reported the use of a ZIP distribution in statistical process control. Also they have studied the performance of various tests for testing a Poisson distribution against the zero inflated Poisson alternative. Murat and Szynal (2003) studied the moments of certain deformed probability distributions. Gupta *et al*., (2004) discussed the score test for the zero inflated generalized Poisson regression model. Min and Agresti (2005) studied a random effect model for repeated measures of zero inflated count data. Patil and Shirke (2007) studied testing a parameter of the power series distribution of a zero inflated power series model. Xue-Dong and Ying-Zi (2010) studied model selection for zero - inflated regression with missing covariates.

In this paper, we proposed Missing Inflated Power Series regression models. We provide missing inflated power series (MIPS) regression in section 2, including a discusion of maximum likelihood estimation of these models. In section 3, we have a simulation study. In section 4, we used to the example data set to illustrate the new model of this paper. Finally, concluding remarks are given.

Corresponding Author: E. Bahrami Samani, Department of Statistics, Faculty of Mathematical Science, Shahid Beheshti University, Tehran, 1983963113, Iran.

2. Mips - Regression Model:

In missing inflated Power Series (MIPS) regression model, the responses $Y_1,...,Y_n$ are independent and

$$Y_i \sim \begin{cases} missing, & with\ probability\ \ \pi_i \\ h(y_i \mid \theta_i), & with\ probability\ \ 1-\pi_i \end{cases}$$

where $h(y_i \mid \theta_i)$ is a member of the power series distribution (PSD) with general form

$$h(y_i \mid \theta_i) = \frac{b_{y_i}\theta_i^{y_i}}{f(\theta_i)}, \quad y_i = 0,1,2,...$$

where $b_{y_i} > 0$, θ_i and $f(\theta_i) = \sum_{y_i=0}^{\infty} b_{y_i}\theta_i^{y_i}$ are positive, finite and differentiable function of θ_i and $0 < \pi_i < 1$. So, there are a large number of missing counts and a mixture assigning a mass of π_i=$P(\{y_i$ is missing $|Y_i\})$ to the extra missing counts and a mass of 1=π_i to the unimodel Power Series distribution is used.

Obviously, Poisson distribution [$P(\lambda)$] and negative binomial [$NB(\theta, \sigma)$] distribution belong to power Series distribution with $b_y = \frac{1}{y!}$, $f(\theta) = e^{\theta}$ and $b_y = \frac{\Gamma(y+\sigma)}{\Gamma(y+1)}$, $f(\theta) = \Gamma(\sigma)(1-\theta)^{-\sigma}$ respectively, where σ =0 is dispersion parameter. The MIPS includes a lot of common distribution such as missing inflated Poisson (MIP) and missing inflated negative binomial (MINB).

We model $\theta = (\theta_1,...,\theta_n)'$ and $\pi = (\pi_1,...,\pi_n)'$ with log-linear and logistic regression models,

$$\log(\theta_i) = G_i'\beta,$$
$$logit(\pi_i) = B_{i'}\gamma + \eta_1 Y_{i1}^* + \eta_2 Y_{i2}^*$$

where β and γ are vectors of regression parameters associated with covariates G and B. η_1 and η_2 are the non-ignorable missing parameters and $Y_{i1}^* = 1$ if y_i is missing and $Y_{i1}^* = 0$ otherwise, $Y_{i2}^* = Y_i$ if Y_i is observed and $Y_{i2}^* = 0$ otherwise.

The log likelihood for regression parameters β, γ, η_1 and η_2 based on all of the data is given by

$$\log L(\beta,\gamma,\eta_1,\eta_2;y)$$
$$= \sum_{i=1}^{n} a_i \log\left(\frac{exp[B_{i'}\gamma + \eta_1 Y_{i1}^* + \eta_2 Y_{i2}^*]}{1+exp[B_{i'}\gamma + \eta_1 Y_{i1}^* + \eta_2 Y_{i2}^*]} + \frac{b_0}{f(exp[G_{i'}\beta])}\frac{1}{1+exp[B_{i'}\gamma + \eta_1 Y_{i1}^* + \eta_2 Y_{i2}^*]}\right)$$
$$+ \sum_{i=1}^{n}(1-a_i)(B_{i'}\gamma + \eta_1 Y_{i1}^* + \eta_2 Y_{i2}^* - \log(1+exp[B_{i'}\gamma + \eta_1 Y_{i1}^* + \eta_2 Y_{i2}^*]))$$
$$+ \sum_{i=1}^{n}(1-a_i)b_{y_i} + \sum_{i=1}^{n}(1-a_i)y_i G_{i'}\beta - \sum_{i=1}^{n}(1-a_i)\log(f(exp[G_{i'}\beta]))$$

where G_i and B_i are the i th rows of G and B and α_i=1 y_1 if is missing and α_i=1 if y_1 is observed. The sum of exponentials in the likelihood function complicates the maximization of $l(\gamma,\beta,\eta_1,\eta_2;y)$. As described in Lambert (1992), the ZIP model can be fit using maximum likelihood via the EM algorithm. I define the unobserved random variable *Zi*=1 when *Yi* is generated from the missing state and *Zi*=0 when *Yi* comes from the Power Series state. If we could observe Z=$(Z_1,...,Z_n)$' then, the complete-data (*Y,Z*) log likelihood would be

$$l_c(\gamma,\beta,\eta;y,z)=\sum_{i=1}^{n}\log[P(Y_i=y_i,Z_i=z_i)]=l_c(\gamma,\eta_1,\eta_2;y,z)+l_c(\beta;y,z)$$

where,

$$l_c(\gamma,\eta;y,z)=\sum_{i=1}^{n}[z_i(B_i'\gamma+\eta_1Y_{i1}^*+\eta_2Y_{i2}^*)-\log(1+\exp(B_{i'}\gamma+\eta_1Y_{i1}^*+\eta_2Y_{i2}^*))]$$

$$l_c(\beta;y,z)=\sum_{i=1}^{n}(1-z_i)[\log b_{y_i}+y_iG_{i'}\beta-\log f(\exp(G_{i'}\beta)].$$

This log-likelihood is easy to maximize, because $l_c(\beta;y,z)$ and $l_c(\gamma,\eta_1,\eta_2;y,z)$ can be maximized separately. The EM algorithm can be used to maximize $l(\gamma, \eta_1, \eta_2, \beta; y)$ by alternating between an E step, in which the missing data Z is estimated by its expectation under the current estimates of $(\gamma,\eta_1,\eta_2,\beta)$ and the expectations Zi,Yi, Zi,Yi^*_1 and Zi,Yi^*_2 given Yi and under the current estimates of $(\gamma, \eta_1, \eta_2, \beta)$. a maximization step, in which the expectation $l(\gamma, \eta_1, \eta_2, \beta; y)$ evaluated at the current (fixed) estimate of z and the expectations $ZiYi$, Zi,Yi^*_1 and Zi,Yi^*_2 given Yi with respect to γ, η_1, η_2 and β. In more detail, iteration $(K+1)$ of the EM algorithm requires three steps.

(1) E -Step. (1.a) Estimate Zi by its posterior mean $Zi^{(K)}$ under the current estimates $\gamma^{(K)}$, $\eta_1^{(K)}$, $\eta_2^{(K)}$, and $\beta^{(K)}$. Here

$$Z_i^{(k)}=E[Z_i\mid Y_i]=P[Z_i=1\mid y_i,\gamma^{(k)},\eta_1^{(k)},\eta_2^{(k)},\beta^{(k)}]$$

$$=\begin{cases}[1+\dfrac{b_0\exp(-B_{i'}\gamma-\eta_1Y_{i1}^*-\eta_2Y_{i2}^*)}{f(\exp(G_{i'}\beta))}]^{-1} & y_i \text{ is missing}\\ 0 & y_i \text{ is observed}\end{cases}$$

(1.b) We calculate the expectations Zi,Yi, Zi,Yi^*_1 and Zi,Yi^*_2 given Yi with respect to γ, η_1, η_2 and β.

$$E(Z_iY_i\mid Y_i,\gamma^{(k)},\eta_1^{(k)},\eta_2^{(k)},\beta^{(k)}]=\begin{cases}\sum_{i=1}^{n}y_iE(Z_i\mid Y_i,\gamma^{(k)},\eta_1^{(k)},\eta_2^{(k)},\beta^{(k)}] & y_i \text{ is missing}\\ 0 & y_i \text{ is observed}\end{cases}$$

$$E(Z_iY_{i1}^*\mid Y_i,\gamma^{(k)},\eta_1^{(k)},\eta_2^{(k)},\beta^{(k)}]=\begin{cases}\sum_{i=1}^{n}E(Z_iY_i\mid Y_i,\gamma^{(k)},\eta_1^{(k)},\eta_2^{(k)},\beta^{(k)}] & y_i \text{ is missing}\\ 0 & y_i \text{ is observed}\end{cases}$$

$$E(Z_iY_{i2}^*\mid Y_i,\gamma^{(k)},\eta_1^{(k)},\eta_2^{(k)},\beta^{(k)}]=\begin{cases}\pi_i & y_i \text{ is missing}\\ 0 & y_i \text{ is observed}\end{cases}$$

(2) M- step for γ, η_1 and η_2 . Maximize

$$E[l_c(\gamma,\eta;y,z)\mid Y_i] = \sum_{i=1}^{n}[E[z_iB_i'\gamma\mid Y_i] + \eta_1 E[Y_{i1}^*\mid Y_i] + \eta_2 E[Y_{i2}^*\mid Y_i])$$

$$-E[\log(1+\exp(B_{i'}\gamma + \eta_1 Y_{i1}^* + \eta_2 Y_{i2}^*))\mid Y_i]]$$

where,

$$E[\log(1+\exp(B_{i'}\gamma + \eta_1 Y_{i1}^* + \eta_2 Y_{i2}^*)\mid Y_i] = \begin{cases} \log(1+\exp(B_{i'}\gamma+\eta_1)) & y_i \textit{ is missing} \\ \log(1+\exp(B_{i'}\gamma+\eta_2 y_i)) & y_i \textit{ is observed} \end{cases}$$

(3) M- step for β. Find $\beta^{(K+1)}$ by maximizing,

$$E[l_c(\beta;y,z^{(k)})\mid Y_i] = \sum_{i=1}^{n} E[(1-z_i)[\log b_{y_i} + y_i G_{i'}\beta - \log f(\exp(G_{i'}\beta)]\mid Y_i].$$

where,

$$E[\log b_{Y_i} + Y_i G_{i'}\beta - \log f(\exp(G_{i'}\beta)\mid Y_i] =$$

$$\begin{cases} E[\log b_{Y_i}] + E[Y_i]G_{i'}\beta - \log f(\exp(G_{i'}\beta) & y_i \textit{ is missing} \\ \log b_{y_i} + y_i G_{i'}\beta - \log f(\exp(G_{i'}\beta) & y_i \textit{ is observed} \end{cases}$$

and $E[Z_i \log b_{Y_i} \mid Y_i]$ as a function of $.E[Z_iY_i \mid Y_i]$

3. Simulation Study:

In this section, simulation study is used to illustrate the application of our proposed model. Firstly, two data sets are generated from the following cases:

Case (I):

$$Y_i \sim \begin{cases} missing, & with\ probability\ \ \pi_i \\ P(\theta_i), & with\ probability\ \ 1-\pi_i \end{cases}$$

Case (II):

$$Y_i \sim \begin{cases} missing, & with\ probability\ \ \pi_i \\ NB(\theta_i,\sigma), & with\ probability\ \ 1-\pi_i \end{cases}$$

where $\log(\theta_i) = \beta_0 + \beta_1 X$ and $logit(\pi_i) = \gamma_0 + \gamma_1 X + \eta_1 Y_{i1}^* + \eta_2 Y_{i2}^*$.

In the simulation study, I set the true value of the parameters to be $\beta = (\beta_0, \beta_1) = (0,1)$ $\gamma = (\gamma_0, \gamma_1) = (0,1)$ and the Poisson means θi range from 2. I set the dispersion parameter σ = 0.5. Throughout, there is one covariate X taking on n uniformly spaced values between 0 and 1. The response Y was obtained by first generating a uniform (0, 1) random vector U of length n and then assigning Yi is missing if $Ui < \pi i$ and y_1~Poisson(θi) for case (I) and y_1~NB (θi,σ) for case (II), otherwise.

Missing data are generated from the following missingness mechanism:

$$logit(\pi_i) = \gamma_0 + \gamma_1 X + \eta_1 Y_{i1}^* + \eta_2 Y_{i2}^*$$

with true parameters $\gamma = (\gamma_0, \gamma_1, \eta_1, \eta_2)$'=(0,1,1,1)' , which implies that this missingness mechanism is NMAR. The response *Y* was obtained by second generating a *uniform*=(0,1) random vector *V* of length *n* and then assigning *Yi* is missing if $Vi<\pi i$ and *Yi* is observed, otherwise. The average proportions of missing data generated in this way are about 25% (for Poisson model) and 24% (for NB model) and we take sample size *n* to be 50, 100 and 1000.

Missing data are generated from the following missingness mechanism:

$$logit(\pi_i) = \gamma_0 + \gamma_1 X$$

with true parameters γ $(\gamma_0, \gamma_1,)$'=(0,1)', which implies that this missingness mechanism is MAR. The response *Y* was obtained by second generating a *uniform*=(0,1) random vector *V* of length *n* and then assigning *Yi* is missing if $Vi<\pi i$ and *Yi* is observed, otherwise. In Table (I) and (II), the parameters estimate by the model (for *n*=50, *n*=100 and *n*=1000) are close to the true values of the parameters. Of course, the more the value of *n* the better the estimates.

Table 1: Results of the simulation study assumption of NMAR.

Model		MIP					
		n=50		n=100		n=1000	
Parameter	True value	Est.	S.D	Est.	S.D.	Est.	S.D.
β_0	0.000	0.044	0.021	0.011	0.012	0.004	0.006
β_1	1.000	1.178	0.127	1.017	0.064	1.006	0.005
γ_0	0.000	0.048	0.088	0.035	0.067	0.003	0.002
γ_1	1.000	1.169	0.111	1.039	0.010	1.001	0.007
η_1	1.000	1.153	0.156	1.012	0.086	1.009	0.008
η_2	1.000	1.176	0.145	1.056	0.098	1.003	0.009
Model		MINB					
Parameter	True value	Est.	S.D	Est.	S.D.	Est.	S.D.
β_0	0.000	0.091	0.049	0.045	0.013	0.001	0.008
β_1	1.000	1.106	0.106	1.067	0.049	1.011	0.007
γ_0	0.000	0.066	0.094	0.023	0.028	0.010	0.011
γ_1	1.000	1.111	0.178	1.032	0.079	1.021	0.019
η_1	1.000	1.178	0.139	1.049	0.099	1.057	0.012
η_2	1.000	1.161	0.199	1.079	0.046	1.011	0.010
σ	0.500	0.450	0.133	0.485	0.087	0.503	0.006

Table 2: Results of the simulation study assumption of MAR.

Model		MIP					
		n=50		n=100		n=1000	
Parameter	True value	Est.	S.D	Est.	S.D.	Est.	S.D.
β_0	0.000	0.054	0.022	0.021	0.013	0.006	0.002
β_1	1.000	1.133	0.134	1.017	0.084	1.009	0.001
γ_0	0.000	0.051	0.046	0.013	0.014	0.003	0.001
γ_1	1.000	1.105	0.166	1.016	0.032	1.001	0.009
Model		MINB					
Parameter	True value	Est.	S.D	Est.	S.D.	Est.	S.D.
β_0	0.000	0.060	0.033	0.021	0.041	0.001	0.001
β_1	1.000	1.144	0.138	1.081	0.092	1.018	0.014
γ_0	0.000	0.075	0.042	0.041	0.015	0.004	0.003
γ_1	1.000	1.179	0.104	1.065	0.035	1.003	0.006
σ	0.500	0.470	0.166	0.495	0.086	0.501	0.005

4. Application:

4.1. Data

The data used in this paper is excerpted from the 13th wave (2003) of the British Household Panel Survey (BHPS). The selected variables which will be used in this application are explaind in the following. Here the variable of interest is the number visits to patients in past year (NV), [where the related question is Q: ``: And since September 1st 2002, approximately how many times have you attended a hospital or clinic as an patient or day patient?"]. Socio-demographic characteristics, namely: Gender (baseline: Male), Marital Status (MS) [MS_1, MS_2 and MS_3 are dummy variables for married or living as couple, widowed and divorced or separated and never married (baseline)], Age and Highest Educational Qualification (HEQ) [HEQ_1, HEQ_2, HEQ_3 are dummy variables for higher or first degree, other higher QF and other QF, and no qualification(baseline)]. In our application, the percentage of missing values of NV is 10.000 % in the 13th wave. We shall try to find answers for some questions, including: (1) Do male's *NV* differ from female's? (2) How does *HEQ* affect the response? (3) Is the missing mechanism for *NV* at random (MAR)?

We consider the model in section 2 considering for models for *NV* [1: Poisson(θ), 2: NB(θ,σ), 3: MIP(θ,π) and 4: MINB(θ,σ)]. We calculated the values of AIC for these models by EM algorithm, the corresponding results are shown in Table (III).

Table 3: AIC of the real data for all considered models.

Models	Poisson	NB	MIP	MINB
AIC	4140.735	4055.345	3860.411	4019.074

It can be easily seen from Table (III) that model 3 is slected as best by AIC (3860.411). So,

$$NV \sim \begin{cases} missing, & with\ probability\ \ \pi \\ Poisson(\theta), & with\ probability\ \ 1-\pi \end{cases}$$

where

$$\begin{aligned} \log(\theta) &= \beta_0 + \beta_1 MS_1 + \beta_2 MS_2 + \beta_3 MS_3 + \beta_4 Age \\ &\quad + \beta_5 HEQ_1 + \beta_6 HEQ_2 + \beta_7 HEQ_3 + \beta_8 Gender \\ logit(\pi_i) &= \gamma_0 + \gamma_1 MS_1 + \gamma_2 MS_2 + \gamma_3 MS_3 + \gamma_4 Age \\ &\quad + \gamma_5 HEQ_1 + \gamma_6 HEQ_2 + \gamma_7 HEQ_3 + \gamma_8 Gender + \eta_1 NV_1^* + \eta_2 NV_2^*. \end{aligned}$$

where NV^*_1, if *NV* is missing and NV^*_1 otherwise, NV^*_2 =*NV* if *NV* is *observed* and NV^*_2 =0 otherwise.

Mle estimates of parameters and their standard error corresponding to best model with all covariates considered and distributed as MIP are reported in Table (IV). The estimated regression models have the following interpretation: (1) $\hat{\beta}_1$ =1.519, $\hat{\beta}_4$ =0.025 , $\hat{\beta}_5$ =0.0135 and $\hat{\beta}_6$ =0.018 indicate that the marital status (Married or Living as Couple) and the increase of the highest educational qualification (higher or first degree, other higher QF and other QF) have a positive impact on the frequency of number visits to patients. (2) $\hat{\beta}_8$ = -1.055 indicates that the increase of age has a negative impact on the frequency of number visits to patients. (3) The estimates of coefficients $\hat{\gamma}_1$ =0.050 and $\hat{\gamma}_8$ =0.200 are significant different from missing response, which futher implies that our considered missing inflated model for this real data set. (4) The analysis of missing mechanism for the number visits to patients shows a significant Poisson states on the probability of nonresponse which indicates a NMAR evidence ($\hat{\eta}_2$ =0.234).

Table 4: Estimation results of the best model of real data, parameter estimates highlighted in bold are significant at 5 % level).

parameter	Est.	S.D.
β_0	0.505	0.210
β_1	1.519	0.040
β_2	1.006	0.540
β_3	0.087	0.077
β_4	0.025	0.001
β_5	0.013	0.001
β_6	0.018	0.004
β_7	0.019	0.016
β_8	-1.055	0.004
γ_0	0.019	0.009

Table 4: Continue

γ_1	0.050	0.020
γ_2	0.035	0.039
γ_3	0.063	0.086
γ_4	0.063	0.078
γ_5	0.018	0.045
γ_6	0.079	0.043
γ_7	0.076	0.078
γ_8	0.200	0.028
η_1	0.089	0.057
η_2	0.234	0.034

Conclusion:

In this paper missing - inflated Power series regression model to analyze count data. The resulting missing -inflated Power series regression models can be fit using the EM algorithm in a manner similar to that described in Lambertâ€™s (1992) paper. Generalization of our model for random effects into the portion of the MIPS models is an ongoing research on our part.

ACKNOWLEDGMENTS

The author thank the Professor Mojtaba Ganjali for helpful comments and suggestions on earlier versions of the paper.

REFERENCES

Diggle, P.J. and M.G. Kenward, 1994. Informative Drop-out in longitudinal data analysis. Journal of Applied Statistics, 43: 49-93.

Gupta, P.L., R.C. Gupta, R.C. Tripathi, 1996. Analysis of zero-adjusted count data. Comput. Statist. Data Anal., 23: 207-218.

Gupta, P.L., R.L. Gupta, R.C. Tripathi, 2004. Score test for zero inflated generalized Poisson regression model, Communications in Statistics. Theory and Methods., 33(1): 47-64.

Kale, B.K., 1998. Optimal estimating equations for discrete data with higher frequencies at a point, Journal of Indian Statistical Association, 36: 125-136.

Kale, B.K., K. Muralidharan, 2000. Optimal estimating equations in mixture distributions accommodating instantaneous or early failures, Journal of Indian Statistical Association, 38: 317-329.

Lambert, D., 1992. Zero-inflated Poisson regression, with an application to defects in manufacturing. Technometrics, 34: 1-14.

Little, R.J. and D. Rubin, 2002. Statistical analysis with missing data. Second edition. New york, Wiley.

Min, Y.Y., A. Agresti, 2005. Random effect models for repeated measures of zero-inflated count data, Statistical Modeling, 5(1)-19.

Murat, M., D. Szynal, 2003. Moments of certain deformed probability distributions. Communications in Statistics. Theory and Methods, 32(2): 291-313.

Patila, M.K., D.T. Shirkeb, 2007. Testing parameter of the power series distribution of a zero inflated power series model. Statistical Methodology, 4: 393-406.

Rubin, D.B., 1976. Inference and missing data. Biometrica, 82: 669-710.

Vandenbrock, J., 1995. A score test for zero inflation in a Poisson distribution, Biometrics, 51(2): 738-743.

Xie, M., B. He, T.N. Goh, 2001. Zero-inflated Poisson model in statistical process control, Computational Statistics and Data Analysis., 38: 191-201.

Xue-Dong, Ch., F. Ying-Zi, 2010. Model Selection for zero - inflated regression with missing covariates. Computational Statistics and Data Analysis, accepted.

Yip, P., 1988. Inference about the mean of a Poisson distribution in the presence of a nuisance parameter. Australian Journal of Statistics, 30: 299-306.