# Selection of Variables in Regression Models Based on Inflated Distributions

Aruna Rao K
Department of Statistics, Mangalore University
Mangalagangotri, Konaje 574199
Karnataka, India
arunaraomu@yahoo.com

Sumathi K
Department of Statistics, Mangalore University
Mangalagangotri, Konaje 574199
Karnataka, India
chaitra_udipi@yahoo.com

## Abstract

Regression models based on zero inflated distributions are often used in exploratory data analysis having excess zeroes. The difficulty faced by many researchers is with regard to the selection of covariates to be included in the model. Following the idea of focused information criterion, observed focused information criterion is proposed for model selection. The motivation for this has its roots in the concept of observed Fisher information. Using this criterion, a forward selection procedure is proposed for selection of variables in regression models based on inflated distributions. The procedure is illustrated using a dataset on decayed missing filled teeth (DMFT) index using the modified observed focused information criterion.

**Keywords:** Regression models, Zero inflated distributions, Focussed information criteria, Observed Fisher information, Forward selection procedure.

## 1. Introduction

In the recent years, regression models based on inflated distributions, in particular, zero inflated distributions, have been widely used in exploratory data analysis. It originated from the land mark paper of Lambert (1992) who first proposed regression model based on zero inflated Poisson distribution. Since then, several zero inflated distributions have been used to identify the factors influencing a count response $Y$. Some of the commonly used distributions for constructing zero inflated distributions are the Poisson, the negative binomial, the beta binomial and the generalized Poisson distributions. For a discussion on regression models based on these distributions and their applications, refer Czado and Min (2005), Sumathi and Rao (2009), Bhattacharya, et. al. (2008) and the references cited therein.

Exploratory data analysis is used when information is collected on a large number of factors which may possibly influence the response variable. The researcher may be interested in developing a parsimonious model consisting of only a fewer number of covariates. In multiple linear regression models, standard criteria are available for selecting the best model. They include the forward selection procedure, the backward elimination procedure, the step wise procedure and the subset selection approach. For a discussion on these methods, refer Montgomery (2006) and Draper and Smith (1998). These procedures are available in almost all statistical software packages dealing with multiple linear regression. These procedures have also been extended to the logistic regression models, and the computer software packages like R, SPSS and STATA offer this facility to the users.

Estimation of the regression parameters is computationally tedious when the models are based on zero inflated distributions. The common difficulty is when there is non-convergence of the iterative procedure used for solving the maximum likelihood equations. Thus the computer softwares like R and STATA provide the modules for fitting these models for a set of covariates. In each of these software packages, there is a restriction on the maximum number of covariates that can be included in the model. Thus an applied researcher faces the difficulty of choosing an appropriate model as in the case of linear regression analysis.

Recently, several other criteria have been proposed for selection of the best model, especially in the non-linear setting. They include Akaike Information Criteria (AIC) (Akaike (1973)), modified AIC, Bayesian Information Criteria (BIC) (Schwarz (1978)) and Focussed Information Criteria (FIC). FIC is of recent origin and is proposed by Claeskens and Hjort (2003). AIC and BIC have been used for model selection in regression models based on zero inflated distributions. FIC has not been employed as a model selection criteria in these cases. Several investigations in time series, survival analysis and other areas have showed the superiority of the FIC over AIC and BIC. See Psaradakis et. al., (2009), Claeskens et. al., (2007) and Claeskens and Hjort (2003) and the references cited therein.

Following the idea in observed Fisher information, we propose the observed FIC and use it to propose a forward selection procedure for regression models based on the zero inflated distributions. The expression for FIC relies on the asymptotic normality of the wide model. See section 2 of this paper for the terminology. In the case of regression models based on zero inflated distributions, the regularity conditions for the existence and asymptotic normality of the maximum likelihood estimators (MLE) are generally not satisfied when the number of covariates is large and thereby the FIC as well as the observed FIC cannot be defined. We propose modified observed FIC for the selection of variables in the regression models based on the zero inflated distribution based on the Poisson, negative binomial and the generalized Poisson distributions. This approach is new and gives a theoretical justification for the adhoc procedure suggested by Sumathi and Rao (2009).

The rest of the paper is organized as follows. In section 2, we propose the observed FIC criterion. In section 3, a forward selection procedure based on the modified observed FIC is proposed for selection of covariates in zero inflated

Poisson regression model. In section 4, we illustrate the procedure with a real life data set. Concluding remarks are provided in section 5.

## 2. Focussed Information Criterion

Let $Y_1, Y_2, ..., Y_n$ be independently distributed random variables following a distribution with parameters $\theta = (\theta_1, \theta_2, ..., \theta_p)'$ and $\gamma = (\gamma_1, \gamma_2, ..., \gamma_q)'$ is the vector of regression coefficients. We refer to this model as a wide model. The model becomes a narrow model when $\gamma = \gamma_0 = (0,0,...,0)'$. Thus $\theta$ refers to the minimum number of parameters that should be included in every model. A particular model $S$ refers to the case when $S = s$ parameters among $\gamma_1, \gamma_2, ..., \gamma_q$ are included in the model and the rest of the parameters are set equal to their corresponding values in $\gamma_0$, $s = (1,2,...,q)$. Let $\mu = (\theta, \gamma)$ be the focussed parameter of interest. Let $\hat{\theta}$ and $\hat{\gamma}$ be the maximum likelihood estimators (MLE) of $\theta$ and $\gamma$ in the wide model and, $\hat{\theta}_s$ and $\hat{\gamma}_s$ be the restricted maximum likelihood estimators of $\theta$ and $\gamma$ in the model $S$ when the parameters not included in the model are set to their values in $\gamma_0$. The FIC is given by

$$\begin{aligned} FIC &= E\left[\sqrt{n}\left(\mu(\hat{\theta},\hat{\gamma}) - \mu(\hat{\theta}_s,\hat{\gamma}_s)\right)\right]^2 \\ &= E\left[\sqrt{n}\left(\mu(\hat{\theta}_s,\hat{\gamma}_s) - E\left[\mu(\hat{\theta}_s,\hat{\gamma}_s) \mid \text{wide model}\right]\right)\right]^2 + E\left[\sqrt{n}E\left(\mu(\hat{\theta}_s,\hat{\gamma}_s) \mid \text{wide model}\right) - \mu(\hat{\theta},\hat{\gamma})\right]^2 \quad (2.1) \end{aligned}$$

The first term on the right hand side of (2.1) corresponds to the variance of $\mu(\hat{\theta}_s, \hat{\gamma}_s)$ and the second term corresponds to the squared bias. Under local misspecification, Claeskens and Hjort (2003) have derived the expressions for the two terms of the right hand side of (2.1). (See expressions (6.1) and (6.2) of Claeskens and Hjort (2007), p-147.)

Fisher information is given by $-E\dfrac{\partial^2 \log L[\theta]}{\partial\theta_i \partial\theta_j}$, where $L$ is the likelihood function of the parameters based on the sample observations and $[.]$ refers to the $(i,j)^{th}$ element in the Hessian matrix. Observed Fisher information is an unbiased estimator of the Fisher information and is given by $-\dfrac{\partial^2 \log L[\theta]}{\partial\theta_i \partial\theta_j}$. In the literature, Fisher information is referred to as the expected Fisher information since it is the expected value of the observed Fisher information. This has motivated us to name $E\left[\sqrt{n}\left(\mu(\hat{\theta},\hat{\gamma}) - \mu(\hat{\theta}_s,\hat{\gamma}_s)\right)\right]^2$ as the expected focussed information criteria. An unbiased estimator of this quantity is given by $\left[\sqrt{n}\left(\mu\left(\hat{\theta},\hat{\gamma}\right)-\mu\left(\hat{\theta}_s,\hat{\gamma}_s\right)\right)\right]^2$. We have named it as the observed FIC.

## 3. Forward Selection Procedure in zero inflated Poisson model

Let $Y_1,Y_2,...,Y_n$ be independently distributed inflated Poisson random variables with parameters $(p,\lambda_i), i=1,2,...,n$. The probability mass function of the zero inflated Poisson (ZIP) distribution is given by

$$P[Y_i=y_i]=\begin{cases} p+(1-p)\exp(-\lambda_i), \text{when } y_i=0 \\ (1-p)\exp(-\lambda_i)\dfrac{\lambda_i^{y_i}}{y_i!}, \text{when } y_i=1,2,... \end{cases} \tag{3.1}$$

Here $p$ is referred to as the inflate parameter and $\lambda_i$'s are the mean parameters. The covariates $X_1,X_2,...,X_k$ are related to $Y_i$'s by the log link function $\log\lambda_i=\beta_0+\beta_1x_{i1}+...+\beta_kx_{ik}$, where $X_j=(x_{i1},x_{i2},...,x_{ik}), j=1,2,...k,$ refers to the observed values of the covariates for the $i^{th}$ individual or item and $\beta_0,\beta_1,...,\beta_k$ are the regression coefficients. Initially, the inflate parameter $p$ is not linked to the covariates, the reason for this would be clear in the sequel.

Let $\mu(p,\beta_0,\beta_1,...,\beta_k)=p$ (3.2)

The reason for choosing the inflate parameter $p$ as the focussed parameter is that at the time of estimation of $p$ in the wide model as well as in the selected model, we would get an indication whether the maximum likelihood estimator is reasonable, that is, the estimated value of $p$ is nearer to the proportion of zeroes in the sample. The observed FIC is given by $OFIC=n\left(\hat{p}_{\text{model}}-\hat{p}_{\text{wide}}\right)^2$ (3.3)

The forward selection procedure is carried out in two stages. In the first stage, only the mean parameter is linked to the covariates. In this stage, the steps involved in the forward selection procedure are as follows.

1. Estimate $\hat{p}$ by linking all the covariates to the mean parameter $\lambda_i$.
2. Estimate $\hat{p}$ by linking only one covariate to the mean parameter. When $X_j$ is the included covariate, denote the estimate of $p$ as $\hat{p}_{(x_j)}$, $j=1,2,...,k$.
3. Compute $n\left(\hat{p}_{(x_j)}-\hat{p}_{\text{wide}}\right)^2$, $j=1,2,...,k$. Choose that covariate which gives the minimum value of $n\left(\hat{p}_{(x_j)}-\hat{p}_{\text{wide}}\right)^2$. (3.4)
4. Estimate $\hat{p}_{\text{model}}$ by linking the selected covariate and each one of the other covariates to the mean parameter $\lambda_i$, that is, two covariates are linked with the mean parameter, the first covariate corresponding to the selected covariate in the first step.

5. Repeat step 3 and select the second covariate that minimizes the value of the $OFIC$.
6. Repeat the procedure until no covariate is included in the model.

It would be helpful if in addition to the minimum $OFIC$, the significance of the regression coefficients is also included as a criteria for taking decision at each stage. Thus in the first stage, we select all the relevant covariates to be linked to the mean parameter $\lambda_i$. Having this as the narrow model or the first stage model, we continue our search by linking the covariates to the inflate parameter $p$ through the logit link function. The focussed parameter is

$$\left(\frac{1}{n}\sum_{i=1}^{n}\log\frac{\hat{p}_{i_{wide}}}{1-\hat{p}_{i_{wide}}}\right)-\left(\frac{1}{n}\sum_{i=1}^{n}\log\frac{\hat{p}_{i_{\text{model}}}}{1-\hat{p}_{i_{\text{model}}}}\right) \tag{3.5}$$

In the second stage of the selection procedures, as in the first stage, we include one covariate at a time. The procedure is terminated if no significant covariate enters the model.

## 3.1 Modified Focussed Information Criteria

While dealing with several data sets, we observed that either

(a) the numerical procedure for the estimation of the maximum likelihood estimators did not terminate or
(b) in case the procedure terminates, then the estimated value of $p$ is far away from the proportion of zeroes in the sample.

The reasons for both of the above situations is that, the regularity conditions for the existence and asymptotic normality of the maximum likelihood estimators are not satisfied. Czado and Min (2005) have established the regularity conditions for the existence and asymptotic normality of the parameters in zero inflated generalized Poisson (ZIGP) models. The ZIP distribution is a particular member belonging to the ZIGP family. The main violation of the regularity condition refers to the Fisher information matrix. If the entries of the Fisher information matrix grows rapidly compared to $n$ when $n \to \infty$, the regularity conditions are not satisfied. In such cases, the previous forward selection procedure breaks down. This motivated us to propose the modified observed FIC. Let $\hat{p}_{\max} = \frac{n_0}{n}$, where $n_0$ refers to the number of zeroes in the sample. From the probability mass function, it is clear that $\hat{p}_{\max}$ is the upper bound for the estimate of $p$ from any model. As pointed out by one of the referees, $\hat{p}_{\max}$ is an over estimate of $p$ since it does not take into account the zero counts from the (non-degenerated) Poisson distribution. $\hat{p}_{wide}$ does not exist when FIC cannot be used for the selection of the variables but $\hat{p}_{\max}$ always exists though an over estimate of $p$. However, any other efficient estimator may also be used instead of $\hat{p}_{\max}$. Therefore, in the first stage, the modified observed FIC is defined as

Modified observed FIC = $n\left(\hat{p}_{\text{model}} - \hat{p}_{\max}\right)^2$ (3.6)

and in the second stage, it is defined as $n\left(\frac{1}{n}\sum_{i=1}^{n}\log\frac{\hat{p}_i}{1-\hat{p}_i} - \log\frac{\hat{p}_{\max}}{1-\hat{p}_{\max}}\right)^2$ (3.7)

Using this criterion, the steps in the forward selection procedure as described previously are repeated to arrive at a forward selection procedure based on the modified observed FIC. In the next section, we apply this procedure on a dental epidemiological data set. The data set is such that the regularity conditions are not satisfied as exhibited either by the non-convergence of the iterative procedure for the estimation of the parameters or by an estimate of $p$ being far away from $\hat{p}_{\max}$.

## 4. Analysis of a DMFT index.

In this section, an exploratory data analysis of a data set from a dental survey is considered. The present data also forms a part of a large scale study conducted by A. B. Shetty Memorial Institute of Dental Sciences of coastal Karnataka, India. The data set consisted of 2000 individuals out of which 1000 (50%) were from rural area and 1000 (50%) were from urban area. There were 1016 (50.8%) males and 984 (49.2%) females. The age of the subjects, the DMFT index indicating the number of teeth decayed, missing or filled and the factors associated with the DMFT index such as caries and periodontitis were obtained. The individuals surveyed were above the age of fifteen years. The explanatory variables such as diet and brushing habits were also collected but have not been considered here. Out of the 2000 individuals, 1362 (68.1%) had DMFT index as zero. This indicates that 68.1% of the population surveyed had no dental problems. The remaining individuals had DMFT index in the range 1 to 27, both values inclusive. Since this is an example of count data with excess zeroes, the zero inflated models have been considered. From literature survey, it has been observed that, ZIP and zero inflated negative binomial (ZINB) regression models were used to develop epidemiological models for similar type of data. (Bohning et.al. (1999), Lewsey and Thomson (2004)).

In exploratory data analysis of the inflated count data, the points to be considered are

(i) which distribution to be used and
(ii) having chosen a particular model, the issue regarding the convergence of maximum likelihood estimates of the regression parameters.

Generally, the first aspect is taken care of by fitting more than one distribution to the data. The second aspect has to be handled with an amount of ingenuity. A non-convergence situation or convergence to a wrong value (a value which is not a reasonable one) of the parameter may arise when there is non-existency of a maximum likelihood estimator or there exists some problem associated with the numerical optimization technique. The analysis was started using ZIP regression. All the above explanatory variables were included in the regression model for the Poisson mean $\lambda$ using the log link function. The inflated parameter $p$ was kept as a constant. The analysis was carried out using the software STATA 10.0. The results are summarized in table 1 below.

**Table 1**

| Variable | Coefficient $\beta$ | Standard Error | p-value | 95% confidence interval for $\beta$ |
|---|---|---|---|---|
| gender | 0.2025 | 0.0349 | <0.0010 | ( 0.1341, 0.2710) |
| location | 0.0661 | 0.0353 | 0.0610 | (-0.0030, 0.1354) |
| age | 0.0577 | 0.0015 | <0.0010 | ( 0.0548, 0.0606) |
| caries | 1.3411 | 0.0479 | <0.0010 | ( 1.2472, 1.4350) |
| periodontitis | 1.3948 | 0.0533 | <0.0010 | ( 1.2903, 1.4993) |
| trauma | -0.8335 | 0.5011 | 0.0960 | (-1.8157, 0.1486) |

The estimate of logit $p$ was -5.8286 which yielded an estimate of the inflation parameter $p$ as 0.0029. Since the data set consisted of 68.1 % zeroes, the results are not satisfactory and casts doubts on the estimates of the regression parameters for the log link model. STATA 10.0 has five optional optimization techniques namely, the usual Newton-Raphson's technique which is the default, STATA's modified version of Newton-Raphson's technique specified as NR, the Berndt-Hall-Hall Hausman algorithm specified as BHHH, the Davidson-Fletcher-Powell algorithm specified as DFP and the Broyden- Fletcher-Goldfarb-Shanno algorithm specified as BFGS. The data was re-analyzed using the other optional optimization techniques. BHHH approach did not converge while the other optimization techniques gave the estimates of $p$ as 0.0029 for NR and DFP techniques and 0.0030 for BFGS. All these estimates of the inflation parameter are not reasonable. This is a common problem encountered by practitioners and they were not clear why such a problem arises. It may be a situation where the existence of MLE is doubtful. In such cases the entries of the fisher information matrix has to be thoroughly examined so as to identify whether the regularity conditions are satisfied.

The problem of obtaining a reasonable estimate for $p$ was resolved by building up a model including one covariate at a time in the log link function. The covariate which yielded the maximum value of the inflation parameter was retained. This amounts to selecting the covariate which minimizes the modified observed FIC. Taking this as the first covariate, the model was then fitted with two covariates, the second covariate was that which yielded a next higher estimate of the inflate parameter $p$. The set of two covariates which yielded the maximum value of the inflation parameter was then retained. The search continued by including one more covariate to this set. The termination procedure involved two criteria, (i) a stable value of the inflation parameter and

(ii) significance of the regression coefficients in the log link model. The table 2 below presents the results for the model that emerged in the end.

**Table 2**

| Variable | Coefficient $\beta$ | Standard Error | p-value | 95% confidence interval for $\beta$ |
|---|---|---|---|---|
| gender | 0.0504 | 0.0346 | 0.1450 | (-0.0174, 0.1181) |
| location | 0.1570 | 0.0349 | <0.0010 | ( 0.0885, 0.2255) |
| caries | -1.1380 | 0.0355 | <0.0010 | (-1.2076,-1.0685) |
| trauma | -0.6783 | 0.5008 | 0.1760 | (-1.6597, 0.3032) |

It is evident from the above table that location of residing and the prevalence of dental caries are the variables which highly influence the DMFT index. The estimate of logit $p$ was 0.7584 which yielded an estimate of the inflation parameter $p$ as 0.6810. This is exactly the proportion of the population who had DMFT index as 0.

A second type of model was developed by including covariates for the inflation parameter $p$ in the logit model. The covariates were included one at a time. The process continued till the variables left out were significant or when the iteration did not converge. The summary of the results is given in the table 3 below for the model that finally emerges.

**Table 3**

| Variable | Coefficient $\beta$ | Standard Error | p-value | 95% confidence interval for $\beta$ |
|---|---|---|---|---|
| Gender (for inflate parameter $p$ ) | -0.2140 | 0.0964 | 0.0260 | (-0.4030, -0.0250) |
| Location(for inflate parameter $p$ ) | 0.2544 | 0.0965 | 0.0080 | ( 0.0652, 0.4435) |
| caries | -0.4632 | 0.0857 | <0.0010 | (-0.6312, -0.2953) |
| periodontitis | 0.7242 | 0.0836 | <0.0010 | ( 0.5604, 0.8880) |
| location | 0.1550 | 0.0354 | <0.0010 | ( 0.0856, 0.2243) |

The estimate of logit $p$ was 0.6957 which yielded an estimate of the inflation parameter $p$ as 0.6700. This estimate is slightly lower than the estimate obtained by the first model. The prevalence of dental caries and the location of residence are again highly significant in addition to the gender and periodontitis, in influencing the DMFT index.

## 5. Conclusion

In this paper, we have proposed the observed focused information criteria. The focused information criteria or the observed focused information criteria break down when we cannot obtain the maximum likelihood estimators under the wide model. This has motivated us to propose modified focused information criteria with reference to regression models based on zero inflated distributions. This intuitive idea of the modified focused information criteria was available in Sumathi and Rao (2009). They proposed this criterion based on reasoning. The present paper gives a theoretical justification for their work.

The forward selection procedure is described with reference to zero inflated Poisson regression models. This can be easily extended to the models based on the zero inflated generalized Poisson distribution or the zero inflated negative binomial distribution by carrying out the search in three stages. In the first stage, the covariates relating to the mean parameter are identified. In the second stage, the covariates linked to the dispersion parameter are identified. In the last stage, the search is carried out for the covariates relating to the inflate parameter. In the second stage, the model arrived at the first stage is taken as the narrow model. In the last stage, the model arrived at the second stage is taken as the narrow model.

In this paper, we have not compared the performance of the focused information criteria and the proposed procedures with reference to the Akaike information criteria or the Bayesian information criteria. Carrying out simulations is quite tedious in the regression models based on zero inflated distributions. This can be attempted as a future research problem. The performance of the observed focused information criterion as compared to the expected focused information criterion is yet to be investigated. As in the case of observed fisher information, we conjecture that the observed focused information criteria performs at par with the expected focused information criteria for model selection. The advantage with the observed focused information criteria is that it is simple to compute. We hope that this paper would lead to the use of focused information criteria as a selection criterion in regression models based on inflated distributions.

## References

1. Akaike, H. (1973). Information theory and an extension of the maximum likelihood principle in '*Second international symposium on information theory*', Academiai Kiado, Budapest, pp. 267-281.
2. Bhattacharya, A., Clarke, B.S. and Datta, G.S. (2008). A Bayesian test for excess zeros in a zero-inflated power series distribution. *IMS collections. Beyond Parametrics in Interdisciplinary Research: Festschrift in Honor of Professor Pranab K Sen.* Vol.1, 89-104.
3. Bohning, D., Dietz, E., Schlattmann, P., Mendonca, L. and Kirchner, U. (1999). The zero-inflated Poisson model and the decayed, missing and filled teeth index in dental epidemiology. *Journal of the Royal Statistical Society, Series A* 162, 195-209.
4. Claeskens, G. and Hjort, N. L. (2003). The focused information criterion. *Journal of American Statistical Association,* 98, 900-916.

5. Claeskens, G. and Hjort, N. L. (2006). *Model Selection and Model Averaging.* Cambridge: Cambridge University Press.
6. Claeskens, G., Croux, C., and Van Kerckhoven., J. (2007): Prediction-focused model selection for auto regressive models. *Aust. N. Z. J. Stat.* 49(4), *359-379*.
7. Czado, C and Min, A. (2005). Consistency and asymptotic normality of the maximum likelihood estimator in a zero-inflated generalized Poisson regression. *Sonderforschungsbereich 386, Paper 423.* (http://epub.ub.uni-muenchen.de/)
8. Draper, N. R. and Smith, H. (1998). *Applied Regression Analysis*. Wiley-Interscience. 3rd edition.
9. Kale, B. K. (1998). Optimal estimating equations for discrete data with higher frequencies at a point. *Journal of the Indian Statistical Association*, 36, 125-136.
10. Kale, B. K. (1999). *A First Course on Parametric Inference*. Narosa Publishing House, New Delhi.
11. Lambert, D. (1992). Zero Inflated Poisson Regression with an application to defects in manufacturing. *Technometrics*, 34, 1-14.
12. Lewsey, J. D. and Thomson, W. M. (2004). The utility of the zero-inflated Poisson and zero-inflated negative binomial models: a case study of cross-sectional and longitudinal DMF data examining the effect of socio-economic status. *Community Dentistry and Oral Epidemiology*, 32, 183-189.
13. Montgomery, D. C. (2006). *Introduction to Linear Regression Analysis*. Wiley. 4th edition.
14. Psaradakis, Z., Sola, M., Spagnolo, F. and Spagnolo, N.(2009): Selecting non-linear time series models using information criteria. *Journal of Time series Analysis*, 30(4), 369-394.
15. Schwarz, G. (1978). Estimating the dimension of a model. *Annals of Statistics,* 6, 461-464.
16. Sumathi, K and Rao A. K. (2009). On Estimation and Tests for Zero Inflated Regression Models. *INTERSTAT.*

## Acknowledgement

This paper is dedicated to Professor B. K. Kale. It was he who constantly insisted on looking at the regularity conditions for the existence and asymptotic normality of the estimators. This has motivated us to look at the regularity conditions for the estimation of the maximum likelihood estimators in the inflated distributions. We are thankful to Mr. Vikas for providing us some of the research papers. We are also grateful to the referees for their valuable suggestions for this paper.