TWMS J. App. and Eng. Math. V.16, N.5, 2026, pp. 552-562

ADJUSTED EXPLAINED VARIATION MEASURE FOR ZERO-INFLATED POISSON REGRESSION MODEL

YOUNUS AL-TAWEEL[1*], ZAKARIYA ALGAMAL [2], §

Abstract. Explained variation measures are very popular in regression models. They explain the amount of variation in the dependent variable by the explanatory variables. In some situations, a small sample size is available in comparison to the number of explanatory variables. In this case, explained variation measures may become significantly inflated. In order to tackle this problem, adjusted explained variation measures are used. In this work, we propose adjusted explained variation measures depend on the deviance for zero-inflated Poisson regression model (ZIPRM). We compare the performance of the suggested measures with that of the unadjusted explained variation measures for the ZIPRM using a Monte Carlo simulation experiment and real datasets.

Keywords: Poisson distribution, zero-inflated Poisson regression model, adjusted R^2 measure, deviance.

AMS Subject Classification: 83-02, 99A00.

1. Introduction

The explained variation measure, also named coefficient of determination, is commonly used in a linear regression model under the standard assumptions of normal errors with homoscedasticity. It explains the variation amount of the dependent variable by the explanatory variables of the regression model. The explained variation measure, denoted by R^2, tries to provide a value between zero, for no fit by the explanatory variables, and one, for a perfect fit by the explanatory variables.

Regression models have been utilized in several science and technology disciplines, such as clinical science, engineering, climate and environmental science. In some situations, the sample size, n, is limited and hence only a limited number of observations is available with a large number of explanatory variables, p. In this case, the value of R^2 may be substantially inflated which introduces some risks in its interpretation. This is because when $p > n$, the entire linear regression fitting falls apart since the sum of squares of the

[1] University of Mosul - College of Education for Pure Science - Department of Mathematics - Iraq.
e-mail: younus.altaweel@uomosul.edu.iq; ORCID: https://orcid.org/0000-0001-7167-8079.

[2] University of Mosul - College of Computers Sciences and Mathematics - Department of Statistics and Informatics - Iraq.
e-mail: zakariya.algamal@uomosul.edu.iq; ORCID: https://orcid.org/0000-0002-0229-7958.

[*] Corresponding author.

§ Manuscript received: April 02, 2025; accepted: Jul 03, 2025.
TWMS Journal of Applied and Engineering Mathematics, Vol.16, No.5;

residual is supposed to have degrees of freedom equal to $(n-p)$ and so it becomes negative. For linear regression models, adjusted explained variation measures are used to tackle the inflation problem of the R^2 measure when the model is no longer valid with $p > n$. The adjusted explained variation measure, denoted by R^2_{adj}, takes into account the number of variables that are utilized in the model.

In generalized linear models (GLMs), the maximum likelihood estimate (MLE) is usually used for estimating the regression model parameters. The, R^2 measures that are based on the reduction proportion in the maximized log-likelihood have been used by many authors. Recently, R^2 and R^2_{adj} measures have been used by several authors for Poisson regression models [1, 2]. The aim has been to investigate the performance of R^2 and R^2_{adj} measures that are based of different criteria. In this paper, we propose adjusted explained variation measures for the R^2 measure that depend on the deviance for ZIPRM. The behavior of the suggested measures is investigated using a Monte Carlo simulation experiment as well as real datasets where the behavior of R^2_{adj} measures is compared with that of R^2 measures.

This paper is constructed as follows. Section 2 explains the concept of zero-inflation in count data models and reviews the methodology of the Poisson regression model and the zero-inflated Poisson regression model. In Section 3, the concept and the mathematical definition of the deviance are explained. Section 4 reviews various unadjusted and adjusted explained variation measures. Section 5, presents a Monte Carlo experiment to see the behavior of proposed measures. In Section 6, the behavior of the proposed measures is investigated by applying them to real datasets. The conclusion is finally presented in Section 7.

2. Zero-inflation in count data models

In several science and technology disciplines, it is common to see that the interested outputs have generally non-negative values and typically include a huge number of zeros and overdispersion (i.e. higher than expected variability). Zero values are not permitted in several distributions For example, this can be seen in social psychological, biomedical studies, environmental economics, political science and health related research. In such research studies, these data with abundant zeros are practically common when the occurrence of events is counted [3]. For example, number of smoked cigarettes, number of unhealthy people, or number of absences in a particular school. For such type of count data, it is typical to analyze them using a Poisson regression model with the log link function. However, ZIPRM is frequently used to model the counts in the case of the number of zeros is greater than the capacity of the probability mass at zero under a Poisson distribution [4, 5, 6].

2.1. **Poisson regression model.** Poisson regression models are used for analyzing count data. Suppose that $y_i, i = \ldots, n$ represent counts of events that occur in a time period. Therefore, the response variables, y_i, follow the Poisson distribution given by

$$p(y_i, \lambda_i) = \frac{e^{-\lambda_i}\lambda_i^{y_i}}{y_i!}, \quad y_i = 0, 1, \ldots \tag{1}$$

with $\mathrm{E}(y_i) = \mathrm{Var}(y_i) = \lambda_i$. This means that when the mean is large, the variability of the data will be large. However, the mean will have a smaller value than the variance in case of there is a large number of zeros in the data [7]. The log-likelihood function of the Poisson distribution is given by

$$l(\boldsymbol{\lambda}) = \sum_{i=1}^{n} \{y_i \ln(\lambda_i) - \lambda_i - \ln(y_i!)\}. \tag{2}$$

Now suppose we have a design matrix, $\mathbf{X} = (1, \mathbf{x}_1, \ldots, \mathbf{x}_p)$, where $\mathbf{x}_1, \ldots, \mathbf{x}_p$ are the explanatory variables, and a $(p+1) \times 1$ vector of unknown coefficient parameters, $\boldsymbol{\beta}$. Using the log link function

$$\ln(\lambda_i) = \eta_i = \mathbf{x}_i^T \boldsymbol{\beta}, \tag{3}$$

we can obtain the Poisson regression model. Fisher scoring method can be utilized to obtain the MLE, $\hat{\boldsymbol{\beta}}$, of $\boldsymbol{\beta}$ [8, 9].

2.2. **Zero-inflated Poisson regression model (ZIPRM).** The ZIPRM was suggested by [5] to model zero-inflation in count data. It is a combination model for count data with extra zeros. This means that the ZIPRM is a combination of the Poisson distribution and a degenerate distribution at zero. The data in this model include two kinds of zeros. The first type is called structural zeros that come from a non-susceptible group (groups which have no experience or attribute of concern, for instance, those who are healthy without any disease). This kind of structural zeros has a probability w_i. The second type is called random zeros that come from susceptible group, for example, people with a disease in a health research who may misrepresent a zero rating). This kind of zeros has a probability $(1-w_i)e^{-\lambda_i}$ and results in a Poisson distribution [10].

The probability mass function of ZIPRM is defined as

$$p(y_i) = \begin{cases} w_i + (1-w_i)e^{-\lambda_i}, & \text{if } y_i = 0 \\ (1-w_i)\frac{e^{-\lambda_i}\lambda_i^{y_i}}{y_i!} & \text{if } y_i = 1, 2, \ldots, \end{cases} \tag{4}$$

where λ_i is the expected count for the observation ith and $w_i \in [0,1]$ represents probability of additional zeros [11]. The mean of y_i is $\mathrm{E}(y_i) = (1-w_i)\lambda_i$ and the variance is $\mathrm{Var}(y_i) = \lambda_i(1-w_i)(1+w_i\lambda_i) = \sigma_i + \left(\frac{w_i}{1-w_i}\right)\sigma_i^2$ where $\sigma_i = (1-w_i)\lambda_i$. The ZIPRM will be reduced to a Poisson distribution if $w_i = 0$. However, when $w_i > 0$, the marginal distribution of y_i shows overdispersion and will have zero-inflation.

The ZIPRM has a log-likelihood function given by

$$\begin{aligned} l(\mathbf{y}) = l(\lambda, w, \mathbf{y}) &= \sum_{i=1}^{n} I_{\{y_i=0\}} \ln\left[w_i + (1-w_i)e^{-\lambda_i}\right] \\ &+ I_{\{y_i>0\}}\left[\ln(1-w_i) - \lambda_i + y_i\ln(\lambda_i) - \ln(y_i!)\right]. \end{aligned} \tag{5}$$

The $\ln(\lambda_i) = \mathbf{x}_i^T\boldsymbol{\beta}$ can be used for λ_i to obtain the ZIPRM [12].

3. The deviance of ZIPRM

The deviance refers to the amount of deviation of the model of interest from another model. The deviance is modeled as twice of the difference in the log-likelihoods of the full and the fitted models. Therefore, it is given by

$$D = 2\left[l(\mathbf{y}) - l(\boldsymbol{\beta})\right]), \tag{6}$$

where $l(\mathbf{y})$ and $l(\boldsymbol{\beta})$ represent the maximized log-likelihood of the full and the fitted models respectively. The fitted global deviance is a special case of the deviance when $l(\mathbf{y}) = 0$. Hence, it is defined as

$$GD = -2l(\boldsymbol{\beta}). \tag{7}$$

The D and GD are very important tools that are used for testing between different models. The deviance for Poisson regression model is defined as

$$D = 2 \times \sum_{i=1}^{n}\left\{y_i \ln\left(\frac{y_i}{\hat{\lambda}_i}\right) - (y_i - \hat{\lambda}_i)\right\}. \tag{8}$$

For the ZIPRM, the parameters cannot be modeled separately without knowing which zero counts come from the Poisson distribution or the Bernoulli distribution [13]. Hence, for binary data in the ZIPRM, we have

$$l(\boldsymbol{\mu}) = (1 - y_i)\log(1 - \mu_i) + y_i \log(\mu_i) \tag{9}$$

and

$$l(\mathbf{y}) = (1 - y_i)\log(1 - y_i) + y_i \log(y_i), \tag{10}$$

where $\mu_i = (1 - w_i) - w_i \exp(-\lambda_i)$. Hence, the deviance for binary data is given by

$$D = \begin{cases} -2\sum_i \ln(\mu_i), & \text{if } y_i = 1 \text{ and} \\ -2\sum_i \ln(1 - \mu_i) & \text{if } y_i = 0. \end{cases} \tag{11}$$

For non-zero data in the ZIPRM, we have

$$l(\boldsymbol{\theta}) = -\theta_i + y_i \log(\theta_i) - \log(y_i!) - y_i \log(1 - \exp(\theta_i)) \tag{12}$$

where $\mu_i = \frac{\theta_i}{1-\exp(-\theta_i)}$ and $\theta_i = h^{-1}(\mu_i)$.

$$l(\boldsymbol{\mu}) = -h^{-1}(\mu_i) + y \log(h^{-1}(\mu_i)) - \log(y_i!) - \log(1 - \exp(-h^{-1}(\mu_i))) \tag{13}$$

and

$$l(\mathbf{y}) = -h^{-1}(y_i) + y_i \log(h^{-1}(y_i)) - \log(y_i!) - \log(1 - \exp(-h^{-1}(y_i))). \tag{14}$$

So, we have

$$D = 2\sum_i \left[-h^{-1}(y_i) + \hat{\theta}_i + y_i \log\left(\frac{h^{-1}(y_i)}{\hat{\theta}_i}\right) - \log\left(\frac{1 - \exp(-h^{-1}(y_i))}{1 - \exp(-\hat{\theta}_i)}\right)\right], \tag{15}$$

where θ_i in $\theta_i = h^{-1}(\mu_i)$ can be obtained by Newton methods [14].

4. Explained variation measures

Suppose $(y_i, \mathbf{x}_i)$, for $i = 1, \ldots, n$, is a sample of n observations where y_i is the observed value of the dependent variable and $\mathbf{x}_i$ is the vector of the corresponding explanatory variables. The general explained variation measure, R^2, is given by

$$R^2 = \left[\sum_i d(y_i) - \sum_i d(y_i|\mathbf{x}_i)\right] / \sum_i d(y_i), \tag{16}$$

where $d(y_i)$ and $d(y_i|\mathbf{x}_i)$ represent distance measures of y_i from unconditional and conditional central location parameter. The $d(y_i)$ and $d(y_i|\mathbf{x}_i)$ can have different specifications which result in different R^2 measures.

4.1. **Deviance R_D^2 measure.** The distance in this measure is based on the deviance. A deviance R^2 measure can be defined as

$$R_D^2 = 1 - \frac{D_{full}}{D_{null}}, \tag{17}$$

where D_{full} is the deviance of the full model and D_{null} is the deviance of the null model (intercept only). For the fitted global deviance, we have

$$\begin{aligned} \sum_i d(y_i) &= -2l(\bar{y}) \\ \sum_i d(y_i|\mathbf{x}_i) &= -2l(\hat{\boldsymbol{\beta}}), \end{aligned}$$

where $l(\bar{y})$ and $l(\hat{\boldsymbol{\beta}})$ represent the log-likelihood of the null and the full model respectively. Hence, the global deviance R^2 measure, R^2_{GD}, is given by

$$R^2_{GD} = 1 - \frac{l(\hat{\boldsymbol{\beta}})}{l(\bar{y})}. \tag{18}$$

4.2. **Adjusted explained variation measures.** For linear regression models, when the sample size is small and the number of explanatory variables is large, the R^2 measures become inflated. In order to tackle this inflation problem, adjusted measures, R^2_{adj}, are used. In this work we propose various adjusted measures for the ZIPRM based on the deviance. The first R^2 measure is calculated using the log-likelihood:

$$R^2_D = 1 - \frac{l(\mathbf{y}) - l(\hat{\boldsymbol{\beta}})}{l(\mathbf{y}) - l(\bar{y})}. \tag{19}$$

The R^2 measure can be adjusted based on their degrees of freedom to have [15]

$$R^2_{D.adj} = 1 - \frac{(n-p-1)^{-1}[l(\mathbf{y}) - l(\hat{\boldsymbol{\beta}})]}{(n-1)^{-1}[l(\mathbf{y}) - l(\bar{y})]}. \tag{20}$$

Alternative measures are given by

$$\begin{aligned} R^2_{D.adj.1} &= 1 - \frac{l(\mathbf{y}) - [l(\hat{\boldsymbol{\beta}}) - p/2]}{l(\mathbf{y}) - l(\bar{y})} \\ &= 1 - \frac{l(\mathbf{y}) - l(\hat{\boldsymbol{\beta}}) + p/2}{l(\mathbf{y}) - l(\bar{y})} \end{aligned} \tag{21}$$

and

$$\begin{aligned} R^2_{D.adj.2} &= 1 - \frac{l(\mathbf{y}) - [l(\hat{\boldsymbol{\beta}}) - (p+1)/2]}{l(\mathbf{y}) - [l(\bar{y}) - 1/2]} \\ &= 1 - \frac{l(\mathbf{y}) - l(\hat{\boldsymbol{\beta}}) + (p+1)/2}{l(\mathbf{y}) - l(\bar{y}) + 1/2}, \end{aligned} \tag{22}$$

where $(p+1)/2$ and $1/2$ are the expected optimism correction of $l(\hat{\boldsymbol{\beta}})$ and $l(\bar{y})$ respectively [2]. The global deviance for this measure is

$$R^2_{GD.adj} = 1 - \frac{l(\hat{\boldsymbol{\beta}}) - (p+1)/2}{l(\bar{y}) - 1/2}, \tag{23}$$

Therefore, in this work, we propose the $R^2_{GD.adj}$ $R^2_{D.adj}$ $R^2_{D.adj1}$ and the $R^2_{D.adj2}$ measures and compare them with the unadjusted variation measures for the ZIPRM.

5. Simulation Experiment

In this section, we investigate and compare the behavior of unadjusted R^2-measures and adjusted R^2-measures with different conditions for the ZIPRM via a Monte Carlo simulation experiment.

5.1. **The Simulation design of experiment.** We generated fractional factorial designs with sample sizes of $n = 16, 32, 64, 128$ and 4096 $[= 2^{12}]$. The number of explanatory variables was set to be $p = 2, 3$ and 4. In order to generate ZIPRM random variables, the linear predictor was evaluated to get the expected values of y at each value of $\mathbf{x}$. Then, using equation (4), ZIPRM random variables were generated with $\ln(\lambda_i) = \mathbf{x}_i^T \boldsymbol{\beta}$. The sum of $\boldsymbol{\beta}$ was set to be equal to 1. The zero-inflation parameter (the vector of probabilities at zero) was chosen to be 0.4.

5.2. **Simulation results.** This section presents the Monte Carlo simulation results of the unadjusted R^2 measures and adjusted R^2 measures with different combinations of p and n. The number of replications was set to be $B = 1000$ for different sample sizes. Then, the mean of the different unadjusted R^2 measures and adjusted R^2 measures for the 1000 replications was calculated. Table 1 shows the number of explanatory variables (p), the sample size (n) and the estimated mean values for the different unadjusted and adjusted R^2 measures.

TABLE 1. The mean values of unadjusted R^2 measures and adjusted R^2 measures with 1000 replications for the ZIPRM.

p	n	R^2_{GD}	R^2_D	$R^2_{GD.adj}$	$R^2_{D.adj}$	$R^2_{D.adj1}$	$R^2_{D.adj2}$
2	16	0.1678	0.0329	0.0398	-0.1158	0.0209	0.0207
	32	0.1217	0.0257	0.0611	-0.0415	0.0197	0.0196
	64	0.0999	0.0207	0.0704	-0.0114	0.0177	0.0177
	128	0.0931	0.0193	0.0786	0.0037	0.0178	0.0178
	4096	0.0866	0.0178	0.0862	0.0173	0.0178	0.0178
3	16	0.2029	0.038	0.0037	-0.2025	0.0194	0.0193
	32	0.1173	0.0237	0.0227	-0.0809	0.0145	0.0145
	64	0.0882	0.0179	0.0426	-0.0312	0.0133	0.0133
	128	0.0737	0.015	0.0513	-0.0088	0.0127	0.0127
	4096	0.0631	0.0128	0.0624	0.012	0.0127	0.0127
4	16	0.2565	0.046	-0.0139	-0.3009	0.0208	0.0207
	32	0.1296	0.026	0.0007	-0.1183	0.0136	0.0136
	64	0.09	0.0182	0.0283	-0.0483	0.0121	0.0121
	128	0.069	0.0141	0.0387	-0.018	0.011	0.011
	4096	0.0551	0.0111	0.0542	0.0101	0.011	0.011

It can be shown from Table 1 that the values of the unadjusted measures R^2_{GD} and R^2_D increase when the number of explanatory variables increases. In addition, the values of these measures decrease as the sample size increases.

For the adjusted measures, we can see that the $R^2_{GD.adj}$ values decrease when the number of explanatory variables increases, but they increase as n increases. The $R^2_{D.adj}$ values decrease when the number of explanatory variables increases. Moreover, the values of this measure increase when the sample size increases. It can be shown that many negative values of this measure are seen in the table. This may happen when the values of R^2_D is small, negative values of the adjustments of R^2_D may be obtained. Negative values for some measures are also reported and found in several studies for GLMs [1, 16].

For the $R^2_{D.adj1}$ and $R^2_{D.adj2}$ measures, we can see that they provide smallest values among the other measures and their values are very similar. We can also see that the values of these measures decrease when p is increased and also as n increases. Therefore, these two adjusted measures are an improvement over the unadjusted measures.

The sample size $n = 4096$ is chosen to provide an indication of true performance of the unadjusted and adjusted measures over the input space. Thus, the computed values for the measures were considered to be the true values. We can see that in general the values of the adjusted measures are small. However, the values of R^2_{GD} and $R^2_{GD.adj}$ are large in comparison to the other measures. Figure 1 shows the box-plots of the values of unadjusted R^2 measures and adjusted R^2 measures for the simulation experiment when $p = 2$ and $n = 4096$.

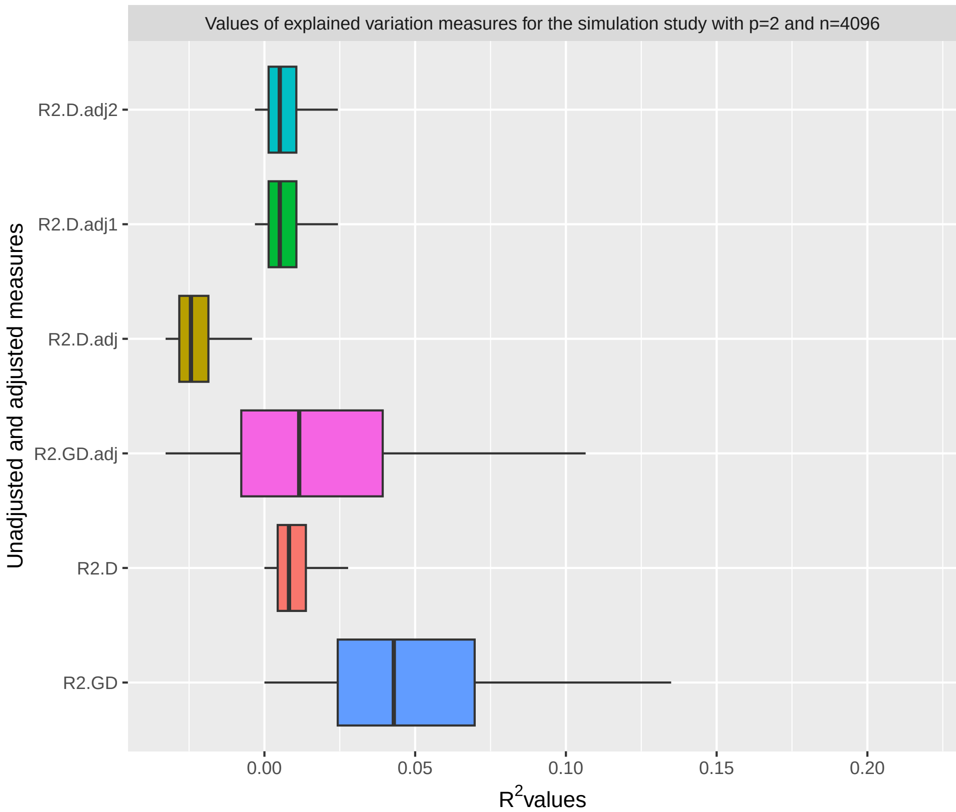

FIGURE 1. The box-plots of the values of unadjusted and adjusted R^2 measures for the simulation experiment when $p = 2$ and $n = 4096$. The values of $R^2_{D.adj1}$ and $R^2_{D.adj2}$ are very small.

6. APPLICATIONS

We investigate in this section the behavior of the adjusted measures for real datasets that have been commonly used for fitting the zero-inflated Poisson regression model. We consider the salamander dataset, collected by [17], and the bioChemists dataset, presented by [18] as examples of real datasets.

6.1. **Example 1.** In this example, we consider the Salamander data that consist of counts of salamanders with the site of the explanatory variables and sampling explanatory variables. The Salamander data have $n = 644$ observations that were observed by [17] four times at 23 sites in streams. The response variable is the counts of salamanders in streams. The $p = 8$ variables that may have an impact on the response variable are described in Table 2.

Table 2. Salamander data variables and their descriptions

Variables	Description
site	represents a location name of the samples.
mined	indicates whether the site was affected by coal mining.
cover	represents the amount of cover objects in the stream.
sample	represents the repeated sample.
DOP	represents the days since precipitation.
Wtemp	represents the temperature of the water.
DOY	represents the day of the year.
spp	represents the name of the abbreviated species, possibly also life stage.

We fitted the ZIPRM, equation (4), to the Salamander data. Then, the unadjusted and adjusted R^2 measures were calculated and the results are shown in Table 3.

Table 3. The unadjusted and adjusted R^2 measures for the ZIPRM.

R^2_{GD}	R^2_D	$R^2_{GD.adj}$	$R^2_{D.adj}$	$R^2_{D.adj1}$	$R^2_{D.adj2}$
0.021	0.0039	0.0157	-0.0016	0.0036	0.0036

It can be shown from Table 3 that the unadjusted measures work well as the values of the all unadjusted measures are positive. The R^2_D measure has the smallest value among the other unadjusted measures. For the adjusted measures, it can be seen that the $R^2_{D.adj}$ measure that is based on the degrees of freedom has a negative value that makes no sense in reality. This is may be because, as reported before, the value of the unadjusted R^2_D measure is small. The $R^2_{D.adj1}$ and $R^2_{D.adj2}$ measures produce similar values and they have the best performance as their values are the smallest. The $R^2_{D.adj1}$ and $R^2_{D.adj2}$ measures were able to avoid having negative values.

6.2. **Example 2.** In this example, we consider the bioChemists dataset, presented by [18], that consists of 915 rows of observations. The bioChemists data have $n = 915$ rows of observations. The response variable is the Articles that represents the number of articles that are published during the last 3 years of Ph.D. The explanatory variables and their descriptions are given in Table 4.

Table 4. BioChemists data variables and their descriptions.

Variable	Description
Female	represents the gender of student, 1 if female and 0 if male.
MentorArts	represents the number of articles published by Ph.D. during last 3 years.
Prestige	represents prestige of student in the Ph.D study.
Married	indicates status of marital, 1 refers to married and 0 refers to single.
Children	represents the number of the child of aged 5 or younger.

We fitted ZIPRM, equation (4), to bioChemists data. Then, the unadjusted and adjusted R^2 measures were calculated and their values are shown in Table 5.

Table 5. Unadjusted and adjusted R^2 measures for the ZIPRM.

R^2_{GD}	R^2_{D}	$R^2_{GD.adj}$	$R^2_{D.adj}$	$R^2_{D.adj1}$	$R^2_{D.adj2}$
0.1526	0.0065	0.1433	-0.0044	0.0064	0.0064

It can be shown from Table 5 that the results of the unadjusted and adjusted measures are quite similar to those of the previous example. We can see that the unadjusted measures also work well for this example as their values are positive. Again, the R^2_D measure also has the smallest value among the other measures. For the adjusted measures, the $R^2_{D.adj}$ measure that is based on the degrees of freedom also has a negative value. The $R^2_{D.adj1}$ and $R^2_{D.adj2}$ measures also produce similar values and they have the best performance as their values are positive and they are smallest.

We can conclude from examples 1 and 2 that the results of the both unadjusted and adjusted R^2 measures for the ZIPRM are very similar. Moreover, the results of these two examples agree with the results of the simulation experiment.

7. Conclusion

In this article, we have suggested several adjusted explained variation measures for zero-inflated Poisson regression model (ZIPRM). The behavior of the suggested measured have been investigated using a Monte Carlo simulation study and also real datasets. According to the results for the ZIPRM, the performance of the adjusted R^2 measures is better than that of the unadjusted R^2. This is because they provide values of the measures closer to zero than the unadjusted measures. Moreover, it was clear that the adjusted measures are preferable to unadjusted measures. In addition, among the adjusted measures, the $R^2_{D.adj1}$ and $R^2_{D.adj2}$ measures are preferable than the other as they always had positive values. Therefore, we recommend to use the adjusted $R^2_{D.adj1}$ and $R^2_{D.adj2}$ measures that are based on deviance as they consider the number of explanatory variables in the ZIPRM.

Acknowledgement. The authors are very grateful to the University of Mosul/ College of Education for Pure Science and College of Computers Sciences and Mathematics for their provided facilities, which helped to improve the quality of this work.

References

[1] Cameron, A. C. and Windmeijer, F. A., (1996), R-squared measures for count data regression models with applications to health-care utilization, Journal of Business and Economic Statistics, 14(2) pp. 209-220.

[2] Mittlböck, M., and Waldhör, T., (2000), Adjustments for R^2-measures for Poisson regression models, Computational statistics and data analysis, 34(4) pp. 461-472.

[3] Mohammadi, A., Shadrokh, A., and Yarmohammadi, M., (2025), Stochastic expectation maximization algorithm for exponential-poisson distribution under type-I progressive interval censoring, TWMS Journal of Applied and Engineering Mathematics, 15(8), pp. 2019-2030.

[4] King, G., (1989), Event count models for international relations: Generalizations and applications, International Studies Quarterly, 33(2), pp. 123-147.

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

[6] Greene, W. H., (1994), Accounting for excess zeros and sample selection in Poisson and negative binomial regression models, NYU working paper no. EC-94-10.

[7] Rashad, N. and Hammoud, N., (2024), Biased Estimators in Zero-Inflated Poisson Regression Model in the Presence of Multicollinearity: Subject Review, College of Basic Education Research Journal, 20(1) pp. 819-842.

[8] Alangood, H. N., Algamal, Z. Y., and Khaleel, M. A., (2024), Variable selection in Poisson regression model based on chaotic meta-heuristic search algorithm, BIO Web of conferences, 97, pp. 00161.

[9] Abd Ulghani, F. and Al-Hamdani, R., (2022), Shrinkage estimators in inverse Gaussian regression model: Subject review, Iraqi Journal of Statistical Sciences, 19(1) pp. 46-53.

[10] Jansakul, N., and Hinde, J. P., (2002), Score tests for zero-inflated Poisson models, Computational statistics and data analysis, 40(1) pp. 75-96.

[11] Li, C-S, (2011), A lack-of-fit test for parametric zero-inflated Poisson models, Journal of Statistical Computation and Simulation, 81(9) pp. 1081-1098.

[12] Al-Sinjary, A. and Raheem, A., (2022), Gauss-Hermite Cubature Method to estimate parameters of a multivariate GLMM, Journal of Education and Science, 31(2) pp. 29-41.

[13] Al-Taweel, Y. and Algamal, Z., (2022), Almost Unbiased Ridge Estimator in the Inverse Gaussian Regression Model, Electronic Journal of Applied Statistical Analysis, 15(3), pp. 510-526.

[14] Al-Taweel, Y. and Algamal, Z., (2022), Almost unbiased ridge estimator in the zero-inated poisson regression model, TWMS Journal of Applied and Engineering Mathematics, 12(1), pp. 235.

[15] Waldhör, T., and Haidinger, G., and Schober, E., (1998), Comparison of R^2 measures for Poisson regression by simulation, Biostatist, 3(11) pp. 209-215.

[16] Heinzl, H., and Mittlböck, M., (2008), Adjusted R^2 Measures for the Inverse Gaussian Regression Model, Computational Statistics, 17(4) pp. 525-544.

[17] Price, S. J., Muncy, B., Bonner, S. J., Drayer, A. N., and Barton, C. D., (2016), Effects of mountaintop removal mining and valley filling on the occupancy and abundance of stream salamanders, Journal of Applied Ecology, 53(2) pp. 459-468.

[18] Long, J. S., (1990), The origins of sex differences in science, Social forces, 68(4) pp. 1297-1316.

Younus Al-Taweel is an Assistant Professor in the Department of Mathematics, College of Education for Pure Science, University of Mosul, Iraq. He got his Ph.D. in Statistics from the University of Sheffield, UK in 2018,. His research interests are in Bayesian statistics, machine learning, uncertainty quantification for computer models, and regression models.

Zakariya Algamal graduated from Department of Mathematical Scienced of Universiti Teknologi Malaysia in 2016. He is working as a Professor of Statistics in the department of Statistics and Informatics of University of Mosul-Iraq. His research areas include high dimensional data, sparse methods, generalized linear model, machine learning.