AL-Bahir
Journal for Engineering and Pure Sciences
BJEPS

Manuscript 1117

Seemingly Unrelated Exponetiated Exponential Geometric Regression Model

Oluwaseun Michael Famoni

Bamidele Mustapha Oseni

ORIGINAL STUDY

Seemingly Unrelated Exponentiated Exponential Geometric Regression Model

Oluwaseun M. Famoni*, Bamidele M. Oseni

Department of Statistics, Federal University of Technology, Akure, Ondo State, Nigeria

Abstract

In the class of seemingly unrelated regression models, the dispersion nature of the dependent variable can greatly impact the efficiency and reliability of the parameter estimates for the model. Despite this, the seemingly unrelated Poisson regression model and seemingly unrelated negative binomial model are two most commonly used count data models for these class of regression models. This study introduces the seemingly unrelated exponentiated exponential geometric regression (SUEEGR) for modelling count data which might be equi-, under, or over-dispersed. Parameters estimation for the model was carried out using the method of maximum likelihood. A simulation study was carried out to assess the performance of the model under various conditions using certain evaluation criteria for the under-dispersed, over dispersed and when SUEEGR reduces to SU-geometric regression. The model was applied to datasets from Demography and Health Survey (DHS), the descriptive with scattered plot of the dataset were obtained and the results was compared with those obtained from seemingly unrelated Poisson regression. From the findings, it was observed that there is a significant correlation (0.1594) between the response variables and SUEEGR model appears to provide a better overall fit of the data than SUGPR because of the higher Log-likelihood, lower AIC and BIC of the model.

Keywords: Regression, Seemingly unrelated regression, Exponentiated exponential geometric regression, Equi-dispersion, Under-dispersion, Over-dispersion

1. Introduction

Multiple regression analysis that follows Poisson distribution for count data was first considered by Jorgenson [4]. Example of Poisson endogenous variables [4] are number of failures over a time period, quantity demanded, number of purchases over a time, the number of triplet born in Norway [6], the annual number of presidential appointment to the supreme court [10], the number of suicides per month, the number of spells of un-employment. Non-linear regression for count data for Poisson distribution first consider by Frome [5] where the sample mean and sample variance are equal.

When the sample mean and variance of the count data are equal, the model exhibits an equi-dispersion scenario. If the variance is smaller than the mean, the model has under-dispersion, and if the variance is greater than the mean, the model has an over-dispersion scenario.

Moreover, regression models for count data that account for under dispersion or over dispersion, as well as other special cases, can be addressed through several modified forms. Some of the example are: the categorized count data regression model that can handle categorized response variable, the zero-inflated regression models that can handle count data that have several many zero. The zero-truncated regression models that can handle count data that are truncated at zero.

Count data as an excessive number of zeros can also influence the choice of model. Zero-inflated model is one of the account modification of count

Received 26 July 2025; revised 26 September 2025; accepted 19 October 2025.
Available online 29 December 2025

* Corresponding author.

E-mail address: ofamoni@gmail.com (O.M. Famoni).

https://doi.org/10.55810/2313-0083.1117

data models, this zero-inflated model has been widely used for various count distributions since it has been proposed [11]. Currently [12], a fitting of cluster count data with excessive zeros with zero-inflated generalized Poisson regression model based on generalized estimating equation has been proposed. The model which is able to handle cases of over- and under-dispersions was said to outperform some well-known zero-inflated models like zero-inflated Poisson, generalized Poisson and negative binomial models.

Clusterwise regression [15] based on finite mixture models, is useful for analyzing data from heterogeneous populations, with Gaussian clusterwise linear regression commonly applied for continuous responses. A novel multivariate Gaussian clusterwise regression model has been proposed to handle correlated responses and mild outliers while allowing different covariates for each response, showing improved performance in both simulations and real data applications.

The inverse Gaussian regression (IGR) model is suited for positively skewed response variables but suffers from inflated variance of the Maximum Likelihood (ML) estimator in the presence of multicollinearity [17]. To overcome this limitation, a new estimator has been proposed that combines two parameter estimators. Simulation studies and real-data applications show that the proposed estimator consistently outperforms the ML, ridge, Liu, Kibria-Lukman, and modified ridge-type estimators in terms of mean squared error.

Class of parametric regression models designed for count data exhibiting under- or over-dispersion [3]. These models are based on squared polynomial expansions around the Poisson baseline distribution. They applied their model to health service utilization data, using the number of consultations with non-doctor health professionals in the past four weeks as the response variable. In their study, they employed a fifth-order model, which demonstrated superior performance compared to the Negative Binomial Regression (NBR) model.

Subsequently, other researchers analyzed the data using zero-inflated regression models due to the high proportion of zeros (over 90 %). The authors noted that for this particular application, successful modeling does not require a high-order expansion, although the negative binomial model offers greater parsimony.

Bivariate Negative Binomial Regression (BNBR) and Bivariate Poisson Regression (BPR) model to fits bivariate counts observed under a health-care utilization dataset obtained from the Journal of Applied Economterics 1997 Data Archive and number of defectives in samples of textile fibers [8]. The observed value of the score statistic is 455.52. The BNBR model provides a better fit than the BPR model. It was shown that the dataset exhibited substantial over-dispersion. The estimates generated from BNBR and the BPR model are very close, but the standard errors are different.

Efficient method of estimating a seemingly unrelated Poisson regression model (SUPREME) [10]. This study introduced new estimator for analysis of two contemporaneously correlated response event count variables. The SUPREME was applied on presidential vetoes dataset from 1946 to 1984 by presenting a full information maximum likelihood method for simultaneously estimating seemingly unrelated Poisson regression models.

In another development [2], dispersion insensitive truncated models for zero-inflated regression models is recognized for handling datasets that are truncated, zero-inflated truncated, zero- and k-inflated exponentiated exponential geometric model. The concept behind constructing these distributions has been adapted to model specific types of data that exhibit inflation at the upper boundary of the support. A notable example is the zero and k-inflated truncated Poisson regression model, which is built upon the Poisson distribution. However, due to the Poisson distribution's inherent equi-dispersion property, this model is less effective for data exhibiting either under-dispersion or over-dispersion.

This research work proposes an efficient method of estimating a seemingly unrelated exponentiated exponential geometric regression model for analysis of two correlated explained event count variables. This new model or estimator is intended to solve two problems of event count data analysis. The first limitation is that estimating equations individually with exponentiated exponential geometric models rules out potential efficiency improvements obtainable through [14] “Seemingly Unrelated Regression” framework. In contract, applying Zellner's full information maximum likelihood method necessitates assuming normally distributed disturbances and linear or log-linear functional form for each equation. For event count data, these assumptions are implausible and may result in substantial efficiency losses, as well as bias and inconsistency.

A second limitation involves conducting cross-equation hypothesis tests with event count data, such as testing the equality of coefficients across different equations.

2. Models

2.1. Bivariate exponentiated exponential geometric regression model

The first derived the bivariate exponentiated exponential geometric regression (BEEGR) [1,7] this model follow a bivariate exponentiated exponential geometric distribution (BEEGD) which could be defined using the method of bivariate Sarmanov distribution [13]. The BEEGD probability mass function is given by

$$\mathrm{P}(\mathrm{y}_1, \mathrm{y}_2) = \prod_{i=1}^{2} \left[\left(1-\theta_i^{y_1+1}\right)^{b_t} - \left(1-\theta_i^{y}\right)^{b_t} \right] \times \left[1+\lambda(e^{-y_1}-c_1)(e^{-y_2}-c_2)\right] \tag{1}$$

Where $c_i = \mathrm{E}\ (e^{-y_i})$ for i = 1, 2, where c_i can be found using moment-generating function [9]

2.2. Seemingly unrelated exponentiated exponential geometric regression model

Suppose there are two regression equations

$$\mathrm{Y}_{1\mathrm{t}} = \beta_{11}X_{1t} + \beta_{12}X_{2t} + \ldots\beta_{1k}X_{nt} + u_{1t} \tag{2}$$

$$\mathrm{Y}_{2\mathrm{t}} = \beta_{21}X_{1t} + \beta_{22}X_{2t} + \ldots\beta_{2k}X_{nt} + u_{2t} \tag{3}$$

Where t = 1, 2 … n. is an individual observation and there is possibility of exclusion of any independent variable from the equations.

In matrix form,

$$\begin{bmatrix} \mathrm{Y}_{1t} \\ \mathrm{Y}_{2t} \end{bmatrix} = \begin{bmatrix} \beta_{11} & \beta_{12} & 0 \\ \beta_{21} & 0 & \beta_{23} \end{bmatrix} \begin{bmatrix} X_{1t} \\ X_{2t} \\ X_{3t} \end{bmatrix} \tag{4}$$

Specifically,

$$\mathrm{Y}_{1\mathrm{t}} = \beta_{11}X_{1t} + \beta_{12}X_{2t} + u_{1t} \tag{5}$$

$$\mathrm{Y}_{2\mathrm{t}} = \beta_{21}X_{1t} + \beta_{23}X_{3t} + u_{2t} \tag{6}$$

$$\mathrm{Y} = \beta' X \tag{7}$$

$\mathrm{Y}_{1\mathrm{t}}$, $\mathrm{Y}_{2\mathrm{t}}$ are the endogenous variables for each equation.

X_{1t}, X_{2t} are the covariates variables for each equation and

β's are the parameter or coefficient vectors

To define the proposed Seemingly Unrelated Exponentiated Exponential Geometric Regression (SUEEGR). Then, taking $f(x_{it}, \beta_\mathrm{i})$ to be logit function $f(x_{it}, \beta_\mathrm{i})$,

$$\theta_i(x_{it}) = \theta_i = f(x_{it}, \beta_\mathrm{i}) = \exp\left(x'_{it}\beta_\mathrm{i}\right) / \left[1+\left(x'_{it}\beta_\mathrm{i}\right)\right] = 1 / \left[1+\exp-\left(x'_{it}\beta_\mathrm{i}\right)\right] \tag{8}$$

$$\frac{\theta_i - 1}{\theta_i} = exp\left(x'_{it}\beta_\mathrm{i}\right) \tag{9}$$

$$\theta_1 = f(x_{1t}, \beta_1) = \frac{1}{1-\exp\left(x'_{1t}\beta_{1)}\right.} \tag{10}$$

$$\theta_2 = f(x_{2t}, \beta_2) = \frac{1}{1-\exp\left(x'_{2t}\beta_{2)}\right.} \tag{11}$$

let θ_{1t} and θ_{2t} vary over t observations (t = 1, …,n). Then also assume that

$\mathrm{Y}_{1\mathrm{t}} \sim \mathrm{EEGR}\ (\theta_{1t})$

$\mathrm{Y}_{2\mathrm{t}} \sim \mathrm{EEGR}\ (\theta_{2t})$

$\mathrm{U} \sim \mathrm{EEGR}\ (\xi)$

$$\text{Where } \xi = \mathrm{Cov}(\mathrm{Y}_{1\mathrm{t}}, \mathrm{Y}_{2\mathrm{t}}) = 1+\lambda(e^{-y_{1t}}-c_1)(e^{-y_{2t}}-c_2) \tag{12}$$

θ_{1t} and θ_{2t} are the expected value of the endogenous variables at observation t, be the exponential functions of separated linear combinations of the independent variables.

$\mathrm{Y}_{1\mathrm{t}}$, $\mathrm{Y}_{2\mathrm{t}}$, and U are independent at observation t, observations t and j(t $\neq j$) such that all the three random variables are uncorrelated among each other and themselves. By using the useful property of covariance before the bivariate exponentiated exponential geometric by factoring the product of two marginal EEG distributions, under the condition that $\xi = 1$.

Thus Y_{1t} and Y_{2t} are distributed as seemingly unrelated exponentiated exponential geometric with parameter θ_i in (8) is a function of (X_{1t}, X_{2t}) given that $\theta_i(x_{it}) = f(x_{it}, \beta_\mathrm{i})$, where 0< $f(x_{it}, \beta_\mathrm{i})$ < 1 is differentiable with respect to β_i, a $(1 \times k)$ vector parameter. Let Y_{it} (i = 1, 2; t = 1, 2, …n, where n is the sample size) be count endogenous (response) variable, and let $x_{it} = ((x_{i0} = 1, x_{i1}, x_{i2}, \ldots, x_{i(k-1)})$ be a column vector with $k\ -1$ predictors. It is very difficult to determine the exogenous that affect the endogenous variable Y_{1t} and Y_{2t} But we assume that the same exogenous affect the response variable and the vector parameters β_1 and β_2 are not assumed to be same.

The SUEEGR model is given by

$$P(Y_1, Y_2, \beta_1\beta_2|x_{1t}, x_{2t}) = \left[\left(1-[\theta_{1t}(x_{it})]^{y_{1t}+1}\right)^b - \left(1-[\theta_{1t}(x_{1t})]^{y_{1t}}\right)^b\right] \left[\left(1-[\theta_{2t}(x_{2t})]^{y_{2t}+1}\right)^c - \left(1-[\theta_{2t}(x_{2t})]^{y_{2t}}\right)^c\right] \times \left[1+\lambda(e^{-y_1}-c_1)(e^{-y_2}-c_2)\right] \tag{13}$$

Where θ_i is given by (3.14), the expected value of the endogenous variables at observation t, be the exponential functions of separated linear combinations of the independent variables.

x'_{1t} and x'_{2t} are vectors of k_1 and k_2 independent variables, β_1 and β_2 are coefficient vectors.

$$c_i = E\left(e^{-y_i}\right)$$

b and c are the shape parameter been taken as nuisance parameter which is also the dispersion parameter.

3. Estimation

The method of maximum likelihood will be used for the estimation of SUEEGR model parameter. By finding the partial derivative of the model with respect to the parameters. Suppose a random sample of size n is to be considered from SUEEGR model in (13). The log-likelihood function of SUEEGR model in (13) is given by:

$$\mathfrak{l}(\theta_{1t}, \theta_{2t}, b, c, \lambda, c_1, c_2) = \mathfrak{l} = \prod_{t=1}^{n} P(Y_{1t}, Y_{2t}, |\beta_1, \beta_2, \xi)$$

$$= \left[\left(1-\theta_{1t}^{y_{1t}+1}\right)^b - \left(1-\theta_{1t}^{y}\right)^b\right]\left[\left(1-\theta_{2t}^{y_{2t}+1}\right)^c - \left(1-\theta_{2t}^{y_{2t}}\right)^c\right] \times \xi$$

$$\mathrm{Log}\mathfrak{l} = \sum_{t=1}^{n}\left[\log\left[\left(1-\theta_{1t}^{y_{1t}+1}\right)^b - \left(1-\theta_{1t}^{y}\right)^b\right] + \log\left[\left(1-\theta_{2t}^{y_{2t}+1}\right)^c - \left(1-\theta_{2t}^{y_{2t}}\right)^c\right] + \log\xi\right] \tag{14}$$

Where $\xi = \mathrm{cov}(Y_1,Y_2) = 1+\lambda(e^{-y_{1t}} - c_1)(e^{-y_{2t}} - c_2)$

Taking the first partial derivatives of the log-likelihood function for θ_{1t}, θ_{2t}, b, c, c_1, c_2, λ

$$\frac{\partial ln\,\mathfrak{l}}{\partial\theta_{1t}} = \sum_{t=1}^{n}\frac{\left[b y_{1t}\theta_{1t}{}^{y_{1t}-1}\left(1-\theta_{1t}{}^{y_{1t+1}}\right)^{b-1} - b(y_{1t}+1)\,\theta_{1t}{}^{y_{1t}}\left(1-\theta_{1t}{}^{y_{1t}+1}\right)^{b-1}\right]}{\left[\left(1-\theta_{1t}{}^{y_{1t}+1}\right)^b - \left(1-\theta_{1t}{}^{y_{1t}}\right)^b\right]} \tag{15}$$

$$\frac{\partial ln\,\mathfrak{l}}{\partial\theta_{2t}} = \sum_{t=1}^{n}\frac{\left[c y_{2t}\theta_{2t}{}^{y_{2t}-1}\left(1-\theta_{2t}{}^{y_{2t+1}}\right)^{c-1} - c(y_{2t}+1)\,\theta_{2t}{}^{y_{2t}}\left(1-\theta_{2t}{}^{y_{2t}+1}\right)^{c-1}\right]}{\left[\left(1-\theta_{2t}{}^{y_{2t}+1}\right)^c - \left(1-\theta_{2t}{}^{y_{2t}}\right)^b\right]} \tag{16}$$

$$\frac{\partial ln\,\mathfrak{l}}{\partial b} = \sum_{t=1}^{n}\frac{\left(1-\theta_{1t}{}^{y_{1t}+1}\right)^b \ln\left(1-\theta_{1t}{}^{y_{1t}+1}\right)\Big] - \left(1-\theta_{1t}{}^{y_{1t}}\right)^b \ln\left(1-\theta_{1t}{}^{y_{1t}}\right)}{\left(1-\theta_{1t}{}^{y_{1t}+1}\right)^b - \left(1-\theta_{1t}{}^{y_{1t}}\right)^b} \tag{17}$$

$$\frac{\partial ln\,\mathfrak{l}}{\partial c} = \sum_{t=1}^{n}\frac{\left(1-\theta_{2t}{}^{y_{2t}+1}\right)^c \ln\left(1-\theta_{2t}{}^{y_{2t}+1}\right)\Big] - \left(1-\theta_{2t}{}^{y_{2t}}\right)^c \ln\left(1-\theta_{2t}{}^{y_{2t}}\right)}{\left(1-\theta_{2t}{}^{y_{2t}+1}\right)^c - \left(1-\theta_{2t}{}^{y_{2t}}\right)^c} \tag{18}$$

$$\frac{\partial \mathfrak{l}}{\partial c_1}=\frac{-\lambda(e^{-y_2}-c_2)}{1+\lambda(e^{-y_1}-c_1)(e^{-y_2}-c_2)} \tag{19}$$

$$\frac{\partial \mathfrak{l}}{\partial c_2}=\frac{(e^{-y_1}-c_1)(e^{-y_2}-c_2)}{1+\lambda(e^{-y_1}-c_1)(e^{-y_2}-c_2)]} \tag{20}$$

Setting the partial derivatives of (15-20) to zero and solving simultaneously for each parameter give maximum likelihood estimates of the parameters. Unfortunately, the closed form solution for the parameters does not exist hence an iterative solution has been carried out in method “Optim” of the R development software by R core Team [16] is to be considered. The fisher's information matrix may be used to find the second partial derivatives of the log-likelihood of (15–20) [2,9].

4. Simulation study

Here, the maximum likelihood estimates of the parameters will be examined using the Monte Carlo simulation procedure. The simulation is carried out for different values of the likelihood parameter and different sample size as n = 20, 50, 100 and 200. The experiment is replicated 500 times of each of the sample size. The value of $\beta_{11}, \beta_{12}, \beta_{21}, and\ \beta_{22}$ used for the study are 0.5, −0.3, −0.2 and 0.4 respectively for both under and over dispersed. Where b and c are set to be (0.5, 1.5); (1.2, 1.8) and (1.8, 0.5) for over-dispersion. Dispersed parameter b and c are set to be (4.5, 7.5); (5.0, 10.0) and (4.5, 10.0) for under-dispersion. Lastly, dispersed parameter b and c are set to be (1.0, 1.0) for special case when SUEEGR reduce to SU-geometric regression. The model performance is obtained using the total mean square error (TMSE) and total absolute bias (TAB) with standard errors of each of the parameter estimate and the total bias is given as

$$\mathrm{TAB}=\frac{1}{r}\sum_{i=1}^{r}\left(|\beta_{1t}-\widehat{\beta_{1t}}\,|+|\beta_{2t}-\widehat{\beta_{2t}}|\right) \tag{21}$$

Where t = 1, 2, …n and β_{t1}, β_{t2} are true values of the regression parameter. $\widehat{\beta_{1t}}$, $\widehat{\beta_{2t}}$ are the estimated value of the regression parameter with replication and r is the replication number since the data are generated when the regression parameter are fixed. The standard error give the amount of variability that exist from each parameter. This is obtained by taking the square of the leading diagonal of the inverse of the fisher's information matrix.

The simulation results are presented in Tables 1–3, the estimates, bias, mean square error, average standard error, TABS, TMSE of the regression parameter are generated for the SUEEGR for over dispersed model are given in Table 1. Simulation results for over-dispersion indicate the performance of parameter estimation for the SUEEGR under different shape parameter set. When n = 20, 50, 100, 200 the estimate of b_{12} and b_{21} are negative for all the value of dispersed parameter b and c, the remaining parameter estimates are positive.

When n = 20 for the over dispersed biases, all the beta's parameter are negatives across the true shape parameter b and c except for bias estimate for b_{12} and b_{21} where b = 1.2 and c = 1.8 but the biases for the estimate shape parameter are positive. For n = 100, across the true value of b and c, there are negatives biases estimate for b_{11} and b_{21}. When n = 200, across the true value of b and c, there are negatives biased estimate values of b_{11}, b_{12} and b_{21}.

Total absolute biases show that for all sample size expect for n = 50; the true shape parameter 1.2 and 1.8 had the minimum absolute biased estimate, follow by the true shape parameter (0.5, 1.5) and (1.8, 0.5) but when n = 50 the true shape dispersed with (0.5, 1.5) has the minimum absolute estimate follow by the (1.2, 1.8) and (1.8, 0.5).

Total mean square error show that when n = 20 and 200 the minimum TSME estimate is when the true parameter is (1.2, 1.8), follow by (0.5, 1.5) and (1.2 and 1.8). But when n = 50 and 100 the minimum (TSME) estimate is when the dispersed parameter is (0.5, 1.5). When comparing average standard errors and biases, the parameter estimates show that biased estimates are more efficient across all sample sizes. For all dispersed values, the bias estimates are smaller than the corresponding standard errors.

In Table 2. Simulation result for under dispersed indicate the performance of parameter estimation for the SUEEGR under different shape parameter set. The estimate b_{12} and b_{21} for all the dispersed value set for all the sample sizes are negatives but the remaining estimate for beta's and dispersion are positive.

For biased estimates when n = 20, the biased estimate for b_{11}, b_{12}, b_{21} are negatives except for when the dispersed value are (4.5, 10.0), the remaining biased estimates are positive. When n = 100, b_{11}, b_{21} are negatives while b_{12}, b_{22} are positive with the dispersed estimate when the true value of dispersion are (4.5, 7.5), when the true dispersed value are (5.0, 10.0) b_{11}, b_{12} and b_{21} are

Table 1. Simulation output of seemingly unrelated exponentiated exponential geometric regression for over-dispersion.

N	b	c	$\widehat{b_{11}}$	$\widehat{b_{12}}$	$\widehat{b_{21}}$	$\widehat{b_{22}}$	$\widehat{b}$	$\widehat{c}$	TABS	TMSE
20	0.5	1.5	0.0083	−0.3128	−0.4749	0.3999	13.4761	2.2814	14.5371	27014.3
		Bias	−0.4917	−0.0128	−0.2749	−0.0001	12.9762	0.7814		
		Mse	1.2595	1.0819	0.7670	0.5550	27007.37	3.2701		
		Avg.se	0.7139	0.7924	0.5383	0.5319	70.1427	1.3524		
	1.2	1.8	0.2762	−0.2953	−0.4576	0.4175	1.6151	2.7073	1.8260	7.5205
		Bias	−0.2238	0.0047	−0.2576	0.0175	0.4151	0.9073		
		Mse	0.3987	0.5702	0.6898	0.4970	0.9157	4.4492		
		Avg.se	0.4303	0.4849	0.4911	0.4825	0.7382	1.5185		
	1.8	0.5	0.3258	−0.3072	−1.6844	0.3926	2.3880	179.5381	181.2992	726789.2
		Bias	−0.1741	−0.0072	−1.4844	−0.0074	0.5879	179.0381		
		Mse	0.3485	0.4746	11.4563	3.3842	2.19239	726771.3		
		Avg.se	0.3655	0.3991	2.1197	1.1090	1.0364	1368.2209		
50	0.5	1.5	0.3596	−0.3039	−0.3134	0.3865	0.5881	1.7677	0.6270	2.1573
		Bias	−0.1404	−0.0040	−0.1137	−0.013	0.0881	0.2677		
		Mse	0.4060	0.5237	0.3609	0.2942	0.0655	0.5070		
		Avg.se	0.3902	0.4476	0.3189	0.3143	0.2136	0.5973		
	1.2	1.8	0.4254	−0.3088	−0.2990	0.3843	1.3449	2.1025	0.6453	2.2221
		Bias	−0.0745	−0.0088	−0.0990	−0.0157	0.1449	0.3025		
		Mse	0.3465	0.4154	0.3413	0.2662	0.1900	0.6627		
		Avg.se	0.2624	0.2931	0.2945	0.2892	0.3680	0.6754		
50	1.8	0.5	0.4548	−0.3056	−0.5323	0.3662	1.9370	2.9942	3.0480	2521.945
		Bias	−0.0452	−0.0056	−0.3322	−0.0338	0.1370	2.4942		
		Mse	0.3397	0.3852	0.9942	0.5204	0.2785	2519.427		
		Avg.se	0.2263	0.2469	0.5721	0.5372	0.4992	15.4240		
100	0.5	1.5	0.4258	−0.2796	−0.2370	0.4166	0.5410	1.5892	0.2783	1.4203
		Bias	−0.0741	0.0204	−0.0370	0.0166	0.0410	0.0892		
		Mse	0.3512	0.3946	0.2552	0.2426	0.0200	0.1566		
		Avg.se	0.2709	0.3093	0.2223	0.2215	0.1357	0.3701		
	1.2	1.8	0.4650	−0.2910	−0.2365	0.4142	1.2624	1.9103	0.2676	1.4649
		Bias	−0.0350	0.0090	−0.0365	0.0143	0.0624	0.1103		
		Mse	0.3398	0.3546	0.2504	0.2372	0.07113	0.2118		
		Avg.se	0.1846	0.2056	0.2062	0.2043	0.2418	0.4251		
	1.8	0.5	0.4641	−0.2933	−0.3101	0.4274	1.9131	0.5793	0.3726	1.4495
		Bias	−0.0359	0.0067	−0.1101	0.0275	0.1130	0.0793		
		Mse	0.1131	0.3487	0.4309	0.3581	0.1354	0.0633		
		Avg.se	0.1584	0.1713	0.3717	0.3713	0.3451	0.2082		
200	0.5	1.5	0.4518	−0.3058	−0.2176	0.4018	0.5277	1.5488	0.1498	1.2061
		Bias	−0.0481	−0.0059	−0.0176	0.0018	0.0277	0.0488		
		Mse	0.3210	0.3782	0.2119	0.2066	0.0098	0.0695		
		Avg. se	0.1895	0.2154	0.1558	0.1549	0.0927	0.2527		
	1.2	1.8	0.4802	−0.3010	−0.2146	0.4022	1.2340	1.8513	0.1229	1.1994
		Bias	−0.0197	−0.0010	−0.0146	0.0022	0.0340	0.0514		
		Mse	0.3253	0.3440	0.2100	0.2019	0.0290	0.0891		
		Avg. se	0.1297	0.1441	0.1448	0.1433	0.1661	0.2886		
	1.8	0.5	0.4816	−0.3027	−0.2488	0.4091	1.8612	0.5351	0.1755	1.2760
		Bias	−0.0183	−0.0027	−0.0488	0.0091	0.0612	0.0351		
		Mse	0.3174	0.3390	0.2761	0.2585	0.0644	0.0205		
		Avg.se	0.1118	0.1208	0.2570	0.2575	0.2362	0.1315		

negatives but the rest estimates are positives, when the true dispersed value are (4.5, 10.0) the biased estimate of beta's are negatives except b_{12} and the dispersed value are positives. When n = 200, the estimate for b_{11} and b_{21} are negatives for all true dispersed values, but when dispersed true value are (5.0, 10.0) the biased estimate for b_{12} was also negative.

Total absolute biased, when n = 20, 50, 100 and 200; with the true shape parameter is (4.5, 7.5) have the smallest or minimum absolute biased estimate but the results follow the same sequence when n = 20, 100, and 200.

Total mean square error for under-dispersed scenarios, when n = 50, 100 and 200 the minimum. TMSE estimate was when the true parameter is (4.5, 7.5), but when n = 20 the minimum TMSE estimate is when the true parameter was (4.5, 10.0). When comparing the average standard error and biased it is clearly shown that the biased estimate are more

Table 2. Simulation output of seemingly unrelated exponentiated exponential geometric regression for under-dispersion.

N	b	c	$\widehat{b_{11}}$	$\widehat{b_{12}}$	$\widehat{b_{21}}$	$\widehat{b_{22}}$	$\hat{b}$	$\hat{c}$	TABS	TMSE
20	4.5	7.5	0.3376	−0.3083	−0.3695	0.4234	6.8732	13.2961	8.5328	261.7042
		bias	−0.1624	−0.0083	−0.1695	0.0233	2.3732	5.7961		
		Mse	0.3095	0.3983	0.4233	0.2893	25.1158	235.1679		
		Avg. se	0.2909	0.2690	0.3372	0.2715	3.4465	8.8316		
	5.0	10.0	0.3468	−0.2903	−0.3695	0.4175	7.5843	18.9604	11.8945	616.0758
		bias	−0.1533	−0.1695	−0.1695	0.0175	2.5843	8.9604		
		se	0.3152	0.3731	0.4211	0.2562	30.5343	584.1759		
		Avg. se	0.2856	0.2572	0.3273	0.2454	3.8627	13.56901		
	4.5	10.0	0.3376	−0.3082	−0.3694	0.4175	6.8735	18.9601	11.6912	610.3519
		bias	−0.1624	−0.0083	−0.1695	0.0175	2.3735	8.9601		
		Mse	0.3095	0.3983	0.4211	0.2562	25.1231	583.8437		
		Avg. se	0.2909	0.2690	0.3273	0.2454	3.4468	13.5676		
50	4.5	7.5	0.4453	−0.2988	−0.2645	0.3997	5.2099	8.7988	2.1311	15.7928
		Bias	−0.0546	0.0012	−0.0663	−0.0003	0.7099	1.2988		
		Mse	0.3182	0.3406	0.2670	0.2083	3.2773	11.3814		
		Avg. se	0.1803	0.1692	0.2050	0.1681	1.4733	2.9764		
	5.0	10.0	0.4454	−0.2988	−0.2645	0.3989	5.2100	12.2275	3.0591	31.8801
		Bias	−0.0547	0.0012	−0.0645	−0.0011	0.7100	2.2275		
		Mse	0.3182	0.3406	0.2637	0.1986	3.2782	27.481		
		Avg,se	0.1803	0.1695	0.1994	0.1520	1.4734	4.4560		
	4.5	10.0	0.4519	−0.3062	−0.2645	0.3989	5.7525	12.2273	3.0997	33.0155
		Bias	−0.0481	−0.0062	−0.0645	−0.0013	0.7525	2.2272		
		Mse	0.3197	0.3609	0.2637	0.1986	4.3958	27.4768		
		Avg. se	0.1777	0.1635	0.1994	0.1520	1.6634	4.4559		
100	4.5	7.5	0.4739	−0.2970	−0.2294	0.4098	4.8033	8.1004	0.9720	6.5416
		Bias	−0.0261	0.0030	−0.0293	0.0098	0.3033	0.6004		
		Mse	0.3174	0.3222	0.2193	0.2009	1.1166	4.3652		
		Avg. se	0.1269	0.1197	0.1435	0.1189	0.9326	1.8701		
	5.0	10.0	0.4717	−0.3009	−0.2256	0.4050	5.3757	10.8882	1.3238	11.9156
		Bias	−0.0283	−0.0009	−0.0256	0.0050	0.3757	0.8882		
		Mse	0.3175	0.3357	0.2153	0.1959	1.4950	9.3562		
		Avg. se	0.1251	0.1150	0.1397	0.1083	1.0690	2.6908		
	4.5	10.0	0.4739	−0.2970	−0.2256	0.4050	4.8034	10.8885	1.2515	10.6038
		Bias	−0.0260	0.0030	−0.0256	−0.0256	0.3034	0.8885		
		Mse	0.3174	0.3220	0.2153	0.1959	0.1959	9.3572		
		Avg. se	0.1269	0.1197	0.1397	0.1083	0.9327	2.6909		
200	4.5	7.5	0.4879	−0.2989	−0.2106	0.4029	4.6268	7.8089	0.4624	3.2352
		Bias	−0.0121	0.0011	−0.0106	0.0028	0.1268	0.3089		
		Mse	0.3222	0.3280	0.1970	0.1881	0.4565	1.7434		
		Avg. se	0.0896	0.0846	0.1007	0.0835	0.6268	1.2510		
	5.0	10.0	0.4871	−0.3038	−0.2094	0.4034	5.1813	10.3724	0.5832	5.1395
		Bias	−0.0128	−0.0038	−0.0094	0.0033	0.1813	0.3724		
		Mse	0.3194	0.3312	0.1965	0.1921	0.5657	3.5345		
		Avg. se	0.0882	0.0812	0.0983	0.0763	0.7179	1.7750		
	4.5	10.0	0.4879	−0.2989	−0.2094	0.4034	4.6268	10.37204	0.5248	5.0296
		Bias	−0.0121	0.0011	−0.0094	0.0034	0.1268	0.3720		
		Mse	0.3222	0.3280	0.1965	0.1921	0.4565	3.5343		
		Avg. se	0.0896	0.0846	0.0982	0.0763	0.0763	0.6268		

efficient than the average standard error because it has the least estimate than the average standard error.

In Table 3. Simulation result for SU-geometric regression indicate the performance of parameter estimation when the shape parameter set b and c are 1.

The simulation output when SUEEGR is been reduced to SU-geometric regression, when $n = 20$, 50, 100 and 200 the estimate of b_{12} and b_{21} are negatives, the remaining estimates for beta's and dispersion parameter are positive. Across the biased estimates, b_{11}, b_{12} and b_{21} are negative except when $n = 100$ which b_{12} is positive, when $n = 50$ all the biased estimate are negative but the dispersed biased estimate are positive.

By comparing the estimates of bias, mean square errors (MSE), and average standard errors across all sample sizes, it was observed that the biased beta estimates and dispersed estimates showed

Table 3. Simulation output when SUEEGR reduces to SU-geometric regression.

N	b	c	$\widehat{b_{11}}$	$\widehat{b_{12}}$	$\widehat{b_{21}}$	$\widehat{b_{22}}$	$\widehat{b}$	$\widehat{c}$	TABS	TMSE
20	1	1	0.2441	−0.3017	−0.6403	0.4243	1.3860	7.5282	7.6363	9456.605
		Bias	−0.2559	−0.0017	−0.4402	0.0243	0.3860	6.5281		
		Mse	0.4463	0.6160	1.5064	0.8689	0.7891	9452.378		
		Avg. se	0.4663	0.5296	0.7123	0.6632	0.6632	67.8415		
50	1	1	0.4207	−0.3057	−0.3532	0.3764	1.1170	1.2234	0.6022	2.0722
		Bias	−0.0793	−0.0057	−0.1532	−0.0237	0.1172	0.2234		
		Mse	0.3537	0.4261	0.4551	0.34351	0.1374	0.3566		
		Avg. se	0.2836	0.3198	0.3860	0.3802	0.3191	0.4807		
100	1	1	0.4569	−0.2853	−0.2520	0.4245	1.0582	1.0744	0.2671	1.4124
		Bias	−0.0431	0.0148	−0.0520	0.0245	0.0582	0.0745		
		Mse	0.3409	0.3577	0.2936	0.2748	0.0506	0.0949		
		Avg. se	0.1986	0.2229	0.2659	0.2663	0.2114	0.2859		
200	1	1	0.4752	−0.3010	−0.2209	0.4021	1.0330	1.0346	0.1163	1.1786
		Bias	−0.0209	−0.0010	−0.0209	0.0021	0.0330	0.0346		
		Mse	0.3266	0.3483	0.2225	0.2198	0.0222	0.0393		
		Avg. se	0.1397	0.1563	0.1857	0.1859	0.1451	0.1920		

consistent patterns. The biased estimates generally had the minimum values, except when n = 200, where the MSE for the dispersed parameter b was lower. For total absolute bias (TABS) and total mean square error (TMSE), the estimates decreased as the simulation sample size increased, reaching their minimum values at n = 200.

5. Real life data application and results

The data set represent the survey conducted by the Demographic and Health Survey (DHS) 2018, through interviews of 33,924 women who were under age 15years to 44 years old residing across all the 36 states in Nigeria including FCT. The count data response variable are Antenatal Care visits (the number of antenatal visit during pregnancy) with 12,132 missing variable and child vaccination (Polio inactive received) with 15,632 missing variable. After removing all the cases of missing observation, we have 16,291 observations.

There are six (6) independent variables used from the data set as follows: maternal education level, which are in four levels (from 0 "no education" to 4 "higher education"); household wealth index with five levels (from 1 "poorest" to 5 "richest"); residences with two levels (1 for "urban" and 2 for "rural"); mother's age with six levels (from 1 "15—19" to 6 "40—44"); access to health information that is household that has radio (ranging from 0 "no" to 1 "yes") and number of living children. All the missing variable are excluding in the analysis. These independent variables capture key social, economic and environmental factors that can influence both antenatal care visits and vaccination rates, using this variable in two-equation SUR model allows for exploring how each factor influences maternal and child healthcare behaviours.

The descriptive statistics for the variables used in the analysis is shown in Table 4. The two dependent variables exhibit over-dispersion suggesting the sample variance are greater than the mean which make the dataset for the dependent variables to have a high variability.

However, the data set has a higher level of kurtosis within the independent variable compare to the kurtosis of the dependent variable. Only the variable of the residence is negative, showing that it is left-skewed, which other variables are positive suggesting a right-skewed. A simple correlation for the two response variable Antenatal care visits (the number of antenatal visit during pregnancy) and the child vaccination (Polio inactive received) is computed, and it is 0.1594 with the p-value less than 0.0000 at 0.05 level of significance which shown that the two variables are significant correlated.

Table 5 shows the parameters, estimates, standard error, t-statistic, p-value and significance level of each estimate when maternal education level, household wealth Index, urban/rural residences, mother's age, access to health information and number of living children as covariates in the models for SUEEGR. At the 0.05 level of significance, all parameter estimates are statistically significant, confirming that the independent variables are generally associated with the dependent variable. However, maternal education and urban/rural residence were not significant when included as covariates in the SUEEGR model, indicating that these factors exert a substantial influence on the model's estimates.

In Table 6, this table presents Seemingly Unrelated Geometric Poisson Regression (SUGPR) model for the Demographic and Health Survey (DHS), the maternal education level, household

Table 4. Descriptive statistics for the real life dataset (n = 16,291).

variable	Description of variable	mean	Standard deviation	proportion	skewness	kurtosis
Antenatal	Antenatal Care visit	6.0199	12.1374		6.5439	49.5590
Vchild	Child Vaccination	0.7252	0.9827		2.2754	14.8305
Education	Maternal education level			0.4537 (no education) 0.1555 (primary) 0.3131 (secondary) 0.0777 (higher)	0.3872	1.7082
Wealth	Household wealth Index			0.2378 (poorest) 0.2282 (poorer) 0.2113 (middle) 0.1817 (richer) 0.1408 (richest)	0.2109	1.8179
Residences	Urban/rural residences			0.6551 (rural)	−0.6528	1.4261
Age	Mother's Age	3.5848	1.3964		0.3511	2.5771
Health	Access to health information			0.4318 (no access)	4.7548	35.3546
Lchild	Number of living Children	3.6624	2.1179		0.7734	3.2849

Table 5. Seemingly Unrelated Exponentiated Exponential Regression (SUEEGR) model for the Demographic and Health Survey (DHS).

Parameter	estimate	Standard error	t-statistic	p-value	Significance (5 %)
$\widehat{b_{11}}$	0.0128	0.0119	1.08	0.280	Not significant
$\widehat{b_{12}}$	0.3002	0.0073	41.04	0.0001	Significant
$\widehat{b_{13}}$	0.0153	0.0101	1.52	0.128	Not significant
$\widehat{b_{14}}$	0.1388	0.0127	10.91	0.0001	Significant
$\widehat{b_{15}}$	0.3556	0.0110	32.20	0.0001	Significant
$\widehat{b_{16}}$	0.0118	0.0055	2.13	0.033	Significant
$\widehat{b_{21}}$	0.2212	0.0155	14.26	0.0001	Significant
$\widehat{b_{22}}$	−0.0262	0.0097	−2.70	0.007	Significant
$\widehat{b_{23}}$	0.0162	0.0124	1.30	0.194	Not significant
$\widehat{b_{24}}$	−0.4115	0.0172	−23.93	0.0001	Significant
$\widehat{b_{25}}$	0.2536	0.0145	17.44	0.0001	Significant
$\widehat{b_{26}}$	−0.0850	0.0072	−11.77	0.0001	Significant

Table 6. Seemingly unrelated geometric Poisson regression (SUGPR) model for the Demographic and Health Survey (DHS).

Parameter	estimate	Standard error	t-statistic	p-value	Significance (5 %)
$\widehat{b_{11}}$	0.0962	0.0088	10.88	0.0001	Significant
$\widehat{b_{12}}$	0.2753	0.0074	37.04	0.0001	Significant
$\widehat{b_{13}}$	0.0119	0.0100	1.18	0.238	Not significant
$\widehat{b_{14}}$	0.0953	0.0101	9.49	0.0001	Significant
$\widehat{b_{15}}$	0.3360	0.0111	30.30	0.0001	Significant
$\widehat{b_{16}}$	−0.0341	0.0062	−5.47	0.0001	Significant
$\widehat{b_{21}}$	0.0416	0.0116	3.60	0.0001	Significant
$\widehat{b_{22}}$	−0.0477	0.0101	−4.72	0.0001	Significant
$\widehat{b_{23}}$	0.0165	0.0125	1.32	0.186	Not significant
$\widehat{b_{24}}$	−0.5269	0.0163	−32.33	0.0001	Significant
$\widehat{b_{25}}$	0.2644	0.0148	17.88	0.0001	Significant
$\widehat{b_{26}}$	−0.0393	0.0082	−4.79	0.0001	Significant

wealth Index, urban/rural residences, mother's age, access to health information and number of living children are the covariates of the models for SUGPR.

The parameter are significant when p-value is less than 0.05 level of significance for the SUGPR which indicate that there is a meaningful relationship with the dependent variables expect when urban/rural residence was used as a covariates in the SUGPR that is not significant in the SUGPR model. This also indicate that urban/rural residence covariate has a very strong effect in SUGPR model estimates.

In Table 7, this table presents dispersion parameter with standard error, log-likelihood, AIC and BIC for both SUEEGR and SUGPR. For the comparison the dispersed parameter standard error of both model it appears that the SUGPR and SUEEGR are not far apart they are closed together which indicated that they both measure over-dispersion.

Based on the fit statistics (Log-likelihood, AIC, BIC), the SUEEGR model appears to provide a better overall fit of the data because of the higher Log-likelihood, lower AIC and BIC of the model.

6. Discussion

Based on the findings, the Seemingly Unrelated Exponentiated Exponential Geometric Regression (SUEEGR) model was developed. A simulation study was conducted for under-dispersion, over-dispersion, and the special case where the SUEEGR model reduced to the SU-geometric regression (SUGPR) model. The results showed that the bias estimates were more efficient than the average standard errors for all sample sizes across the three scenarios. The 2018 Domestic and Health Survey data was used to apply the model. It was observed that the majority of the dataset exhibited over-dispersion, indicating that most of the variable sample variances were greater than their means, resulting in high variability. Additionally, the dependent variables displayed high kurtosis compared to the independent variables, suggesting that the response variables had wider tails than the explanatory variables. The regression parameters were estimated, and most of the parameter estimates for both the SUEEGR and SUGPR models were significant. However, in the SUEEGR model, the parameters $\widehat{b_{11}}(0.280), \widehat{b_{13}}(0.128), \widehat{b_{23}}$ (0.194) that are not significant in the SUEEGR model and only the parameter of $\widehat{b_{13}}(0.238)$ $\widehat{b_{23}}$ (0.186) are not significant in the SUGPR. A comparative analysis was conducted using BIC, AIC, and log-likelihood criteria. The results indicated that the SUEEGR model provided a better fit compared to the SUGPR model.

Table 7. Dispersion parameter with standard error in parentheses of the model and model fit.

Parameter	SUEEGR estimate	SUGPR estimate
$\widehat{b}$	1.0344 (0.0138)	1.0309 (0.0137)
$\widehat{c}$	1.5303 (0.0343)	1.4500 (0.0318)
Log-likelihood	−61955.25	−337035.2
AIC	123938.5	675974.3
BIC	124046.3	676082.1

7. Conclusion

In conclusion, the study successfully developed the Seemingly Unrelated Exponentiated Exponential Geometric Regression (SUEEGR) model, demonstrating its effectiveness in handling scenarios of under-dispersion and over-dispersion. The application of the model to the 2018 Domestic and Health Survey data revealed over-dispersion in the dataset, with most parameter estimates being significant. However, the simulation results showed that the bias estimates were more efficient than the average standard errors for all sample sizes across the three scenarios, a few parameters in both models were not statistically significant. Furthermore, the SUEEGR model outperformed the SUGPR model based on real life application and comparative analyses using BIC, AIC, and log-likelihood.

Source of Funding

No funding received.

Conflict of Interest

None.

Ethical Approval

Not applicable.

Data Availability

The data supporting the findings of this study are available upon request from the corresponding author.

Author Contributions

Oluwaseun Michael Famoni wrote the original draft and implemented the codes. Bamidele Mustapha Oseni conceived the idea, wrote the codes, supervised the work and edited the initial draft. All authors have read and approved the final manuscript.

Acknowledgments

The authors appreciate the DHS program for permission to access data utilized. Useful comments from reviewers that help in improving on the earlier draft are also acknowledged.

References

[1] Alzaatreh A, Lee C, Famoye F. On the discrete analogues of continuous distributions. Stat Methodol 2012;9(6):589–603. https://doi.org/10.1016/j.stamet.2012.03.003.
[2] Oseni BM, Makinde OS, Adepetun AO. Dispersion insensitive truncated models for inflated count data with upper bound, communications in statistics – simulation and computation 2023. https://doi.org/10.1080/03610918.2023.2202843.
[3] Cameron AC, Johansson P. Count data regression on using series expansion: with applications. J Appl Econom 1997; 12(3):203–23. https://doi.org/10.1002/(SICI)1099-1255(199705)12:3<203::AID-JAE446>3.0.CO;2-2.
[4] Jorgenson DW. Multiple regression analysis of a poisson process. Am Stat Assoc 1961;56:235–45.
[5] Frome EL, Kurtner MH, Beauchamp JJ. Regression analysis of Poisson-distributed data. J Am Stat Assoc 1973;68:935–40.
[6] El-sayyad OM. Bayesian and classical analysis of Possion regression. J R Stat Soc Ser B Methodol 1973;35(3):445–51.
[7] Famoye F, Lee Carl. Exponentiated-exponential geometric regression model. J Appl Stat 2017;44(16):2963–77. https://doi.org/10.1080/02664763.2016.1267117.
[8] Famoye F. On the bivariate negative binomial regression model. J Appl Stat 2010;37(6):969–81. https://doi.org/10.1080/02664760902984618.
[9] Famoye F. Bivarate exponentiated-exponential geometric regression model. Stat Neerl 2019;77(3):434–50. https://doi.org/10.1111/stan.12177.
[10] King G. "Discovering statistical patterns in the presidency": seemingly unrelated passion regression model. Sociol Method Res 1989 February;17:235–55.
[11] Lambert D. Zero-inflated poisson regression, with an application to defects in manufacturing. Technometrics 1992;34(1):1–14. https://doi.org/10.2307/1269547.
[12] Sarvi F, Moghimbeigi A, Mahjub H. GEE-based zero-inflated generalized poisson model for clustered over or under-dispersed count data. J Stat Comput Simulat 2019; 89(14):2711–32. https://doi.org/10.1080/00949655.2019.1632857.
[13] Sarmanov OV. Generalized normal correlation and two-dimensional frechet classes. Doklady Akaaemii Nauk 1966; 168:32–5.
[14] Zellner A. An efficient estimation of seemingly unrelated regression equations and tests of aggregation bias. J Am Stat Assoc 1962;57:500–9.
[15] Perrone G, Soffritti G. Seemingly unrelated clusterwise linear regression for contaminated data. Stat Pap 2023;64(3): 883–921. https://doi.org/10.1007/s00362-022-01344-6.
[16] R Core Team. R: A language and environment for statistical computing. Vienna, Austria: R Foundation for Statistical Computing; 2024. https://www.R-project.org/.
[17] Dawoud I. New two parameter inverse gaussian regression estimators: applications and simulation. Jordan J Math Stat 2025;18(1):1–15. Retrieved from, https://jjms.yu.edu.jo/index.php/jjms/article/view/660.