


Generalized Poisson Hurdle Model for Count Data and Its Application in Ear Disease

Guoxin Zuo ¹ , Kang Fu ¹, Xianhua Dai ^{2,3,*} and Liwei Zhang ¹

¹ School of Mathematics and Statistics, Central China Normal University, Wuhan 430079, China; zuogx@mail.ccnu.edu.cn (G.Z.); fukang@mails.ccnu.edu.cn (K.F.); mclimeng@gmail.com (L.Z.)

² School of Public Administration, Central China Normal University, Wuhan 430079, China

³ Center for Labor and Social Security Research, Central China Normal University, Wuhan 430079, China

* Correspondence: xhdai@mail.ccnu.edu.cn

Abstract: For count data, though a zero-inflated model can work perfectly well with an excess of zeroes and the generalized Poisson model can tackle over- or under-dispersion, most models cannot simultaneously deal with both zero-inflated or zero-deflated data and over- or under-dispersion. Ear diseases are important in healthcare, and falls into this kind of count data. This paper introduces a generalized Poisson Hurdle model that work with count data of both too many/few zeroes and a sample variance not equal to the mean. To estimate parameters, we use the generalized method of moments. In addition, the asymptotic normality and efficiency of these estimators are established. Moreover, this model is applied to ear disease using data gained from the New South Wales Health Research Council in 1990. This model performs better than both the generalized Poisson model and the Hurdle model.

Keywords: over-dispersion and under-dispersion; excess of zero; generalized Poisson regression; Hurdle model; generalized Poisson Hurdle model; generalized method of moments



Citation: Zuo, G.; Fu, K.; Dai, X.; Zhang, L. Generalized Poisson Hurdle Model for Count Data and Its Application in Ear Disease. *Entropy* **2021**, *23*, 1206. <https://doi.org/10.3390/e23091206>

Academic Editor: Pentti Nieminen

Received: 22 August 2021
Accepted: 10 September 2021
Published: 13 September 2021

Publisher's Note: MDPI stays neutral with regard to jurisdictional claims in published maps and institutional affiliations.



Copyright: © 2021 by the authors. Licensee MDPI, Basel, Switzerland. This article is an open access article distributed under the terms and conditions of the Creative Commons Attribution (CC BY) license (<https://creativecommons.org/licenses/by/4.0/>).

1. Introduction

Count data are common in various areas, such as public health, insurance, traffic, and epidemiology. The Poisson model and negative binomial model are usually applied to handle count data. When the sample variance is either larger or smaller than the sample mean, this means that it is over-dispersed or under-dispersed, respectively. The Poisson model and negative binomial model cannot handle over- or under-dispersed count data. Hence, the use of generalized Poisson distribution is proposed for over- or under-dispersed count data [1]. The generalized Poisson regression (GPR) model based on generalized Poisson distribution has been widely studied [2,3]. Additionally, the generalized Poisson Regression model was estimated by the maximum likelihood method and the method of moments [4]. Some measures, such as the Pearson chi-squared test, deviance, the likelihood ratio test, Akaike Information Criteria (AIC), and Bayesian Information Criteria (BIC), have been proposed for testing the goodness of fit of a model. Generalized Poisson Regression performs better than other regression models [5]. Zero inflated Waring distribution (ZIW) has been proposed to solve the problem of overdispersion in the data and the reduction of its mean [6].

For count data with an excess of zeros, zero-inflated regression models, such as zero-inflated Poisson (ZIP) and zero-inflated negative binomial (ZINB), have been proposed [7,8]. The zero-inflated Poisson (ZIP) model deals with count data with an excess of zeros [9], while the zero-inflated negative binomial (ZINB) model deals with over-dispersed count data with excess zeros, for example, the zero-inflated generalized Poisson (ZIGP) model [10–15], and the Zero-inflated Bell regression model [16,17]. These models have been used to analyze the relationship between the number of mothers that receive antenatal care visits and other factors, such as maternal education, partner education level, age of

mothers, religion of mothers, and wealth index. In particular, these models are compared using AIC and BIC. In addition, the Zero-Inflated Hierarchical Poisson Model has been applied to data with an excess of zeros to analyze maternal mortality data from 2010 to 2013 in health facilities in four regions of Ghana [18].

Although these models deal with zero-inflated data, they cannot simultaneously deal with both zero-deflated and over-/under-dispersed data. An alternative model that accounts for zero-inflated or zero-deflated is the Hurdle model [19]. This model specifies two processes that generate zero counts and positive counts. Feng compared the zero-inflated and Hurdle models to determine the differences and performances of both models in a simulation [20]. The extended negative binomial Hurdle (ENBH) model deals with zero-inflation in addition to under-dispersion in non-zero counts [21]. Bocci et al. generalized the usual Hurdle regression model by specifying a multiple inflated truncated negative binomial distribution for the positive responses and applied it to the tourism behavior of Italian residents [22]. Park and Kim proposed a tree-structured hierarchical model for count data of both excessive zeros and over-dispersion [23]. Hasanah et al. estimated a Hurdle model using the Bayesian method with the non-closed form of posterior distributions by specified non-information priors for parameters [24].

Parameter estimation is important. Estimation methods include the maximum likelihood (ML), the method of moments, the generalized method of moments (GMM), and Bayesian estimation. The GMM method is widely used in parameter estimation due to its good statistical properties. Chen and Cheng proposed a partially linear additive spatial error model (PLASEM), estimated its parameters using GMM, and derived consistency and asymptotic normality for some estimators [25]. Muris investigated missing data using GMM, derived a set of moment conditions, and obtained the efficiency [26]. For count data, Sarvi et al. studied the use of the GEE-based generalized Poisson Regression model for over- and under-dispersed clustered count data with excess zeros [27]. Mahpolah et al. used GMM to estimate a Poisson regression model; the result gained with the use of GMM is better than that gained by the use of ML [28]. Allo et al. estimated the parameters of the generalized Poisson regression model using GMM and applied it to diarrhea in infants in Pasuruan Regency, East Java [29]. Yogita and Kirtee estimated parameters of the zero-inflated model using a probability estimated method based on a moment estimator of the mean parameter [30].

Ear diseases are important in healthcare. For ear disease, Lee et al. used a meta-analysis method for the incidence of ear diseases caused by swimming without ear protection [31]. Sanchez et al. investigated the impact of swimming in water containing saline chloride on the occurrence of ear disease [32]. Subtil et al. investigated whether water precautions reduce the rate of ear diseases [33]. Sanchez et al. measured the impact of 4 weeks of daily swimming on rates of ear discharge with a tympanic membrane perforation and on the middle ear, and found swimming not to be associated with increased risk of ear disease [34].

Count data, for example in ear disease, usually contains too many or too few zeros and exhibits dispersion characteristics meanwhile. This paper proposes the generalized Poisson Hurdle model (GPHR), which simultaneously deals with count data that are both zero-inflated/zero-deflated and over-/under-dispersed, and utilizes GMM to estimate the parameters. Furthermore, the asymptotic normality and efficiency for GMM estimators are established for the GPHR model under certain conditions. As high-order moments are not easy to calculate, the bootstrap method is used to estimate the variance of the GMM estimator in real data. In its application to ear disease using the data gained from New South Wales Health Research Council in 1990, it can be seen that the GMM estimators perform better than the ML estimators of the GPHR model.

The paper is organized as follows. Section 2 introduces the generalized Poisson Hurdle model. Section 3 carries out GMM estimation and establishes its asymptotic normality and efficiency. Section 4 introduces the Nelder mean algorithm in detail. Section 5 discusses the

application to real data in ear disease. Section 6 concludes the paper. Technical proofs are given in Appendix A.

2. Basic Model

For count data, the generalized Poisson regression (GPR) model performs better with over-dispersion or under-dispersion, while the Hurdle regression model performs better with zero excess data. This paper introduces the generalized Poisson Hurdle regression (GPHR) model to simultaneously deal with both over- or under-dispersion and zero-inflated data.

2.1. Generalized Poisson Regression

To handle data with over- or under-dispersion upon generalized Poisson distribution, two main versions of generalized Poisson distribution are used, denoted by GP_1 and GP_2 .

Suppose that random variable Y follows a GP_1 distribution [1]. The probability density function of Y is:

$$\Pr(Y = y) = \begin{cases} \frac{\lambda_1(\lambda_1 + y\lambda_2)^{y-1} e^{-(\lambda_1 + y\lambda_2)}}{y!}, & \lambda_1 + y\lambda_2 > 0, \\ 0, & \lambda_1 + y\lambda_2 \leq 0, \end{cases}$$

where $y = 0, 1, 2, \dots$ and λ_1 and λ_2 are unknown parameters.

Another example is generalized Poisson distribution GP_2 [35]. For random variable Y , the probability density function is:

$$\Pr(Y = y) = \left[\frac{\mu}{1 + \alpha\mu} \right]^y \frac{(1 + \alpha y)^{y-1}}{y!} \exp \left[-\frac{\mu(1 + \alpha y)}{1 + \alpha\mu} \right], \quad y = 0, 1, 2, \dots \quad (1)$$

where α is a dispersion parameter. If $\alpha > 0$, then the variance is greater than the mean, which is known as overdispersion; if $\alpha < 0$, then the variance is less than the mean, which is known as under-dispersion; and if $\alpha = 0$, then the generalized Poisson distribution degenerates to a Poisson distribution. Hence, the dispersion characteristics of data depends on dispersion parameter α .

As can be seen, GP_1 is identical to GP_2 , when $\lambda_1 = \frac{\mu}{1 + \alpha\mu}$, $\lambda_2 = \frac{\alpha\mu}{1 + \alpha\mu}$.

If $Y \sim GP_1$, the mean and variance are $E[Y] = \frac{\lambda_1}{1 - \lambda_2}$ and $\text{Var}[Y] = \frac{E[Y]}{(1 - \lambda_2)^2}$, respectively.

Similarly, if $Y \sim GP_2$, the mean and variance are:

$$E[Y] = \mu \quad (2)$$

And:

$$\text{Var}[Y] = \mu(1 + \alpha\mu)^2 \quad (3)$$

The generalized Poisson regression (GPR) model runs as follows:

$$\ln(\mu) = \mathbf{x}^T \boldsymbol{\beta} \quad (4)$$

where \mathbf{x} is the $(p + 1)$ -dimension vector from the predictor variable (with a 1 in the first element) and $\boldsymbol{\beta}$ is the $(p + 1)$ -dimension vector from the regression parameter.

2.2. Hurdle Model

The Hurdle regression model, also named the two-part structure, is an effective model in dealing with zero-inflated and zero-deflated data [19]. This model separates the generation of zero data from that of positive data and regards these as two separate processes. The first process judges whether the zero events happen; this is denoted by 0 when zero events happen with probability w and denoted by 1 when zero events do not happen with probability $1 - w$. When the studied event has happened, we enter into the

second process, i.e., how many times the event happens. In this process, the occurrence of the event conforms to, for example, a Poisson distribution, negative binomial distribution, or general Poisson distribution. However, in this process, the event's value must take a positive value and the events must occur at least once, which is based on the first process. Therefore, the event is conditionally distributed and zero truncated.

This leads us to the Hurdle model:

$$\Pr(Y = j) = \begin{cases} f_1(0) = w, & j = 0, \\ \frac{1-f_1(0)}{1-f_2(0)} f_2(j), & j > 0, \end{cases} \quad (5)$$

where $f_1(y)$ and $f_2(y)$ are the density functions of the first process and second process, respectively. $f_1(0) = w$ is the probability that zero occurs and $\frac{f_2(y)}{1-f_2(0)}$ is the truncated density. The model allows for excess zeros if $f_1(0) > f_2(0)$ and can model too-few zeros if $f_1(0) < f_2(0)$. Obviously, the distribution reduces to the simple f distribution only if $f_1 = f_2 = f$.

The moments of the Hurdle model can be easily calculated. The mean is:

$$E[Y] = \Pr[Y > 0] \cdot E_{Y>0}[Y|Y > 0] = \frac{1-w}{1-f_2(0)} \cdot v_2, \quad (6)$$

where v_2 is the untruncated mean for the density $f_2(y)$. Similarly, the 2-nd moment is:

$$E[Y^2] = \frac{1-w}{1-f_2(0)} \cdot \sigma_2^2$$

where σ_2^2 is the untruncated variance for the density $f_2(y)$. Hence, the variance of the Hurdle model is:

$$\text{Var}[Y] = \frac{1-w}{1-f_2(0)} \cdot \sigma_2^2 - \left[\frac{1-w}{1-f_2(0)} \cdot v_2 \right]^2 \quad (7)$$

2.3. Generalized Poisson Hurdle Regression Model

This subsection combines the advantage of the generalized Poisson regression model for dispersed data and the hurdle model for zero-inflated and zero-deflated data. According to the basic principle of the Hurdle model, if the second process is in a zero-truncated generalized Poisson distribution, then the Generalized Poisson Hurdle Regression model can be proposed as follows:

$$\Pr(Y = j) = \begin{cases} w, & j = 0, \\ (1-w) \frac{g}{1-g(0)}, & j > 0, \end{cases} \quad (8)$$

where $g = g(y; \mu, \alpha) = \left[\frac{\mu}{1+\alpha\mu} \right]^y \frac{(1+\alpha y)^{y-1}}{y!} \exp \left[-\frac{\mu(1+\alpha y)}{1+\alpha\mu} \right]$

Using Equations (2), (3), (6) and (7), the mean and variance are:

$$E[Y] = \frac{1-w}{1 - \exp \left(-\frac{\mu}{1+\alpha\mu} \right)} \mu \quad (9)$$

And:

$$\text{Var}[Y] = \frac{1-w}{1 - \exp \left(-\frac{\mu}{1+\alpha\mu} \right)} \cdot \left[\mu(1 + \alpha\mu)^2 + \mu^2 \right] - \left[\frac{1-w}{1 - \exp \left(-\frac{\mu}{1+\alpha\mu} \right)} \cdot \mu \right]^2 \quad (10)$$

Obviously, the GPHR model changes to a simple Poisson Hurdle regression model when $\alpha = 0$. Then, the GPHR model can be followed by Equations (11) and (12).

$$\ln(\mu) = \mathbf{x}^T \boldsymbol{\beta}, \quad (11)$$

$$\log \text{it}(w) = \mathbf{z}^T \boldsymbol{\delta}. \quad (12)$$

where \mathbf{x} and \mathbf{z} are $(p + 1)$ -dimension vector as predictor variables (with a 1 in the first element), and $\boldsymbol{\beta}$ and $\boldsymbol{\delta}$ are $(p + 1)$ -dimension vector as regression parameters. In general, \mathbf{x} and \mathbf{z} may be the same in the model, the reason why the estimated coefficients of explanatory variables appear twice in real data analysis.

3. Estimation

In this section, we estimate the parameters using the generalized method of moments (GMM) and establish their asymptotically normal properties and efficiencies [36,37].

3.1. Parameter Estimation

In simple notation, let $X = (\mathbf{x}^T, \mathbf{z}^T)^T$, $\boldsymbol{\zeta} = (\boldsymbol{\beta}^T, \boldsymbol{\delta}^T)^T$, where T indicates the transposition for matrix or vector; then, Equations (11) and (12) are simplified to:

$$\log \left(\frac{\mu}{1-w} \right) = \begin{pmatrix} \mathbf{x}^T \boldsymbol{\beta} \\ \mathbf{z}^T \boldsymbol{\delta} \end{pmatrix} = X^T \boldsymbol{\zeta}. \quad (13)$$

For the GPHR model, the 1st and 2nd moments can be easily obtained from Equations (9) and (10):

$$\begin{aligned} E[Y] &= \frac{1-w}{1-\exp\left(-\frac{\mu}{1+\alpha\mu}\right)} \mu =: g_1(X, \boldsymbol{\theta}), \\ E[Y^2] &= \frac{1-w}{1-\exp\left(-\frac{\mu}{1+\alpha\mu}\right)} \cdot [\mu(1 + \alpha\mu) + \mu^2] =: g_2(X, \boldsymbol{\theta}), \end{aligned}$$

where $\boldsymbol{\theta} = (\alpha, \boldsymbol{\zeta}^T)^T$ is an unknown parameter. Hence, when X is non-random, the moment condition for the GPHR model can be followed by Equation (14):

$$\mathbf{h}(Y, X, \boldsymbol{\theta}) = \begin{bmatrix} X(Y - g_1(X, \boldsymbol{\theta})) \\ X(Y^2 - g_2(X, \boldsymbol{\theta})) \end{bmatrix}. \quad (14)$$

As can easily be seen, $E[\mathbf{h}(Y, X, \boldsymbol{\theta}_0)] = 0$, where $\boldsymbol{\theta}_0$ is the vector of the true parameters. Using Equation (14), the sample moment condition can be obtained.

$$\mathbf{h}_n(\mathbf{Y}, \mathbf{X}, \boldsymbol{\theta}) = \frac{1}{n} \sum_{i=1}^n \mathbf{h}(Y_i, X_i, \boldsymbol{\theta}). \quad (15)$$

Then, we can obtain the objective function:

$$Q_n(\mathbf{Y}, \mathbf{X}, \boldsymbol{\theta}) = \left[\frac{1}{n} \sum_{i=1}^n \mathbf{h}(Y_i, X_i, \boldsymbol{\theta}) \right]^T W(\boldsymbol{\theta}) \left[\frac{1}{n} \sum_{i=1}^n \mathbf{h}(Y_i, X_i, \boldsymbol{\theta}) \right], \quad (16)$$

where $W(\boldsymbol{\theta})$ is a weight matrix. There are several common forms for the weight matrix. for example, an identity matrix and the covariance matrix of the estimating equation vector. Without ambiguity, let $Q_n(\boldsymbol{\theta}) = Q_n(\mathbf{Y}, \mathbf{X}, \boldsymbol{\theta})$, $\mathbf{h}_n(\boldsymbol{\theta}) = \mathbf{h}_n(\mathbf{Y}, \mathbf{X}, \boldsymbol{\theta})$ and $W = W(\boldsymbol{\theta})$. Let $\hat{\boldsymbol{\theta}}$ be a minimizer of Equation (16); then, $\hat{\boldsymbol{\theta}}$ is called the GMM estimator.

3.2. Asymptotic Property and Efficiency

In this subsection, we investigate the efficiency and asymptotic normality of the GMM estimator. Let the true value of the parameter $\theta_0 \in \Theta$, where Θ is a compact set. For a matrix A , let $\|A\| = (\sum a_{ij}^2)^{\frac{1}{2}}$. Some assumptions are required.

Assumption 1.

- (i) The covariate X is a non-random variable;
- (ii) The weight matrix W is positive definite matrix;
- (iii) $WE[h(Y, X, \theta)] = 0 \Leftrightarrow \theta = \theta_0$.

Assumption 1 ensures the identification of the GMM estimator by Lemma 1 and this is a condition for the existence of the GMM estimator [38].

Lemma 1: If W is a positive semi-definite matrix, let $h_0(\theta) = E[h(Y, X, \theta)]$. Assume that $h_0(\theta_0) = 0$ and $h_0(\theta) \neq 0$ for $\theta \neq \theta_0$. Then, $Q_0(\theta) = [h_0(\theta)]^T W [h_0(\theta)]$ has a unique minimum value at $\theta = \theta_0$.

To rigorously establish the consistency and asymptotic normality of the estimator, the following regular assumptions are required.

Assumption 2.

- (i) $g_1(X, \theta)$ and $g_2(X, \theta)$ are continuous functions in $\theta \in \Theta$ with probability one;
- (ii) $g_1(X, \theta)$ and $g_2(X, \theta)$ are continuously differentiable in a neighborhood $\mathcal{N}(\theta_0)$ of θ_0 .

As can easily be seen, Assumption 2 is equivalent to the statement that $h(Y, X, \theta)$ is continuously differentiable in neighborhood $\mathcal{N}(\theta_0)$ of θ_0 .

Assumption 3. There exist $C_1 > 0$ and $C_2 > 0$ such that

$$\sup_{\theta \in \Theta} \|g_1(X, \theta)\| < C_1, \sup_{\theta \in \Theta} \|g_2(X, \theta)\| < C_2.$$

Assumption 4. $Var[Y^k], k = 1, 2$ are bounded away from both zero and infinity uniformly.

The above assumptions ensure the consistency of the estimator. Hence, the following theorem holds.

Theorem 1. Let observed data $\{(Y_i, X_i), i = 1, 2, \dots, n\}$ be i.i.d. If assumptions 1-4 hold, then

$$\hat{\theta} \xrightarrow{P} \theta_0,$$

where $\hat{\theta} = \underset{\theta}{\operatorname{argmin}} Q_n(\theta)$.

To establish the asymptotic normality of the proposed estimator, some strict conditions are required on the moment and on the parameter space Θ .

Assumption 5. $\sqrt{n}h_n(\theta_0) \xrightarrow{d} N(0, \Sigma)$ and $E[||h_n(\theta_0)||^2] < \infty$.

Assumption 6. There exist $G(\theta) = E[\nabla h_n(\theta)]$ and $G \equiv G(\theta_0)$, such that $G^T W G$ is nonsingular matrix.

Assumptions 5 and 6 are required for the asymptotic variance and its asymptotic normality. Similar to the conditions of theorem 3.4 in [38], asymptotic normality holds.

Theorem 2 (Asymptotic normality). Let the observed data $\{(Y_i, X_i), i = 1, 2, \dots, n\}$ be i.i.d. If assumptions 1-6 hold, then:

$$\sqrt{n}(\hat{\theta} - \theta_0) \xrightarrow{d} N(0, \Gamma),$$

where $\Gamma = (G^T W G)^{-1} G^T W \Sigma W G (G^T W G)^{-1}$.

As can be seen from Theorem 2, the asymptotic variance can be simplified to $(G^T \Sigma^{-1} G)^{-1}$ when $W = \Sigma^{-1}$. In some cases, the efficiency of the estimator is established—that is, an estimator has minimum variance. For an unbiased estimator, an unbiased estimator (MVUE) with a minimum variance can be easily defined. In fact, if the estimator is unbiased, then it is asymptotically unbiased; hence, we concentrate on asymptotically unbiased estimators with minimum variance. For matrix P and Q , $P > Q$ and $P \geq Q$, where $P - Q$ is a positive definite matrix and a positive semi-definite matrix. As a special case, the following theorem holds:

Theorem 3. Let the conditions of Theorem 2 hold. If

$$\Sigma = E \left[\mathbf{h}(Y, X, \theta_0)^T \mathbf{h}(Y, X, \theta_0) \right]$$

is nonsingular matrix and $W = \Sigma^{-1}$, then:

$$\sqrt{n}(\hat{\theta} - \theta_0) \xrightarrow{d} N \left(0, (G^T \Sigma^{-1} G)^{-1} \right).$$

Hence, the $\hat{\theta}$ is the estimator of minimum variance.

As can be seen from Theorem 3, the GMM estimator is most efficient when the weight matrix is identical to the inverse of the covariance of estimating function $\mathbf{h}(Y, X, \theta)$ in all families of GMM.

4. Algorithm

Indeed, it is difficult to calculate the first derivative (16). However, GMM estimation can be calculated by the Nelder–Mead algorithm [39], which is widely known as the Nelder–Mead simplex algorithm, to find a minimum of the function of several variables. In particular, for complex functions, this algorithm is a good choice, since it does not require differentiation.

The Nelder–Mead algorithm runs as follows.

1. Choose $m + 1$ point, $\theta_1, \theta_2, \dots, \theta_m, \theta_{m+1}, \theta_j \in \mathbb{R}^m$ and $Q_n(\theta_1) \leq \dots \leq Q_n(\theta_{m+1})$.
2. Calculate $\bar{\theta} = \frac{1}{m} \sum_{i=1}^m \theta_i$.
3. Compute $\theta_k = (1 + \eta)\bar{\theta} - \eta\theta_{m+1}$, where $\eta > 0$ is a reflection coefficient. If $Q_n(\theta_1) \leq Q_n(\theta_k) \leq Q_n(\theta_m)$, then $\theta_k \rightarrow \theta_{m+1}$, else go as follows:
 - a. If $Q_n(\theta_k) \leq Q_n(\theta_1)$, then calculate $\theta^* = (1 - \gamma)\bar{\theta} + \gamma\theta_k$, where $\gamma > 1$ is an expansion coefficient. Next, if $Q_n(\theta^*) \leq Q_n(\theta_k)$, then $\theta^* \rightarrow \theta_{m+1}$, else $\theta_k \rightarrow \theta_{m+1}$.
 - b. If $Q_n(\theta_m) \leq Q_n(\theta_k) \leq Q_n(\theta_{m+1})$, then calculate $\theta^{**} = (1 - \xi)\bar{\theta} + \xi\theta_k$, where $0 < \xi < 1$ is a contraction coefficient. Next, if $Q_n(\theta^{**}) \leq Q_n(\theta_k)$, then $\theta^{**} \rightarrow \theta_{m+1}$, else go to Step 4.
 - c. If $Q_n(\theta_k) \geq Q_n(\theta_{m+1})$, then calculate $\theta^{***} = (1 + \xi)\bar{\theta} - \xi\theta_k$. Next, if $Q_n(\theta^{***}) \leq Q_n(\theta_{m+1})$, then $\theta^{***} \rightarrow \theta_{m+1}$, else go to Step 4.
4. $(1 - \kappa)\theta_1 + \kappa\theta_i \rightarrow \theta_i$, where $0 < \kappa < 1$ and $2 \leq i \leq m + 1$.

In the algorithm above, symbol $A \rightarrow B$ indicates replacing B with A . For the Nelder–Mead algorithm, the standard settings of $\eta, \gamma, \zeta, \kappa$ are $\eta = 1, \gamma = 2, \zeta = \frac{1}{2}, \kappa = \frac{1}{2}$. Usually, the iteration stops when $\left\{ m^{-1} \sum_{i=1}^m [Q_n(\theta_i) - \bar{Q}_n(\theta)] \right\}^{\frac{1}{2}} < \varepsilon$, or when the number of iterations reaches a fixed value.

In practice, the Nelder–Mead algorithm is widely used to optimize target functions, since it simply and easily achieves its minimization. It does not require continuous or differentiable target functions. This algorithm is significantly improved in the first few iterations and quickly provides satisfactory results. However, there are two disadvantages of the Nelder–Mead algorithm. The algorithm is very sensitive to the initial values. Additionally, the convergence of the algorithm is difficult to guarantee globally, even for smooth and well-behaved functions. See, for example, for special conditions [40–43]. The Nelder–Mead algorithm is modified to improve the worst-case performance of the algorithm in terms of convergence, but retains some or most of its efficiency in best-case scenarios [44,45].

5. Real Data Analysis

As a healthy sport, swimming can improve human energy metabolism and maintain respiratory health. However, frequent swimming may lead to excessive moisture in the ear and inflammation of the external auditory canal. Moisture can cause ear eczema. Skin damage caused by repeated scratching of eczema can make bacteria or fungi invade the ear canal tissue and cause infection. However, swimming in bacterially-contaminated water is a common cause of swimming ear disease. Therefore, in health care, it is significant in practice to explore the relationship between the frequency of ear disease and other factors such as swimming. On the other hand, people who often swim may not be infected with ear diseases.

In this section, we analyze real data in ear disease using the method proposed above, and compare results between different methods.

5.1. Data Description

The real data in this application relate to the incidence of ear diseases and come from the investigation carried out by the New South Wales Health Research Council in 1990. The data gathered include the frequency of swimming, the place of swimming, and the frequency of ear diseases, from a total of 190 observation data points. The number of ear diseases is recorded by a type of count data—i.e., how many times the self-diagnosis of ear infection has occurred. The frequency of swimming is a categorical variable—i.e., how often the swimmer swims in the ocean—and takes two values: “Often” or “Occas”; quantitative values of “Often” and “Occas” are 1 and 2, respectively. The place of swimming is also a categorical variable, relating to the usual swimming location, and takes the values “Beach” or “Non-beach”; quantitative values of “Beach” and “Nonbeach” are 1 and 4, respectively. In this paper, we use the frequency of swimming and place of swimming as explanatory variable x in our model and the number of ear diseases as response variable Y . Figure 1 is a histogram of the frequency distribution of ear disease, where the value interval of the number of ear disease is (0, 17) and 92 cases occurred in 0, accounting for 48.4% of the total; 27 occurred in 1, accounting for 14.2% of the total; 26 occurred in 2, accounting for 13.7% of the total; 21 occurred in 3, accounting for 11.1% of the total; and 12 occurred in 4, accounting for 6.3% of the total. As the number of occurrences increases, the proportion of the number of ear disease becomes smaller and smaller. As shown in Figure 1, the number of cases of ear disease has a larger probability accumulation at zero. Therefore, if there is zero excess in this set of data, the use of the Hurdle model in this paper may provide a better fit with the data. Information relating to the number characters for variable Y is shown in Table 1. As can be seen in Table 1, the sample data contain more zeros, the expected number of occurrences is 1.6, and the variance is 6.5. The variance is larger than expected—that is, the sample data suffer from over-dispersion. Therefore, the generalized Poisson Hurdle regression model is applied in this real data.

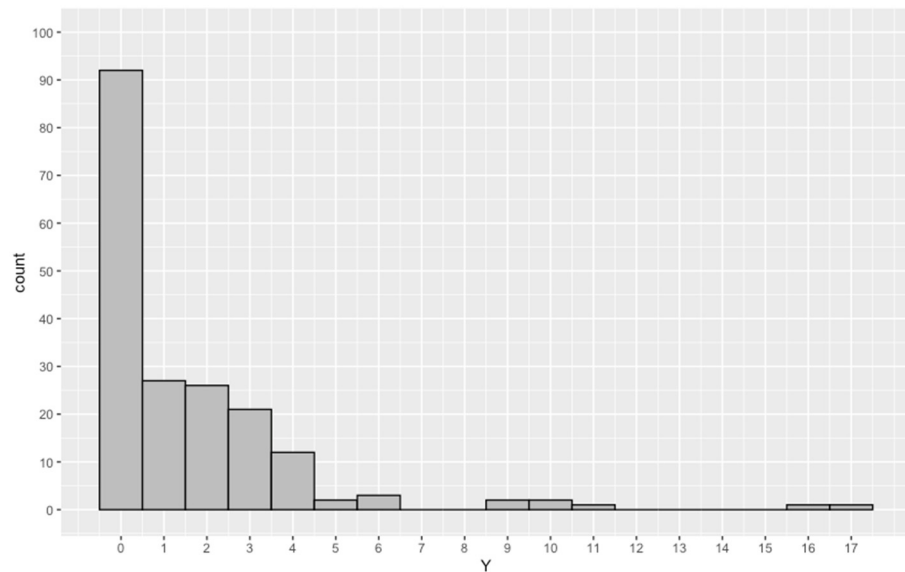


Figure 1. Histogram of number of ear diseases.

Table 1. Some number characters for ear disease (NED).

Variable	Min.	1st Qu.	Med.	Mean	3rd Qu.	Max.	Var.
NED	0.0	0.0	1.0	1.6	2.0	17.0	6.5

Note: Number characters of the number of ear disease (NED) include the minimum value (Min.), first quartile (1st Qu.), median (Med.), third quartile (3rd Qu.), maximum value (Max.) and variance (Var.).

5.2. Empirical Analysis

In this subsection, we utilize the generalized Poisson (GP) regression model, the Poisson Hurdle (PH) regression model, and the GPHR model for ear disease. We estimate parameters using the ML method and GMM. For GMM, the initial value used is ML estimation. The Akaike Information Criteria (AIC) are introduced to compare the fitting effects of these three models, where the form of the AIC statistics is $AIC = -2l + 2p$, l is the log-likelihood value, and p is the number of free parameters. In general, the smaller the AIC values, the better the fitting effect of the model.

In order to make statistical inferences, the variance of the regression coefficients is estimated. It is natural to attempt to derive a consistent estimator, but as seen from the proof of the theorem, such an estimator is too complex and not practical. Hence, we use the bootstrap method to estimate variance. As shown in Figure 2, the basic schematic of bootstrapping proceeds. We denote the training set by $Z = (z_1, z_2, \dots, z_N)$, randomly draw data sets with replacements from training set Z , and name these bootstrap data sets. The size of the bootstrap data sets may be not equal to that of the original training set. We repeat B times and produce B bootstrap data sets. Then, we refit the model to each of the bootstrap data sets and examine the behavior of the fits over the i -th bootstrap data set.

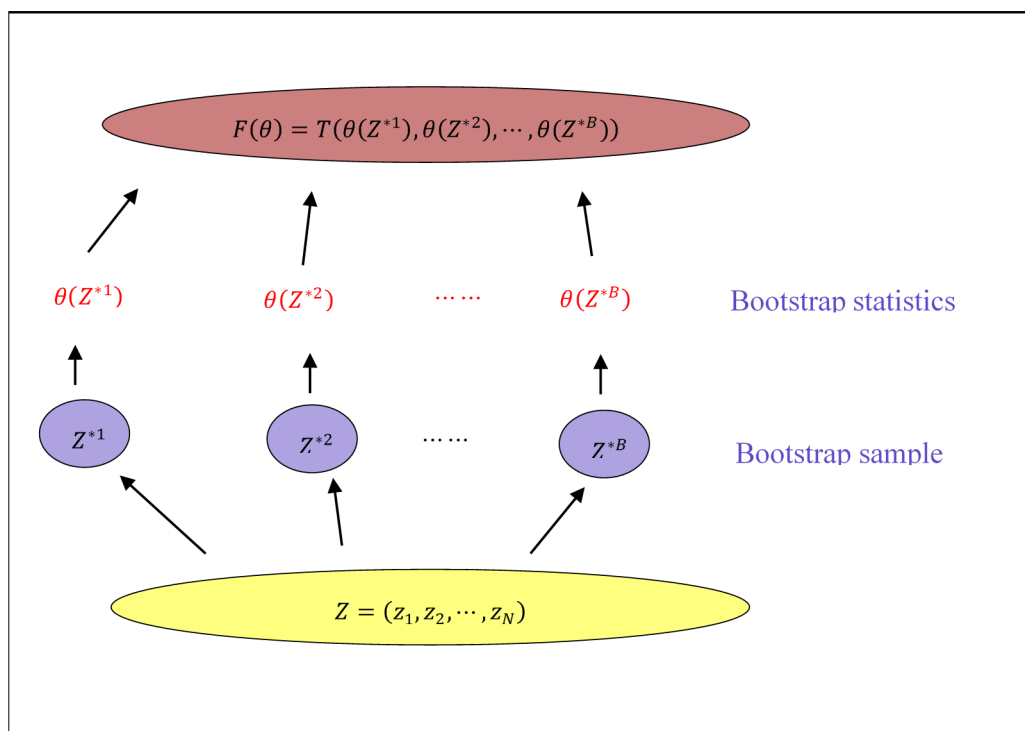


Figure 2. Schematic of the bootstrap process.

In Figure 2, $\theta(Z)$ represents any statistics computed from the data set Z . Using bootstrap sampling, we can estimate any aspect $F(\theta)$ of the distribution of $\theta(Z)$, for example, its variance:

$$\hat{\text{Var}}[\theta(Z)] = \frac{1}{B-1} \sum_{i=1}^B \left[\theta(Z^{*i}) - \bar{S}^* \right]^2,$$

where $\bar{S}^* = \frac{1}{B} \sum_{i=1}^B \theta(Z^{*i})$. For the bootstrap method used, see, for example, Efron and Tibshirani [46]. In this paper, all programs and algorithms are calculated by R in version 3.6.3 and Rstudio in version 1.2.5042.

Table 2 shows the results for three models. For the generalized Poisson Hurdle regression model and the Poisson hurdle model, each model includes two parameters related to the explanatory variables, and the coefficients of explanatory variable are estimated twice, i.e., intercept, frequency and place appear twice in the results. The Wald confidence intervals at the significance of 5% and p -values are also shown in Table 2. To estimate the variance of GMM, bootstrap methods are used where $B = 50$ times. The GMM estimation and ML estimation produce similar results for the three models. The PH model obtains the worst results and the results of the AIC are relatively large, which may be due to the over-dispersion. The GP regression model and the GPHR model both have over-dispersed characteristics, meaning that the fitting effect is improved. On the other hand, the GPHR model improves the results in the case of zero excess, as can be seen from the AIC value, compared with the GP regression model. The variances of the GMM estimator are less than the variances of the ML estimator for three models. Additionally, it can be seen that people who often swim are more likely to suffer from ear disease. Furthermore, it is evident that swimming at the beach has a negative impact on ear disease. Under the null hypothesis, this statistic is asymptotically normally distributed. At the 5% significance level, the GPHR model fits the data for ear disease better than both the GP regression model and the PH regression model.

Table 2. Estimation results for real data.

Variables	GPHR				PH				GP			
	MLE		GMM		MLE		GMM		MLE		GMM	
	Coefficient	SE	Coefficient	SE	Coefficient	SE	Coefficient	SE	Coefficient	SE	Coefficient	SE
DIS	0.28 (0.1, 0.4)	0.09 (0.002)	0.17 (0.07, 0.3)	0.05 (<0.001)	–	–	–	–	0.59 (0.4, 0.8)	0.10 (<0.001)	0.52 (0.4, 0.6)	0.06 (<0.001)
INT1	−0.48 (−1.8, 0.8)	0.66 (0.474)	−0.42 (−0.6, −0.3)	0.07 (<0.001)	−0.08 (−0.8, 0.7)	0.38 (0.837)	0.09 (−0.2, −0.01)	0.04 (0.002)	−1.56 (−2.8, −0.3)	0.64 (0.016)	−1.43 (−1.5, −1.3)	0.06 (<0.001)
FRE1	0.63 (0.1, 1.1)	0.26 (0.016)	0.67 (0.5, 0.8)	0.08 (<0.001)	0.55 (0.2, 0.9)	0.16 (<0.001)	0.41 (0.3, 0.6)	0.08 (<0.001)	0.75 (0.2, 1.3)	0.26 (0.004)	0.71 (0.6, 0.8)	0.06 (<0.001)
PLA1	0.07 (−0.1, 0.2)	0.10 (0.513)	0.11 (0, 0.2)	0.06 (0.06)	0.06 (−0.05, 0.2)	0.06 (<0.001)	0.03 (−0.1, 0)	0.05 (0.543)	0.23 (0.03, 0.4)	0.10 (0.017)	0.20 (0.1, 0.3)	0.04 (<0.001)
INT2	2.29 (0.6, 4.0)	0.85 (0.007)	2.35 (2.2, 2.4)	0.07 (<0.001)	−2.29 (−3.9, −0.6)	0.85 (0.007)	2.08 (−2.2, −2.0)	0.05 (<0.001)	–	–	–	–
FRE2	−0.71 (−1.3, −0.02)	0.35 (0.040)	−0.72 (−1, −0.4)	0.14 (<0.001)	0.71 (0.02, 1.3)	0.35 (0.004)	0.83 (0.7, 0.9)	0.05 (<0.001)	–	–	–	–
PLA2	−0.37 (−0.6, −0.1)	0.13 (0.004)	−0.29 (−0.4, −0.1)	0.07 (<0.001)	0.37 (0.1, 0.6)	0.13 (0.004)	0.57 (0.5, 0.6)	0.03 (<0.001)	–	–	–	–
AIC		639.90				745.64				643.31		

Note: DIS, INT, FRE and PLA are abbreviations for dispersion parameter α , model intercept, frequency, and place, respectively; generalized Poisson Hurdle regression model (GPHR), Poisson hurdle regression model (PH), and generalized Poisson regression model (GP) for ear disease data. The coefficients are estimated by maximum likelihood (MLE) method and generalized method of moment (GMM), and standard error (SE) is calculated by Fisher information matrix and bootstrap method, respectively. The numbers in the parentheses below coefficient and standard error (SE) are Wald confidence intervals and p -values, respectively. For GPHR and PH model, the coefficients of INT1, FRE1 and PLA1 are count model coefficients β , while the coefficients of INT2, FRE2 and PLA2 are zero Hurdle model coefficients δ .

6. Conclusions

This paper introduces the GPHR model for count data. It combines the generalized Poisson Regression (GPR) model with the Hurdle Regression model. Compared with other models, such as the PH regression model and the GP regression model, the GPHR model can simultaneously deal with both zero-inflated or zero-deflated data and over- or under-dispersion in count data. GMM is used to estimate the parameters of the models. In some cases, the GMM performs better than ML. Since it is difficult to calculate the first derivative of the objective function, the Nelder–Mead method is used to estimate the parameters.

Ear diseases are important in healthcare. The proposed model is applied to real data on ear diseases, which is another purpose of this paper. Compared with other zero-inflated models, the GPHR model has a minimum AIC value. Given a significant level (5%), the Wald test is used to measure the significance of the factors incorporated in each model. As shown, the GPHR model fits well with the real data. In fact, one count data is not enough to validate advantages and disadvantages of the purposed model for both zero-inflated or zero-deflated data and over- or under-dispersion. Although this model performs well in the data for ear disease, other count data are better at checking this model. Indeed, we check consumer health information for medical treatment, and this model also performs well. We do not report in detail, since this paper aimed to purpose the GPHR model for count data of both zero-inflated or zero-deflated data and over- or under-dispersion, and apply this to ear disease data.

In this paper, we only studied data that were zero-inflated. The model may be extended for other types of count data, such as zero-deflated data. In addition, for the first process of the Hurdle model, the occurrence probability w is used. In fact, this may be extended to the Poisson distribution at zero when the first process occurs, with the same parameter μ used as one in the second process. For the second process of the Hurdle model, this may be extended to the case of truncation.

Author Contributions: Conceptualization, G.Z. and X.D.; methodology, X.D. and G.Z.; analysis, K.F. and L.Z.; investigation, X.D., K.F. and L.Z.; data curation, K.F.; writing—original draft preparation, K.F.; writing—review and editing, X.D.; supervision, G.Z. and X.D.; project administration, X.D.; funding acquisition, X.D. All authors have read and agreed to the published version of the manuscript.

Funding: This research was funded in part by the Humanities and Social Science Research Fund of the Ministry of Education in China, grant 18YJA790018, in part by the Fundamental Research Funds of the Central Universities, grant CCNU19TS047, and in part by the Philosophical and Social Science Research Key Fund of the Department of Education in Hubei Province, grant 17ZD018. The APC was funded by Central China Normal University, China.

Institutional Review Board Statement: Ethical review and approval were waived for this study, due to publically open data from the investigation carried out by the New South Wales Health Research Council in 1990.

Informed Consent Statement: Patient consent was waived due to publically open data from the investigation carried out by the New South Wales Health Research Council in 1990.

Data Availability Statement: Data used in this paper can be found from <https://search.r-project.org/CRAN/refmans/GLMsData/html/earinf.html>.

Acknowledgments: X.D. is grateful to the support from Beijing International Center for Mathematical Research (BICMR) for the work in part carried out at BICMR.

Conflicts of Interest: The authors declare no conflict of interest.

Appendix A

To prove the main theorems, we need the following lemma:

Lemma A1. [47]: Let observed data $\{X_i, i = 1, 2, \dots, n\}$ be i.i.d random variable, and the parametric space compact set. Function $G(X, \theta)$ is continuous for $\theta \in \Theta$ with probability 1. If there exists

a function $B(X)$ such that $G(X, \theta) \leq B(X)$ for all $\theta \in \Theta$, and $E[B(X)] < \infty$, then $E[G(X, \theta)]$ is continuous and

$$\sup_{\theta \in \Theta} \left\| \frac{1}{n} \sum_{i=1}^n G(X_i, \theta) - E[G(X, \theta)] \right\| \xrightarrow{P} 0.$$

Proof of Theorem 1. Using lemma 1 and assumption 1, we can determine that $Q_0(\theta)$ has a unique minimum value θ_0 . Next, we prove that the $Q_n(\theta)$ uniformly converges to $Q_0(\theta)$ with a high probability.

Based on assumptions 2 and 3:

$$\|\mathbf{h}(Y, X, \theta)\| \leq \|X^T X\| \cdot \left(\|Y - g_1(X, \theta)\| + \|Y^2 - g_2(X, \theta)\| \right).$$

i.e., $E \sup_{\theta \in \Theta} \|\mathbf{h}(Y, X, \theta)\| < \infty$. Then, based on Lemma A1, $\mathbf{h}_0(\theta)$ is continuous and:

$$\sup_{\theta \in \Theta} \|\mathbf{h}_n(\theta) - \mathbf{h}_0(\theta)\| \xrightarrow{P} 0.$$

In addition, $Q_0(\theta)$ is continuous.

Due to triangle inequality and Cauchy–Schwartz inequality:

$$\begin{aligned} & |Q_n(\theta) - Q_0(\theta)| \\ &= \left| [\mathbf{h}_n(\theta)]^T W [\mathbf{h}_n(\theta)] - [\mathbf{h}_0(\theta)]^T W [\mathbf{h}_0(\theta)] \right| \\ &\leq \left| [\mathbf{h}_n(\theta) - \mathbf{h}_0(\theta)]^T W [\mathbf{h}_n(\theta) - \mathbf{h}_0(\theta)] - [\mathbf{h}_n(\theta)]^T (W + W^T) [\mathbf{h}_n(\theta) - \mathbf{h}_0(\theta)] \right| \\ &\leq \|\mathbf{h}_n(\theta) - \mathbf{h}_0(\theta)\|^2 \cdot \|W\| + 2\|W\| \cdot \|\mathbf{h}_n(\theta)\| \cdot \|\mathbf{h}_n(\theta) - \mathbf{h}_0(\theta)\|, \end{aligned}$$

Then $\sup_{\theta \in \Theta} |Q_n(\theta) - Q_0(\theta)| \xrightarrow{P} 0$, i.e., $Q_n(\theta)$ uniformly converges to $Q_0(\theta)$ with probability. Recall that Θ is a compact set; for any given $\varepsilon > 0$:

$$Q_n(\hat{\theta}) < Q_n(\theta_0) + \frac{\varepsilon}{3}, Q_0(\hat{\theta}) < Q_0(\hat{\theta}) + \frac{\varepsilon}{3},$$

And:

$$Q_n(\theta_0) < Q_0(\theta_0) + \frac{\varepsilon}{3}$$

With probability of 1. Hence:

$$Q_0(\hat{\theta}) < Q_n(\hat{\theta}) + \frac{\varepsilon}{3} < Q_n(\theta_0) + \frac{2\varepsilon}{3} < Q_0(\theta_0) + \varepsilon.$$

Thus, with a probability of 1:

$$Q_0(\hat{\theta}) < Q_0(\theta_0) + \varepsilon.$$

Set \mathcal{M} as any open set, such that $\mathcal{M} \subseteq \Theta$ and $\theta_0 \in \mathcal{M}$. Additionally, $\Theta \cap \mathcal{M}^c$ is a compact set. Thus:

$$Q_0(\tilde{\theta}) \equiv \inf_{\theta \in \Theta \cap \mathcal{M}^c} Q_0(\theta) > Q_0(\theta_0),$$

where $\tilde{\theta} \in \Theta \cap \mathcal{M}^c$. Let $\varepsilon = Q_0(\tilde{\theta}) - Q_0(\theta_0)$, then:

$$Q_0(\hat{\theta}) < Q_0(\tilde{\theta}).$$

Additionally, $\hat{\theta} \in \mathcal{M}$ with a probability of 1. Due to the arbitrariness of \mathcal{M} :

$$\hat{\theta} \xrightarrow{P} \theta_0. \square$$

Proof of Theorem 2. Based on assumption 2, $\mathbf{h}(Y, X, \theta)$ and $\mathbf{h}_n(\theta)$ are continuous in $\mathcal{N}(\theta_0)$. Under the condition of the extreme minimum value:

$$\nabla Q_n(\theta) = 2[\nabla_{\theta} \mathbf{h}_n(\theta)]^T W \mathbf{h}_n(\theta) = 0 \tag{A1}$$

With a probability of 1. Denote $G_n(\theta) = \nabla_{\theta} \mathbf{h}_n(\theta)$. Applying Taylor expansion:

$$\mathbf{h}_n(\hat{\theta}) = \mathbf{h}_n(\theta_0) + G_n(\tilde{\theta})(\hat{\theta} - \theta_0), \tag{A2}$$

where $\tilde{\theta}$ lies between θ_0 and $\hat{\theta}$. By multiplying $G_n(\hat{\theta})^T W$ (A2) from the left:

$$G_n(\hat{\theta})^T W \mathbf{h}_n(\hat{\theta}) = G_n(\hat{\theta})^T W \mathbf{h}_n(\theta_0) + G_n(\hat{\theta})^T W G_n(\tilde{\theta})(\hat{\theta} - \theta_0).$$

Based on (A1):

$$\sqrt{n}(\hat{\theta} - \theta_0) = -\sqrt{n} [G_n(\hat{\theta})^T W G_n(\tilde{\theta})]^{-1} G_n(\hat{\theta})^T W \mathbf{h}_n(\theta_0).$$

Based on assumptions 1–4, Theorem 1 holds—i.e., $\hat{\theta} \xrightarrow{P} \theta_0$. In addition, since X is nonrandom:

$$G_n(\hat{\theta}) \xrightarrow{P} G, \text{ and } G_n(\tilde{\theta}) \xrightarrow{P} G.$$

Hence:

$$- [G_n(\hat{\theta})^T W G_n(\tilde{\theta})]^{-1} G_n(\hat{\theta})^T W \xrightarrow{P} - [G^T W G]^{-1} G^T W.$$

Based on the Slutsky theorem and assumption 5, the asymptotic variance can be established. Thus, theorem 2 holds. \square

Proof of Theorem 3. Based on Theorem 2, the covariance matrix of the estimator is $(G^T \Sigma^{-1} G)^{-1}$ when $W = \Sigma^{-1}$. Hence, $\sqrt{n}(\hat{\theta} - \theta_0) \xrightarrow{d} N(0, (G^T \Sigma^{-1} G)^{-1})$.

Next, the estimator has minimum variance in this case. Indeed, let $\Sigma = E[\mathbf{h}(Y, X, \theta_0)^T \mathbf{h}(Y, X, \theta_0)]$, $\eta = G^T W \mathbf{h}(Y, X, \theta_0)$, $\bar{\eta} = G^T \Sigma^{-1} \mathbf{h}(Y, X, \theta_0)$. Thus:

$$G^T W G = E[\eta \eta^T], \quad G^T \Sigma^{-1} G = E[\bar{\eta} \bar{\eta}^T], \text{ and } G^T W \Sigma W G = E[\eta \eta^T].$$

By some simple algebraic manipulation:

$$(G^T W G)^{-1} G^T W \Sigma W G (G^T W G)^{-1} - (G^T \Sigma^{-1} G)^{-1} = (G^T W G)^{-1} E \left\{ \eta - E[\eta \eta^T] \left(E[\bar{\eta} \bar{\eta}^T] \right)^{-1} \bar{\eta} \right\} (G^T W G)^{-1}.$$

Since $E \left\{ \eta - E[\eta \eta^T] \left(E[\bar{\eta} \bar{\eta}^T] \right)^{-1} \bar{\eta} \right\}$ is a positive semi-definite matrix:

$$(G^T W G)^{-1} G^T W \Sigma W G (G^T W G)^{-1} - (G^T \Sigma^{-1} G)^{-1}$$

This is also a positive semi-definite matrix—i.e.,:

$$\left(G^T \Sigma^{-1} G\right)^{-1} \leq \left(G^T W G\right)^{-1} G^T W \Sigma W G \left(G^T W G\right)^{-1}.$$

Hence, $\hat{\theta}$ is the estimator of minimum variance. \square

References

1. Consul, P.C.; Jain, G.C. A generalization of the Poisson distribution. *Technometrics* **1973**, *15*, 791–799. [\[CrossRef\]](#)
2. Consul, P.C.; Famoye, F. Generalized Poisson regression model. *Commun. Stat.-Theory Methods* **1992**, *21*, 89–109. [\[CrossRef\]](#)
3. Famoye, F. Restricted generalized Poisson regression model. *Commun. Stat.-Theory Methods* **1993**, *22*, 1335–1354. [\[CrossRef\]](#)
4. Noriszura, I.; Jemain, A.A. Handling overdispersion with negative binomial and generalized Poisson regression models. *Casualty Actuar. Soc. Forum* **2007**, *2007*, 103–158.
5. Obubu, M.; Babalola, A.; Ikediwu, U.C.; Peace, A. Modelling count data; a generalized linear model framework. *Am. J. Math. Stat.* **2018**, *8*, 179–183.
6. Rivas, L. Zero inflated waring distribution zero inflated waring distribution. *Commun. Stat.-Simul. Comput.* **2021**, *50*, 1–16. [\[CrossRef\]](#)
7. Cheung, Y. Zero-inflated models of regression analysis of count data: A study of growth and development. *Stat. Med.* **2002**, *21*, 1361–1469. [\[CrossRef\]](#) [\[PubMed\]](#)
8. Lambert, D. Zero-Inflated Poisson Regression, with an Application to Defects in Manufacturing. *Technometrics* **1992**, *34*, 1–14. [\[CrossRef\]](#)
9. Truong, B.-C.; Pho, K.-H.; Dinh, C.-C.; McALEER, M. Zero-inflated poisson regression models: Applications in the sciences and social sciences. *Ann. Financ. Econ.* **2021**, *16*, 1–19. [\[CrossRef\]](#)
10. Bekalo, D.B.; Kebede, D.T. Zero-Inflated Models for Count Data: An Application to Number of Antenatal Care Service Visits. *Ann. Data Sci.* **2021**. [\[CrossRef\]](#)
11. Czado, C.; Erhardt, V.; Min, A.; Wagner, S. Zero-inflated generalized Poisson models with regression effects on the mean, dispersion and zero-inflation level applied to patent outsourcing rates. *Stat. Model. Int. J.* **2007**, *7*, 125–153. [\[CrossRef\]](#)
12. Famoye, F.; Preisser, J.S. Marginalized zero-inflated generalized Poisson regression. *J. Appl. Stat.* **2017**, *45*, 1247–1259. [\[CrossRef\]](#)
13. Famoye, F.; Singh, K. On inflated generalized Poisson regression models. *Adv. Appl. Stat.* **2003**, *3*, 145–158.
14. Famoye, F.; Singh, K.P. Zero-Inflated Generalized Poisson Regression Model with an Application to Domestic Violence Data. *J. Data Sci.* **2006**, *4*, 117–130. [\[CrossRef\]](#)
15. Kamalja, K.K.; Wagh, Y.S. Estimation in zero-inflated Generalized Poisson distribution. *J. Data Sci.* **2021**, *16*, 183–206. [\[CrossRef\]](#)
16. Amin, M.; Akram, M.N.; Majid, A. On the estimation of Bell regression model using ridge estimator. *Commun. Stat.-Simul. Comput.* **2021**, 1–14. [\[CrossRef\]](#)
17. Lemonte, A.J.; Moreno-Arenas, G.; Castellares, F. Zero-inflated Bell regression models for count data. *J. Appl. Stat.* **2019**, *47*, 265–286. [\[CrossRef\]](#)
18. Tawiah, K.; Iddi, S.; Lotsi, A. On Zero-Inflated Hierarchical Poisson Models with Application to Maternal Mortality Data. *Int. J. Math. Math. Sci.* **2020**, *2020*, 1–8. [\[CrossRef\]](#)
19. Mullahy, J. Specification and testing of some modified count data models. *J. Econ.* **1986**, *33*, 341–365. [\[CrossRef\]](#)
20. Feng, C.X. A comparison of zero-inflated and hurdle models for modeling zero-inflated count data. *J. Stat. Distrib. Appl.* **2021**, *8*, 1–19. [\[CrossRef\]](#)
21. Noh, M.; Lee, Y. Extended negative binomial hurdle models. *Stat. Methods Med Res.* **2018**, *28*, 1540–1551. [\[CrossRef\]](#)
22. Bocci, C.; Grassini, L.; Rocco, E. A multiple inflated negative binomial hurdle regression model: Analysis of the Italians' tourism behaviour during the great recession. *Stat. Methods Appl.* **2020**. [\[CrossRef\]](#)
23. Park, M.H.; Kim, J.H.T. Hierarchical mixture-of-experts models for count variables with excessive zeros. *Commun. Stat.-Theory Methods* **2020**, 1–25. [\[CrossRef\]](#)
24. Hasanah, S.; Abdullah, S.; Lestari, D. Bayesian method for hurdle regression. *ICSA-Int. Conf. Stat. Anal.* **2019**, *2019*, 143–154.
25. Chen, J.; Cheng, S. GMM Estimation of a Partially Linear Additive Spatial Error Model. *Mathematics* **2021**, *9*, 622. [\[CrossRef\]](#)
26. Muris, C. Efficient GMM Estimation with Incomplete Data. *Rev. Econ. Stat.* **2020**, *102*, 518–530. [\[CrossRef\]](#)
27. Sarvi, F.; Moghimbeigi, A.; Mahjub, H. GEE-based zero-inflated generalized Poisson model for clustered over or under-dispersed count data. *J. Stat. Comput. Simul.* **2019**, *89*, 2711–2732. [\[CrossRef\]](#)
28. Mahpolah, M.; Suharto, S.; Wibowo, A.; Otok, B.W. The Estimation of Generalized Method Moment Poisson Regression Model on the Prevalence of Acute Respiratory Tract Infection (RTI) in South Kalimantan. *CAUCHY* **2018**, *5*, 161–168. [\[CrossRef\]](#)
29. Allo, C.B.G.; Otok, B.W. Purhadi Estimation Parameter of Generalized Poisson Regression Model Using Generalized Method of Moments and Its Application. *IOP Conf. Ser. Mater. Sci. Eng.* **2019**, *546*, 052050. [\[CrossRef\]](#)
30. Yogita, S.; Kirtee, K. Zero-inflated models and estimation in zero-inflated Poisson distribution. *Commun. Stat.-Simul. Comput.* **2018**, *47*, 2248–2265.
31. Lee, D.; Youk, A.; Goldstein, N.A. A Meta-Analysis of Swimming and Water Precautions. *Laryngoscope* **1999**, *109*, 536–540. [\[CrossRef\]](#) [\[PubMed\]](#)

32. Sanchez, L.; Carney, A.; Esterman, A.; Sparrow, K.; Turner, D. Do esaccess to saltwater swimming pools reduce ear pathology and hearing loss in school children of remote arid zone aboriginal communities? A prospective three year cohort study. *Clin. Otolaryngol.* **2019**, *44*, 736–742. [[CrossRef](#)] [[PubMed](#)]
33. Subtil, J.; Jardim, A.; Araujo, J.; Moreira, C.; Eça, T.; McMillan, M.; Dias, S.S.; Cruz, P.V.; Voegels, R.; Paço, J.; et al. Effect of Water Precautions on Otorrhea Incidence after Pediatric Tympanostomy Tube: Randomized Controlled Trial Evidence. *Otolaryngol. Neck Surg.* **2019**, *161*, 514–521. [[CrossRef](#)] [[PubMed](#)]
34. Sanchez, A.; Arom, G.; Perez, H.A.; Royal, L.; O-Lee, T. Are water precautions necessary after tympanostomy tube placement? A cadaver study. *Int. J. Pediatr. Otorhinolaryngol.* **2021**, *143*, 110632. [[CrossRef](#)] [[PubMed](#)]
35. Wang, W.; Famoye, F. Modeling household fertility decisions with generalized Poisson regression. *J. Popul. Econ.* **1997**, *10*, 273–283. [[CrossRef](#)] [[PubMed](#)]
36. Hansen, L.P. Large Sample Properties of Generalized Method of Moments Estimators. *Econometrica* **1982**, *50*, 1029. [[CrossRef](#)]
37. Hansen, L.P.; Heaton, J.; Yaron, A. Finite-sample properties of some alternative Gmm estimators. *J. Bus. Econ. Stat.* **1996**, *14*, 262–280.
38. Newey, W.; Mcfadden, D. Large sample estimation and hypothesis testing. In *Handbook of Econometrics*; Elsevier: Amsterdam, The Netherlands, 1994; Volume 4, pp. 2111–2245.
39. Nelder, J.A.; Mead, R. A Simplex Method for Function Minimization. *Comput. J.* **1965**, *7*, 308–313. [[CrossRef](#)]
40. Galántai, A. A convergence analysis of the Nelder-Mead simplex method. *Acta Polytech. Hung.* **2021**, *18*, 93–105. [[CrossRef](#)]
41. Han, L.; Neumann, M. Effect of dimensionality on the Nelder—Mead simplex method. *Optim. Methods Softw.* **2006**, *21*, 1–16. [[CrossRef](#)]
42. Lagarias, J.C.; Reeds, J.A.; Wright, M.H.; Wright, P.E. Convergence Properties of the Nelder-Mead Simplex Method in Low Dimensions. *SIAM J. Optim.* **1998**, *9*, 112–147. [[CrossRef](#)]
43. McKinnon, K.I.M. Convergence of the Nelder-Mead Simplex Method to a Nonstationary Point. *SIAM J. Optim.* **1998**, *9*, 148–158. [[CrossRef](#)]
44. Búrmen, Á.; Puhán, J.; Tuma, T. Grid restrained nelder-mead algorithm. *Comput. Optim. Appl.* **2006**, *34*, 359–375. [[CrossRef](#)]
45. Price, C.; Coope, I.; Byatt, D. A Convergent Variant of the Nelder—Mead Algorithm. *J. Optim. Theory Appl.* **2002**, *113*, 5–19. [[CrossRef](#)]
46. Efron, B.; Tibshirani, R. *An Introduction to the Bootstrap*; Chapman and Hall: London, UK, 1993.
47. Zhou, Y. *Estimation Method of Generalized Estimation Equation*; Science Press: Beijing, China, 2013; pp. 27–62.