Discrete Weibull Regression Model For Count Data

A thesis submitted for the degree of
Doctorate of Philosophy
by

Hadeel Saleh Kalktawi
B.Sc., M.Sc.

Supervised by
Prof. Keming Yu

Brunel
UNIVERSITY
LONDON

Department of Mathematics and Computing
College of Engineering, Design and Physical Sciences

Abstract

Data can be collected in the form of counts in many situations. In other words, the number of deaths from an accident, the number of days until a machine stops working or the number of annual visitors to a city may all be considered as interesting variables for study.

This study is motivated by two facts; first, the vital role of the continuous Weibull distribution in survival analyses and failure time studies. Hence, the discrete Weibull (DW) is introduced analogously to the continuous Weibull distribution, (see, Nakagawa and Osaki (1975) and Kulasekera (1994)). Second, researchers usually focus on modeling count data, which take only non-negative integer values as a function of other variables.

Therefore, the DW, introduced by Nakagawa and Osaki (1975), is considered to investigate the relationship between count data and a set of covariates. Particularly, this DW is generalised by allowing one of its parameters to be a function of covariates. Although the Poisson regression can be considered as the most common model for count data, it is constrained by its equi-dispersion (the assumption of equal mean and variance). Thus, the negative binomial (NB) regression has become the most widely used method for count data regression. However, even though the NB can be suitable for the over-dispersion cases, it cannot be considered as the best choice for modeling the under-dispersed data. Hence, it is required to have some models that deal with the problem of under-dispersion, such as the generalized Poisson regression model (Efron (1986) and Famoye (1993)) and COM-Poisson regression (Sellers and Shmueli (2010) and Sáez-Castillo and Conde-Sánchez (2013)). Generally, all of these models can be considered as modifications and developments of Poisson models. However, this thesis develops a model based on a simple distribution with no modification. Thus, if the data are not following the dispersion system of Poisson or NB, the true structure generating this data should be detected. Applying a model that has the ability to handle different dispersions would be of great interest. Thus, in this study, the DW regression model is introduced.

Besides the flexibility of the DW to model under- and over-dispersion, it is a good model for inhomogeneous and highly skewed data, such as those with excessive zero counts, which are more disperse than Poisson. Although these data can be fitted well using some developed models, namely, the zero-inflated and hurdle models, the DW demonstrates a good fit and has less complexity than these modified models.

However, there could be some cases when a special model that separates the probability of zeros from that of the other positive counts must be applied. Then, to cope with the problem of too many observed zeros, two modifications of the DW regression are developed, namely, zero-inflated discrete Weibull (ZIDW) and hurdle discrete Weibull (HDW) models.

Furthermore, this thesis considers another type of data, where the response count variable is censored from the right, which is observed in many experiments. Applying the standard models for these types of data without considering the censoring may yield misleading results. Thus, the censored discrete Weibull (CDW) model is employed for this case.

On the other hand, this thesis introduces the median discrete Weibull (MDW) regression model for investigating the effect of covariates on the count response through the median which are more appropriate for the skewed nature of count data. In other words, the likelihood of the DW model is re-parameterized to explain the effect of the predictors directly on the median. Thus, in comparison with the generalized linear models (GLMs), MDW and GLMs both investigate the relations to a set of covariates via certain location measurements; however, GLMs consider the means, which is not the best way to represent skewed data.

These DW regression models are investigated through simulation studies to illustrate their performance. In addition, they are applied to some real data sets and compared with the related count models, mainly Poisson and NB models.

Overall, the DW models provide a good fit to the count data as an alternative to the NB models in the over-dispersion case and are much better fitting than the Poisson models. Additionally, contrary to the NB model, the DW can be applied for the under-dispersion case.

Dedicated to

my beloved late father

SALEH KALKTAWI

whose final dream was for me to
obtain a PhD degree

I miss him everyday

Acknowledgements

After many challenges and difficulties, first and above all thanks to Allah for giving me this opportunity, the strength and the patience to complete my thesis. Special appreciation goes to my supervisor, Prof. Keming Yu, for his supervision, constant support and expert guidance throughout my studies. I am particularly grateful for his understanding, patience and support when I experienced difficulties. Also my special thanks go to Dr. Veronica Vinciotti for all her support and for helping me significantly with my research as a second supervisor.

Sincere thanks go to the staff and fellow students who made my time at Brunel University a thoroughly enjoyable experience and helped me a lot during my PhD program.

A very special thanks to my parents, who sacrificed considerably to help me to complete this work and left everything behind to move with me to the UK. All their prayers, encouragement, love, support and the strength they have provided me over the years was the greatest gift anyone has ever given me. Words cannot express my appreciation for all your help. I am very thankful to my dear mother, who has always stood by me like a pillar in times of need and to whom I owe my life for her constant love, encouragement, moral support and blessings. My deepest thanks to my dear late father, who suffered a lot to help me with everything he could, until his death before the thesis was finished; his final wish was for me to complete this study. I miss you so much dad, and I hope you are in a better place.

I want to express my deepest love and heartfelt thanks to my husband, for his endless patience and support during the past several years of my studying. I do not know how to thank him enough for providing me with the opportunity to be where I am today. I am truly honored to know him and thankful for having him in my life.

I am also greatly indebted to my children, who stuck with me through all the good and bad times and put up with my moods and absences at times. I hope all the time I spent on this instead of being with you was worth it. I appreciate all of your love, caring, patience and support during mommys PhD studies and

wish for you all to be better than me in the future.

Thanks are also extended to my brothers and sisters for their constant prayers, support, encouragement and the belief that they have shown in me and my work.

Finally, I wish to thank many other people whose names are not mentioned here but who have helped me immeasurably; I have not forgotten your contributions.

Contents

List of Figures

List of Tables

Acronyms

iid identically independent distributed

cdf cumulative distribution function

pmf probability mass function

RV random variable

MLE maximum likelihood estimator

MSE mean squared error

SE standard error

CI confidence interval

LRT likelihood ration test

AIC Akaike information criterion

Dip index of dispersion relative to Poisson

VR variance ratio

GLM generalized linear model

DW discrete Weibull

NB negative binomial

ZIP zero-inflated Poisson

ZINB zero-inflated negative binomial

ZIDW zero-inflated discrete Weibull

HP hurdle Poisson

HNB hurdle negative binomial

HDW hurdle discrete Weibull

CDW censored discrete Weibull

CP censored Poissin

CNB censored negative binomial

CZIP censored zero-inflated Poisson

CZINB censored zero-inflated negative binomial

CZIDW censored zero-inflated discrete Weibull

CHP censored hurdle Poisson

CHNB censored hurdle negative binomial

CHDW censored hurdle discrete Weibull

MDW median discrete Weibull

UPB unwanted pursuit behavior

Chapter 1

Introduction

Count data, which refers to the number of times an item or an event occurs within a fixed period of time, is essential in many fields. Indeed, examples of count data include the number of heart attacks or hospitalisation days in medical studies, the number of students absence during a period of time in education research or the number of times parents perpetrate domestic violence against their child(ren) in social science investigations. Count data can be found in many practical lifetime studies, such as the number of days before death in certain diseases or the number of cycles (runs) until a machine stops working and so on. Hence, a number of statistical distributions has been applied to model the case of a random variable (RV) with a non-negative integer value. A good overview of these distributions can be found in Johnson et al. (2005).

On the other hand, there is now a great deal of interest in the literature in investigating the relationship between a count response variable and other variables. In other words, count data is explained in terms of a set of covariates, for instance, how the education level of parents can affect the incidence of domestic violence against their children. Methods for addressing these questions fall into the general area of regression analysis for count data. Since these data are not normally distributed, ordinary least squares regression models are not appropriate and may yield misleading estimates for the impact of covariates.

Some statistical distributions, such as Poisson, geometric and NB are usually applied to model these count data. However, in some cases, these models cannot be considered as the best for fitting these RVs. Therefore, this thesis focuses on finding an appropriate model for the analysis of count data. DW, which is an analogue of the standard (continuous) Weibull distribution, is considered due to the vital role of its corresponding continuous Weibull distribution in modeling lifetime data and non-negative integers.

This chapter discusses an overview of some previous research and topics on the knowledge that related to our study.

1.1 Regression

According to Montgomery and Peck (1982), regression analysis can be simply defined as a statistical process that attempts to describe the relationship between a dependent variable and one or more corresponding value(s) of other RVs. That is, the process of building a statistical model represents the mathematical explanation of a RV based on other variables. Hence, the main objective of regression analysis is to find a function to forecast the change of the response variable on the basis of change in one or more predictors. This function involves the following variables:

1. **Dependent variable (Y):** it is the variable whose value is affected by and conditional upon other variables. This variable can also be called a response, measured, explained, outcome, experimental or output variable.

2. **Independent variable (X):** this value is not affected or dependent on other variables. It may also be known as the regressor or the controlled, concomitant, manipulated, explanatory or input variable.

3. **Unknown parameters (α):** any regression function is defined in terms of a finite number of unknown parameters, then the objective of the regression analysis is to estimate these parameters, based on the observed pairs of Y and X, to form the regression equation that measures the covariate effects on the data.

Specifically, the regression model relates a response variable Y to the function of $\boldsymbol{X}$ and $\boldsymbol{\alpha}$, as:

$$Y \approx f(\boldsymbol{X}, \boldsymbol{\alpha}).$$

1.1.1 Linear Model

This is a traditional regression model, depending on the normal distribution, and can be considered as the most familiar and widely used class of statistical models, relating the response variable Y to a linear combination of predictors. This model can be formulated for a sample, $(y_i, x_{i1}, x_{i2}, \ldots, x_{iP}; i = 1, 2, \ldots, n)$, in a mathematical equation as:

$$y_i = \alpha_0 + \sum_{p=1}^{P} x_{ip}\alpha_p + \epsilon_i$$

For simplification, it can be written as:

$$y_i = \boldsymbol{x}_i'\boldsymbol{\alpha} + \epsilon_i$$

where, y_i is the response variable, that is, the variable that is to be predicted or explored, $\boldsymbol{x}_i'$ is the vector of P predictors, $\boldsymbol{\alpha}$ refers to the P vector of regression parameters, which will be estimated and ϵ_i is a random error term or residual term, which reflects that this relationship between y_i and $\boldsymbol{x}_i$ is deterministic and not exact.

This linear regression model has a normally distributed random error ϵ and thus the outcome observations have normal distribution. That is, y_i can be distributed as $N(E(y_i), \sigma^2)$, where:

$$E(y_i|x_i) = \alpha_0 + \sum_{p=1}^{P} x_{ip}\alpha_p$$

On the other hand, real-life data are often not expected to be normally distributed. For instance, the general linear model is inadequate for count data for multiple reasons. First, this regression may result in the prediction of some non-positive counts. Additionally, the count data are usually highly skewed or have many zero counts, thus conflicting with the the variance assumption for the normal regression model (Gardner et al. (1995)). For instance, Hutchinson and Holtman (2005) compared the use of linear model, Poisson and NB regressions for analyzing the number of pregnancies experienced by sexually experienced adolescent females from some schools recorded by the National Longitudinal Survey of Adolescent Health. Their experiment showed the inappropriate use of normal linear regression, whereas Poisson and NB provided a better fit.

1.1.2 Generalized linear models

The normal linear regression is extended to a general class of statistical models, called GLMs, for relating a response variable Y to a linear combination of explanatory variable(s) $\boldsymbol{X}$. In other words, the normal distribution for the response variable Y is generalized to be any distribution from the exponential family. Generally, theses models consist of three components:

- A distribution that must be a member of the exponential family of distributions.
- The linear function of covariates, called linear predictors

$$\eta_i = \alpha_0 + \sum_{p=1}^{P} x_{ip}\alpha_p$$

- A link function, g, relates the predictor with the expected value μ of the response variable Y as:

$$g(\mu_i) = \eta_i = \boldsymbol{x_i'\alpha}$$

These approaches model only the conditional mean (μ) of the response variable, as a function of the explanatory variable, via the link function, as:

$$\mu_i = g^{-1}(\eta_i)$$

This class includes many of the statistical models from which the data in many applications may arise. For example, the normal linear model, which could be a good model for analyzing continuous data, usually equates the expected value of the response variable to a linear combination of the covariates and the regression parameters, that is, the link function here is the identity.
In addition, the Poisson regression for modeling discrete count data can be considered as a special case of GLMs, in which log(mean) is modeled as a linear function of the covariates. For more details, see Nelder and Wedderburn (1972).

1.1.3 Generalized additive models for location, scale and shape

Rigby and Stasinopoulos (2005) introduced this general class of statistical models to be more general than the GLM, in which the distribution of the response variable Y is not limited to the exponential family. Additionally, it can model not only the mean (location) but also any other parameter (θ) of the distribution of Y. The fully parametric class of this model could be described as:

$$g(\theta_i) = \boldsymbol{x_i'\alpha}$$

where θ could be any parameter of the population distribution, such as shape, location or scale parameter.

1.2 Maximum likelihood estimation method

Maximum likelihood is a very general technique for parameter estimation and inference in statistics. Suppose we have a density function $f(y;\underline{\theta})$, characterized by some unknown but fixed parameters $\underline{\theta}$, which could be a parameter θ or a vector of parameters $\underline{\theta} = (\theta_1, \theta_2, \ldots, \theta_P)$, where P is the number of parameters to be estimated. Then, the maximum likelihood method estimates these parameters by finding the values of θ that maximize the likelihood of Y and $\underline{\theta}$.

Due to the fact that the likelihoods are all positive and the logarithm is an increasing function, the log-likelihood is equivalent to the likelihood, and they have their maximum at the same point. Therefore, it would be easier to maximize the log-likelihood instead of the likelihood since the summation is easier than the product. In other words, this method of estimation can be briefly applied according to the following three steps:

1. **Likelihood Function:** the likelihood function for an observed sample $(y_1, y_2, \ldots, y_n)$ of size n, which is identically independent distributed (iid) as $f(y; \underline{\theta})$ and regarded as a function of $\underline{\theta}$ given the sample data, can be defined to be the joint probability function, as follows:

$$L(\underline{\theta}; y) = \prod_{i=1}^{n} f(y_i; \underline{\theta})$$

2. **Log-Likelihood Function:** the log-likelihood function is the natural logarithm of the likelihood function which is defined as follows:

$$\ell(\underline{\theta}; y) = \log(L(\underline{\theta}; y)) = \sum_{i=1}^{n} \log(f(y_i; \underline{\theta}))$$

3. **Maximum Likelihood Estimator:** an MLE $\hat{\theta}_{ML}$ of θ maximizes the likelihood, $L(\theta; y)$, or typically, the log-likelihood $\ell(\theta; y)$:

$$\hat{\theta}_{ML} = \arg\max_{\theta} \ell(\theta; y)$$

Optimizing the likelihood (or equivalently log-likelihood) functions can be found analytically by differentiating the log-likelihood function $\ell(\underline{\theta}; y)$ with respect to the parameter θ and setting the results equal to zero. However, for some complicated cases this may result in non-linear equations, which might require the application of numerical solutions using several algorithms. The complexity of the MLEs depends on the form of the probability function $f(y; \underline{\theta})$. In other words, the MLEs for the parameters of a normal distribution can be simply obtained by setting these derivatives of $\ell(\underline{\theta}; y)$ and solving for μ and σ^2. On the other hand, there is another cases where it is more difficult to find such explicit solutions, and thus numerical techniques are required.

The maximum likelihood method is the most commonly applied method of classical inference. This is due to its useful standard large sample properties, such as consistency and asymptotic normality.
Numerous studies have applied this technique to estimate parameters, especially for coefficient regression. In other words, the maximum likelihood approach can be applied to the traditional normal linear regression to estimate its parameters. Moreover, the maximum likelihood approach is used to fit most of the GLMs and Generalized additive models for location, scale and shape.

1.3 Count data modeling

The word *count* is generally used as a verb denoting the enumeration of some units or events that occurred within a period of time or in a specific place. Examples include the number of people that died in a disaster last year, how many items were purchased in the last week from a shop, the number of patients that were cured and left the hospital within the last three days, the number of days that a student was absent in the last year and so on. Then, count data is referring to the number of such enumerated items or events. Therefore, mathematically, this count data is represented by an RV that takes on only positive integer values because events cannot occur in negative numbers of times. The number of events taking place can take any positive number up to positive infinity; thus, there is no upper limits for counts. Additionally, there is a chance of having zero counts when the event is not experienced. That is, for any count RV, the range can be from zero to infinity (usually to some inferior distinct number, which is the maximum number that occurred in this dataset).

1.3.1 Dispersion for count data

It is important to clearly define the context of dispersion due to its essential role in modeling count data, and distributions for modeling these data should take into account the data's dispersion. Generally, the dispersion for any data can be described as the variability or spread of the data. In other words, dispersion refers to the stretch or the squeeze of a data's distribution. Specifically, dispersion in count data is formally defined in relation to a specified model being fitted to the data (Cameron and Trivedi (2013) and Hilbe (2014)). In this context, the variance ratio (VR) can be defined as the ratio between the observed variance from the data and the theoretical variance from the model fit, as:

$$\mathrm{VR} = \frac{\text{observed variance}}{\text{theoretical variance}} \tag{1.1}$$

Accordingly, modeling any count data might exhibit three types of dispersion; namely, over-dispersion, under-dispersion and equi-dispersion. Over-dispersion refers to the case when the observed variance of the count data is greater than the expected variance specified by the fitted model. Under-dispersion describes the opposite case, where the observed variance is less than that theorized by the model. Equi-dispersion refers to the case of equal variances. Then, a model that fails to capture the over- or under-dispersion in the data and shows different variance than that observed is called an over- or under-dispersed model. Therefore, the definition of dispersion through the VR can be helpful in studying the dispersion of a model.

Moreover, the dispersion of count data can be defined in relation to the Poisson model. Hence, it is common with these data to refer to the dispersion as being relative to Poisson. In such a case, the variance of the model is estimated by the sample mean. Thus, over-, equi- or under- dispersion relative to Poisson refers to cases where the sample variance (observed variance) is greater, equal or smaller than the sample mean (theoretical variance), respectively. Therefore, the dispersion of a dataset, under this definition, can be identified with regard to the Dip, or the dispersion coefficient, which is defined as the ratio of the variance to the mean (variance-to-mean relation):

$$Dip = \frac{\sigma^2}{\mu} \tag{1.2}$$

Then, data is over-, equi- or under-dispersion relative to Poisson if $Dip > 1$, $Dip = 1$ or $Dip < 1$, respectively. Dispersion has been commonly defined in the literature using the variance-to-mean ratio, namely, Dip. Specifically, the over- or under-dispersion relative to Poisson can be found in the data when the variance is greater or less than the mean (Cameron and Trivedi (2013), Hilbe (2014)). Accounting for over-dispersion and under-dispersion in modeling count data is essential because failing to cope with these cases can cause biased parameter estimates and thus lead to false conclusions and decisions.

A considerable amount of literature has been published on the regression analysis of the count response variable. For such data, Poisson and NB are the most popular and the most widely applied models for investigating the relationship between the outcome count variable and a set of covariates. Additionally, zero-inflated and hurdle models are applicable in the case where many zeros are counted in the data.

1.3.2 Poisson model

Generally, the GLM with the Poisson distribution is the classical and first choice to model any count data (Cameron and Trivedi (2013)). This regression model can be obtained based on the Poisson distribution with pmf as:

$$f(y) = \frac{\lambda^y e^{-\lambda}}{y!} \quad , \quad y = 0, 1, 2, \ldots \tag{1.3}$$

The parameter $\lambda(> 0)$, is the mean (and also the variance) of this Poisson distribution. Then, for a sample y_i; $i = 1, 2, ..., n$, and within the framework of GLM discussed previously, this distribution is generalized by allowing λ to be related to a set of covariates $\boldsymbol{x}_i = (x_{i_1}, x_{i_2}, \ldots, x_{i_P})$ with corresponding parameters $\boldsymbol{\alpha}$, through the log link function, as:

$$\lambda_i = e^{\boldsymbol{x}_i'\boldsymbol{\alpha}} \tag{1.4}$$

The exponential of $\boldsymbol{x}_i'\boldsymbol{\alpha}$, which makes the Poisson regression into a non-linear regression model, was chosen to ensure that λ_i remains positive and to guarantee that its predicted values will always be positive. Thus, the response variable Y represents the frequencies of an event of interest and $\boldsymbol{\alpha}$ is the vector of linearly independent predictors that are supposed to affect Y. In this regression model, $P+1$ parameters need to be estimated, that is, the regression coefficient $\boldsymbol{\alpha} = \alpha_0, \alpha_1, \ldots, \alpha_P$.
Although the Poisson model is widely considered to be the most basic model for analyzing count data in many disciplines, the reliance of this model on a single parameter often restricts its use on real data. This is due to the violated feature of Poisson distribution, which is the identical mean and variance, called "equi-dispersed" (Hilbe (2014)).

One common way to handle the issue of over-dispersion is to fit a parametric model that is more dispersed than the Poisson. A reasonable choice could be the NB model.

1.3.3 Negative binomial model

The NB model belongs to the GLM and relaxes the assumption of the equi-dispersion of Poisson regression by adding a dispersion parameter (heterogeneity or ancillary parameter) for considering the variability and allowing the variance to exceed the mean. In particular, this makes it possible to cope with the over-dispersion and the unobserved heterogeneity that might result from not considering some predictors for the count data. For the regression's purposes, the NB distribution can be derived as:

$$f(y) = \frac{\Gamma(y+k)}{\Gamma(k)y!}\Big(\frac{k}{k+\mu}\Big)^k\Big(\frac{\mu}{k+\mu}\Big)^y \tag{1.5}$$

where $\mu > 0$ is the mean for the NB, k is a scale or dispersion parameter and the variance of this model is:

$$\sigma^2 = \mu(\frac{k+\mu}{k})$$

which can be re-written as:

$$\sigma^2 = \mu + \frac{1}{k}\mu^2 \tag{1.6}$$

Since the theoretical variance of an NB is always greater than its mean, this regression model is the most commonly used for count data with over-dispersion relative to the Poisson. In other words, if the observed outcome is supposed to have a larger variance than the mean, then the NB regression is more appropriate. For the NB regression model, μ is associated with a set of covariates $\boldsymbol{X}$ with some corresponding parameters $\boldsymbol{\alpha}$ through the log-linear link function. That is, for a sample $(y_i, x_{i1}, x_{i2}, \ldots, x_{iP}; i = 1, 2, \ldots, n)$, μ_i is defined as :

$$\mu_i = e^{\boldsymbol{x}_i'\boldsymbol{\alpha}} \tag{1.7}$$

Again, the exponential of $\boldsymbol{x}_i'\boldsymbol{\alpha}$ was chosen to ensure that μ_i is positive. For more details, see for example, Lawless (1987) and Hilbe (2011).

A considerable amount of literature has been published on the NB. For example, Abdel-Aty and Radwan (2000) used it to fit the accident frequency in Central Florida. Additionally, Byers et al. (2003) applied an NB model in aging research for a clinical trial designed to assess the performance of a medical program for elderly people. In this study, the variance was much greater than mean, hence NB present better fit than Poisson regression model.

Even though the NB model is mainly suitable for over-dispersion relative to the Poisson situations, it is not appropriate for modeling the under-dispersion data relative to the Poisson (see for example Sellers and Shmueli (2010)). Hence, it is necessary to have some models to cope with the cases of under-dispersed data relative to the Poisson.

1.3.4 Censored count models

The censoring from below or above commonly occurs for count data. A popular example includes data that come from answering a question regarding a specific event, with possible responses of $0, 1, 2, 3$ or $4+$, which means four or more. This structure of censoring results from the pattern of the experiment. In addition, the censoring might be required in some cases, where the response takes large values or outliers affecting its mean and variance, causing over-dispersion relative to Poisson. Thus, cutting the large values of this response can control this over-dispersion.

The most common type of censoring associated with count data is right censoring, where a point is considered to cut the observed counts from the right. That is, for some values of the response variable Y that are greater than a fixed value C, it is recorded as greater than or equal to C, with no knowledge regarding the exact value of Y. This case can be found in many applications, for example, if a study is interested in investigating the relationship between heavy smokers and their income. Such a study may define a person as being a heavy smoker if he or she smokes 10 cigarettes or more per a day. Then, the response variable, which is the number of daily cigarettes consumed, can be recorded as, $0, 1, 2, \ldots, 10$ or more, even for individuals who smoke 11 or more cigarettes. Thus, the response variable here is censored at $C = 10$. However, the independent variable, which is the income, is exactly recorded for the whole sample, even for the censored respondents who smoke more than 10 cigarettes daily. A considerable amount of literature has been published on censored count data. For example, Terza (1985) analyzed censored data on the number of times individuals shopped in an area in a given period of time. The observed number of times in this experiment were 0, 1, 2 and 3 or more. Thus, even if there were people who shopped four or more times, they would be listed under the category of 3 or more. Another example is in fertility, where Caudill and Mixon Jr (1995) considered the censoring case for their dependent variable, that is, the number of children in the family.

For the case of censored count data, it is necessary to have special models that account for this restriction. Otherwise, if the regular count models (full or uncensored models) are applied without considering this censoring, the resulting inferences might be inappropriate. Thus, a variety of studies have applied regression models for such censored count data. The Poisson regression model, which is typically used for count data, was applied by Terza (1985) and Brännäs (1992). However, the Poisson model has the disadvantage of its equi-dispersion assumption, which is the identical mean and variance of the data, as mentioned previously. Thus, Caudill and Mixon Jr (1995) applied the censored negative binomial (CNB) model, which is more capable for modeling over-dispersion with censored data, and it provided better fit than the censored Poissin (CP) regression model. Additionally, a number of studies have modeled censored data using models that can handle both cases for over- and under-dispersion. For example, Famoye and Wang (2004) and Mahmoud and Alderiny (2010) used generalized Poisson regression for censored data, and Sellers and Shmueli (2010) applied COM-Poisson regression for censored data.

Generally, for regular uncensored count data, the type of underlying dispersion can be noted from the relation between the sample mean and the sample variance for the response. However, for the censored data this relationship is not completely known. That is, for the right censored data the observed mean will be smaller than the true mean, and the observed variance could be less than the true one. However, if the observed variance is greater than the observed mean, the true variance could be less than or greater than the true mean. Then, the dispersion type will not be clear for the censored data. Consequently, misleading inferences may result if the underlying dispersion is not taken into account. Hence, it would be very useful to utilize a model that can handle a variety of dispersion types for this censored data where the type of dispersion is unknown.

The concept of censoring for count data can be summarized into the case when the data $(Y, \boldsymbol{X})$ are available for some range of Y, but all the values of $\boldsymbol{X}$ are observed. Thus, censoring should be taken into account, as it includes some loss of information, which might result in misleading estimates. Thus, for a count variable Y^* and some fixed positive integer C, the censoring count model can be defined as follows:

- **Right censoring**

  C is the biggest observed value in the model. Then, any response in the data greater than C is considered to be greater than or equal to C.

- **Left censoring**

  C is the smallest value in the model. Hence, any y^* less than C is considered as less than or equal to C.

However, the most common censoring type, as mentioned earlier, is right censoring at C; hence, this type is considered in this study. Subsequently, some values for y^* are incompletely observed, as they will be recorded as greater than or equal to C, and their real values are unspecified. In other words, under the right censoring scheme, the observed response variable y_i can be defined as:

$$y_i = \begin{cases} y_i^* & \text{if} \quad y_i^* < C \\ C & \text{if} \quad y_i^* \geq C \end{cases}$$

Therefore, if Y^* has a pmf $f(Y^*|x)$ and cumulative distribution function (cdf) $F(Y^*|x)$, we have:

- For $Y^* < C$, y is the observed and the pmf of y will be the usual $f(y|x)$
- For $Y^* \geq C$, C is observed with a probability $Pr(Y^* \geq C)$, where

$$Pr(y^* \geq C) = \sum_{j=C}^{\infty} f(y = j|x) = 1 - \sum_{j=0}^{C-1} f(y = j|x) = 1 - F(C - 1|x) \quad (1.8)$$

Thus, a binary indicator for the censoring can be defined to combine these two terms as:

$$\delta_{c_i} = \begin{cases} 0 & \text{if} \quad y_i^* < C \\ 1 & \text{if} \quad y_i^* \geq C \end{cases} \quad (1.9)$$

Then, the likelihood function can be defined as:

$$L = \prod_{i=1}^{n} \left[f(y_i|x_i)\right]^{1-\delta_{c_i}} \left[Pr(y_i^* \geq C|x_i)\right]^{\delta_{c_i}} \tag{1.10}$$

Hence, the log-likelihood function is:

$$\ell = \sum_{i=1}^{n} (1-\delta_{c_i}) \log\left[f(y_i|x_i)\right] + \sum_{i=1}^{n} \delta_{c_i} \log\left[Pr(y_i^* \geq C|x_i))\right]$$

Then, from Equation 1.8:

$$\ell = \sum_{i=1}^{n} (1-\delta_{c_i}) \log\left[f(y_i|x_i)\right] + \sum_{i=1}^{n} \delta_{c_i} \log\left[1 - F(C-1|x_i)\right]$$

where $f(.)$ and $F(.)$ are, respectively, the pmf and cdf for y^*. For more details, see, for example, Cameron and Trivedi (2013) and Hilbe (2014).

1.3.4.1 Censored Poisson regression model

From the pmf of the Poisson regression in Equation 1.3 with λ as in Equation 1.4, then, from Equation 1.10, the likelihood for the right CP regression model can be obtained as:

$$L = \prod_{i=1}^{n} \left[\frac{\lambda_i^{y_i} e^{-\lambda_i}}{y_i!}\right]^{1-\delta_{c_i}} \left[1 - e^{-\lambda_i} \sum_{j=0}^{C-1} \frac{\lambda^j}{j!}\right]^{\delta_{c_i}} \tag{1.11}$$

where C is the censored point.

1.3.4.2 Censored negative binomial regression model

For the NB regression with the pmf defined as Equation 1.5 and μ in Equation 1.7, the likelihood for the right CNB regression model can be obtained from Equation 1.10 as:

$$\begin{aligned} L = \prod_{i=1}^{n} \left[\frac{\Gamma(y_i+k)}{\Gamma(k) y_i!} \left(\frac{k}{k+\mu_i}\right)^k \left(\frac{\mu_i}{k+\mu_i}\right)^{y_i}\right]^{1-\delta_{c_i}} \times \\ \left[1 - \frac{1}{\Gamma k} \left(\frac{k}{k+\mu_i}\right)^k \sum_{j=0}^{C-1} \frac{\Gamma(j+k)}{j!} \left(\frac{\mu_i}{k+\mu_i}\right)^j\right]^{\delta_{c_i}} \end{aligned} \tag{1.12}$$

where C is the censored point.

For more details on these models see, for example, Terza (1985), Caudill and Mixon Jr (1995), Cheol Jung et al. (2006) and Raciborski et al. (2011), among others.

Count data may possess a huge amount of zeros in many experiments. The excess number of observed zeros results when many individuals fail to experience the event of interest. To illustrate, in a study investigating smoking, the count variable for the number of cigarettes smoked during the last two hours could have excess zeros because the zeros from the population could be recorded from two cases: first, from participants who are smokers but chose not to smoke during the last two hours, and second, from non-smoking individuals.
Hence, in these cases, the zero counts can be categorized as having two different origins:

- The first observed zeros (sampling zeros or false zero), obtained by chance in the sample, are due to the usual Poisson or NB. Thus, in the smoking study above, the smokers who did not smoke during the last two hours are supposed to be modeled as Poisson or NB, which include both zero counts (sampling zeros) and non-zero counts.

- The second zeros (structural zeros or true zero) are caused by some structure in the data when the event cannot be exhibited for some reasons. That is, in the smoking example, the zeros observed for non-smokers are structural zeros because they cannot exhibit a non-smoking condition in any period of time. This arises from a binary process: smoking.

The over-dispersion in the count data can arise for the dataset with too many response zero counts, which the Poisson and NB regressions cannot predict correctly. In other words, the zero count is greater than that predicted by the Poisson or NB models.
Therefore, a modified method should be applied to address this issue of many zeros in the count dataset. Particularly, the zero-inflated and hurdle models were suggested for this condition of zero inflation. Generally, these modified models can be described as having two parts, one procedure for zero counts and a different one for positive counts. In other words, the first part defines a binary process, commonly logistic models, for having zero or count values. The second part considers a discrete distribution, called the parent count model, conditional

on a count value, such as the Poisson or NB for zero-inflated models or their zero-truncated formulas for the hurdle models. For more details, see Cameron and Trivedi (2013) and Hilbe (2014).

1.3.5 Zero-inflated models

Two structures for generating the observations can be considered for these models: the first group generates only zero counts and the other one a positive integer outcome (including the zero counts). Some Bernoulli trials can be used to direct which process is observed. Then,

$$Y \sim \begin{cases} 0 & \text{with probability } \pi \\ g(y;\underline{\theta}) & \text{with probability } (1-\pi) \end{cases}$$

According to Cameron and Trivedi (2013), the zero-inflated models can be derived as a two-component mixture models, that is, mixing a point mass at zero and a count distribution, $f_p(y)$, such as Poisson or NB; namely, parent count model. Consequently, the zero count could be observed due to two different origins: from the point mass and from the count component. Thus, the zero-inflated regression models have pmf, as follows:

$$f(y_i) = \begin{cases} \pi_i + (1-\pi_i)f_p(0) & \text{for y=0} \\ (1-\pi_i)f_p(y_i) & \text{for y=1, 2, 3, \ldots} \end{cases} \tag{1.13}$$

where y is the count variable, $0 < \pi < 1$ is a zero-inflation parameter (the probability or proportion of a structural zero) and $f_p(.)$ is the pmf of the parent count model with some vector of parameters $\underline{\theta}$.

In the zero-inflated regression model, the proportion parameter, π, and some parameter in the vector $\underline{\theta}$, can be related to some sets of covariates z_i and x_i, respectively, in which these predictors could be the same, $z_i = x_i$, or different predictors could affect the data, $z_i \neq x_i$.

To obtain the likelihood of the zero-inflated models, a binary indicator needs to be defined, as follows:

$$\delta_{z_i} = \begin{cases} 1 & \text{if} \quad y_i = 0 \\ 0 & \text{if} \quad y_i > 0 \end{cases} \tag{1.14}$$

Then, based on the regression structure, the likelihood can be described as:

$$L = \prod_{i=1}^{n} \left\{ \pi(z_i) + (1 - \pi(z_i)) f_p(0|x_i) \right\}^{\delta_{z_i}} \left\{ (1 - \pi(z_i)) f_p(y_i|x_i) \right\}^{1-\delta_{z_i}} \tag{1.15}$$

Although the proportion π can take any link function that transforms π from the probability scale to the interval $[-\infty, +\infty]$, this study assumes the logit link function for π as it is the most common link for similar models. Hence, π can be related to a set of covariates, as follows:

$$logit(\pi(z_i)) = \log\left(\frac{\pi(z_i)}{1 - \pi(z_i)}\right) = \boldsymbol{z}_i'\boldsymbol{\gamma}$$

Then, this proportion can be rewritten as:

$$\pi_i \equiv \pi(\boldsymbol{z}_i) = \left(e^{-\boldsymbol{z}_i'\boldsymbol{\gamma}} + 1\right)^{-1} \tag{1.16}$$

For more details, see Cameron and Trivedi (2013) and Staub and Winkelmann (2013).

1.3.5.1 Zero-inflated Poisson model

The ZIP regression model was introduced by Lambert (1992) for analyzing manufacturing data and investigating the number of defects in equipment. Poisson models are mixed with zeros to allow for the excessive zeros in the data, which is commonly encountered in real life. Let the parent distribution $f_p(.)$ in Equation 1.13 be the Poisson with pmf in Equation 1.3; then, the ZIP can be derived as follows:

$$f(y_i|x_i,z_i) = \begin{cases} \pi_i + \big(1-\pi_i\big)e^{-\lambda_i} & \text{for } y_i = 0 \\ (1-\pi_i)\dfrac{\lambda_i^{y_i}e^{-\lambda_i}}{y_i!} & \text{for } y_i = 1,2,3,\ldots \end{cases} \tag{1.17}$$

where λ_i is defined in Equation 1.4, π_i is defined in Equation 1.16.

The $\boldsymbol{x}$ and $\boldsymbol{z}$ are the set of covariates with $\boldsymbol{\alpha}$ and $\boldsymbol{\gamma}$ as regression coefficients, respectively.

1.3.5.2 Zero-inflated negative binomial model

The ZINB regression model has more flexibility in its variance due to the additional parameter, compared to ZIP. Let the parent distribution $f_p(.)$ in Equation 1.13 be NB with pmf in Equation 1.5; then, the ZINB can be defined as follows:

$$f(y_i|x_i,z_i) = \begin{cases} \pi_i + \big(1-\pi_i\big)\Big(\dfrac{k}{k+\mu_i}\Big)^k & \text{for } y_i = 0 \\ (1-\pi_i)\frac{\Gamma(y_i+k)}{\Gamma(k)y_i!}\Big(\dfrac{k}{k+\mu_i}\Big)^k\Big(\dfrac{\mu_i}{k+\mu_i}\Big)^{y_i} & \text{for } y_i = 1,2,3,\ \ldots \end{cases} \tag{1.18}$$

where, μ_i is defined in Equation 1.7, and π_i is defined in Equation 1.16. The $\boldsymbol{x}$ and $\boldsymbol{z}$ are the set of covariates with $\boldsymbol{\alpha}$ and $\boldsymbol{\gamma}$ as regression coefficients, respectively.

A considerable amount of literature has been published on zero-inflated models in different fields. For example, Kong et al. (2015) used some modifications of the ZINB to analyze the dental caries of children in Iowa. In public health, Lam et al. (2006) applied ZIP for medical research.

1.3.6 Hurdle models

The hurdle model was proposed by Mullahy (1986) for count data modeling, but the term was first used by Cragg (1971). These models were developed to cope with response variables that have excessive zero outcomes, alternatively with the zero-inflated models. The idea behind the hurdle model is to separate the statistical process that governs zero and non-zero counts and divides the model into two parts: first, a binary process for generating the zero counts only, using a binary model, for example, the logit, probit or complementary log-log models; and second, a process that generates positive non-zero counts only, using a truncated count model such as the truncated Poisson or NB.

Therefore, compared to the zero-inflated models, hurdle models and zero-inflated models both are used in the case of zero inflation; however, they are different in how they analyze the zeros. In other words, unlike zero-inflated models, hurdle models assume that all zero counts arise from the structural zeros or true zero. The positive non-zero counts come from the sampling structure, that is, the truncated Poisson or NB. Therefore, in the hurdle model, zero counts are not allowed in the second step, whereas zero count could arise in either part in the zero-inflated models.

Hence, in the study on smoking, for example, if the count number of the cigarettes smoked during the last three months, is considered. Then, the zeros would be observed only as a structural origin for the non-smoking individuals. Thus, it would be better to choose the hurdle model for this analysis.

Suppose a count RV Y takes the values from zero to some positive number. Then, assume the zero counts are generated by some binary process, π, and the non-zero positive counts are observed with a probability based on a truncated parent count model $f_{p_{tr}}(y) = \dfrac{f_p(y)}{1-f_p(0)}$. The hurdle regression model can then be described as:

$$f(y_i) = \begin{cases} \pi_i & \text{for y=0} \\ (1-\pi_i)\dfrac{f_p(y_i)}{1-f_p(0|x_i)} & \text{for y=1, 2, 3, \ldots} \end{cases} \tag{1.19}$$

The truncated part is multiplied by $1-\pi$ to ensure that the summation of the probabilities is one. Then, for example, in the smoking study mentioned earlier, the $\pi(0)$ includes both zeros, those that come from non-smokers and those that come from smokers who chose not to smoke during that period. Conversely, $f_p(0)$ is for smokers who choose not to smoke in the last two hours. Subsequently, the truncated function $\dfrac{f_p(y)}{1-f_p(0)}$ corresponds only to smokers who smoked y (at least one) cigarettes in the last two hours.

The parent count model $f_p(.)$ in Equation 1.19, depending on the parameter(s) $\underline{\theta}$, is the pmf of the positive non-zero response count variable Y. This function can be Poisson or NB, with some parameters in $\underline{\theta}$ depending on some covariates $\boldsymbol{X}$ with regression parameter α. In addition, the logit transformation is considered in this study for π to model the binary outcome $Y = 0$ versus $Y > 0$ conditioning in $\boldsymbol{Z}$, as in Equation 1.16.

Thus, the regression parameter can be different for θ than π, and the covariates $\boldsymbol{X}$ and $\boldsymbol{Z}$ might be the same or different.

Then, using the indicater variable δ_{z_i} in Equation 1.14, the likelihood function can be written as follows:

$$L = \prod_{i=1}^{n} \left[\pi(z_i)\right]^{\delta_{z_i}} \left[(1 - \pi(z_i)) \frac{f_p(y_i|x_i)}{1 - f_p(0|x_i)}\right]^{1-\delta_{z_i}}$$

which can be rewritten as:

$$L = L_1 \times L_2 \tag{1.20}$$

where

$$L_1 = \prod_{i=1}^{n} \left[\pi(z_i)\right]^{\delta_{z_i}} \left[(1 - \pi(z_i))\right]^{1-\delta_{z_i}}$$

and

$$L_2 = \prod_{i=1}^{n} \left[\frac{f_p(y_i|x_i)}{1 - f_p(0|x_i)}\right]^{1-\delta_{z_i}}$$

Thus, theses two likelihoods could be maximized separately with regard to π and θ, respectively.

1.3.6.1 Hurdle Poisson model

The hurdle Poisson (HP) can be defined using Equation 1.19 as follow:

$$f(y_i|x_i, z_i) = \begin{cases} \pi_i & \text{for } y_i = 0 \\ (1 - \pi_i)\dfrac{\lambda^{y_i}}{(e^{\lambda_i} - 1)y_i!} & \text{for } y_i = 1, 2, 3, \ldots \end{cases}$$

where λ_i and π_i are defined in Equation 1.4 and Equation 1.16, respectively.

1.3.6.2 Hurdle negative binomial model

According to Equation 1.19, the hurdle negative binomial (HNB) model can be written as:

$$f(y_i|x_i,z_i) = \begin{cases} \pi_i & \text{for } y_i = 0 \\ (1-\pi_i)\dfrac{\Gamma(y_i+k)}{\Gamma(k)\Gamma(y_i+1)}\left(\dfrac{\mu_i}{k+\mu_i}\right)^{y_i}\dfrac{1}{\left(\dfrac{k+\mu_i}{k}\right)^k - 1} & \text{for } y_i = 1,2,3,\ldots \end{cases}$$

where μ_i and π_i are defined in Equation 1.7 and Equation 1.16, respectively.

A number of researchers have applied hurdle models. For example, Silva and Covas (2000) applied a modification of hurdle models to a fertility study. Additionally, Bilgic and Florkowski (2007) used an HNB model to analyze the demand for a bass fishing trip in the southeastern United States.

For more details about the zero-inflated and hurdle models, one can refer to Hilbe (2014), Cameron and Trivedi (2013), Zorn (1996) and Hu et al. (2011).

Generally, all the above models can be applied in many standard statistical packages, such as, R, Zeileis et al. (2007). For more details about the regression analysis of count data, see Cameron and Trivedi (2013).

The type of data considered in the following sections is a combination of the previous two types, that is, censored count responses with excessive zero counts. The motivation behind introducing these models is based on two objectives. The first objective is to handle data that show excessive zero counts together with censoring, as mentioned previously. Second, censoring some values from the right might overcome the over-dispersion in the count data caused by including too many zeros (Saffari et al. (2012)). The models for censored count responses with excessive zero counts can be described as follows:

1.3.7 Censored zero-inflated models

From Equation 1.10 and Equation 1.13, the likelihood of the zero-inflated models with right censored data can be defined as:

$$
\begin{aligned}
L = \prod_{i=1}^{n} & \Big\{ \big[\pi(z_i) + (1-\pi(z_i)) f_p(0|x_i)\big]^{\delta_{z_i}} \big[(1-\pi(z_i)) f_p(y_i|x_i)\big]^{1-\delta_{z_i}} \Big\}^{1-\delta_{c_i}} \\
& \Big\{ 1 - \Big[\pi(z_i) + (1-\pi(z_i)) f_p(0|x_i) + (1-\pi(z_i)) \sum_{j=1}^{C-1} f_p(y=j|x_i)\Big] \Big\}^{\delta_{c_i}}
\end{aligned}
$$

which can be rewritten, using the cdf of the parent count model, $F_p(.)$ as:

$$
\begin{aligned}
L = \prod_{i=1}^{n} & \left\{ [\pi(z_i) + (1-\pi(z_i)) f_p(0|x_i)]^{\delta_{z_i}} [(1-\pi(z_i)) f_p(y_i|x_i)]^{1-\delta_{z_i}} \right\}^{1-\delta_{c_i}} \\
& \{1 - [\pi(z_i) + (1-\pi(z_i))\, F_p(C-1|x_i)]\}^{\delta_{c_i}}
\end{aligned}
\tag{1.21}
$$

Then, the log-likelihood would be as follows:

$$
\begin{aligned}
\ell = & \sum_{i=1}^{n} (1-\delta_{c_i})\delta_{z_i} \log\left[\pi(z_i) + (1-\pi(z_i)) f_p(0|x_i)\right] + \\
& \sum_{i=1}^{n} (1-\delta_{c_i})(1-\delta_{z_i}) \log\big(1-\pi(z_i)\big) + \sum_{i=1}^{n} (1-\delta_{c_i})(1-\delta_{z_i}) \log\left(f_p(y|x_i)\right) + \\
& \sum_{i=1}^{n} \delta_{c_i} \log\left\{1 - [\pi(z_i) + (1-\pi(z_i))\, F_p(C-1)]\right\}
\end{aligned}
\tag{1.22}
$$

1.3.7.1 Censored zero-inflated Poisson

As discussed in Saffari and Adnan (2011) and Saffari et al. (2013), the likelihood of the censored zero-inflated Poisson (CZIP) model can be defined, using Equation 1.3 and Equation 1.21, as:

$$\begin{aligned} L = \prod_{i=1}^{n} & \left\{ \left[\pi_i + (1-\pi_i)e^{-\lambda_i}\right]^{\delta_{z_i}} \left[(1-\pi_i)\left(\frac{\lambda_i^{y_i}e^{-\lambda_i}}{y_i!}\right)\right]^{1-\delta_{z_i}} \right\}^{1-\delta_{c_i}} \times \\ & \left\{1 - \left[\pi_i + (1-\pi_i)e^{-\lambda_i}\sum_{j=0}^{C-1}\frac{\lambda_i^{j}}{j!}\right]\right\}^{\delta_{c_i}} \end{aligned} \tag{1.23}$$

where C is a censored point and λ_i and π_i, are defined in Equation 1.4 and Equation 1.16, respectively.

1.3.7.2 Censored zero-inflated negative binomial

The censored zero-inflated negative binomial (CZINB) has been considered in Saffari and Adnan (2011), then the likelihood of this model can be defined, using Equation 1.5 and Equation 1.21, as:

$$\begin{aligned} L = \prod_{i=1}^{n} & \left\{ \left[\pi_i + (1-\pi_i)\left(\frac{k}{k+\mu}\right)^{k}\right]^{\delta_{z_i}} \left[(1-\pi_i)\frac{\Gamma(y_i+k)}{\Gamma(k)y_i!}\left(\frac{k}{k+\mu_i}\right)^{k}\left(\frac{\mu_i}{k+\mu_i}\right)^{y_i}\right]^{1-\delta_{z_i}} \right\}^{1-\delta_{c_i}} \\ & \times \left\{1 - \left[\pi_i + (1-\pi_i)\frac{1}{\Gamma(k)}\left(\frac{k}{k+\mu_i}\right)^{k}\sum_{j=0}^{C-1}\frac{\Gamma(j+k)}{j!}\left(\frac{\mu_i}{k+\mu_i}\right)^{j}\right]\right\}^{\delta_{c_i}} \end{aligned} \tag{1.24}$$

where C is a censored point and μ_i and π_i, are defined in Equation 1.7 and Equation 1.16, respectively.

1.3.8 Censored hurdle models

From Equation 1.10 and Equation 1.20, the likelihood of the hurdle models with right censoring can be defined as:

$$L = \prod_{i=1}^{n} \left\{ [\pi(z_i)]^{\delta_{z_i}} \left[1-\pi(z_i)\right]^{1-\delta_{z_i}} \left[\frac{f_p(y_i|x_i)}{1-f_p(0|x_i)}\right]^{1-\delta_{z_i}} \right\}^{1-\delta_{c_i}} \left\{1-\left[\pi(z_i)+\frac{(1-\pi(z_i))}{1-f_p(0|x_i)}\sum_{j=1}^{C-1} f_p(y=j|x_i)\right]\right\}^{\delta_{c_i}}$$

which can be rewritten, using the cdf of the parent count model, $F_p(.)$, as:

$$L = \prod_{i=1}^{n} \left\{ [\pi(z_i)]^{\delta_{z_i}} \left[1-\pi(z_i)\right]^{1-\delta_{z_i}} \left[\frac{f_p(y_i|x_i)}{1-f_p(0|x_i)}\right]^{1-\delta_{z_i}} \right\}^{1-\delta_{c_i}} \left\{1-\left[\pi(z_i)+\frac{(1-\pi(z_i))}{1-f_p(0|x_i)}\left(F_p(C-1|x_i)-f_p(0|x_i)\right)\right]\right\}^{\delta_{c_i}} \tag{1.25}$$

Then, the log-likelihood would be as follows:

$$\ell = \sum_{i=1}^{n}(1-\delta_{c_i})\delta_{z_i}\log\left(\pi(z_i)\right) + \sum_{i=1}^{n}(1-\delta_{c_i})(1-\delta_{z_i})\log\left(1-\pi(z_i)\right) + \sum_{i=1}^{n}(1-\delta_{c_i})(1-\delta_{z_i})\log\left(f_p(y_i|x_i)\right) - \sum_{i=1}^{n}(1-\delta_{c_i})(1-\delta_{z_i})\log\left(1-f_p(0|x_i)\right) + \sum_{i=1}^{n}\delta_{c_i}\log\left\{1-\left[\pi(z_i)+\frac{(1-\pi(z_i))}{1-f_p(0|x_i)}\left(F_p(C-1|x_i)-f_p(0|x_i)\right)\right]\right\} \tag{1.26}$$

1.3.8.1 Censored hurdle Poisson

The censored hurdle Poisson (CHP) has been studied previously in SAFFAR et al. (2012).The likelihood of this model can be defined, using Equation 1.3 and Equation 1.25, as:

$$
\begin{aligned}
L = \prod_{i=1}^{n} & \left\{ [\pi_i]^{\delta_{z_i}} \, [1-\pi_i]^{1-\delta_{z_i}} \left[\frac{\lambda_i^{y_i} e^{-\lambda_i}}{y_i! \, (1-e^{-\lambda_i})} \right]^{1-\delta_{z_i}} \right\}^{1-\delta_{c_i}} \\
& \left\{ 1 - \left[\pi_i + (1-\pi_i) \frac{e^{-\lambda_i}}{1-e^{-\lambda_i}} \left(\sum_{j=0}^{C-1} \frac{\lambda^j}{j!} - 1 \right) \right] \right\}^{\delta_{c_i}}
\end{aligned}
\tag{1.27}
$$

where C is a censored point and λ_i and π_i, are defined in Equation 1.4 and Equation 1.16, respectively.

1.3.8.2 Censored hurdle negative binomial

According to Saffari et al. (2012), the likelihood of the censored hurdle negative binomial (CHNB) model can be defined, using Equation 1.5 and Equation 1.25, as:

$$
\begin{aligned}
L = \prod_{i=1}^{n} & \left\{ [\pi_i]^{\delta_{z_i}} \, [1-\pi_i]^{1-\delta_{z_i}} \left[\frac{\Gamma(y_i+k)}{\Gamma(k) y_i! \left[1 - \left(\frac{k}{k+\mu_i} \right)^{k} \right]} \left(\frac{k}{k+\mu_i} \right)^{k} \left(\frac{\mu_i}{k+\mu_i} \right)^{y_i} \right]^{1-\delta_{z_i}} \right\}^{1-\delta_{c_i}} \\
& \left\{ 1 - \left[\pi_i + (1-\pi) \frac{\left(\frac{k}{k+\mu_i} \right)^{k}}{1 - \left(\frac{k}{k+\mu_i} \right)^{k}} \left(\frac{1}{\Gamma(k)} \sum_{j=0}^{C-1} \frac{\Gamma(j+k)}{j!} \left(\frac{\mu_i}{k+\mu_i} \right)^{j} - 1 \right) \right] \right\}^{\delta_{c_i}}
\end{aligned}
\tag{1.28}
$$

where C is a censored point and μ_i and π_i, are defined in Equation 1.7 and Equation 1.16, respectively.

1.4 Programming language and software for computations

All the analyses and computations in this thesis have been carried out using the R programming language (R Core Team (2014)); some of the applied packages and functions are as follows:

- *"optim"* function with the derivative-free optimization routine, *Nelder-Mead*, to optimize the log-likelihood function and obtain the MLEs numerically.
- *glm* function to fit the Poisson model.
- *"glm.nb"* function in the *MASS* package to fit the NB model.
- *DiscreteWeibull* package (Barbiero (2015)) to sample from and fit by DW distribution.
- *pscl* package for the zero-inflated and hurdle models.
- *Ecdat* (Croissant (2015)) and *COUNT* (Hilbe (2014)), for data on which to apply the method

1.5 Motivations and contributions

Due to the essential role of the continuous Weibull distribution in modeling lifetime data, the DW distribution is introduced analogously by Nakagawa and Osaki (1975) for the positive discrete RVs. Although there is some previous studies have examined this distribution for count data with over-dispersion case, no research has been found that investigated the DW for an under-dispersion situation. Moreover, so far this distribution has only been applied to the univariate count data analysis and no previous research considered this model to a response count data in a regression structure.

Hence, the overall aim of this thesis is to introduce the DW regression model as a unified model for capturing different levels of dispersion in count data. In addition, some modifications for this DW model have been discussed in this study. In other words, the CDW model is suggested for the censored count data, where the type of dispersion is not easily identified and a model that has the ability to model different type of dispersion is highly recommended to apply. Furthermore, ZIDW and HDW are proposed for the zero-inflation count data, in which the existence of excessive zeros may increase the over-dispersion of the data and hide the under-dispersion within the sub-groups of this data. Thus, it could be better to consider a model that can handle a variety dispersion levels for this type of count data.

Even though some research has been carried out on finding some appropriate models that can cope with different type of dispersion, most of them are based on some extensions from the Poisson model and no simple model exists for these different type of count data. For instance, some modifications from the simple Poisson model included; quasi-Ppoisson, generalized Poisson, double Poisson, COM-Poisson and hyper-Poisson are all has the ability to handle the under-dispersion case for count data. These models have some limitations such as, the lack of a likelihood, complexity and the intensive computations. However, this thesis considers a simple and basic DW regression model for the analysis of count data.

1.6 Thesis outline

This thesis is divided into chapters as follows: chapter 2 discusses the DW distribution, its properties and its ability to handle different type of data dispersions. Then, chapter 3 introduces the DW regression model, one of the parameters of which is considered to be a function of explanatory variables. After that, the DW regression model is modified for a different type of count data, that is, censored, in chapter 4. Additionally, two different types of count data are considered in chapter 5: those with excessive zero counts and those with excessive zero counts and right censoring. Two modifications of the DW model are developed to cope with the excessive zero counts in the dataset, namely, the ZIDW and the HDW. Consequently, CZIDW and CHDW are considered for the excessive zero counts with right censoring. chapter 6 re-parameterizes the DW regression model through its median to obtain a likelihood, where the direct effect of the covariates on a location measurement is of interest.

For each chapter, from chapter 2 to chapter 6, the models are fitted using maximum likelihood estimation. Then, the performance of these MLEs is evaluated via Monte Carlo simulation studies, and the models are applied to real data sets and compared to their corresponding related Poisson and NB regression models. Finally, chapter 7 concludes this thesis by discussing and summarizing the main results of the research. Additionally, some recommendations for possible future research directions are suggested.

Chapter 2

Discrete Weibull distribution

2.1 Introduction

This chapter discusses the DW distribution, presented by Nakagawa and Osaki (1975), and some of its properties. The motivation behind considering the DW distribution, stems from the vital role played by the continuous Weibull distribution in the survival analysis and failure time studies. The estimation and inference for parameters of a DW distribution have been investigated in a small number of studies. Khan et al. (1989) proposed the method of proportion whereas Kulasekera (1994) suggested MLEs of the DW parameters based on type I censored samples. The count data application examples of DW include Englehardt and Li (2011) and Englehardt et al. (2012), who showed that the counts of living microbes (pathogen) in water are highly skewed and can be efficiently modeled using a DW distribution.

2.2 Discrete Weibull distribution

The DW introduced with relation to the continuous Weibull distribution for lifetime data, in three different types in the literature namely; type I, typeII and type III. Type I and type II have been obtained by starting from the continuous Weibull distribution; in which type I retains the form of the continuous cdf while type II retains the form of the continuous hazard rate. However, type II defined only for a limited range for the RV, which is not very applicable since it is not really known when the end of a lifetime RV would be. Type III dose not start from the continuous Weibull distribution but tries to generalize the notions of hazard rate and mean residual life to the discrete case. For more details see Bracquemond and Gaudoin (2003) and Rinne (2008).

The type II DW (c, β) introduced by Stein and Dattero (1984) with a hazard rate:

$$h(y) = \begin{cases} cy^{\beta-1} & \text{for } y = 1, 2, 3, \ldots, m \\ 0 & \text{otherwise} \end{cases}$$

where $\beta > 0$ and $0 < c \leq 1$ and

$$m = \begin{cases} int\{c^{-[1/(\beta-1)]}\} & \text{if } \beta > 1 \\ +\infty & \text{if } \beta \leq 1 \end{cases}$$

Furthermore, type III DW was proposed by Padgett and Spurrier (1985) with pmf, as follows:

$$f(y) = e^{-c\sum_{i=1}^{y} i^{\beta}}(1 - e^{-c(y+1)^{\beta}}) \qquad , \qquad y = 0, 1, 2, \ldots$$

where, $c > 0$ and $-\infty < \beta < +\infty$.

This study focuses on the most common type in the literature, which is the first type of DW distribution, that will simply be denoted as DW distribution. This type has been defined by Nakagawa and Osaki (1975), as follows:

If Y follows a type I DW distribution, then the cdf of Y is giving by:

$$F(y) = \begin{cases} 1 - q^{(y+1)^{\beta}} & \text{for } y = 0, 1, 2, 3, \ldots \\ 0 & \text{otherwise} \end{cases} \tag{2.1}$$

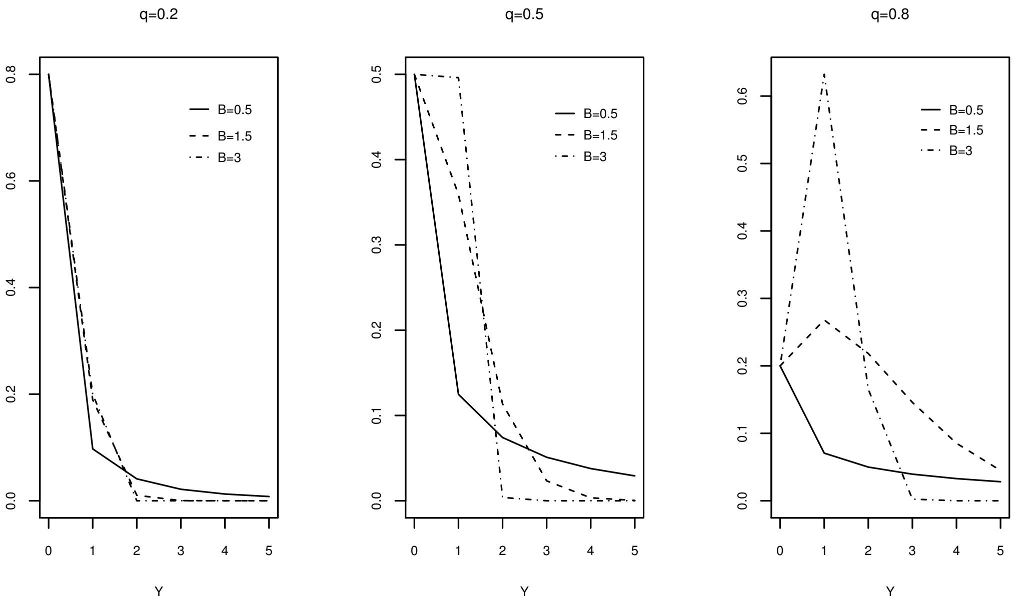
and its pmf by

$$f(y) = \begin{cases} q^{y^{\beta}} - q^{(y+1)^{\beta}} & \text{for } y = 0, 1, 2, 3, \dots \\ 0 & \text{otherwise} \end{cases} \tag{2.2}$$

with the parameters $0 < q < 1$ and $\beta > 0$.

The parameter β is the shape parameter, and it affects the pmf, as shown in Figure 2.1.

Figure 2.1: The effect of β on the pmf of the DW.



Furthermore, β can be considered as controlling the range of values of the DW RV. In other words, this parameter controls the skewness of the DW distribution. To investigate, Figure 2.2 plots the frequency distributions for some samples of DW with a fixed parameter $q = 0.6$ and some different values of β. From this plot, it can be seen that the DW distribution is more skewed: $\beta \longrightarrow 0$. The range of the DW counts would be smaller: $\beta \longrightarrow \infty$.

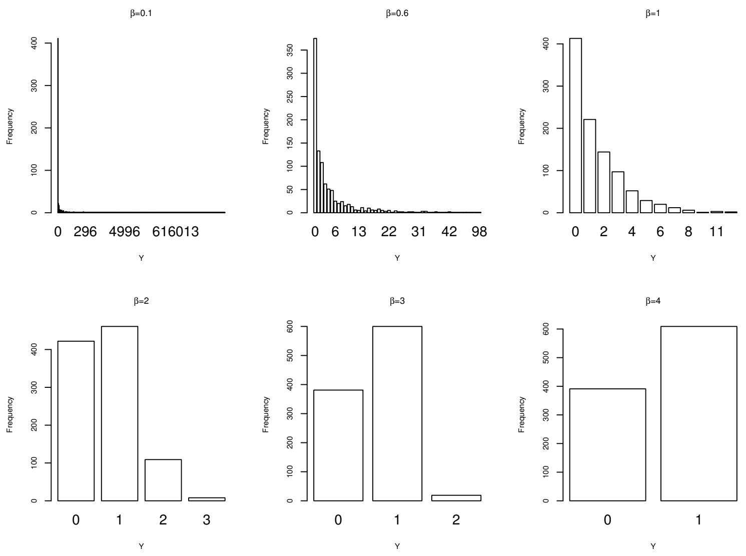


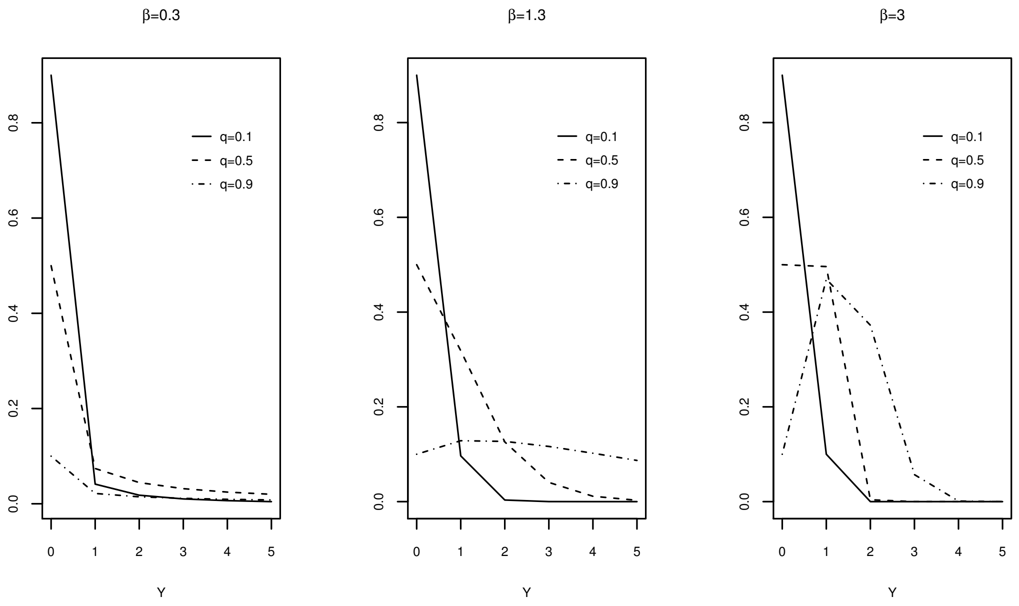
Figure 2.2: The effect of β on the frequency distribution for DW samples with $q = 0.6$.

In addition, q is the probability of Y being greater than zero. To illustrate, from Equation 2.2,

$$p(0) = 1 - q$$

Additionally, the parameter q can be considered as another shape parameter for the DW distribution, and it affects the shape of the pmf, as shown in Figure 2.3.

Figure 2.3: The effect of q on the pmf of the DW.



In comparison to the continuous Weibull distribution whose cdf can be obtained as:

$$F(y) = 1 - e^{-\lambda y^{\beta}} \qquad , \qquad y \geq 0$$

the parameter β from the DW is analogous to the shape parameter β in the continuous Weibull distribution. On the other hand, q from the DW is equivalent to $e^{-\lambda}$, where λ is the scale parameter in the continuous Weibull distribution.

2.3 Properties of the discrete Weibull distribution

2.3.1 Mean and variance

The mean of the DW in Equation 2.2 can be derived as follows:

$$\begin{aligned} E(Y) = \mu &= \sum_{y=0}^{\infty} y f(y) \\ &= \sum_{y=1}^{\infty} q^{y^{\beta}} \end{aligned} \tag{2.3}$$

Additionally, the variance, can be calculated as follows:

$$\sigma^2 = E(Y^2) - (E(Y))^2$$

Then,

$$\begin{aligned} E(Y^2) &= \sum_{y=0}^{\infty} y^2 f(y) \\ &= \sum_{y=1}^{\infty} (2y-1)\, q^{y^{\beta}} \\ &= 2\sum_{y=1}^{\infty} y q^{y^{\beta}} - E(Y) \end{aligned}$$

Therefore, the variance of the DW can be derived as:

$$\sigma^2 = 2\sum_{y=1}^{\infty} yq^{y^\beta} - \mu - \mu^2 \tag{2.4}$$

These can be calculated numerically by the approximated moments of the DW using the truncated support in Barbiero (2015).

2.3.2 Quantile function

The $\tau-$quantile function for any RV is the value $Q(\tau)$ that satisfies:

$$P\left(Y \leq Q(\tau)\right) \geq \tau \qquad \text{and} \qquad P\left(Y \geq Q(\tau)\right) \leq 1-\tau$$

Or, specifically, the quantile can be

$$Q(\tau) = F^{-1}(\tau) = inf\left\{y : F(y) \geq \tau\right\} \tag{2.5}$$

Then, from Equation 2.1, the quantile $Q(\tau)$ for a DW RV can be obtained as follows:

$$1 - q^{(Q(\tau)+1)^\beta} \geq \tau$$

which is equivalent to

$$\left(Q(\tau)+1\right)^\beta \log(q) \leq \log(1-\tau)$$

divided both sides by $\log(q)$, which is a negative value,

$$\left(Q(\tau)+1\right)^\beta \geq \frac{\log(1-\tau)}{\log(q)}$$

Then,

$$Q(\tau) \geq \left(\frac{\log(1-\tau)}{\log(q)}\right)^{\frac{1}{\beta}} - 1$$

Thus, the DW distribution has a nice property of being its τ^{th} $(0 < \tau < 1)$ quantile has a closed form. That is, the smallest value of y for which $F(y) \geq \tau$

has the following expression:

$$Q(\tau) = \left\lceil \left(\frac{\log(1-\tau)}{\log(q)} \right)^{\frac{1}{\beta}} - 1 \right\rceil \tag{2.6}$$

These definitions are derived from defining the quantiles or distribution functions for the continuous RVs, which have been previously discussed extensively in the literature. However, limited work exists in the literature in the area of quantile functions for discrete RVs. This might be due to the non-uniqueness of a specific quantile of a discrete distribution, which can be noted from the above definitions. Consequently, a few studies on this topic limit the quantiles to only integer values. Nevertheless, similar to the mean of a discrete RV, quantiles of discrete distributions could be non-integer values, which will not only satisfy the general definition of quantile function but also make the research more convenient. Hence, the quantile for a discrete RV can be defined as follows:

$$Q(\tau) = F^{-1}(\tau) = \{y : F(y) = \tau\} \tag{2.7}$$

Then, the quantile of DW would not be limited to integers, and a closed form for this quantile could take the following form:

$$Q(\tau) = \left(\frac{\log(1-\tau)}{\log(q)} \right)^{\frac{1}{\beta}} - 1 \tag{2.8}$$

However, since $Y \geq 0$, this definition is valid only for $\tau \geq 1 - q$.

2.4 Special cases of a discrete Weibull distribution

- It can be seen from Equation 2.2 that:

$$f(0) = 1 - q$$

Thus, when q is small, an excessive zero case occurs.

- The discrete Rayleigh distribution in Roy (2004) is a special case of a DW with $\beta = 2$ and $q = \theta$.

- The geometric distribution is a special case of a DW, with $\beta = 1$ and $q = 1 - p$. It can be noted that for the geometric distribution, the variance is always greater than its mean. Therefore, a DW with $\beta = 1$ is a case of over-dispersion relative to Poisson, regardless of the value of q.

- When $\beta = 1$ and $q = e^{-\lambda}$, the distribution is the discrete exponential distribution introduced by Sato et al. (1999).

- As $\beta \to \infty$, the DW approaches a Bernoulli distribution with probability q.

Some of the above special cases can be seen clearly in Figure 2.2.

In the next section, we discuss a property of DW that is particularly advantageous as a model for count data.

2.5 Discrete Weibull accounts for different types of dispersion

The Poisson and NB models are capable for data that are equi- and over-dispersed relative to the Poisson, as mentioned previously. In contrast, an extensive study has been done in this section to show how a DW distribution can handle data that are both over- and under- dispersed relative to Poisson, in different ways. The first method considers the dispersion of the DW model itself, whereas the reminder are for investigating the data generating from DW, as follows:

2.5.1 Using the variance ratio

Figure 2.4 shows the VR values in Equation 1.1 for data simulated by DW with a sequence of values for β and a small value of $q = 0.3$ on the left side and a large value of $q = 0.7$ on the right, and fitted by Poisson and NB distributions. The VRs from Poisson fitting show values greater than one and less than one, indicating cases of over- and under-dispersed data, respectively, relative to the Poisson. That is, the Poisson is over- and under-dispersed model to fit this data. While NB can fit well to data that are over-dispersed relative to the Poisson (i.e. VR close to 1), this does not happen for under-dispersed data where both the Poisson and NB are inappropriate and they consider as under-dispersed model fitting this case of data. Thus, it can be seen how DW, a single distribution with as many parameters as NB, can capture both cases of under- and over- dispersion relative to the Poisson.

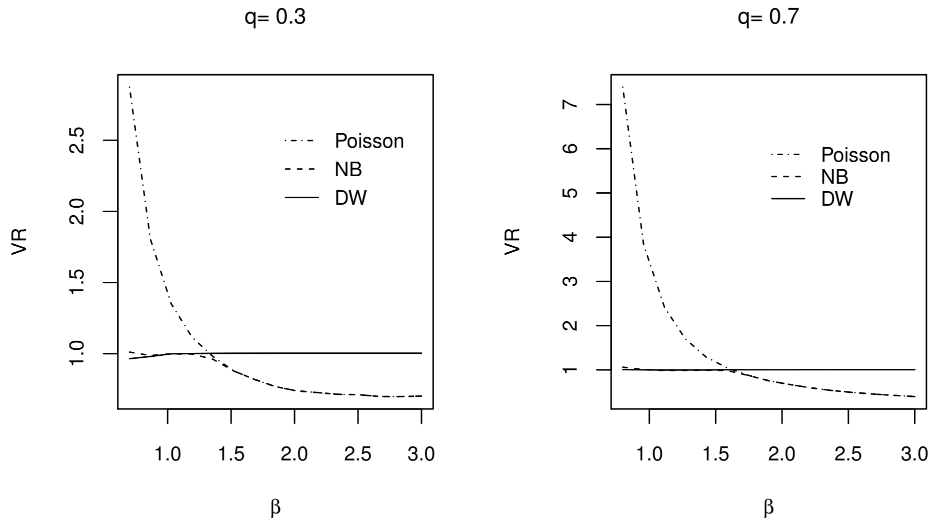


Figure 2.4: Ratio of the observed and theoretical variance of data simulated by DW distribution and fitted by the Poisson, NB and DW distributions.

2.5.2 Using the index of dispersion numerically

Figure 2.5 considers more closely the case of dispersion relative to the Poisson and shows how DW can produce both cases of under- and over- dispersed data relative to the Poisson. In other words, a simulation study is conducted for samples from DW with a sequence of values for β and q. For each sample, the Dip from Equation 1.2, is calculated. This represents the ratio of the observed variance from the data to the observed mean. Figure 2.5 examines the Dips with values less than, equal to or greater than one, implying that DW samples can include under-, equi- and over-dispersed data relative to the Poisson distribution. In other words, the white area corresponds to data under-dispersed relative to the Poisson, whereas the grey shaded area corresponds to over-dispersion.

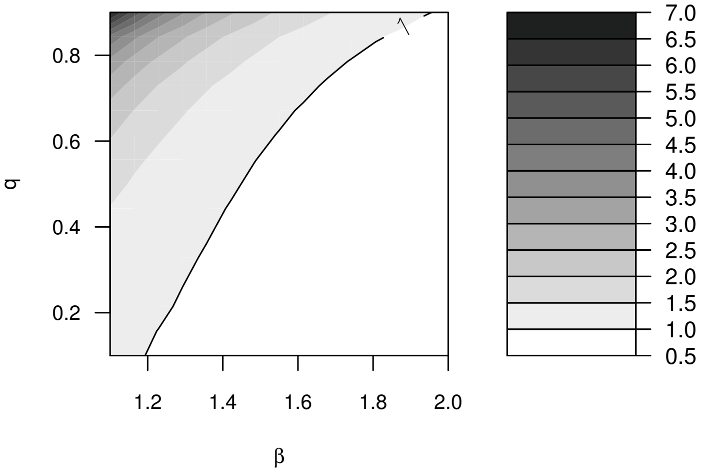


Figure 2.5: Level and contour plots for the ratios of the observed and theoretical variance of data simulated by DW distribution and fitted by the Poisson.

2.5.3 Using the index of dispersion theoretically

The mean and variance of the DW can be computed using the moments approximation by Barbiero (2015). Thus, Figure 2.6 shows the ratio between these means and variances, Dip, for a range of values for parameters q and β. The under- and over-dispersion can be captured, with the Dip < 1 and Dip > 1, respectively.

Figure 2.6 is very similar to Figure 2.5, except the latter is based on a sample descriptive while the former is based on the model quantity.

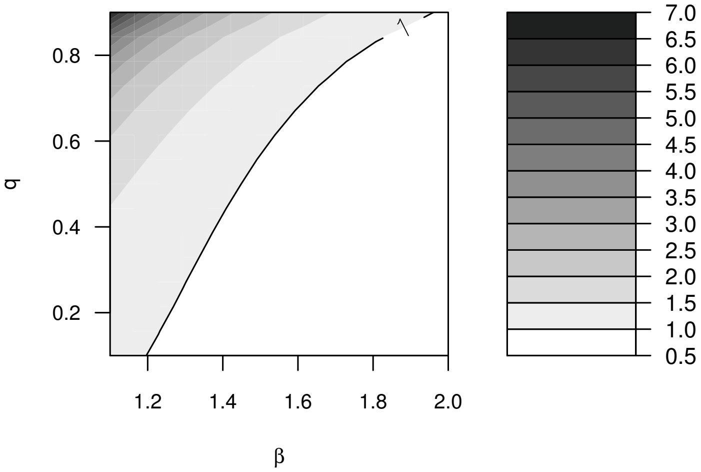


Figure 2.6: Level and contour plots for the Dips for data simulated by DW distribution based on its numerical moments.

2.5.4 Using the index of dispersion for an approximation of the mean and variance

This depends on approximating the means and variances for the DW distribution and was found in Englehardt et al. (2009), Englehardt and Li (2011) and Englehardt et al. (2012). They reported that the approximation for the mean and variance can be obtained by the integral for large $y > L$, as follows:

$$\sum_{y=1}^{\infty} q^{y^\beta} \simeq \sum_{y=1}^{L} q^{y^\beta} + \int_{L+1}^{\infty} q^{y^\beta} dy \tag{2.9}$$

Here we use $L = 1000$, as it has shown a convergent value for the mean in previous studies, see Englehardt et al. (2009), Englehardt and Li (2011) and Englehardt et al. (2012) for details regarding the accuracy for this approximation. Additionally, we can approximate the following:

$$\sum_{y=1}^{\infty} yq^{y^\beta} \simeq \sum_{y=1}^{L} yq^{y^\beta} + \int_{L+1}^{\infty} yq^{y^\beta} dy \tag{2.10}$$

Then, the integrals can be found as follows:

$$\int_{L+1}^{\infty} q^{y^\beta} dy = \int_{L+1}^{\infty} e^{-y^\beta\big(-\log(q)\big)} dy$$

using the transformation, $z = y^\beta\big(-\log(q)\big)$

for $y = L+1$, $z = \big(L+1\big)^\beta\big(-\log(q)\big)$,

for $y = \infty$, $z = \infty$,

$$\int_{L+1}^{\infty} q^{y^\beta} dy = \frac{1}{\beta\big(-\log(q)\big)^{\frac{1}{\beta}}} \int_{\big(L+1\big)^\beta\big(-\log(q)\big)}^{\infty} z^{\frac{1}{\beta}-1} e^{-z} dz$$

Then, the integral can be solved using the incomplete gamma, where:

$$\Gamma(s,x) = \int_{x}^{\infty} t^{s-1} e^{-t} dt$$

Therefore:

$$\int_{L+1}^{\infty} q^{y^\beta} dy = \frac{\Gamma\Big(\frac{1}{\beta}, \big(L+1\big)^\beta\big(-\log(q)\big)\Big)}{\beta\big(-\log(q)\big)^{\frac{1}{\beta}}}$$

Similarly:

$$\begin{aligned}\int_{L+1}^{\infty} y q^{y^\beta} dy &= \frac{1}{\beta\big(-\log(q)\big)^{\frac{2}{\beta}}} \int_{\big(L+1\big)^\beta\big(-\log(q)\big)}^{\infty} z^{\frac{2}{\beta}-1} e^{-z} dz \\ &= \frac{\Gamma\Big(\frac{2}{\beta}, \big(L+1\big)^\beta\big(-\log(q)\big)\Big)}{\beta\big(-\log(q)\big)^{\frac{2}{\beta}}}\end{aligned}$$

Then, from Equation 2.9:

$$\begin{aligned}\mu = E(y) &\simeq \sum_{y=1}^{L} q^{y^\beta} + \frac{\Gamma\Big(\frac{1}{\beta}, \big(L+1\big)^\beta\big(-\log(q)\big)\Big)}{\beta\big(-\log(q)\big)^{\frac{1}{\beta}}} \\ &= \frac{\beta\big(-\log(q)\big)^{\frac{1}{\beta}} \sum_{y=1}^{L} q^{y^\beta} + \Gamma\Big(\frac{1}{\beta}, \big(L+1\big)^\beta\big(-\log(q)\big)\Big)}{\beta\big(-\log(q)\big)^{\frac{1}{\beta}}}\end{aligned}$$

From Equation 2.10:

$$
\begin{aligned}
2\sum_{y=1}^{\infty} yq^{y^\beta} &\simeq 2\Big[\sum_{y=1}^{L} yq^{y^\beta} + \frac{\Gamma\Big(\frac{2}{\beta}, (L+1)^\beta\big(-\log(q)\big)\Big)}{\beta\big(-\log(q)\big)^{\frac{2}{\beta}}}\Big] \\
&= \frac{2\beta\big(-\log(q)\big)^{\frac{2}{\beta}}\sum_{y=1}^{L} yq^{y^\beta} + 2\Gamma\Big(\frac{2}{\beta}, (L+1)^\beta\big(-\log(q)\big)\Big)}{\beta\big(-\log(q)\big)^{\frac{2}{\beta}}}
\end{aligned}
$$

Then, to study the dispersion relatively to the Poisson, the Dip, following Equation 1.2, for the DW can be found as follows:

$$
\begin{aligned}
\frac{\sigma^2}{\mu} &= 2\sum_{y=1}^{\infty} yq^{y^\beta}\frac{1}{\mu} - 1 - \mu \\
&= \frac{\frac{2\beta\big(-\log(q)\big)^{\frac{2}{\beta}}\sum_{y=1}^{L} yq^{y^\beta} + 2\Gamma\Big(\frac{2}{\beta}, (L+1)^\beta\big(-\log(q)\big)\Big)}{\beta\big(-\log(q)\big)^{\frac{2}{\beta}}}}{\frac{\beta\big(-\log(q)\big)^{\frac{1}{\beta}}}{\beta\big(-\log(q)\big)^{\frac{1}{\beta}}\sum_{y=1}^{L} q^{y^\beta} + \Gamma\Big(\frac{1}{\beta}, (L+1)^\beta\big(-\log(q)\big)\Big)}} - 1 \\
&\quad - \frac{\beta\big(-\log(q)\big)^{\frac{1}{\beta}}\sum_{y=1}^{L} q^{y^\beta} + \Gamma\Big(\frac{1}{\beta}, (L+1)^\beta\big(-\log(q)\big)\Big)}{\beta\big(-\log(q)\big)^{\frac{1}{\beta}}} \\
&= \frac{1}{\big(-\log(q)\big)^{\frac{1}{\beta}}}\left[\frac{2\beta\big(-\log(q)\big)^{\frac{2}{\beta}}\sum_{y=1}^{L} yq^{y^\beta} + 2\Gamma\Big(\frac{2}{\beta}, (L+1)^\beta\big(-\log(q)\big)\Big)}{\beta\big(-\log(q)\big)^{\frac{1}{\beta}}\sum_{y=1}^{L} q^{y^\beta} + \Gamma\Big(\frac{1}{\beta}, (L+1)^\beta\big(-\log(q)\big)\Big)}\right. \\
&\quad \left. - \frac{\beta\big(-\log(q)\big)^{\frac{1}{\beta}}\sum_{y=1}^{L} q^{y^\beta} + \Gamma\Big(\frac{1}{\beta}, (L+1)^\beta\big(-\log(q)\big)\Big)}{\beta}\right] - 1
\end{aligned}
$$

Let

$$
\begin{aligned}
A = &\frac{1}{\big(-\log(q)\big)^{\frac{1}{\beta}}}\left[\frac{2\beta\big(-\log(q)\big)^{\frac{2}{\beta}}\sum_{y=1}^{L} yq^{y^\beta} + 2\Gamma\Big(\frac{2}{\beta}, (L+1)^\beta\big(-\log(q)\big)\Big)}{\beta\big(-\log(q)\big)^{\frac{1}{\beta}}\sum_{y=1}^{L} q^{y^\beta} + \Gamma\Big(\frac{1}{\beta}, (L+1)^\beta\big(-\log(q)\big)\Big)}\right. \\
&\left. - \frac{\beta\big(-\log(q)\big)^{\frac{1}{\beta}}\sum_{y=1}^{L} q^{y^\beta} + \Gamma\Big(\frac{1}{\beta}, (L+1)^\beta\big(-\log(q)\big)\Big)}{\beta}\right]
\end{aligned}
$$

Hence,

- Data might be over-dispersed relative to the Poisson for $A > 2$,
- Data might be under-dispersed relative to the Poisson for $A < 2$,
- Data might be equi-dispersed relative to the Poisson for $A = 2$,

To demonstrate, A has been calculated based on a sequence of values for β and q in order to investigate the Dip through this A. Figure 2.7 examines these values of A, which take values of less than, equal to or greater than two, implying that DW can show under-, equi- and over-dispersed data relative to the Poisson distribution.

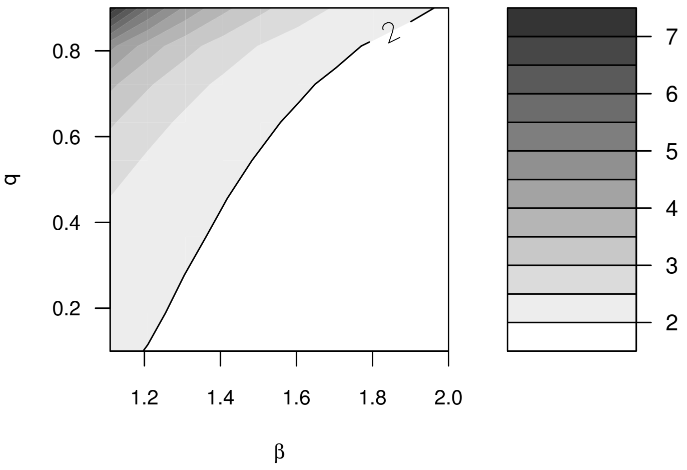


Figure 2.7: Level and contour plots for the Dips for data simulated by DW distribution based on its approximated mean and variance.

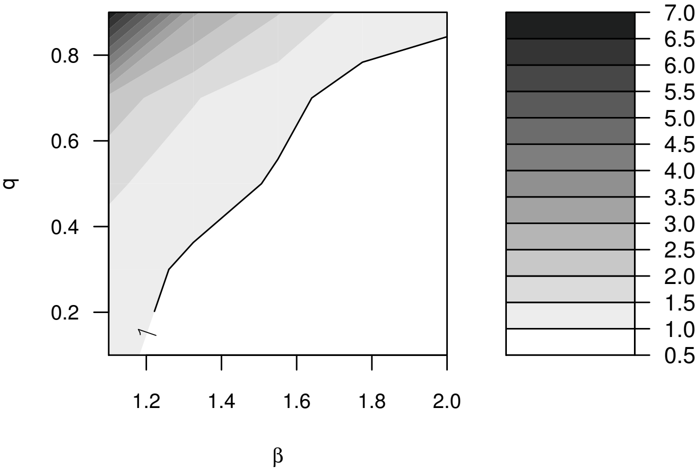
2.5.5 Using a dispersion parameter

There are some models, such as the COM-Poisson and quasi-Poisson which have a dispersion parameter that can define the dispersion for the data, depending on the Dip. In this simulation study, the dispersion parameter of the quasi-Poisson has been considered. This approach is based on the quasi-likelihood, where the variance is adjusted to be smaller or larger than the mean as, $\sigma^2 = \phi\mu$, where ϕ is a scale or dispersion parameter and can be defined as:

$$\phi \begin{cases} > 1 & \text{for over-dispersion} \\ = 1 & \text{for equi-dispersion} \\ < 1 & \text{for under-dispersion} \end{cases}$$

To implement this approach in R Core Team (2014), *glm* can be used with specifying *family=quasipoisson*. Then, using simulated samples from DW with different values for β and q, Figure 2.8 shows the level and contour plot for the quasi-Poisson dispersion parameter with values less than or greater than one. That is, the white area corresponds to under-dispersed data whereas the grey shaded area corresponds to over-dispersion. Thus, these DW data can display under- and over-dispersion.

Figure 2.8: Level and contour plots for the dispersion parameter ϕ of data simulated by DW distribution and fitted by the quasi-Poisson.



In particular, these numerical analyses have approximately shown that:

- $0 < \beta \leq 1$ is a case of over-dispersion, regardless of the value of q.
- $\beta \geq 2$ is a case of under-dispersion, regardless of the value of q. In fact, DW approaches the Bernoulli distribution with mean p and variance $p(1-p)$ for $\beta \to \infty$.

- $1 < \beta < 2$ leads to both cases of over and under-dispersion depending on the value of q.

2.6 Parameter estimation

Given $y_1, y_2, \ldots, y_n$, from a DW distribution in Equation 2.2, the log-likelihood can be written as:

$$\ell = \sum_{i=1}^{n} \log\left(q^{y_i^{\beta}} - q^{(y_i+1)^{\beta}}\right) \tag{2.11}$$

from which the MLEs of q and β can be easily obtained by directly maximizing this log-likelihood using any non-linear optimization tool. Additionally, this can be obtained using the *estdweibull* function with (method = "ML") in the *DiscreteWeibull* package (Barbiero (2015)).

2.7 Model selection

Based on the maximum likelihood approach, numerous studies have suggested the use of information measures, such as the AIC, as good criteria for measuring a models fit and selecting the best fitting model for nested and non-nested models, for example, Posada and Buckley (2004), Dayton (2003) and Kuha (2004) among others.

These criteria can be calculated as follows:

$$AIC = -2\ell + 2P \tag{2.12}$$

where, ℓ is the log-likelihood and P is the number of parameters to be estimated in the model. This computation measures how faraway the fitted model is from the observed data. Then, the better fit is the one with the smaller AIC.

There is another measurements can be used for model selection based on the maximum likelihood approach, such as, Bayesian information criterion (BIC). In this study this BIC has been calculated and its results show similar conclusions to those results from the AIC. However, several studies have used the AIC to compare the model fit for count data, for instance, Chipeta et al. (2014), Aa and Naing (2012), Loeys et al. (2012), Sellers and Shmueli (2010) and Sáez-Castillo and Conde-Sánchez (2013), among others. Then, the results from BIC have not been included in this study and here the results for the AIC only have been considered.

2.8 Numerical examples

In order to investigate the flexibility and adequacy of the DW distribution, it is applied to fit some real count data sets. In addition, theses models are compared to some related models, particularly the Poisson and NB. Thus, after considering these models to fit the data, the AIC is computed and the model with the smallest values of this indicator can be chosen as the best fit for this dataset. Additionally, the expected frequency is compared with the observed frequency via the histogram, as it is the best visual descriptive for count data.

2.8.1 Under-dispersed data relative to the Poisson

In linguistics studies, the number of words in some texts is commonly considered as being under-dispersed data relative to the Poisson. The following example is for two data sets count the number of articles ("the", "a" and "an") in groups of words in literary essays (Bailey (1990)) and available in data number "486" in Hand et al. (1993), also in `http://www.statsci.org/data/oz/wdcount.html`. The first dataset is for five-word samples with $Dip = 0.5926$, indicating to under-dispersion case, while the second under-dispersed dataset is for ten-word samples with $Dip = 0.6229$.

Table 2.1: AIC from the Poisson, NB and DW distributions fitted to the word count data sets.

Data	Poisson	NB	DW
Articles count in 5-word samples	192.6219	194.6232	179.7142
Articles count in 10-word samples	248.5483	250.5496	239.6654

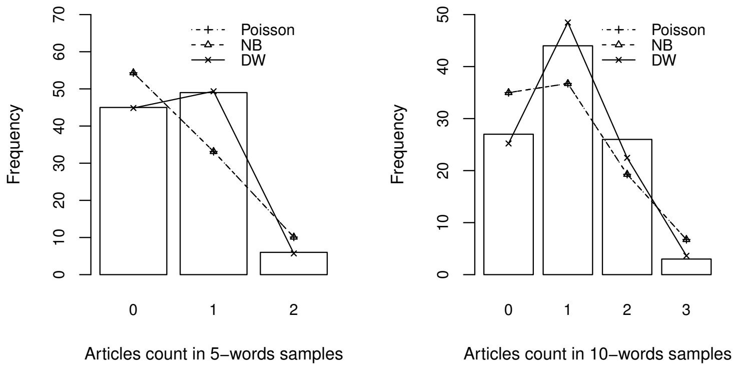


Figure 2.9: Observed and expected frequencies for the word count data sets fitted by the Poisson, NB and DW distributions.

2.8.2 Over-dispersed data relative to the Poisson

The number of visits count variable can be found in many fields. For example, in tourism studies, researchers seek to analyze the number of visitors to a country in a given year. In addition, in health studies, the number of patients who visit a hospital is studied as a health measurement (e.g. health demand).

Data representing the number of visits have been considered here. This data, from Hosmer Jr and Lemeshow (2004) and available under the “COUNT” package in R, includes the number of visits to a doctor by pregnant women in the first three months of their pregnancies, ranging from 0 to 6, with $Dip = 1.4138$, indicating to an over-dispersion case. This variable is modeled by the MLE for the Poisson, NB and DW.

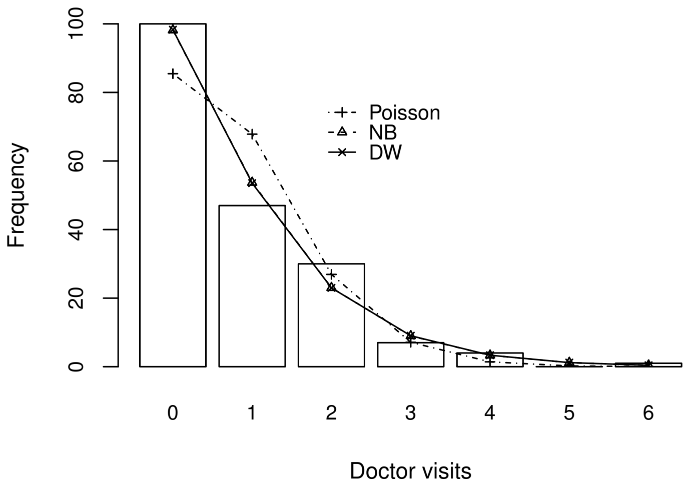


Figure 2.10: Observed and expected frequencies for the doctor visits by pregnant women dataset fitted by the Poisson, NB and DW distribution.

Table 2.2: AIC from the Poisson, NB and DW distributions fitted to the doctor visits by pregnant women dataset.

Poisson	NB	DW
476.5899	466.8534	466.8447

2.8.3 Excessive zero counts

Besides the flexibility of the DW to handle under- and over-dispersed data relative to the Poisson, it can also be considered as a good model for skewed data with an excessive number of observations with zero counts. This is because β controls the range of Y and q can be defined as the probability of 0; then, for $\beta \longrightarrow 0$ and $q \longrightarrow 0$, skewed data and too many zeros might be obtained from this DW distribution, as mentioned previously in section 2.4 and Figure 2.2. The data with too many zeros can be described as being more dispersed than the Poison and cannot be fitted properly by it. Hence, a modified from Poisson model has been developed to cope with the problem of over-dispersion due to the many zeros, that is, ZIP. In this adjusted model, the probability of zeros is considered separately of the other positive counts. However, the DW model, alternatively with NB, demonstrates a better fit for such data than this developed model, which is more complicated, harder to interpret and contains more parameters that need to be estimated than in the conventional NB and DW.

If there are data with too many zero counts generated by the same system, then a single distribution model should be applied. Then, it might be sensible to consider one part model, NB or DW for these cases, when there is no need to apply the modified models.

Generally count data with small means can be fitted well using the basic Poisson model, and for count data with more over-dispersion, that is, small means but large variances, then the DW or NB can provide a good fit. Therefore, these simple and one-part models with fewer parameters should be considered before attempting to use the more complex zero-inflated models (Englehardt and Li (2011), Allison (2012) and Xie et al. (2013)).

For example, Englehardt and Li (2011) and Englehardt et al. (2012) suggested the DW for excessive zeros and highly skewed data in pathogen counts of treated water over time. Additionally, the following example demonstrates the ability of DW to handle this type of data. This dataset includes the number of doctor visits for 5190 patients from the Australian Health Survey in $1977 - 1988$, as analyzed by Cameron and Trivedi (1986) and cited in Cameron and Trivedi (2013). Additionally, the *Dip* of this RV is 2.1112 indicating to an over-dispersion case.

Table 2.3: AIC from the Poisson, NB and DW distributions fitted to the doctor visits from the Australian health dataset.

Poisson	NB	DW	ZIP
7968.389	7175.983	7164.983	7432.471

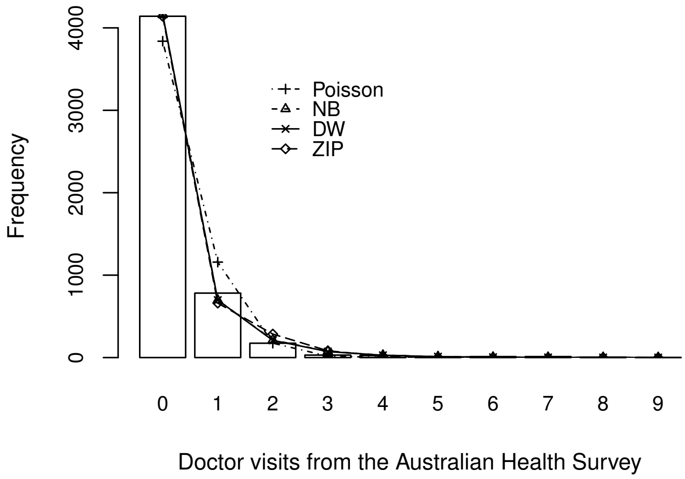


Figure 2.11: Observed and expected frequencies for the doctor visits from the Australian health dataset fitted by the Poisson, NB, DW and ZIP distribution.

As can be seen from Table 2.1, Table 2.2 and Table 2.3, the lowest AIC is for the DW and the expected frequencies from DW are close to the observed frequencies. Therefore, DW can be considered as the best model to fit these data among the Poisson and NB.

2.9 Concluding remarks

This chapter discussed some of the properties of the DW, focusing on the dispersion characteristics. Using a variety of methods, the capability of this distribution to handle different types of data dispersion was investigated. Thus, in contrast to the Poisson and NB models that can respectively capture equi- and over-dispersed data relative to Poisson, DW has the ability to handle over- and under-dispersed data relative to the Poisson. Additionally, it works well alternatively with NB for modeling excessive zero counts, instead of using the ZIP, which is more complicated. To illustrate, some real data examples representing different type of dispersion have been applied, and the results suggested that the DW can be considered as the best model to fit these data among the Poisson and NB.

Chapter 3

Discrete Weibull Regression Model

3.1 Introduction

On the one hand, estimations and inferences for the unknown parameters for the DW distribution have been discussed in previous studies, as mentioned earlier. On the other hand, researchers often seek to investigate the effect of other variables on the frequency of events and explore these counts as a function of covariates, that is, to consider a regression analysis for count data. The response variable in this analysis is considered to be discrete with a pmf for non-negative integer RVs. Hence, this study suggests DW for this kind of analysis.

3.2 Discrete Weibull regression model

As discussed in chapter 2, the parameter q affects the shape of the pmf of the DW, as shown in Figure 2.3. This study introduces a count regression model for the discrete response RV, Y based on the DW distribution, by relating this parameter q to some covariates. Some different points with regard to this relation, which develops the regression for $Y_i|X_i$, for $i = 1, 2, \ldots, n$, can be considered:

3.2.1 Introducing the regression in relation to the continuous Weibull model

It has been stated that q is equivalent to $e^{-\lambda}$, where λ is the scale parameter in the continuous Weibull distribution. Then, to develop a regression model for the continuous Weibull regression, it is often assumed that this parameter λ is related to predictors (Lee and Wang (2003) and Da Silva et al. (2008)) as follows:

$$\log(\lambda_i) = \boldsymbol{x}_i'\boldsymbol{\alpha}$$

Analogously with this continuous Weibull regression, the covariates can be incorporated for the DW regression model. Assume that the response variable Y, has a DW conditional distribution $f(y_i, q(\boldsymbol{x}_i), \beta)$, where $q(\boldsymbol{x}_i)$ is related to the explanatory variables $\boldsymbol{X}$ via the link function:

$$\log\left(-\log(q_i)\right) = \boldsymbol{x}_i'\boldsymbol{\alpha} \qquad , \qquad \boldsymbol{x}_i'\boldsymbol{\alpha} = \alpha_0 + x_{i1}\alpha_1 + \ldots + x_{iP}\alpha_P \tag{3.1}$$

This transforms q from the probability scale (i.e. the interval $[0,1]$) to the interval $[-\infty, +\infty]$ and ensures that the parameter q remains in $[0,1]$. Indeed, from Equation 3.1, q_i can be expressed as:

$$q_i \equiv q(\boldsymbol{x}_i) = e^{-e^{\boldsymbol{x}_i'\boldsymbol{\alpha}}} \tag{3.2}$$

from which the pmf of y_i conditional on $\boldsymbol{x_i}$ and β can be obtained as follows:

$$f(y_i|x_i) = \left(e^{-e^{\boldsymbol{x}_i'\boldsymbol{\alpha}}}\right)^{y_i^\beta} - \left(e^{-e^{\boldsymbol{x}_i'\boldsymbol{\alpha}}}\right)^{(y_i+1)^\beta} \tag{3.3}$$

3.2.2 Introducing the regression in relation to the geometric model

The geometric regression can be defined as:

$$f(y_i|x_i) = p_i\left(1 - p_i\right)^{y_i} \tag{3.4}$$

with the expected value:

$$\mu_i = \frac{1-p_i}{p_i}$$

Then, analogously with the GLMs, the regression structure for the geometric regression can be built as:

$$\frac{1-p_i}{p_i} = e^{\boldsymbol{x}_i'\boldsymbol{\alpha}}$$

in which the parameter p_i can be defined as:

$$p_i = \left(1 + e^{\boldsymbol{x}_i'\boldsymbol{\alpha}}\right)^{-1} \tag{3.5}$$

Note that the geometric model is a special case from the NB model, with $k = 1$ in Equation 1.5; see, for example, Zeileis et al. (2008).

As a result of being the geometric model is a special case of the DW with $q = 1-p$, as mentioned previously then, the DW regression might be introduced analogously with the geometric regression in Equation 3.4, by assuming q is a function of covariates with the *logit* link function:

$$q_i = \left(1 + e^{-\boldsymbol{x}_i'\boldsymbol{\alpha}}\right)^{-1} \tag{3.6}$$

Thus, the conditional pmf of the response variable Y_i given X_i can be derived as:

$$f(y_i|x_i) = \left(1 + e^{-\boldsymbol{x}_i'\boldsymbol{\alpha}}\right)^{-y_i^{\beta}} - \left(1 + e^{-\boldsymbol{x}_i'\boldsymbol{\alpha}}\right)^{-(y_i+1)^{\beta}} \tag{3.7}$$

Alternatively, one can choose different transformations to link q with a set of covariates, for example, *probit*. This study has utilized the *logit* link function, and it provides similar results to the link function in Equation 3.2. Then, the results reported in this thesis are based on the link function in Equation 3.1.

Commonly, the type of dispersion should be taken into account when the covariates $\boldsymbol{X}$ are considered. Thus, the different types of dispersion are investigated for the regression structure for the DW model.

3.3 Maximum likelihood estimation

Estimation of the unknown parameters is performed using the maximum likelihood approach. Thus, given a sample of n independent observations, (x_i, y_i), $i = 1, 2, \ldots, n$ from Equation 3.3, the likelihood function based on this observed sample, is given by:

$$L = f(y_1, y_2, \ldots, y_n|\boldsymbol{x}) = \prod_{i=1}^{n} \left[\left(e^{-e^{\boldsymbol{x}_i'\boldsymbol{\alpha}}}\right)^{y_i^\beta} - \left(e^{-e^{\boldsymbol{x}_i'\boldsymbol{\alpha}}}\right)^{(y_i+1)^\beta} \right]$$

Then, the corresponding log-likelihood function can be obtained as:

$$\ell = \sum_{i=1}^{n} \log \left[w_i(\beta, \boldsymbol{\alpha}) \right] \tag{3.8}$$

where

$$w_i(\beta, \boldsymbol{\alpha}) = \left(e^{-e^{\boldsymbol{x}_i'\boldsymbol{\alpha}}}\right)^{y_i^\beta} - \left(e^{-e^{\boldsymbol{x}_i'\boldsymbol{\alpha}}}\right)^{(y_i+1)^\beta}$$

Then, to obtain the MLEs, the first partial derivatives with respect to each unknown parameter in $\underline{\theta}$ are obtained and set equal to zero. The first partial derivative of ℓ with respect to parameter β is obtained as follows:

$$\frac{\partial \ell}{\partial \beta} = -\sum_{i=1}^{n} \frac{e^{\boldsymbol{x}_i'\boldsymbol{\alpha}}}{w_i(\beta, \boldsymbol{\alpha})} \Bigg[\left(e^{-e^{\boldsymbol{x}_i'\boldsymbol{\alpha}}}\right)^{y_i^\beta} y_i^\beta \log(y_i) - \\ \left(e^{-e^{\boldsymbol{x}_i'\boldsymbol{\alpha}}}\right)^{(y_i+1)^\beta} (y_i+1)^\beta \log(y_i+1) \Bigg]$$

The first partial derivative of ℓ with respect to parameter α_p, $(p = 1, \ldots, P)$ is obtained as follows:

$$\frac{\partial \ell}{\partial \alpha_p} = -\sum_{i=1}^{n} \frac{x_{ip} e^{x_{ip}\alpha_p - e^{x_{ip}\alpha_p}}}{w_i(\beta, \alpha_p)} \left[y_i^\beta \left(e^{-e^{x_{ip}\alpha_p}}\right)^{y_i^\beta - 1} - (y_i+1)^\beta \left(e^{-e^{x_{ip}\alpha_p}}\right)^{(y_i+1)^\beta - 1} \right]$$

As can be seen, the system of these likelihood equations cannot be solved simultaneously, and hence they do not have an analytic solution. Thus, the common problem of the maximum likelihood approach, where no closed form solution exists, has been experienced here. Therefore, the MLEs of $\boldsymbol{\alpha}$ and β can be found by directly maximizing the log-likelihood function in Equation 3.8 numerically. This can be performed easily using any iterative numerical optimization tool, such as the *optim* function in R.

Given that the parameter inferences are performed using the maximum likelihood method, then under some regularity conditions (Serfling (1980) or Greene (2003)) these estimators enjoy standard asymptotic properties. In other words, the MLEs; $\hat{\beta}_{ML}$ and $\hat{\boldsymbol{\alpha}}_{ML}$ have certain characteristics, as follows:

- They are asymptoticly consistent (unbiased)
- They asymptotically have an variance-covariance matrix obtained from the inverse of the expected Fisher information matrix:

$$I_{exp} = \begin{pmatrix} I_{\alpha\alpha} & I_{\alpha\beta} \\ I_{\beta\alpha} & I_{\beta\beta} \end{pmatrix}$$

  where

$$I_{\theta_i\theta_j} = -E\left(\frac{\partial^2 \ell}{\partial\theta_i\theta_j}\right) \qquad i,j = 1,2 \tag{3.9}$$

- The MLEs $\hat{\beta}_{ML}$,$\hat{\boldsymbol{\alpha}}_{ML}$ are asymptotically normal distributed:

$$\sqrt{n}\begin{pmatrix} \hat{\beta}_{ML} - \beta \\ \hat{\boldsymbol{\alpha}}_{ML} - \boldsymbol{\alpha} \end{pmatrix} \sim N\left(\begin{pmatrix} 0 \\ 0 \end{pmatrix} , \quad I^{-1}\right) \tag{3.10}$$

where β and $\boldsymbol{\alpha}$ are the true values of β and $\boldsymbol{\alpha}$, respectively, and I is the Fisher information matrix.

Hence, this expected Fisher information matrix is obtained in the following steps. Note that, for a sample of n independent observations, the log-likelihood sample in Equation 3.8 can be re-written as follows:

$$\ell(\boldsymbol{\alpha}, \beta) = \sum_{i=1}^{n} \ell(q_i, \beta) \tag{3.11}$$

where

$$\ell(q_i,\beta) = \log\left[q_i^{y_i{}^\beta} - q_i^{(y_i+1)^\beta}\right] \qquad , \qquad q_i \equiv q(\boldsymbol{x}_i)$$

Then, for $p = 1, 2, \ldots, P$, the score functions $\dfrac{\partial \ell(\boldsymbol{\alpha},\beta)}{\partial \alpha_p}$ and $\dfrac{\partial \ell(\boldsymbol{\alpha},\beta)}{\partial \beta}$ can be re-obtained as follows:

$$\begin{aligned}\frac{\partial \ell(\boldsymbol{\alpha},\beta)}{\partial \alpha_p} &= \frac{\partial}{\partial \alpha_p}\left[\sum_{i=1}^{n} \ell(q_i,\beta)\right] \\ &= \sum_{i=1}^{n} \frac{\partial \ell(q_i,\beta)}{\partial q_i}\frac{\partial q_i}{\partial \eta_i}\frac{\partial \eta_i}{\partial \alpha_p}\end{aligned} \tag{3.12}$$

where

$$\frac{\partial \ell(q_i,\beta)}{\partial q_i} = \frac{y_i^\beta q_i^{y_i^\beta - 1} - (y_i+1)^\beta q_i^{(y_i+1)^\beta - 1}}{q_i^{y_i^\beta} - q_i^{(y_i+1)^\beta}} \tag{3.13}$$

and

$$\eta = \boldsymbol{x}'\boldsymbol{\alpha} \qquad \Rightarrow \qquad \frac{\partial q_i}{\partial \eta_i} = -e^{\eta_i}e^{-e^{\eta_i}}, \qquad \frac{\partial \eta_i}{\partial \alpha_p} = x_{ip}$$

Hence, back to Equation 3.12, the score function for $\boldsymbol{\alpha}$ is given by

$$\frac{\partial \ell(\boldsymbol{\alpha},\beta)}{\partial \alpha_p} = \frac{y_i^\beta q_i^{y_i^\beta - 1} - (y_i+1)^\beta q_i^{(y_i+1)^\beta - 1}}{q_i^{y_i^\beta} - q_i^{(y_i+1)^\beta}}\frac{\partial q_i}{\partial \eta_i}x_{ip} \tag{3.14}$$

and similarly, the score function for β is obtained as follows:

$$\begin{aligned}\frac{\partial \ell(\boldsymbol{\alpha},\beta)}{\partial \beta} &= \sum_{i=1}^{n} \frac{\partial \ell(q_i,\beta)}{\partial \beta} \\ &= \sum_{i=1}^{n} \log(q_i)\left[\frac{q_i^{y_i^\beta} y_i^\beta \log(y_i) - q_i^{(y_i+1)^\beta}(y_i+1)^\beta \log(y_i+1)}{q_i^{y_i^\beta} - q_i^{(y_i+1)^\beta}}\right]\end{aligned} \tag{3.15}$$

After that, the elements for the Fisher information matrix are obtained.

From Equation 3.12, the second derivative of $\ell(\boldsymbol{\alpha},\beta)$ with respect to αs, for $l = 1, 2, \ldots, P$, is given by:

$$\begin{aligned}\frac{\partial^2 \ell(\boldsymbol{\alpha},\beta)}{\partial\alpha_p\partial\alpha_l} =& \frac{\partial}{\partial\alpha_l}\left[\sum_{i=1}^{n}\frac{\partial \ell(q_i,\beta)}{\partial q_i}\frac{\partial q_i}{\partial\eta_i}\frac{\partial\eta_i}{\partial\alpha_p}\right]\\ & \sum_{i=1}^{n}\frac{\partial}{\partial q_i}\left[\frac{\partial \ell(q_i,\beta)}{\partial q_i}\frac{\partial q_i}{\partial\eta_i}\right]\frac{\partial q_i}{\partial\eta_i}x_{ip}\frac{\partial\eta_i}{\partial\alpha_l}\\ &= \sum_{i=1}^{n}\left[\frac{\partial^2 \ell(q_i,\beta)}{\partial q_i^2}\frac{\partial q_i}{\partial\eta_i}+\frac{\partial}{\partial q_i}\frac{\partial q_i}{\partial\eta_i}\frac{\partial \ell(q_i,\beta)}{\partial q_i}\right]\frac{\partial q_i}{\partial\eta_i}x_{ip}x_{il}\end{aligned}$$

Next, we will take the expectation and note that, under regularity condition, $E\left(\frac{\partial \ell(q_i,\beta)}{\partial q_i}\right) = 0$, then:

$$\begin{aligned}E\left(\frac{\partial^2 \ell(\boldsymbol{\alpha},\beta)}{\partial\alpha_p\partial\alpha_l}\right) &= \sum_{i=1}^{n}\left[E\left(\frac{\partial^2 \ell(q_i,\beta)}{\partial q_i^2}\right)\frac{\partial q_i}{\partial\eta_i}+0\right]\frac{\partial q_i}{\partial\eta_i}x_{ip}x_{il}\\ &= \sum_{i=1}^{n}E\left(\frac{\partial^2 \ell(q_i,\beta)}{\partial q_i^2}\right)\left(\frac{\partial q_i}{\partial\eta_i}\right)^2 x_{ip}x_{il}\end{aligned}$$

where $\frac{\partial^2 \ell(q_i,\beta)}{\partial q_i^2}$ is given in the Appendix. Therefore, for the Fisher information matrix, we have the following:

$$I_{\alpha\alpha} = -E\left(\frac{\partial^2 \ell(\boldsymbol{\alpha},\beta)}{\partial\alpha_p\partial\alpha_l}\right) = -X^T DX \tag{3.16}$$

where $D = diag(d_1, d_2, \ldots, d_n)$ with $d_i = \frac{\partial^2 \ell(q_i,\beta)}{\partial q_i^2}\left(\frac{\partial q_i}{\partial\eta_i}\right)^2$

Similarly, from Equation 3.12, the second derivative of $\ell(\boldsymbol{\alpha},\beta)$ with respect to α_p and β is:

$$\begin{aligned}\frac{\partial^2 \ell(\boldsymbol{\alpha},\beta)}{\partial\alpha_p\partial\beta} =& \frac{\partial}{\partial\beta}\left[\frac{\partial \ell(\boldsymbol{\alpha},\beta)}{\partial\alpha_p}\right]\\ &= \sum_{i=1}^{n}\frac{\partial}{\partial\beta}\left[\frac{\partial \ell(q_i,\beta)}{\partial q_i}\right]\frac{\partial q_i}{\partial\eta_i}\frac{\partial\eta_i}{\partial\alpha_p}\\ & \sum_{i=1}^{n}\frac{\partial^2 \ell(q_i,\beta)}{\partial q_i\partial\beta}\frac{\partial q_i}{\partial\eta_i}x_{ip}\end{aligned} \tag{3.17}$$

The expectation can be taken as follows:

$$E\left(\frac{\partial^2\ell(\boldsymbol{\alpha},\beta)}{\partial\alpha_p\partial\beta}\right)=\sum_{i=1}^{n}E\left(\frac{\partial^2\ell(q_i,\beta)}{\partial q_i\partial\beta}\right)\frac{\partial q_i}{\partial\eta_i}x_{ip}$$

where $\dfrac{\partial^2\ell(q_i,\beta)}{\partial q_i\partial\beta}$ is provided in the Appendix. Then, for the Fisher information matrix

$$I_{\alpha\beta}=-E\left(\frac{\partial^2\ell(\boldsymbol{\alpha},\beta)}{\partial\alpha_p\partial\beta}\right)=-X^TS \tag{3.18}$$

where $S=diag(s_1,s_2,\ldots,s_n)$ with $s_i=\dfrac{\partial^2\ell(q_i,\beta)}{\partial q_i\partial\beta}\dfrac{\partial q_i}{\partial\eta_i}$

Finally, $\dfrac{\partial^2\ell(\boldsymbol{\alpha},\beta)}{\partial\beta^2}$ can be derived by differentiating Equation 3.15 with respect to β. Then, we have the following:

$$\frac{\partial^2\ell(\boldsymbol{\alpha},\beta)}{\partial\beta^2}=\sum_{i=1}^{n}\frac{\partial}{\partial\beta}\left[\frac{\partial\ell(\boldsymbol{\alpha},\beta)}{\partial\beta}\right]=\sum_{i=1}^{n}\frac{\partial^2\ell(\boldsymbol{\alpha},\beta)}{\partial\beta^2}$$

and after taking the expectation, we get

$$E\left(\frac{\partial^2\ell(\boldsymbol{\alpha},\beta)}{\partial\beta^2}\right)=\sum_{i=1}^{n}E\left(\frac{\partial^2\ell(\boldsymbol{\alpha},\beta)}{\partial\beta^2}\right)$$

where $\dfrac{\partial^2\ell(\boldsymbol{\alpha},\beta)}{\partial\beta^2}$ is given in the Appendix. Thus, for the Fisher information matrix

$$I_{\beta\beta}=-E\left(\frac{\partial^2\ell(\boldsymbol{\alpha},\beta)}{\partial\beta^2}\right)=-Tr(V) \tag{3.19}$$

where $V=diag(v_1,v_2,\ldots,v_n)$ with $v_i=\dfrac{\partial^2\ell(\boldsymbol{\alpha},\beta)}{\partial\beta^2}$.
Consequently, the Fisher information matrix is structured by combining the elements in Equation 3.16, Equation 3.18 and Equation 3.19 with $I_{\alpha\beta}=I_{\beta\alpha}^T$.

For many cases calculating these expectations in Equation 3.9 is impractical. Hence, the expected Fisher information matrix might be replaced by the observed Fisher information matrix, which is composed of the negative second partial derivatives of the log-likelihood function $\ell(\underline{\theta};t)$ evaluated at $\theta=\hat{\theta}_{ML}$. That is, each element in this matrix would be as follows:

$$I_{\theta_i\theta_j} = -\frac{\partial^2 \ell}{\partial\theta_i\theta_j} \quad i,j=1,2 \tag{3.20}$$

Then, the asymptotic confidence intervals (CIs) based on the asymptotic normal distribution of MLEs, as mentioned earlier, can be obtained by inverting the Fisher information matrix, to obtain the asymptotic variance-covariance matrix, given by:

$$\begin{aligned}\sum = I^{-1} &= \begin{pmatrix} -\frac{\partial^2\ell}{\partial\beta^2} & -\frac{\partial^2\ell}{\partial\beta\partial\boldsymbol{\alpha}} \\ -\frac{\partial^2\ell}{\partial\beta\partial\boldsymbol{\alpha}} & -\frac{\partial^2\ell}{\partial\boldsymbol{\alpha}^2} \end{pmatrix}^{-1} \\ &= \begin{pmatrix} AVar(\hat{\beta}_{ML}) & ACov(\hat{\beta}_{ML}\hat{\boldsymbol{\alpha}}_{ML}) \\ ACov(\hat{\beta}_{ML}\hat{\boldsymbol{\alpha}}_{ML}) & AVar(\hat{\boldsymbol{\alpha}}_{ML}) \end{pmatrix}\end{aligned}$$

This matrix can be easily obtained by inverting the *Hessian* matrix from the *optim* function in R. The *Hessian* is the second derivative of the objective function (log-likelihood), that is, the *Hessian* matrix is the observed Fisher information matrix. Then, the two-sided approximate CI for the parameters can be conducted as

$$\left[\hat{\beta_{ML}} \pm Z_{\frac{\tau}{2}}\sqrt{AVar(\hat{\beta}_{ML})}\right] \qquad , \qquad \left[\hat{\boldsymbol{\alpha}}_{ML} \pm Z_{\frac{\tau}{2}}\sqrt{AVar(\hat{\boldsymbol{\alpha}}_{ML})}\right] \tag{3.21}$$

3.4 Fitted values

After a DW regression model has been applied, the following can be obtained:

- The fitted values for the central trend of the conditional distribution can be obtained using one of two methods:
  - Fitted mean: Equation 2.3, as mentioned earlier, can be calculated numerically using the approximated moments of the DW (Barbiero, 2015).
  - Fitted median: commonly, count data are skewed and include some outliers, and hence the median is more appropriate than the mean for

these situations. Then, the quantile formula provided in Equation 2.6 can be applied and the fitted conditional median can be obtained easily from the closed form expression of quantiles for the DW, as follows:

$$M = \left\lceil \left(-\frac{\log(2)}{\log(\hat{q}(\boldsymbol{x}))} \right)^{\frac{1}{\hat{\beta}}} - 1 \right\rceil \tag{3.22}$$

- The conditional quantile for any τ can be obtained from Equation 2.6.

3.5 Coefficient interpretation

In the regression analysis for count data, it is common to consider the effect of the covariates on the mean of the response variable, such as the Poisson and NB models. In this study, however, there is no closed form for the expected value of the response, and hence the effect of the covariates can be examined in two ways:

- Investigate the effect of these predictors on the parameter q based on the applied link function in Equation 3.2 or Equation 3.6. Then, examine the relationship between the fitted values and the set of covariates. Therefore, in comparison to the related regression models from the Poisson and NB, the regression coefficients for the DW model are on a different scale for this case, and hence they are not directly comparable. Regardless this difference, it might be expected to obtain regression coefficients from DW with opposite signs compared to those from GLMs.

- The relationship between certain explanatory variables and the median can also be obtained. As mentioned before, the median is more appropriate for count data. Additionally, from Equation 2.8, the median has a closed form for the DW model. To illustrate, as previously stated in Equation 2.5, the median for discrete distributions can be defined as any value of y that satisfies $F(y) \geq 0.5$. Then, although this median is un-uniquely defined, a special case for this median's definition will be considered here to be

$$\left\{y : F(y) = \frac{1}{2}\right\} \quad \subseteq \quad \left\{y : F(y) \geq \frac{1}{2}\right\} \tag{3.23}$$

Then, the median would not be limited to being an integer and can be defined to have a closed form:

$$M = \left(-\frac{\log(2)}{\log(\hat{q}(x))}\right)^{\frac{1}{\hat{\beta}}} - 1$$

However, this form is valid only for $q \geq \frac{1}{2}$ and $M = 0$ can be considered $\forall\, q < \frac{1}{2}$. Thus, $M + 1$ might be examined for this case:

$$M + 1 = \left(-\frac{\log(2)}{\log(\hat{q}(x))}\right)^{\frac{1}{\hat{\beta}}} \tag{3.24}$$

Then, by substituting Equation 3.2 in Equation 3.24 and taking the log, we have the following:

$$\log\left(M+1\right) = \frac{1}{\beta}\log\big(\log(2)\big) - \frac{1}{\beta}\boldsymbol{x}'\boldsymbol{\alpha}. \tag{3.25}$$

Thus, the regression parameter $\boldsymbol{\alpha}$ can be interpreted in relation to the log of the median. This is analogous with the Poisson and NB models, where the parameters are linked to the mean as follows:

$$\log(\mu) = \boldsymbol{x}'\boldsymbol{\alpha} \tag{3.26}$$

In particular, the part $\dfrac{\log\left(\log(2)\right) - \alpha_0}{\beta}$ can be relatively equivalent to the intercept part for the Poisson and NB. Additionally, it is related to the conditional median when all covariates are set to zero, while the parts $\frac{-\alpha_p}{\beta}$, $p = 1, \ldots, P$, are comparable with the log of the change in Y relative to a one unit change in X. In other words, these parts can be related to the change in the median of the response corresponding to a one unit change of $\boldsymbol{x}_p$, keeping all other covariates constant.

In this work, the first approach is consider to investigate the regression coefficients related directly to the parameter q through the link function in Equation 3.2.

3.6 Measurements for model checking

Throughout this study, different regression models were fitted and compared. Then, to check the adequacy of the models, some measurements have been considered.

3.6.1 Model selection

For the model selection, the AIC in Equation 2.12 is calculated. The model with the minimum AIC is the best.

3.6.2 Mean-to-Variance plot

This plot investigates the relation between mean and variance with regards to the covariates' effect. It considers the relation between the observed mean and observed variance, in addition to the theoretical mean and theoretical variance based on some models. Some note is required here on the calculation of the observed mean and observed variance since they cannot be computed for each individual covariate vector $\boldsymbol{X}$. To illustrate, these calculations divide the data into groups based on the percentiles of the linear predictors $\eta_i = \boldsymbol{X}'\boldsymbol{\alpha}$. Although using the linear predictor of any fitting, such as NB or DW, would provide approximately the same groups, the groups here are created based on the DW linear predictor. Thus, η_i from DW is split into 10 groups of similar size. Then, the observed mean and the observed variance are obtained for each group. Subsequently, the theoretical mean and variance based on each model are also computed for the same groups. Accordingly, the relation between Mean-to-Variance ratios is plotted. The points with greater observed variance than observed mean are considered as over-dispersed relative to the Poisson case, while the points with observed variance less than observed mean are considered to be under-dispersed relative to the Poisson case.

Then, to asses which model might be more appropriate, their variance functions are plotted to check which is closer to the observed relation. For example, the NB and DW models demonstrate different variance functions, that is, different Mean-to-Variance relations. The NB variance can only handle the over-dispersed data relative to the Poisson, as mentioned earlier.

3.6.3 Variances ratio plot

As mentioned before, the data can be over-or under-dispersed relative to some model; if its VR in Equation 1.1, less or greater than one. Hence, it is informative to check whether the data shows any under- or over-dispersion relative to the specified model. Therefore, the same some note, mentioned previously in the Mean-to-Variance plot, is required here for the calculation of this VR. Thus, based on the percentiles of the linear predictors $\eta_i = \boldsymbol{X'\alpha}$ from DW, the observed variance and the theoretical variances based on each model are computed for each group. Subsequently, the box-plots for these VR are plotted, in which the well specified model is the one whose VR is closest to one. In other words, in the case of good fitting, we would expect the VR in Equation 1.1 to be close to one for each X.

The points for the observed means and variances in the Mean-to-Variance plot are approximately equivalent to the box-plot for the VR from Poisson fitting.

3.6.4 Expected frequencies

The coefficients of the DW regression model and the corresponding GLMs coefficients cannot be directly compared since these regression parameters are scaled in different ways, as mentioned earlier. Additionally, it might be irrelevant to investigate the error that measure the differences between the true means and estimated means for count data. This is because the main concern may focus on counts rather than on estimated means. Then, it would be more interesting to consider the performance of a model in regard to its ability to predict the frequency of each count, in other words, the number of zeros, the number of ones, etc.

The expected frequencies for the counts, from a model with pmf f, with a regression structure can be obtained for a count h as follows:

$$\sum_{i=1}^{n} f(Y_i = h|\underline{\theta}_i)$$

where $\underline{\theta}$ are the parameters need to be estimated for the applied model and n is the sample size. Thus, a plot for the observed frequencies for the response variable against its expected (predicted) frequencies from each model can be used to assess the model performance. The best fitting is the one that is closest to the observed frequencies.

3.6.5 Model diagnostics

Following a data fitted using any regression model, it is essential to consider a diagnostics analysis to investigate the appropriateness of the model. Therefore, a residual analysis has been considered to detect the departure from a supposed model and outlying observations. Given that the response is discrete, it is advised to perform a residual analysis based on the randomized quantile residuals, as developed by Dunn and Smyth (1996) and used in many other studies, e.g. ?, Ospina and Ferrari (2012), Vanegas et al. (2013), Schmidt and Hurling (2014) and Spyroglou et al. (2015), among others. In particular, for the DW regression model, let:

$$r_i = \Phi^{-1}(u_i) \qquad , \qquad i = 1, \ldots, n \tag{3.27}$$

where $\Phi(.)$ is the standard normal distribution function and u_i is a uniform RV on the interval:

$$\begin{aligned}(a_i, b_i] &= \left(\lim_{y\uparrow y_i} F(y; \hat{q}, \hat{\beta}), F(y; \hat{q}, \hat{\beta})\right] \\ &\approx \left[F(y_i - 1; \hat{q}_i, \hat{\beta}), F(y_i; \hat{q}_i, \hat{\beta})\right]\end{aligned}$$

These residuals follow the standard normal distribution apart from sampling variability in $\hat{q}_i$ and $\hat{\beta}$. Hence, the validity of a DW model can be assessed using some goodness-of-fit investigations of the normality of the residuals, as follows:

- The histogram or the normal Q-Q plot can be used to visually check the normality of these residuals. Another advantage of the Q-Q plot is its ability to detect the outliers in the dataset.
- A normality test, such as Kolmogorov-Smirnov test, can be considered for testing the null hypothesis that these residuals follow a standard normal distribution ($N(0,1)$).
- A simulated envelope can be added to the Q-Q plot, providing a helpful diagnostic tool (Atkinson, 1985), as in Ferrari and Cribari-Neto (2004), Garay et al. (2011) and Sáez-Castillo and Conde-Sánchez (2013). In these plots, few points fall beyond the envelope's bounds, indicating a good model fit.

The simulated envelope graphical tool can be simply described by adding a simulated envelope that assess whether the observed residuals are consistent with the fitted model. Then, the 95% simulated envelope for the residuals, which is mainly an empirical probability plot for the ordered residuals with their sampling distribution quantiles, against the corresponding quantiles from the standard normal distribution, is proposed using the following steps, assuming the DW is correct:

1. After fitting the DW regression model for the observed data $\boldsymbol{Y}^0$ and obtaining the estimated parameters, $\hat{\beta}_{ML}$ and $\hat{\boldsymbol{\alpha}}_{ML}$, their randomized quantile residuals are calculated, called the observed residuals, $\boldsymbol{r}^0$.

2. It has been suggested in Atkinson (1985) that 19 samples be simulated, and thus the probability of a given residual falling beyond the bounds of the envelope will be approximately be $\frac{1}{20} = 0.05$. For this number of iteration, $h = 1, 2, \dots, H = 19$, the following steps are conducted:

   (i) Generate a sample with size n from DW with the fitted parameters $\hat{\beta}_{ML}$ and $\hat{\boldsymbol{\alpha}}_{ML}$

   (ii) Fit this simulated data using the DW regression model

(iii) Compute and store the ordered residuals for these simulated samples, denoted by $\boldsymbol{r}^h$, as columns in a matrix. Thus, we have a matrix represents the Monte Carlo sampling distribution of the residuals, with a dimension of $(n \times H)$ and each column is the $\boldsymbol{r}^{h^{th}}$ residuals for a sample with size n

3. The 2.5^{th} and 97.5^{th} quantiles, are denoted by, $r_i^{2.5}$ and $r_i^{97.5}$ are calculated for each row.

4. Plot the ordered observed residuals $\boldsymbol{r}^0$ against the normal scores, $\Phi^{-1}\left(i/(n+1)\right), i = 1, \ldots, n.$

5. Add a 95% simulated envelope to the plot by drawing the $\boldsymbol{r}^{2.5}$ and $\boldsymbol{r}^{97.5}$ for the lower and upper bounds, respectively.

3.7 Discrete Weibull regression naturally handles covariate-specific dispersion

It has been shown in chapter 2 how DW can model data that are under- and over-dispersed relative to the Poisson. In this section, we would like to investigate this further within a regression context. Here, it is also possible that the conditional variance is larger than the conditional mean for a specific covariate pattern (over-dispersion), but the conditional variance is smaller than the conditional mean for another covariate pattern (under-dispersion). In the literature, regression models for count data that can capture under-dispersion or both types of over- and under-dispersion simultaneously take the form of Poisson regression, such as the quasi-Poisson, COM-Poisson or hyper-Poisson (Sáez-Castillo and Conde-Sánchez, 2013). In the case of mixed types of dispersion, the dispersion parameter can be assumed to be linked to the covariates. However, a covariate-dependent dispersion increases the complexity of the model significantly and reduces its interpretability. So, in practice, most implementations fix the dispersion parameter and assume that only the mean is linked to the covariates. As the DW distribution naturally accounts for over- and under-dispersion, a DW regression model becomes a simple and attractive alternative to existing regression models for count data.

This point is emphasized by a simple simulation study. A multiple regression with two predictors, $X1 \sim N(0,1)$ and $X2 \sim Uniform(0,10)$, is examined. The true value of the regression parameter is assumed to be $\boldsymbol{\alpha} = (\alpha_0, \alpha_1, \alpha_2) = (0.5, 0.4, -.3)$. In addition, the parameter β of the DW is assumed to be $\beta = 1.6$. Then, a sample with size $n = 300$ from $DW(q_i, \beta)$ is considered, where q_i is calculated as in Equation 3.2, for $i = 1, \ldots, n$.

Figure 3.1 displays the Mean-to-Variance relations for the observed variance and observed mean indicating under-dispersion for some cases, and over-dispersion for other. Then, among the theoretical relations from NB and DW fittings, the DW relations are the closest to the observed.

Figure 3.2 shows the VR plot of the dispersions in Equation 1.1 in the case of the Poisson, NB and DW fitting, which is the model used in the simulation. The Poisson and NB both show under-dispersion in most cases and over-dispersion in two cases. Thus, this simulation shows a simple scenario of a mixed level of dispersion, which cannot be captured well by standard Poisson and NB models.

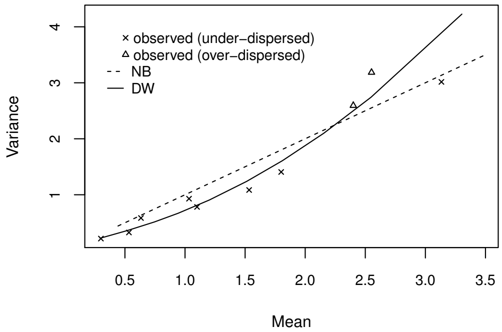


Figure 3.1: The Mean-to-Variance plot for mixed-dispersed simulated data from the DW regression model, with the theoretical Mean-to-Variance fitted by NB and DW.

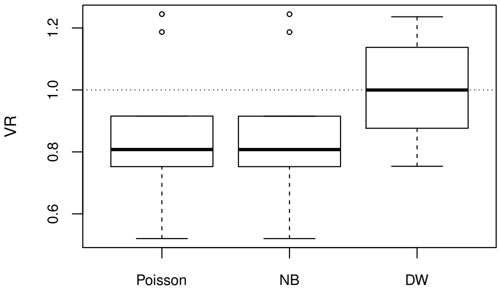


Figure 3.2: Distribution of the ratios of the observed and theoretical conditional variance on mixed-dispersed simulated data from the DW regression model, fitted by the Poisson, NB and DW.

On the other hand, to assess the performance of the estimation under the DW model, this simulation study is carried out for 1000 iterations. Table 3.1 reports the MLEs of the parameters, together with some accuracy measurements that would be explained more in the next section, over 1000 iterations.

Table 3.1: MLEs based on the simulation study for the DW regression model with true parameters $\boldsymbol{\alpha} = (0.5, 0.4, -.3)$ *and* $\beta = 1.6$*, for the mixed-dispersed case.*

	MLE	Bias	MSE	Length
α_0	0.5106	0.0106	0.0184	0.5191
α_1	0.4104	0.0104	0.0046	0.2556
α_2	-0.3038	-0.0038	0.0008	0.1074
β	1.6185	0.0185	0.0091	0.3585

3.8 Simulation study

A simulation study was performed to assess and evaluate the performance of the MLEs for the DW with a regression structure. These estimators can be evaluated using certain accuracy measures.

Different sample sizes $n_1 = 50$, $n_2 = 100$, $n_3 = 250$, $n_4 = 500$ and $n_5 = 1000$ are considered. Additionally, different dispersion types are considered, that is, under-, over-dispersion and zero inflation cases. A multiple regression with three predictors is examined. All the results are based on an average over 1000 repetitions. In each iteration, MLEs and their asymptotic two-sided CIs are computed according to Steenberger (2006), through the following steps:

- **Step 1:** Simulate three random samples with size n to present the covariates from the following distributions:
  - **regressor 1:** normal distribution $N(0, 1)$.
  - **regressor 2:** uniform distribution with parameters $(-0.3, 0.3)$.
  - **regressor 3:** Bernoulli distribution with parameter (0.4).
- **Step 2:** The true values of the parameters are chosen to be:
  - Since the $P(0)$ depends on q, the regression parameters that are as-

sumed to generate some excessive zero data are as follows:

$$\boldsymbol{\alpha} = (\alpha_0, \alpha_1, \alpha_2, \alpha_3) = (0.1, -0.2, 1.6, 0.2)$$

with $\beta = 0.9$.

Additionally, the regression parameters, for the under- and over-dispersion case, are assumed to be:

$$\boldsymbol{\alpha} = (\alpha_0, \alpha_1, \alpha_2, \alpha_3) = (-2.8, 0.01, 0.4, -0.2)$$

  - The shape parameter β of DW has been selected to represent the two cases of under- and over-dispersion. Thus, as mentioned in chapter 2, for the under-dispersed case $\beta = 2.5$ has been selected, while $\beta = 0.9$ for the over-dispersion case.
  - Then, q can be calculated for each $\boldsymbol{X}$ using Equation 3.2.

- **Step 3:** The data fitted by the Poisson regression, in order to obtain some initial values then, minus the regression coefficient from this Poisson fitting is considered for the unknown parameters $\boldsymbol{\alpha}$. In addition, the estimate of β by fitting the DW distribution for the response variable Y is considered for β. Then, for each sample size n, the replications are conducted, where for each iteration $(1 : 1000)$:

  - Generate random samples with sample size n from a population whose pmf is given by Equation 3.3, using the package found in Barbiero (2015). That is, for each $\boldsymbol{x_i}$ and q_i there is a corresponding y_i, where $i = 1, 2, \ldots, n$.
  - Based on this sample and using the above initials, the MLEs$_{itr}$ of the parameters $\boldsymbol{\alpha}$ and β, which is denoted as $\hat{\theta}_{itr}$, are computed by maximizing the log-likelihood function in Equation 3.8, using *"optim"* in R.
  - In addition to the MLEs$_{itr}$, the lower-limit (LL_{itr}) and upper-limit (UL_{itr}) levels of the 95% CIs for each MLE are conducted.

- **Step 4:** The three steps above are repeated 1000 times.

Subsequently, 1000 values are found for the MLEs and CIs bounds. The average of these values is computed to obtain the MLEs, that is, $\hat{\boldsymbol{\alpha}}_{ML}$ and $\hat{\beta}_{ML}$. In other words, for each parameter, the following average is computed:

$$\hat{\theta}_{ML} = \frac{\sum_{itr=1}^{1000} \hat{\theta}_{itr}}{1000} \tag{3.28}$$

In addition, the average of each lower bound and upper bound for the CI is calculated, respectively, as follows:

$$\begin{aligned} LL &= \frac{\sum_{itr=1}^{1000} LL_{itr}}{1000} \\ UL &= \frac{\sum_{itr=1}^{1000} UL_{itr}}{1000} \end{aligned} \tag{3.29}$$

Consequently, the length of the asymptotic CIs are found. Then, the estimators are evaluated using the following measures:

- Bias, which is the difference between the average estimate and true value, and should ideally be close to zero:

$$Bias(\hat{\theta}_{ML}) = \hat{\theta}_{ML} - \theta_{tr} \tag{3.30}$$

- mean squared error (MSE), which measures the average squared distance between the estimate and true value, which also should ideally be close to zero:

$$MSE(\hat{\theta}_{ML}) = \frac{\sum_{itr=1}^{1000} (\hat{\theta}_{itr} - \theta_{tr})^2}{1000} \tag{3.31}$$

  where θ_{tr} is the true value of the parameter θ, $\hat{\theta}_{itr}$ is the MLE of the parameter θ for each iteration and $\hat{\theta}_{ML}$ is the MLE of θ.

To show a description for each case, a histogram for a sample in one of the iterations for $n = 500$ is conducted. Moreover, to investigate the dispersion of this simulated sample, the VR plot is considered. For the over-dispersion case, the VR for Poisson fit is not included, to make a reasonable scale of the figure, as it gives values very far from one.

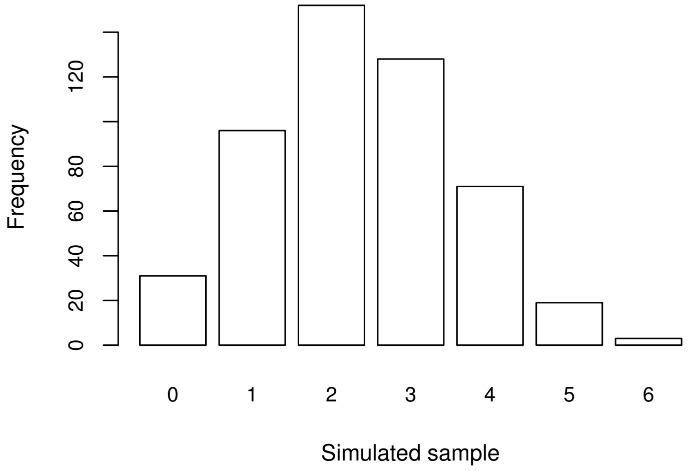


Figure 3.3: Histogram for the under-dispersed simulated data from the DW regression model.

Table 3.2: MLEs based on the simulation study for the DW regression model with true parameters $\boldsymbol{\alpha} = (-2.8, 0.01, 0.4, -0.2)$ *and* $\beta = 2.5$*, for the under-dispersion case.*

n	parameter	MLE	Bias	MSE	Length
50	α_0	-3.0046	-0.2046	0.2844	1.8095
	α_1	0 .0009	-0.0091	0.0333	0.6578
	α_2	0.4576	0.0576	0.9142	3.4338
	α_3	-0.2259	-0.0259	0.1119	1.1909
	β	2.7062	0.2062	0.1734	1.2862
100	α_0	-2.9028	-0.1028	0.1149	1.2179
	α_1	0.003	-0.007	0.014	0.4463
	α_2	0.4639	0.0639	0.427	2.4777
	α_3	-0.2136	-0.0136	0.0518	0.8395
	β	2.6031	0.1031	0.0621	0.8638
250	α_0	-2.855	-0.055	0.0425	0.7526
	α_1	0.0114	0.0014	0.0053	0.2749
	α_2	0.4184	0.0184	0.1553	1.4951
	α_3	-0.1914	0.0086	0.0183	0.5358
	β	2.5471	0.0471	0.0222	0.5305
500	α_0	-2.8252	-0.0252	0.0207	0.5325
	α_1	0.0114	0.0014	0.0023	0.1877
	α_2	0.4017	0.0017	0.0733	1.0813
	α_3	-0.201	-0.001	0.0081	0.3653
	β	2.5238	0.0238	0.0103	0.3704
1000	α_0	-2.8131	-0.0131	0.0097	0.3724
	α_1	0.0111	0.0011	0.0012	0.1334
	α_2	0.4064	0.0064	0.037	0.7259
	α_3	-0.2012	-0.0012	0.0044	0.2613
	β	2.5125	0.0125	0.0046	0.2605

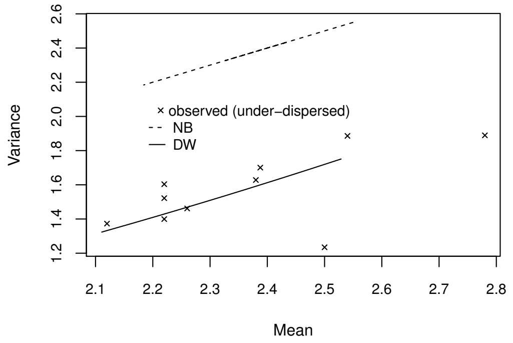


Figure 3.4: The Mean-to-Variance plot for under-dispersed simulated data from the DW regression model, with the theoretical Mean-to-Variance fitted by NB and DW.

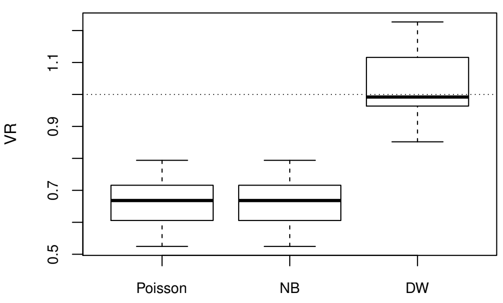


Figure 3.5: Distribution of the ratios of the observed and theoretical conditional variance on simulated data from the DW regression model, for the under-dispersion case, fitted by the Poisson, NB and DW.

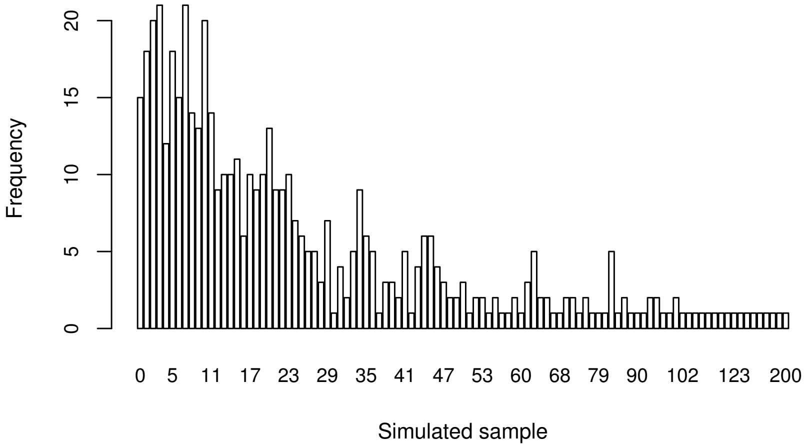


Figure 3.6: Histogram for the over-dispersion simulated data from the DW regression model.

Table 3.3: MLEs based on the simulation study for the DW regression model with true parameters $\boldsymbol{\alpha} = (-2.8, 0.01, 0.4, -0.2)$ and $\beta = 0.9$, for the over-dispersion case.

n	parameter	MLE	Bias	MSE	Length
50	α_0	-2.9859	-0.1859	0.2481	1.72
	α_1	0.0023	-0.0077	0.0306	0.6349
	α_2	0.4448	0.0448	0.8482	3.3255
	α_3	-0.2213	-0.0213	0.1049	1.1511
	β	0.9674	0.0674	0.0189	0.4341
100	α_0	-2.8988	-0.0988	0.1069	1.1652
	α_1	0.0033	-0.0067	0.0131	0.4326
	α_2	0.4434	0.0434	0.4004	2.4021
	α_3	-0.2124	-0.0124	0.0483	0.8142
	β	0.9354	0.0354	0.0072	0.2946
250	α_0	-2.852	-0.052	0.038	0.7208
	α_1	0.0104	0.0004	0.0049	0.2672
	α_2	0.412	0.012	0.1437	1.4512
	α_3	-0.1915	0.0085	0.017	0.5208
	β	0.916	0.016	0.0026	0.1813
500	α_0	-2.8252	-0.0252	0.0188	0.5111
	α_1	0.0115	0.0015	0.0022	0.1824
	α_2	0.3996	-0.0004	0.0699	1.0504
	α_3	-0.2001	-0.0001	0.0076	0.3551
	β	0.9083	0.0083	0.0012	0.1269
1000	α_0	-2.813	-0.013	0.0087	0.3575
	α_1	0.0115	0.0015	0.0011	0.1297
	α_2	0.4064	0.0064	0.0336	0.7057
	α_3	-0.2017	-0.0017	0.004	0.2542
	β	0.9045	0.0045	0.0005	0.0893

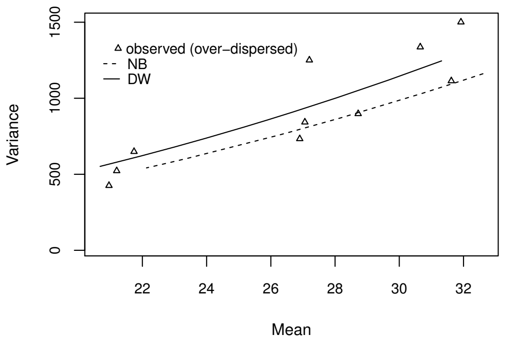


Figure 3.7: The Mean-to-Variance plot for over-dispersed simulated data from the DW regression model, with the theoretical Mean-to-Variance fitted by NB and DW.

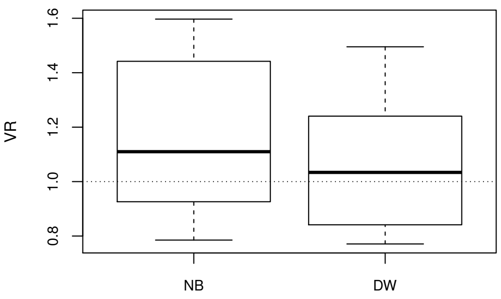


Figure 3.8: Distribution of the ratios of the observed and theoretical conditional variance on simulated data from the DW regression model, for the over-dispersion case, fitted by NB and DW.

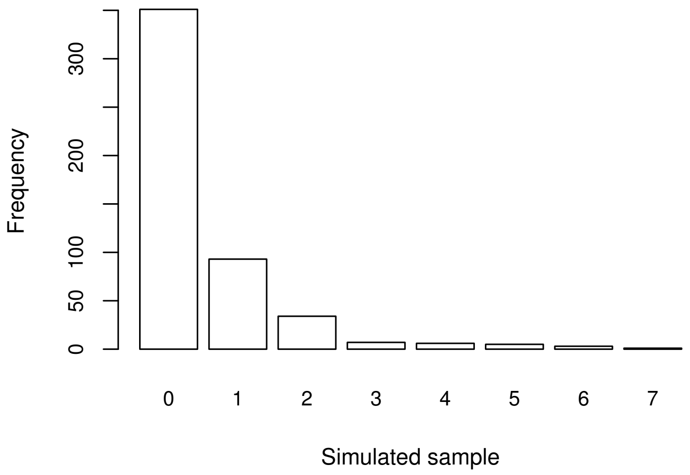


Figure 3.9: Histogram for the excessive zero simulated data from the DW regression model, in which 70.2% of the data is comprised of zeros.

Table 3.4: MLEs based on the simulation study for the DW regression model with true parameters $\boldsymbol{\alpha} = (0.1, -0.2, 1.6, 0.2)$ *and* $\beta = 0.9$*, for the excessive zero case.*

n	parameter	MLE	Bias	MSE	Length
50	α_0	0.1134	0.0134	0.0789	1.0068
	α_1	-0.2515	-0.0515	0.0548	0.778
	α_2	1.902	0.302	1.4137	4.0165
	α_3	0.2179	0.0179	0.1442	1.3091
	β	1.0555	0.1555	0.08	0.8759
100	α_0	0.0985	-0.0015	0.0271	0.64
	α_1	-0.2224	-0.0224	0.0167	0.4859
	α_2	1.7541	0.1541	0.5848	2.7492
	α_3	0.2088	0.0088	0.0591	0.8904
	β	0.9776	0.0776	0.0252	0.5408
250	α_0	0.0922	-0.0078	0.0105	0.3894
	α_1	-0.2058	-0.0058	0.0060	0.2939
	α_2	1.643	0.043	0.1765	1.6191
	α_3	0.2192	0.0192	0.0222	0.5648
	β	0.9304	0.0304	0.0078	0.323
500	α_0	0.0974	-0.0026	0.0052	0.2866
	α_1	-0.2012	-0.0012	0.0027	0.202
	α_2	1.6232	0.0232	0.0889	1.1696
	α_3	0.2052	0.0052	0.0093	0.3841
	β	0.9168	0.0168	0.0038	0.2292
1000	α_0	0.0976	-0.0024	0.0027	0.196
	α_1	-0.2006	0 .0006	0.0013	0.1425
	α_2	1.6208	0.0208	0.0462	0.7839
	α_3	0.2014	0.0014	0.0052	0.2744
	β	0.9105	0.0105	0.0017	0.1579

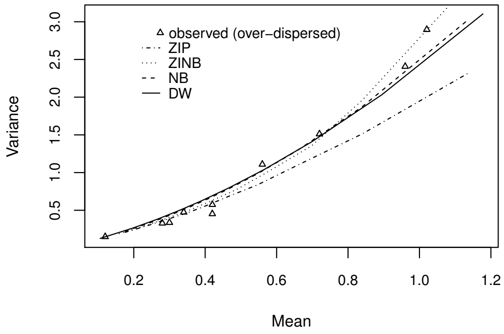


Figure 3.10: The Mean-to-Variance plot for excessive zero simulated data from the DW regression model, with the theoretical Mean-to-Variance fitted by ZIP, ZINB, NB and DW.

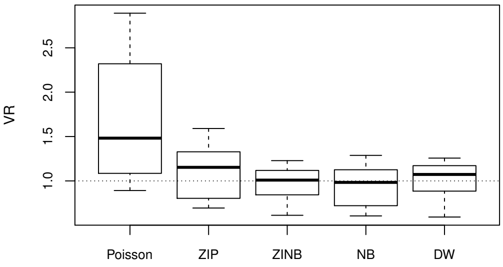


Figure 3.11: Distribution of the ratios of the observed and theoretical conditional variance on simulated data from the DW regression model, for the excessive zero case, fitted by the Poisson, ZIP, ZINB, NB and DW.

It may noted from Table 3.2, Table 3.3 and Table 3.4 that the measurements of accuracy, bias, MSE associated with the MLEs of $\boldsymbol{\alpha}$, and β as well as the length of the CIs are all close to zero, and they decrease as the sample size n increases. Thus, the properties of the MLE are achieved. Additionally, Figure 3.4, Figure 3.7 and Figure 3.10 show that the Mean-to-Variance relations from the DW fitting is the most close to the observed Mean-to-Variance relations. Also, Figure 3.5, Figure 3.8 and Figure 3.11 show that the VR of DW is the closest to one.

3.9 Numerical examples

To demonstrate the ability of the DW regression model to handle over- and under-dispersion automatically, in this section DW regression is applied to different data sets that show various types of dispersions relative to the Poisson. The first part includes an under-dispersed dataset, while the second includes an over-dispersed case. The third subsection focuses on a zero-inflated dataset. Finally, an illustrative example for the mixed level of dispersion is provided. The purpose here is just to demonstrate the DW model and not to test the significance of each covariates. Various popular count data regression models, namely Poisson regression, NB regression, zero-inflated and hurdle models, are applied and compared with DW regression by means of the classical AIC criterion. The model with the smallest values for this criterion is then chosen as the best fitting for this dataset. Additionally, their expected frequencies are compared with the observed frequencies in which the model that provides close expected frequencies to the observed is the best.

3.9.1 The case of under-dispersion: inhaler usage data

For this example, a data from Grunwald et al. (2011) and Canale and Dunson (2012) is used, consisting of 5209 observations and report the daily count of using (albuterol) asthma inhalers for 48 children suffering from asthma, aged between 6 and 13 years, during the school day, for a period of time at the Kunsberg School at the National Jewish Health in Denver, Colorado. The main objective of this analysis is to investigate the relationship between the inhaler use (representing the asthma symptoms) and air pollution, which is recorded by four covariates, as follows:

- The percentage of humidity,
- The barometric pressure (in mmHG/1000),
- The average daily temperature (in Fahrenheit degree/100),
- The morning levels of PM25, which are small air particles less than 25mm in diameter.

The response variable, which is the inhaler use count, has a sample mean of 1.2705 and a variance of 0.8433, thus pointing to a case of under-dispersion relative to the Poisson.

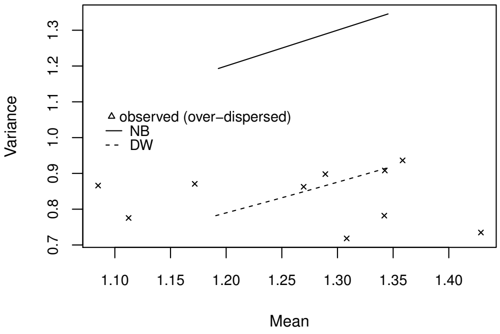


Figure 3.12: The Mean-to-Variance plot for the inhaler use dataset, with the theoretical Mean-to-Variance fitted by NB and DW.

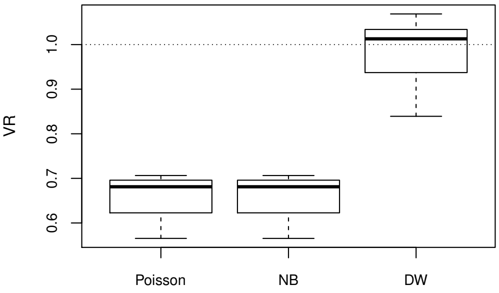


Figure 3.13: Distribution of the ratios of the observed and theoretical conditional variance on the inhaler use dataset, fitted by the Poisson, NB and DW.

Table 3.5: MLEs, SEs (in parentheses) and AIC from the Poisson, NB and DW regression models fitted to the inhaler use dataset.

	intercept	humidity	pressure	temperature	particles	other	AIC
Poisson	-2.2132 (1.7115)	-0.1125 (0.0840)	4.0950 (2.7230)	-0.2035 (0.1293)	0.0225 (0.0129)	-	13915.47
NB	-2.2132 (1.7115)	-0.1125 (0.0840)	4.0950 (2.7230)	-0.2035 (0.1293)	0.0225 (0.0129)	$\hat{k}$=31905.28 (65588.91)	13917.54
DW	1.8112 (1.9731)	0.2233 (0.1013)	-5.6120 (3.1357)	0.3691 (0.1563)	-0.0289 (0.0156)	$\hat{\beta}$=2.1277 (0.0259)	13484.36

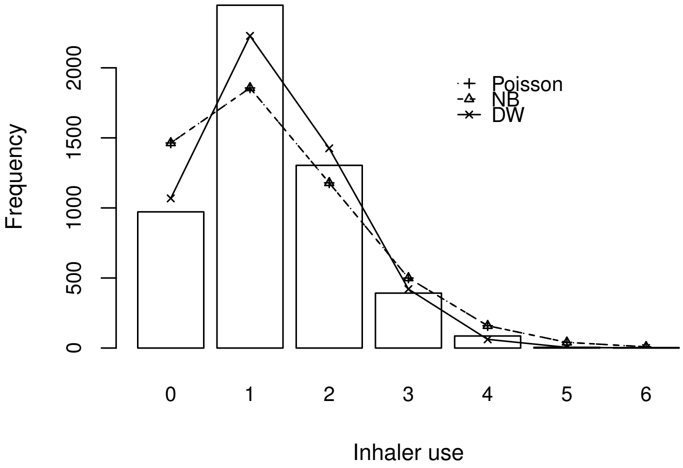


Figure 3.14: Histogram for the observed frequencies and expected frequencies from the Poisson, NB and DW regression models for the inhaler use dataset.

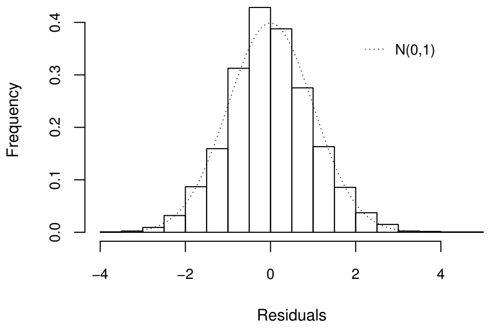


Figure 3.15: Histogram of the randomized quantile residuals from the DW regression model fitted to the inhaler use dataset with superimposed N(0,1) density.

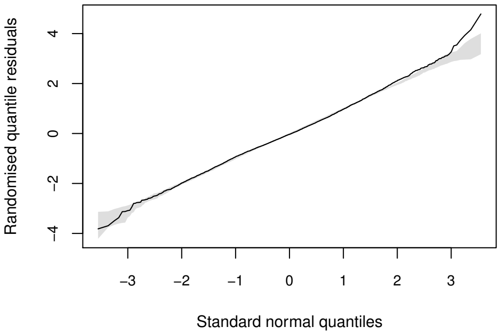


Figure 3.16: Simulated envelope for the randomized quantile residuals from the DW regression model fitted to the inhaler use dataset.

The results in Table 3.5 suggest that DW regression provides better fitting than both the Poisson and NB models, according to the AIC. Figure 3.12 and Figure 3.13 indicate under-dispersion relative to the Poisson and NB across the full range of covariates, and a good fit of DW compared to the other models (theoretical Mean-to-Variance relations close to the observed relations and VR values close to 1). Figure 3.14 compares the observed and expected frequencies for the three models and shows again a good fit for DW. Finally, Figure 3.15 plots the randomized quantile residuals from the DW regression model, which are only moderately departing from normality (p-value of Kolmogorov-Smirnov test: 0.026). Additionally, Figure 3.16 plots the simulated envelopes for theses residuals and no much points falling outside the bounds.

3.9.2 The case of over-dispersion: strikes data

This dataset is available in the Ecdat R package (Croissant (2015)), under the name of *StrikeNb*. The response variable is the number of contract strikes in U.S. manufacturing observed monthly from January 1968 to December 1976. The predictor is the level of economic activity, which is measured as the cyclical departure of aggregate production from its trend level. The response variable has a sample mean of 5.2407 and a variance of 14.0723, suggesting over-dispersion relative to the Poisson. Indeed, a comparison of Poisson and NB distributions solely on the response variable using a likelihood ration test (LRT) (lmtest R package, Zeileis and Hothorn (2002)) shows evidence of over-dispersion, with a chi-square test statistic of 63.372 and a p-value of < 0.001.

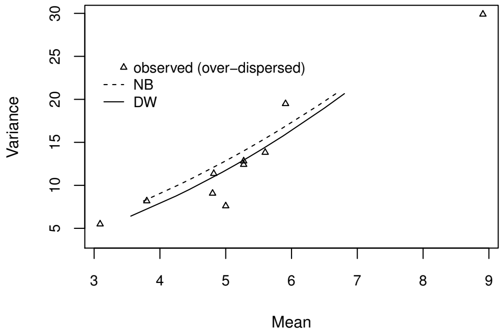


Figure 3.17: The Mean-to-Variance plot for the strikes dataset, with the theoretical Mean-to-Variance fitted by NB and DW.

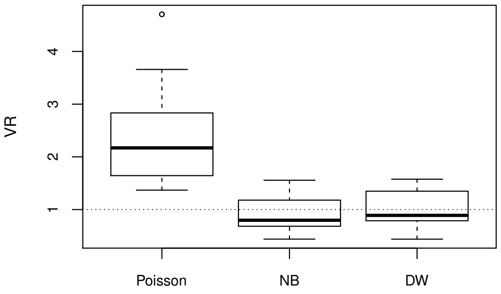


Figure 3.18: Distribution of the ratios of the observed and theoretical conditional variance on the strikes dataset, fitted by the Poisson, NB and DW.

Table 3.6: MLEs, SEs (in parentheses) and AIC from the Poisson, NB and DW regression models fitted to the strikes dataset.

	intercept	economic activity	other	AIC
Poisson	1.6539 (0.0422)	3.1342 (0.8032)	-	627.9689
NB	1.6538 (0.0686)	3.2250 (1.2841)	$\hat{k}$=3.1849 (0.739)	566.5969
DW	-3.0706 (0.2910)	-5.2956 (1.9096)	$\hat{\beta}$=1.6527 (0.1302)	564.157

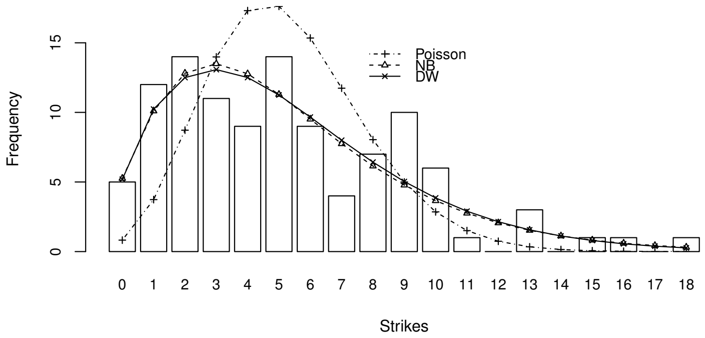


Figure 3.19: Histogram for the observed frequencies and expected frequencies from the Poisson, NB and DW regression models for the strikes dataset.

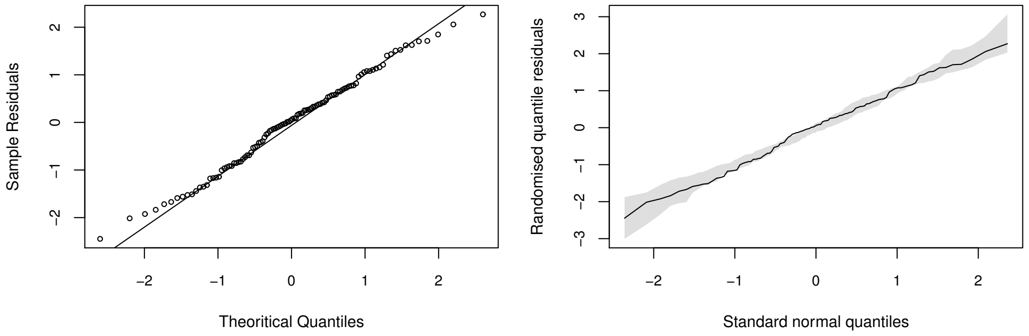


Figure 3.20: Q-Q plot of the randomized quantile residuals from the DW regression model fitted to the strikes dataset.

Figure 3.21: Simulated envelope for the randomized quantile residuals from the DW regression model fitted to the strikes dataset.

After fitting three regression models and comparing them via AIC, Table 3.6 shows that the DW model is only marginally superior to NB, but both DW and NB give much better fit to the data than the Poisson regression model. Figure 3.17 and Figure 3.18 indicate a case of over-dispersion relative to the Poisson across the whole range of covariates. Additionally, they indicate good fitting by NB and DW. Figure 3.19 confirms the good fit of NB and DW. Finally, Figure 3.20 and Figure 3.21 show that the residuals closely follow a normal distribution (Kolmogorov-Smirnov p-value 0.951), with few points falling outside the simulated 95% envelope bounds.

3.9.3 The case of excessive zeros: doctor visits from the German health survey data

The following dataset illustrates the case of excessive zero counts. Thus, besides the Poisson, NB, and DW regressions, we will also include zero-inflated and hurdle models in the comparison. For these, we consider the logit link function for the binomial distribution representing the probability of the extra zeros, using R package pscl ((Zeileis et al., 2008)).

This dataset is available from the *COUNT* R package (Hilbe, 2014) name of *badhealth*, comes from the German health survey and contains 1127 observations for the number of visits to certain doctors during 1998. In addition, the data includes two other variables: an indicator variable representing patients claiming to be in bad health (1) or good health (0) and the age of the patient. The response variable (number of visits) ranges from 0 to 40 visits to doctors throughout 1998, with approximately 32% zeros, and thus it can be considered as a case of excessive zeros. Indeed, the response has a sample mean of 2.3532 and a variance of 11.9818, suggesting over-dispersion. Also, the LRT between the NB and Poisson returns a test statistic of 1165.3 and a p-value of < 0.001.

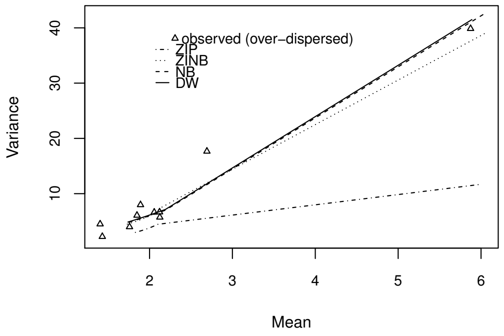


Figure 3.22: The Mean-to-Variance plot for the doctor visits from the German health dataset, with the theoretical Mean-to-Variance fitted by ZIP, ZINB, NB and DW.

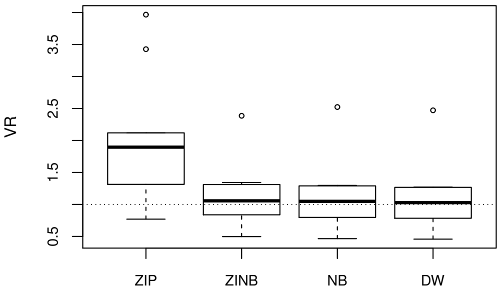


Figure 3.23: Distribution of the ratios of the observed and theoretical conditional variance on the doctor visits from the German health dataset, fitted by the Poisson, ZIP, ZINB, NB and DW.

Table 3.7: MLEs, SEs (in parentheses) and AIC from the Poisson, NB and DW regression models fitted to the doctor visits from the German health dataset.

	intercept	bad health	age	other	AIC
Poisson	0.4470	1.1083	0.0058	-	5638.552
	(0.0714)	(0.0462)	(0.0018)		
NB	0.4041	1.1073	0.0070	$\hat{k}$=0.9975	4475.285
	(0.1308)	(0.1116)	(0.0034)	(0.0693)	
Zero-inflated Models					
Poisson					
count model	0.6852	0.8767	0.0089	-	5110.096
	(0.0767)	(0.0480)	(0.0019)		
logit model	-1.4029	-1.0996	0.0142	-	
	(0.2770)	(0.2947)	(0.0070)		
NB					
count model	0.3482	1.0415	0.0100	log($\hat{k}$)=0.1214	4477.748
	(0.1322)	(0.1142)	(0.0036)	(0.1101)	
logit model	-5.5259	-2.5277	0.0658	-	
	(1.7889)	(5.1685)	(0.0321)		
Hurdle Models					
logit model	0.9678	1.2629	-0.0083	-	-
	(0.2324)	(0.2889)	(0.0060)		
Poisson count model	0.6794	0.8764	0.0090	-	5109.505
	(0.0771)	(0.0479)	(0.0019)		
NB count model	0.1375	1.0710	0.0137	log($\hat{k}$)=-0.0422	4472.075
	(0.1704)	(0.1247)	(0.0041)	(0.1580)	
DW	-0.8999	-0.8481	-0.0004	$\hat{\beta}$=0.9887	4474.974
	(0.1088)	(0.1036)	(0.0028)	(0.0265)	

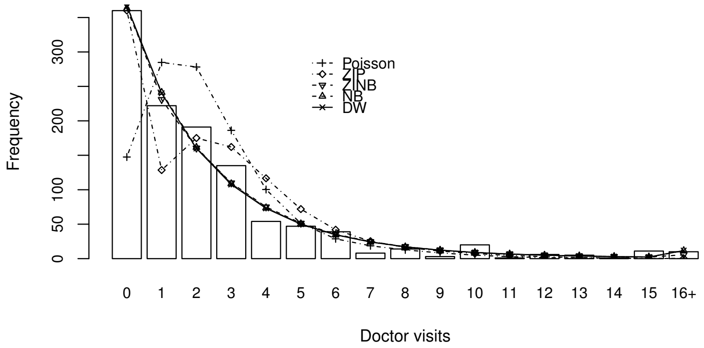


Figure 3.24: Histogram for the observed frequencies and expected frequencies from the Poisson, NB, DW, ZIP and ZINB regression models for the doctor visits from the German health dataset.

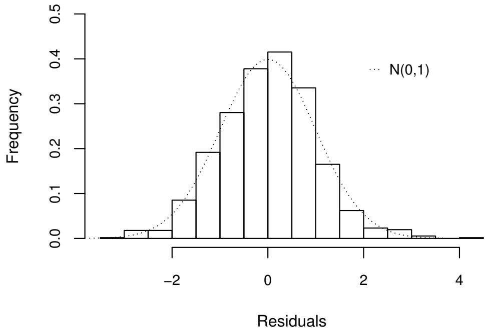


Figure 3.25: Histogram of the randomized quantile residuals from the DW regression model fitted to the doctor visits from the German health dataset with superimposed N(0,1) density.

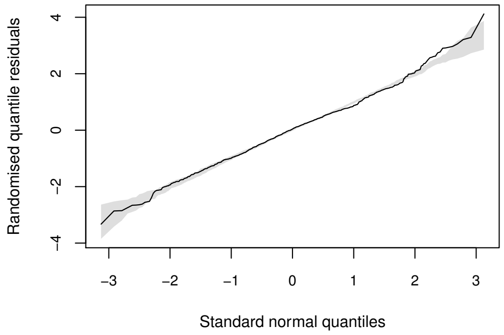


Figure 3.26: Simulated envelope for the randomized quantile residuals from the DW regression model fitted to the doctor visits from the German health dataset.

Table 3.7 shows the best fit for the DW and HNB regression models in terms of their minimum AIC. For the figures, the results from the hurdle regression models are not included, as they provide almost identical results to their corresponding from zero-inflated models. Figure 3.22 and Figure 3.23 show a case of over-dispersion relative to the Poisson across the full range of covariates and a good fit for ZINB, NB and DW. We exclude the Poisson from the plot for visualization purposes, as the VR values are large in this case. Additionally, Figure 3.24 shows that the expected frequencies for ZINB, NB and DW are the closest to those observed, while the Poisson and ZIP are a bit far away. This again confirms the good fit of DW. For visualization purposes, the small number of observations larger than 16 are grouped together in this plot.

As in the previous example, Figure 3.25 shows that the residuals of the DW model are approximated by a normal distribution (Kolmogorov-Smirnov p-value 0.05927). In addition, in the simulated envelope in Figure 3.26, there are few points that fall beyond the envelope bounds. Hence, this example shows how DW can also model cases of excessive zeros, without the need for additional parameters as in the zero-inflated models.

3.9.4 The case of a mixed level of dispersion: bids data

In this section, we report the analysis of a dataset where a mixed level of dispersion was observed, that is, the conditional distribution is over-dispersed relative to the Poisson for some covariate pattern but is under-dispersed for another covariate pattern. The data are taken from Cameron and Johansson (1997) and are available in the Ecdat R package under the name of *Bids*. This dataset records the number of bids received by 126 US firms that were targets of tender offers during a certain period of time. The dependent variable here is the number of bids, with a mean of 1.7381 and a variance of 2.0509. The objective of the study is to investigate the effect of particular variables on the number of bids. For this analysis, we consider the following covariates:

- bid price, taken as the price at a particular week divided by the price 14 working days before the bid,
- size, that is, the total book value of assets measured in billions of dollars,
- regulator, a dummy variable that is equal to 1 if there was an intervention by federal regulators and 0 otherwise.

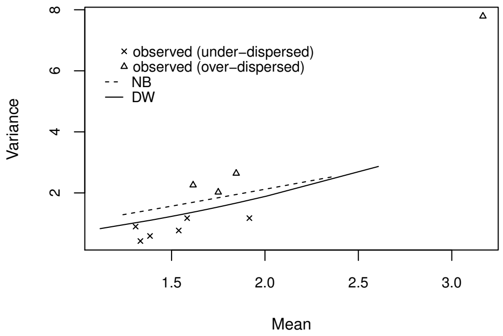


Figure 3.27: The Mean-to-Variance plot for the bids dataset, with the theoretical Mean-to-Variance fitted by NB and DW.

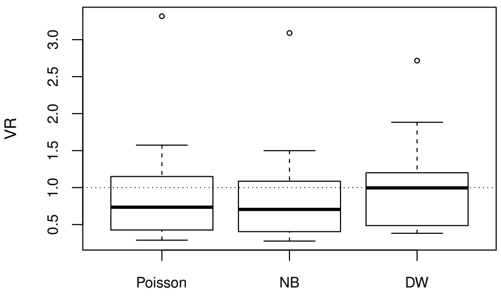


Figure 3.28: Distribution of the ratios of the observed and theoretical conditional variance on the bids dataset, fitted by the Poisson, ZIP, ZINB, NB and DW.

Table 3.8: MLEs, SEs (in parentheses) and AIC from the Poisson, NB and DW regression models fitted to the bids dataset.

	intercept	price	size	regulator	other	AIC
Poisson	1.5318 (0.5043)	-0.7849 (0.3775)	0.0362 (0.0175)	0.0547 (0.1567)	-	402.2602
NB	1.5276 (0.5174)	-0.7824 (0.3870)	0.0369 (0.0183)	0.0544 (0.1610)	$\hat{k}$=33.3289 (63.3334)	403.9481
DW	-3.3933 (0.7257)	1.3119 (0.5006)	-0.1070 (0.0404)	-0.0568 (0.2216)	$\hat{\beta}$=1.9403 (0.1365)	395.1214

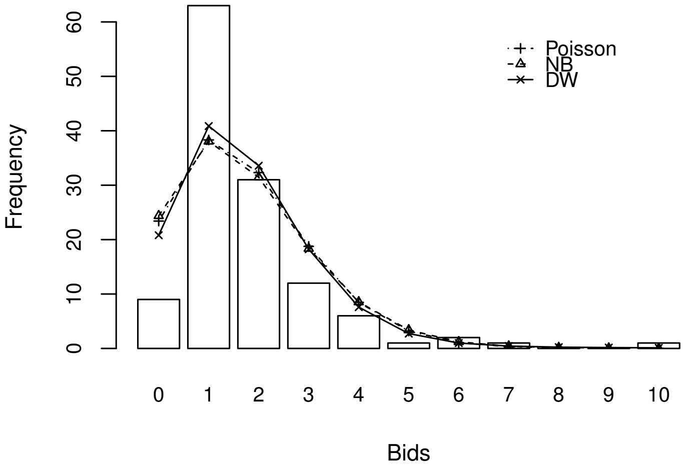


Figure 3.29: Histogram for the observed frequencies and expected frequencies from the Poisson, NB and DW regression models for the bids dataset.

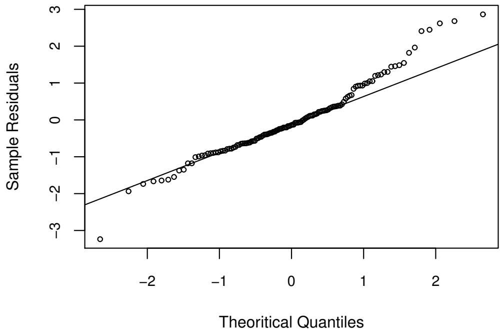


Figure 3.30: Q-Q plot of the randomized quantile residuals from the DW regression model fitted to the bids dataset.

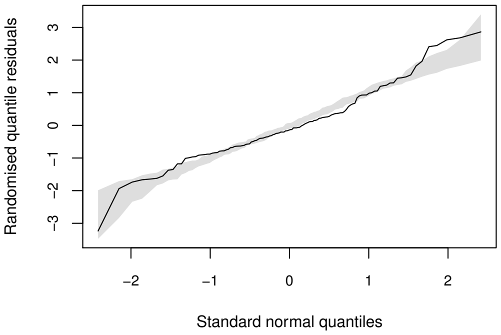


Figure 3.31: Simulated envelope for the randomized quantile residuals from the DW regression model fitted to the bids dataset.

Figure 3.27 and Figure 3.28 show a mixed level of dispersion relative to the Poisson and NB, with most covariate patterns leading to under-dispersion, but with a small number of over-dispersed cases. The DW model has a clearer distribution of VR values around 1. Figure 3.27, Figure 3.28 and Table 3.8 once again show a very good fit of the DW regression model to these data, compared to the Poisson and NB. Finally, from the Q-Q plot of the randomized residuals in Figure 3.30, it can be seen that DW also fits the data well, with a Kolmogorov-Smirnov p-value of 0.137 for the randomized quantile residuals. Additionally, few points fall beyond the envelope bounds in Figure 3.31. However, the expected frequencies in Figure 3.29 for the three different models are not the best for zero and one. Hence, this example will be analyzed further in the following chapters, based on modifications to the DW.

3.10 Concluding remarks

In this chapter a regression model based on a DW distribution is introduced to directly model the count data that is affected by some explanatory variable. Specifically, the pmf of the DW is generalized by allowing its parameter to be a function of covariates through a link function. Thus, compared to the GLMs, in which the conditional mean is central to inference and interpretation, the proposed DW regression model has the advantage of modeling the whole conditional density, including all conditional quantiles and any other properties that can be easily extracted from the fitted model. In other words, for predicting some observation based on the DW regression model fitting, the full conditional fitted distribution is considered instead of relying only on the conditional mean. This is particularly useful since most count data have highly skewed distributions.

Within this regression context, DW model can be seen as a simple and unified model for capturing different levels of dispersion in the data conditional on some covariates, namely, under-dispersion and over-dispersion relative to the Poisson, including the common case of excessive zeros.

A popular model for under-dispersion is the COM-Poisson regression model. However, its pmf is not in a closed form and contains an infinite sum, which requires an approximate computation. In fact, its implementation, which was used for some of the examples in this study, required more computational time than the DW regression model, which uses a straightforward maximum likelihood estimation procedure on a closed form pmf. This is particularly beneficial in the case of large sample sizes.

While NB is the most applied model for over-dispersion, the DW regression model is shown to be an attractive alternative for over-dispersion. In particular, several examples in this study show that DW regression provides the best fitting model, both in cases of over- and under-dispersion, and it is also able to capture situations with a mixed level of dispersion. In addition, the DW regression model can be applied to data with an excessive number of zero counts without requiring additional parameters, as in the case of zero-inflated or hurdle models.

The maximum likelihood approach has been used for the inference of the model. Then, a simulation study was implemented for different cases of dispersion within the regression context, to assess the performance of this model. The results of the study show that the measurements of accuracy, bias and MSE, as well as the length of the CI, are relatively close to zero and they decrease as the sample size n increases. In addition, this model has been applied for different data sets with varying ranges of dispersion. These applications show the well fitting and performance of the DW regression models to these different types of data.

Chapter 4

Discrete Weibull Model for Censored Data

4.1 Introduction

Censored count data can emerge in many applications where recording the response variable is restricted. In other words, the dependent variable is available for a limited range, but the covariate values are always observed. Then, applying the full (standard) regression models discussed in the previous chapter, to this type of data might result in inefficient inferences (see for example Brännäs (1992)). Hence, this chapter is concerned with this case of censored data and develops a CDW regression model for these situations. Particularly, right censoring is considered to cut the observed count data to the right. Then, as a result, some large response values are recorded as small, consequently affecting its distribution. As mentioned before, the exact relation between the mean and variance for the censored data case might be unknown. Hence, it would be useful to consider a model with the ability to handle different levels of dispersion. Consequently, the DW model is adjusted in this chapter for examining the censored data.

4.2 Discrete Weibull regression model with right censoring

The response variable Y_i; $(i = 1, \ldots, n)$ might be censored from the right at a value C for some observations in a sample. Then, the likelihood function of the CDW regression model is given by:

$$L = \prod_{i=1}^{n} \left[\left(e^{-e^{\boldsymbol{x}_i'\boldsymbol{\alpha}}} \right)^{y_i^\beta} - \left(e^{-e^{\boldsymbol{x}_i'\boldsymbol{\alpha}}} \right)^{(y_i+1)^\beta} \right]^{1-\delta_{c_i}} \left[\left(e^{-e^{\boldsymbol{x}_i'\boldsymbol{\alpha}}} \right)^{C^\beta} \right]^{\delta_{c_i}} \tag{4.1}$$

The log-likelihood can be written as follows:

$$\ell = \sum_{i=1}^{n} (1-\delta_{c_i}) \log \left[\left(e^{-y_i^\beta e^{\boldsymbol{x}_i'\boldsymbol{\alpha}}} \right) - \left(e^{-(y_i+1)^\beta e^{\boldsymbol{x}_i'\boldsymbol{\alpha}}} \right) \right] - C^\beta \sum_{i=1}^{n} \delta_{c_i} e^{\boldsymbol{x}_i'\boldsymbol{\alpha}} \tag{4.2}$$

4.3 Maximum likelihood estimation

The parameters of the CDW are estimated in this section using the maximum likelihood approach. Thus, the partial derivatives of the log-likelihood in Equation 4.2 with respect to each unknown parameter are found and then set to zero. Hence, we obtain the following non-linear equations,

The first partial derivative of ℓ with respect to parameter β is obtained as follows:

$$\frac{\partial \ell}{\partial \beta} = \sum_{i=1}^{n} \frac{-e^{\boldsymbol{x}_i'\boldsymbol{\alpha}}(1-\delta_{c_i})}{wc_i(\beta, \boldsymbol{\alpha})} \left[y_i^\beta e^{-y_i^\beta e^{\boldsymbol{x}_i'\boldsymbol{\alpha}}} \log(y_i) - (y_i+1)^\beta e^{-(y_i+1)^\beta e^{\boldsymbol{x}_i'\boldsymbol{\alpha}}} \log(y_i+1) \right] - C^\beta \log(C) \sum_{i=1}^{n} \delta_{c_i} e^{\boldsymbol{x}_i'\boldsymbol{\alpha}}$$

The first partial derivative of ℓ with respect to parameter α_p is obtained as follows:

$$\frac{\partial \ell}{\partial \alpha_p} = \sum_{i=1}^{n} \frac{-x_{ip} e^{x_{ip}\alpha_p}(1-\delta_{c_i})}{wc_i(\beta, \alpha_p)} \left[y_i^\beta e^{-y_i^\beta e^{x_{ip}\alpha_p}} - (y_i+1)^\beta e^{-(y_i+1)^\beta e^{x_{ip}\alpha_p}} \right] - C^\beta \sum_{i=1}^{n} \delta_{c_i} x_{ip} e^{x_{ip}\alpha_p}$$

where

$$wc_i(\beta, \boldsymbol{\alpha}) = e^{-y_i^\beta e^{\boldsymbol{x}_i'\boldsymbol{\alpha}}} - e^{-(y_i+1)^\beta e^{\boldsymbol{x}_i'\boldsymbol{\alpha}}}$$

As can be seen, the equations are not in closed form, that is, the system of these likelihood equations does not have an analytic solution. Therefore, an iterative numerical method is required to find the numerical solutions of these equations to yield the MLEs $\hat{\beta}_{ML}, \hat{\boldsymbol{\alpha}}_{ML}$.

Then, the MLEs for the unknown parameters $\boldsymbol{\alpha}$ and β can be obtained by maximizing the log-likelihood (Equation 4.2) numerically using any standard optimization tool, such as *optim* in (R Core Team (2014)).

4.4 Simulation study

A simulation study is conducted in order to compare the full model, DW in Equation 3.3, with the CDW in Equation 4.1 for censored data. A multiple regression with two covariates is considered. The parameters that need to be investigated in this simulation include the set of regression coefficients $(\alpha_0, \alpha_1, \alpha_2)$ and β, the parameter for the DW.

In chapter 2, the following was reported:

- If $0 < \beta \leq 1 \Rightarrow$ over-dispersion.
- If $\beta \geq 2 \Rightarrow$ under-dispersion.

Thus, two values for the parameter β are selected to represent cases of under-dispersion and over-dispersion, respectively, as, $\beta_1 = 2$ and $\beta_2 = 0.8$. Additionally, the regression parameters are fixed for both cases to be $(\alpha_0 = -2, \alpha_1 = 0.5, \alpha_2 = 0.3)$. Furthermore, the covariates X_1 and X_2 are generated from $unif(0, 1.5)$ and $N(0, 1)$, respectively.

In this simulation study, different sample sizes are considered, specifically, $n_1 = 370$, $n_2 = 500$ and $n_3 = 1200$. Additionally, different censoring constants, C, are assumed. Then, using the above parameter vector and the corresponding independent variables, X_1 and X_2, the sample $y_1, y_2, \ldots, y_n$ is generated from the DW regression model in Equation 3.3, with the following parameterss (q_i, β):

$$q_i = e^{-e^{-2+0.5X_{1i}+0.3X_{2i}}} \tag{4.3}$$

The results of this study were based on 1000 repetitions for the simulation, in which for each iteration, a new sample Y is simulated and the $\boldsymbol{X}$ and parameter vector are fixed.

Three fittings are conducted, as follows:

- *complete*: in this fitting, the complete sample Y is considered and modeled by the DW regression model in Equation 3.3, that is, equivalently to chapter 3.

  Then, to investigate the censoring on a sample, a censored point is considered to cut this simulated sample, in which all the values $y_i \geq C$ are re-valued to be equal to C. Then, two cases are examined as follows:

- *truncated*: in this case, the censored sample $y_i = 0, 1, \ldots, C$ is assumed to be the complete sample and fitted by the standard DW regression model in Equation 3.3, without any consideration for the censoring.

- *censored*: here, the developed model CDW in Equation 4.1 is considered for analyzing the new censored sample $y_i = 0, 1, \ldots, C$.

In each iteration and for each fitting, the parameters $\alpha_0, \alpha_1, \alpha_2$ and β are estimated using the maximum likelihood method, maximizing the log-likelihood in Equation 3.8 for the complete and truncated data and maximizing Equation 4.2 for the censored case. Subsequently, the length of the 95% CIs for these estimated parameters are calculated. In addition, the goodness of fit measurement, AIC, is computed for each fitting in each iteration. Furthermore, the percentage of censored Y observations are found for each iteration.

Afterwards, the 1000 values of the MLEs, CIs, goodness-of-fit measurements and censoring percentages are found. Consequently, these MLEs are averaged, their bias and MSEs are computed, in addition to the length of the 95% CIs are calculated, and all these results are reported in Table 4.1 and Table 4.2 for under- and over-dispersion, respectively.

It is observed from Table 4.1 and Table 4.2 that fitting the complete data using the standard DW regression model and modeling the censored data by the CDW regression model, i.e. the censored and complete cases are the best in terms of bias. However, if the censoring is not taken into account and modeled

Table 4.1: MLEs based on the simulation study for the CDW regression model with true parameters $\boldsymbol{\alpha} = (-2, 0.5, 0.3)$ *and* $\beta = 2$*, for the under-dispersion case.*

			MLE (Bias),(MSE),(Length)			goodness of fit	
n	model	C %	α_0	α_1	α_2	β	AIC
370	truncated	3 18.7984 %	-2.1143 (-0.1143),(0.0415),(0.628)	0.4163 (-0.0837),(0.0209),(0.5743)	0.2566 (-0.0434),(0.0049),(0.2926)	2.3112 (0.3112),(0.1085),(0.4387)	1009.784
	censored	-	-2.0088 (-0.0088),(0.0289),(0.6403)	0.4977 (-0.0023),(0.0193),(0.5349)	0.3067 (0.0067),(0.0042),(0.2518)	2.0134 (0.0134),(0.0136),(0.4487)	965.9587
	truncated	4 6.1722%	-2.0434 (-0.0434),(0.0284),(0.6108)	0.4678 (-0.0322),(0.0161),(0.521)	0.2866 (-0.0134),(0.0034),(0.2538)	2.1036 (0.1036),(0.0199),(0.3862)	1064.163
	censored	-	-2.0113 (-0.0113),(0.0268),(0.6143)	0.4986 (-0.0014),(0.0168),(0.5023)	0.3065 (0.0065),(0.0038),(0.2367)	2.0168 (0.0168),(0.0107),(0.3912)	1049.673
	full	-	-2.0145 (-0.0145),(0.0257),(0.605)	0.4980 (-0.0020),(0.0158),(0.4905)	0.3053 (0.0053),(0.0036),(0.2301)	2.0222 (0.0222),(0.0097),(0.3647)	1084.161
500	truncated	3 18.8506 %	-2.1101 (-0.1101),(0.0331),(0.5442)	0.4118 (-0.0882),(0.0181),(0.5164)	0.2539 (-0.0461),(0.0044),(0.2611)	2.31 (0.31),(0.1048),(0.3769)	1363.477
	censored	-	-2.0070 (-0.0070),(0.0217),(0.557)	0.4983 (-0.0017),(0.0148),(0.4682)	0.3030 (0.0030),(0.0032),(0.2197)	2.0116 (0.0116),(0.0102),(0.3855)	1304.625
	truncated	4 6.2494%	-2.0391 (-0.0391),(0.0212),(0.5294)	0.4663 (-0.0337),(0.0125),(0.4602)	0.2837 (-0.0163),(0.0027),(0.2237)	2.0992 (0.0992),(0.0168),(0.331)	1438.211
	censored	-	-2.0081 (-0.0081),(0.0200),(0.5331)	0.5000 (0.0000),(0.0129),(0.4382)	0.3027 (0.0027),(0.0028),(0.207)	2.0114 (0.0114),(0.0081),(0.3354)	1418.633
	full	-	-2.0104 (-0.0104),(0.0190),(0.5244)	0.4995 (-0.0005),(0.0120),(0.4268)	0.3021 (0.0021),(0.0027),(0.2016)	2.0153 (0.0153),(0.007),(0.3119)	1465.99
1200	truncated	3 18.7651%	-2.1094 (-0.1094),(0.0211),(0.3485)	0.4233 (-0.0767),(0.0103),(0.3541)	0.2525 (-0.0475),(0.0031),(0.1755)	2.3028 (0.3028),(0.0955),(0.2428)	3254.126
	censored	-	-2.0052 (-0.0052),(0.0093),(0.3552)	0.5022 (0.0022),(0.0061),(0.2997)	0.3022 (0.0022),(0.0012),(0.1308)	2.005 (0.005),(0.0043),(0.2481)	3110.334
	truncated	4 6.3187%	-2.0389 (-0.0389),(0.0100),(0.339)	0.4726 (-0.0274),(0.0057),(0.3037)	0.2823 (-0.0177),(0.0012),(0.1411)	2.0947 (0.0947),(0.0119),(0.2135)	3432.122
	censored	-	-2.0052 (-0.0052),(0.0086),(0.3409)	0.5021 (0.0021),(0.0055),(0.2817)	0.3020 (0.0020),(0.0010),(0.1229)	2.0053 (0.0053),(0.0032),(0.2163)	3383.275
	full	-	-2.0056 (-0.0056),(0.0082),(0.3352)	0.5011 (0.0011),(0.0052),(0.2749)	0.3020 (0.0020),(0.0009),(0.1193)	2.0071 (0.0071),(0.0028),(0.2006)	3502.031

Table 4.2: MLEs based on the simulation study for the CDW regression model with true parameters $\boldsymbol{\alpha} = (-2, 0.5, 0.3)$ and $\beta = 0.8$, for the over-dispersion case.

			MLE (Bias),(MSE),(Length)			goodness of fit	
n	model	C %	α_0	α_1	α_2	β	AIC
370	truncated	20 13.4841%	-2.1355 (-0.1355),(0.0452),(0.6029)	0.4039 (-0.0961),(0.0213),(0.5677)	0.2481 (-0.0519),(0.0053),(0.2871)	0.9516 (0.1516),(0.0248),(0.1719)	2196.657
	censored	-	-2.0102 (-0.0102),(0.0265),(0.61)	0.5005 (0.0005),(0.0172),(0.5063)	0.3054 (0.0054),(0.0037),(0.2377)	0.8056 (0.0056),(0.0019),(0.1638)	1945.969
	truncated	30 6.9627%	-2.0735 (-0.0735),(0.0307),(0.5928)	0.4428 (-0.0572),(0.0165),(0.5294)	0.2705 (-0.0295),(0.0036),(0.2609)	0.8779 (0.0779),(0.0075),(0.1556)	2264.579
	censored	-	-2.0097 (-0.0097),(0.0254),(0.5962)	0.4993 (-0.0007),(0.0160),(0.4891)	0.3056 (0.0056),(0.0035),(0.2297)	0.8059 (0.0059),(0.0016),(0.1521)	2122.567
	full	-	-2.0135 (-0.0135),(0.0242),(0.5844)	0.4992 (-0.0008),(0.0148),(0.4739)	0.3045 (0.0045),(0.0032),(0.2216)	0.8082 (0.0082),(0.0014),(0.1396)	2324.01
500	truncated	20 13.6008%	-2.1328 (0.9516),(0.0375),(0.5228)	0.3984 (-0.1016),(0.0192),(0.5122)	0.2457 (-0.0543),(0.0050),(0.2581)	0.9516 (0.1516),(0.0244),(0.1478)	2971.03
	censored	-	-2.0074 (-0.0074),(0.0203),(0.5306)	0.4995 (-0.0005),(0.0131),(0.4433)	0.3030 (0.0030),(0.0028),(0.2077)	0.8041 (0.0041),(0.0014),(0.1408)	2629.767
	truncated	30 7.0422%	-2.0706 (-0.0706),(0.0241),(0.514)	0.4386 (-0.0614),(0.0137),(0.4724)	0.2685 (-0.0315),(0.0031),(0.2318)	0.8773 (0.0773),(0.0071),(0.1337)	3063.657
	censored	-	-2.0080 (-0.0080),(0.0197),(0.518)	0.4998 (-0.0002),(0.0124),(0.4273)	0.3026 (0.0026),(0.0026),(0.201)	0.8044 (0.0044),(0.0012),(0.1306)	2869.941
	full	-	-2.0102 (-0.0102),(0.0184),(0.5069)	0.4991 (-0.0009),(0.0113),(0.4128)	0.3020 (0.0020),(0.0024),(0.1943)	0.8061 (0.0061),(0.001),(0.1196)	3145.584
1200	truncated	20 13.5757%	-2.1300 (-0.1300),(0.0254),(0.3345)	0.4090 (-0.0910),(0.0121),(0.3571)	0.2443 (-0.0557),(0.0038),(0.1765)	0.9484 (0.1484),(0.0226),(0.0951)	7089.972
	censored	-	-2.0051 (-0.0051),(0.0085),(0.3386)	0.5020 (0.0020),(0.0055),(0.2838)	0.3015 (0.0015),(0.0010),(0.1235)	0.8021 (0.0021),(0.0006),(0.0907)	6268.277
	truncated	30 7.1082 %	-2.0700 (-0.0700),(0.0130),(0.3291)	0.4469 (-0.0531),(0.0070),(0.3194)	0.2667 (-0.0333),(0.0019),(0.1523)	0.8753 (0.0753),(0.0061),(0.0862)	7310.443
	censored	-	-2.0054 (-0.0054),(0.0082),(0.3311)	0.5021 (0.0021),(0.0052),(0.2742)	0.3017 (0.0017),(0.0010),(0.1193)	0.8022 (0.0022),(0.0005),(0.0843)	6838.239
	full	-	-2.0059 (-0.0059),(0.0078),(0.3241)	0.5013 (0.0013),(0.0049),(0.2656)	0.3015 (0.0015),(0.0009),(0.115)	0.8029 (0.0029),(0.0004),(0.0769)	7511.817

by the standard model in Equation 3.3, as in truncated cases, the estimates are highly biased, as shown in Table 4.1 and Table 4.2. In regards to the lengths of the CIs, again, the complete and censored fittings provide the shortest length for most of the cases. In addition, the censored models provide the best fit in comparison to other fittings, regarding its minimum AIC. Then, the MLEs from the censored models are much closer than those from the truncated models to their corresponding MLEs from the complete DW case. That is, analyzing the truncated (censored) data with the standard models without considering the censoring may result in misleading fittings.

4.5 Numerical examples

To demonstrate the application of the CDW regression model, it is applied in this section to different data sets that show various types of dispersions. Assorted popular censored count data regression models, namely, CP and CNB, are applied by maximizing the log of the likelihood functions in Equation 1.11 and Equation 1.12, respectively, using *"optim"* in (R Core Team (2014)). The MLEs of these models are compared with that of the CDW regression model by means of the classical AIC criterion. Additionally, the expected frequencies from each model are compared with the observed frequencies. The purpose here is just to demonstrate the CDW model and not to study the significance of the covariates. The results in the tables are for the regression coefficients affecting the parameter q, in Equation 3.2. The results here from censored models can be compared with their correspondents from the full (standard) models in chapter 3.

4.5.1 The case of under-dispersion: inhaler use data

The under-dispersed data relative to the Poisson model in subsection 3.9.1 are applied here. Some censored points are considered and the CDW is applied. The histogram in Figure 3.14 for the observed frequencies of the inhaler use data shows that about 9.33% of the count variables are greater than or equal 3, and about 1.80% are greater than or equal 4. Thus, two cut points are considered in this example, $C_1 = 3$ and $C_2 = 4$, to see the effect of censoring on the dataset. CP, CNB and CDW are applied, and the results are summarized in Table 4.3.

Table 4.4: Observed and expected frequencies for the full, censored and truncated: DW models for the inhaler dataset.

y	0	1	2	3	4	5	6
observed	972	2447	1304	392	86	5	3
full DW	1086.2	2228.0	1425.5	420.6	61.9	4.7	0.2
C=3							
censored	1033.4	2281.5	1448.4	445.6	-	-	-
truncated	986.3	2355.7	1477.2	354.6	-	-	-
C=4							
censored	1060.5	2237.1	1430.7	416.6	64.1	-	-
truncated	1052.7	2247.1	1436.2	411.9	57.1	-	-

4.5.2 The case of over-dispersion: strikes data

The dataset applied in subsection 3.9.2 is used here to show a case of over-dispersed data relative to the Poisson with censoring. It can be seen from the observed frequencies in Figure 3.19, that about 6.48% of the response variables are greater than or equal to 11, and about 2.78% more than or equal to 15. As an example, two censoring points are considered, $C_1 = 11$ and $C_2 = 15$, and the results are reported in Table 4.5.

Table 4.3: MLEs, SEs (in parentheses) and AIC from the truncated and censored: Poisson, NB and DW regression models fitted to the inhaler dataset.

			MLE (SE)					goodness of fit	
C %	model		intercept	humidity	pressure	temperature	particles	other	AIC
3 9.3300 %	CP	truncated	-2.2555 (1.7252)	-0.1117 (0.0847)	4.1279 (2.7449)	-0.1856 (0.1303)	0.0213 (0.0131)	-	13668.36
		censored	-2.1736 (1.7315)	-0.1361 (0.0848)	4.0781 (2.7552)	-0.2157 (0.1305)	0.0227 (0.0131)	-	13327.62
	CNB	truncated	-1.9313 (1.7247)	-0.1118 (0.0847)	3.6110 (2.7440)	-0.1899 (0.1303)	0.0211 (0.0131)	$\hat{k}$=35076.73 (35.5245)	13670.39
		censored	-1.8769 (1.7311)	-0.1356 (0.0847)	3.6044 (2.7545)	-0.2187 (0.1305)	0.0224 (0.0131)	$\hat{k}$=35111.74 (NaN)	13329.64
	CDW	truncated	-0.3525 (1.9967)	0.2573 (0.1021)	-2.3252 (3.1725)	0.3829 (0.1576)	-0.0271 (0.0158)	$\hat{\beta}$=2.2907 (0.0290)	13017.82
		censored	0.6256 (2.0522)	0.2769 (0.1047)	-3.8301 (3.2610)	0.3979 (0.1619)	-0.0292 (0.0163)	$\hat{\beta}$=2.1954 (0.0302)	12913.1
4 1.8046 %	CP	truncated	-2.2125 (1.7130)	-0.1189 (0.0841)	4.1013 (2.7254)	-0.2134 (0.1294)	0.0231 (0.0130)	-	13884.1
		censored	-2.1860 (1.7137)	-0.1218 (0.0841)	4.0716 (2.7266)	-0.2222 (0.1294)	0.0238 (0.0130)	-	13831.42
	CNB	truncated	-2.0473 (1.7127)	-0.1192 (0.0841)	3.8388 (2.7250)	-0.2159 (0.1294)	0.0229 (0.0130)	$\hat{k}$=35099.23 (NaN)	13886.16
		censored	-2.2042 (1.7137)	-0.1216 (0.0841)	4.1017 (2.7265)	-0.2225 (0.1294)	0.0236 (0.0130)	$\hat{k}$=35095.12 (63.1169)	13833.48
	CDW	truncated	-0.6591 (1.9844)	0.2642 (0.1017)	-1.7562 (3.1528)	0.4479 (0.1565)	-0.0308 (0.0156)	$\hat{\beta}$=2.1541 (0.0265)	13420.21
		censored	0.0640 (1.9915)	0.2612 (0.1020)	-2.8971 (3.1642)	0.4454 (0.1572)	-0.0314 (0.0157)	$\hat{\beta}$=2.1406 (0.0268)	13406.72

Table 4.5: MLEs, SEs (in parentheses) and AIC from the truncated and censored: Poisson, NB and DW regression models fitted to the strikes dataset.

			MLE (SE)		goodness of fit	
C %	model		intercept	economic activity	other	AIC
11 6.4815 %	CP	truncated	1.6167 (0.0430)	2.5793 (0.8135)	-	590.1407
		censored	1.6256 (0.0431)	2.6785 (0.8151)	-	582.1582
	CNB	truncated	1.6168 (0.0648)	2.6545 (1.2298)	$\hat{k}$=3.9459 (1.0441)	553.2221
		censored	1.6489 (0.0699)	3.0281 (1.3201)	$\hat{k}$=3.2308 (0.8077)	535.3486
	CDW	truncated	-3.2544 (0.3105)	-4.4024 (1.9057)	$\hat{\beta}$=1.7814 (0.1441)	549.501
		censored	-3.0942 (0.3038)	-4.8661 (1.9471)	$\hat{\beta}$=1.6735 (0.1419)	533.1672
15 2.7778 %	CP	truncated	1.6473 (0.0424)	3.0240 (0.8050)	-	619.8564
		censored	1.6500 (0.0424)	3.0635 (0.8058)	-	617.151
	CNB	truncated	1.6473 (0.0679)	3.1111 (1.2882)	$\hat{k}$=3.3127 (0.7871)	564.1837
		censored	1.6607 (0.0701)	3.3393 (1.3301)	$\hat{k}$=3.0453 (0.7097)	556.2384
	CDW	truncated	-3.1103 (0.2954)	-5.1258 (1.9109)	$\hat{\beta}$=1.6787 (0.1331)	561.4044
		censored	-3.0361 (0.2926)	-5.3884 (1.9246)	$\hat{\beta}$=1.6310 (0.1324)	554.2749

Table 4.6: Observed and expected frequencies for the full, censored and truncated: DW models for the strikes dataset.

y	0	1	2	3	4	5	6	7	8	9	10	11	13	15	16	18
observed	5	12	14	11	9	14	9	4	7	10	6	1	3	1	1	1
full DW	5.2	10.2	12.5	13.1	12.5	11.3	9.7	8.0	6.4	5.0	3.9	2.9	1.6	0.8	0.6	0.3
C=11																
censored	5.0	10.2	12.6	13.3	12.7	11.5	9.8	8.1	6.5	5.0	3.8	9.5	-	-	-	-
truncated	4.3	9.7	12.7	13.9	13.6	12.3	10.5	8.5	6.6	5.0	3.6	2.5	-	-	-	-
C=15																
censored	5.4	10.4	12.5	13.0	12.3	11.1	9.5	7.9	6.4	5.0	3.9	2.9	1.6	2.8	-	-
truncated	5.0	10.1	12.5	13.2	12.7	11.5	9.8	8.1	6.5	5.0	3.8	2.8	1.5	0.7	-	-

4.5.3 The case of excessive zeros: doctor visits from the German health survey data

The dataset in subsection 3.9.3 is considered to investigate the CDW for the too many zeros response case. It can be seen from Figure 3.24 that around 2.4% of the doctor visit numbers were greater than or equal to 13; additionally, around 1.9% were greater than or equal to 15. Thus, these two points, $C_1 = 13$ and $C_2 = 15$, are considered here as cut points to make the data censored from the right. These results are shown in Table 4.7.

Table 4.7: MLEs, SEs (in parentheses) and AIC from the truncated and censored: Poisson, NB and DW regression models fitted to the doctor visits from the German health dataset.

			MLE (SE)			goodness of fit	
C %	model		intercept	bad health	age	other	AIC
13 2.3957%	CP	truncated	0.5042 (0.0728)	1.0138 (0.0486)	0.0037 (0.0019)	-	5238.444
		censored	0.5057 (0.0728)	1.0273 (0.0487)	0.0036 (0.0019)	-	5221.056
	CNB	truncated	0.4701 (0.1256)	1.0109 (0.1074)	0.0046 (0.0033)	$\hat{k}$=1.1085 (0.0818)	4411.506
		censored	0.4670 (0.1302)	1.1380 (0.1171)	0.0050 (0.0034)	$\hat{k}$=1.0218 (0.0743)	4330.218
	CDW	truncated	-0.7486 (0.1107)	-0.9063 (0.1030)	-0.0042 (0.0028)	$\hat{\beta}$=1.0330 (0.0282)	4412.001
		censored	-0.7108 (0.1117)	-1.0033 (0.1096)	-0.0043 (0.0028)	$\hat{\beta}$=0.9995 (0.0281)	4330.194
15 1.8634%	CP	truncated	0.5060 (0.0722)	1.0439 (0.0478)	0.0039 (0.0019)	-	5359.644
		censored	0.5061 (0.0723)	1.0523 (0.0479)	0.0039 (0.0019)	-	5347.35
	CNB	truncated	0.4657 (0.1270)	1.0406 (0.1087)	0.0049 (0.0033)	$\hat{k}$=1.0651 (0.0768)	4435.997
		censored	0.4584 (0.1307)	1.1429 (0.1165)	0.0054 (0.0034)	$\hat{k}$=1.0001 (0.0714)	4370.781
	CDW	truncated	-0.7285 (0.1104)	-0.9232 (0.1031)	-0.0043 (0.0028)	$\hat{\beta}$=1.0164 (0.0275)	4436.323
		censored	-0.6948 (0.1112)	-0.9996 (0.1082)	-0.0045 (0.0028)	$\hat{\beta}$=0.9898 (0.0275)	4370.528

4.6 Concluding remarks

The right censored scheme is applied for the DW regression model. This modification explains a case where some censored point C is considered for all values of response count y greater than or equal to this point. Then, for this type of count data, the CDW regression model is developed in this study. Some simulation studies and numerical examples with different levels of dispersion have been studied to investigate the performance of the CDW.

This developed CDW regression model can be compared with the full DW model,

Table 4.8: Observed and expected frequencies for the full, censored and truncated: DW models for the doctor visits data from the German health dataset.

y	0	1	2	3	4	5	6	7	8	9	10	11	12	13	14	15	16	20	30	40
observed	360	222	191	135	54	47	39	8	14	3	20	1	6	5	1	11	1	6	2	1
full DW	366.9	240.0	159.9	107.6	73.0	50.0	34.6	24.2	17.1	12.3	9.0	6.7	5.0	3.8	3.0	2.4	1.9	0.9	0.2	0.0
C=13																				
censored	364.3	242.0	161.6	108.5	73.3	49.9	34.3	23.9	16.8	12.0	8.7	6.5	4.9	20.5	-	-	-	-	-	-
truncated	355.5	248.7	167.6	112.2	75.1	50.4	34.1	23.2	16.0	11.2	7.9	5.7	4.2	3.1	-	-	-	-	-	-
C=15																				
censored	366.6	240.2	160.0	107.6	72.9	49.9	34.5	24.1	17.1	12.3	8.9	6.6	5.0	3.9	3.0	14.5	-	-	-	-
truncated	359.4	245.4	164.8	110.6	74.5	50.4	34.3	23.6	16.5	11.6	8.3	6.1	4.5	3.4	2.6	2.0	-	-	-	-

in which the DW is applied to the complete data sets without censoring. Then, some censored points are chosen to cut the data from the right. Consequently, the full DW model has been considered to fit this censored dataset, without taking the censoring into account, which is referred to as truncated. Next, the censored data is modeled using the CDW regression model.

In the simulation study, the MLEs for the censored model are very similar to the corresponding MLEs from the full model. Also, from the numerical applications, the AIC from the CDW is smaller than that from the CP and CNB. In terms of the expected frequencies, the CDW provides expected frequencies much closer to the observed in comparison to the truncated fitting, especially for the censored points. On the other hand, applying the full DW model for this case of censored data provides a poor fitting. In other words, in the case with censored data, ignoring the censoring and analyzing it with the standard (full) model may result in misleading estimates. Thus, if there is censoring in the data, a censored model should be applied in the analysis. In such cases, the CDW may provide better results.

Chapter 5

Discrete Weibull Regression with Excess Zero Counts

5.1 Introduction

In this chapter, the issues with the too many zeros response are investigated. Some studies may be more interested in predicting the frequency of zeros, and hence additional care is required to choose the applied model. It seems that the counts in these experiments are generated by two different mechanisms for zeros and non-zeros in the data. Although the DW model can be considered as a good model to fit data with a zero-inflated count, a modified method should be applied when the aim is to distinguish between zero and non-zero data-generating processes, which can be explained using mixture models, especially zero-inflated and hurdle models.

Even though the ZIP, ZINB, HP and HNB are the most commonly applied models for this case of zero-inflation in the response variable, these models are not ideal if the data presents under-dispersion. This is because the Poisson and NB are mainly applied for equi- and over-dispersion cases. However, some data appear to be equi- or over-dispersed, but in fact this may be a mixture of different levels of dispersion. This case is common with zero-inflated data; while this excess of zeros increase the over-dispersion for the data and could hide the fact that in some cases the data are under-dispersed. In other words, the overall structure of the dispersion could be different than the dispersion pattern within the subpopulations and non-zero counts (Tin (2008)). For instance, Sáez-Castillo and Conde-Sánchez (2015) considered the zero-inflated hyper-Poisson, which can handle the over- and under-dispersion within these subclasses. The case of zero inflation with the potential of a mixed level of dispersion within the data is explained further in section 5.2. Then, if a one-part model would be applied for the whole data, a model that can cope with the over-dispersion is suggested to apply. Whereas, if the interest is to have a two-part model, which is one part for zeros and the other for the non-zeros, it is important to apply a zero-inflation (two-part) model with a count modeling process that has the ability to reflect different cases of dispersion. Hence, the DW model is considered and two extensions for this model, ZIDW and HDW, are presented for the case of excessive zeros.

Moreover, this chapter develops two modifications of this ZIDW and HDW, namely, CZIDW and CHDW to consider a case of censored count response with too many zeros. Additionally, censoring from the right might reduced the over-dispersion in the data caused by containing too many zeros.

5.2 Zero-inflated discrete Weibull

Considering the DW distribution in Equation 2.2 as a parent model with a probability of zero $f_p(0) = 1 - q$, that from Equation 1.15 the likelihood of the ZIDW can be found as follows:

$$L = \prod_{i=1}^{n} \left[\pi_i + (1-\pi_i)(1-q_i)\right]^{\delta_{z_i}} \left[(1-\pi_i)\left(q_i^{y_i^{\beta}} - q_i^{(y_i+1)^{\beta}}\right)\right]^{1-\delta_{z_i}} \tag{5.1}$$

Then, using $\pi_i \equiv \pi(\boldsymbol{z}_i)$ in Equation 1.16 and $q_i \equiv q(\boldsymbol{x}_i)$ in Equation 3.2, the log-likelihood can be found as:

$$\begin{aligned} \ell = &\sum_{i=1}^{n} \delta_{z_i} \log\left\{\left(e^{-\boldsymbol{z}_i'\boldsymbol{\gamma}}+1\right)^{-1} + \left[1-\left(e^{-\boldsymbol{z}_i'\boldsymbol{\gamma}}+1\right)^{-1}\right]\left(1-e^{-e^{\boldsymbol{x}_i'\boldsymbol{\alpha}}}\right)\right\} + \\ &\sum_{i=1}^{n} (1-\delta_{z_i}) \log\left\{\left[1-\left(e^{-\boldsymbol{z}_i'\boldsymbol{\gamma}}+1\right)^{-1}\right]\left(e^{-y_i^{\beta}e^{\boldsymbol{x}_i'\boldsymbol{\alpha}}} - e^{-(y_i+1)^{\beta}e^{\boldsymbol{x}_i'\boldsymbol{\alpha}}}\right)\right\} \end{aligned} \tag{5.2}$$

where δ_z is defined in Equation 1.14.

The set of covariates $\boldsymbol{X}_i = \{x_{i1}, x_{i2}, \ldots, x_{iP_1}\}$ that affect q may or may not be the same as the covariates $\boldsymbol{Z}_i = \{z_{i1}, z_{i2}, \ldots, z_{iP_2}\}$ that affect π, where P_1 and P_2 are the number of covariates for q and π, respectively. However, in this study they are assumed to be the same, that is, $\boldsymbol{X} = \boldsymbol{Z}$ with $P_1 = P_2 = P$.

5.2.1 Maximum likelihood estimation

The parameters $(\beta, \boldsymbol{\alpha}, \boldsymbol{\gamma})$ can be estimated using the maximum likelihood method. The partial derivatives of the log-likelihood in Equation 5.2 with respect to each unknown parameter are found and then set to be equal to zero. That is for $p, l = 1, 2, \ldots, P$:

$$\begin{aligned}\frac{\partial \ell}{\partial \beta} = \sum_{i=1}^{n} & \frac{(\delta_{z_i} - 1)\, e^{\boldsymbol{x}_i' \boldsymbol{\alpha}} \left[1 - \left(e^{-\boldsymbol{z}_i' \boldsymbol{\gamma}} + 1\right)^{-1}\right]}{wz2_i(\beta, \boldsymbol{\alpha}, \boldsymbol{\gamma})} \times \\ & \left[y_i^{\beta} e^{-y_i^{\beta} e^{\boldsymbol{x}_i' \boldsymbol{\alpha}}} \log(y_i) - (y_i + 1)^{\beta}\, e^{-(y_i+1)^{\beta} e^{\boldsymbol{x}_i' \boldsymbol{\alpha}}} \log(y_i + 1)\right]\end{aligned} \tag{5.3}$$

$$\begin{aligned}\frac{\partial \ell}{\partial \alpha_p} = & \sum_{i=1}^{n} \frac{\delta_{z_i} \left[1 - \left(e^{-\boldsymbol{z}_i' \boldsymbol{\gamma}} + 1\right)^{-1}\right]}{wz1_i(\boldsymbol{\gamma}, \alpha_p)} \left[x_{ip} e^{x_{ip}\alpha_p - e^{x_{ip}\alpha_p}}\right] + \\ & \sum_{i=1}^{n} \frac{(\delta_{z_i} - 1)\, x_{ip} e^{x_{ip}\alpha_p} \left[1 - \left(e^{-\boldsymbol{z}_i' \boldsymbol{\gamma}} + 1\right)^{-1}\right]}{wz2_i(\beta, \boldsymbol{\gamma}, \alpha_p)} \left[y_i^{\beta} e^{-y_i^{\beta} e^{x_{ip}\alpha_p}} - (y_i + 1)^{\beta}\, e^{-(y_i+1)^{\beta} e^{x_{ip}\alpha_p}}\right]\end{aligned} \tag{5.4}$$

$$\begin{aligned}\frac{\partial \ell}{\partial \gamma_l} = & \sum_{i=1}^{n} \frac{\delta_{z_i} z_{il} e^{-z_{il}\gamma_l} \left(e^{-z_{il}\gamma_l} + 1\right)^{-2}}{wz1(\gamma_l, \boldsymbol{\alpha})} e^{-e^{\boldsymbol{x}_i' \boldsymbol{\alpha}}} + \\ & \sum_{i=1}^{n} \frac{(\delta_{z_i} - 1)\, z_{il} e^{-z_{il}\gamma_l} \left(e^{-z_{il}\gamma_l} + 1\right)^{-2}}{wz2_i(\beta, \gamma_l, \boldsymbol{\alpha})} \left[e^{-y_i^{\beta} e^{\boldsymbol{x}_i' \boldsymbol{\alpha}}} - e^{-(y_i+1)^{\beta} e^{\boldsymbol{x}_i' \boldsymbol{\alpha}}}\right]\end{aligned} \tag{5.5}$$

with

$$\begin{aligned} wz1_i(\boldsymbol{\gamma}, \boldsymbol{\alpha}) &= \left(e^{-\boldsymbol{z}_i' \boldsymbol{\gamma}} + 1\right)^{-1} + \left[1 - \left(e^{-\boldsymbol{z}_i' \boldsymbol{\gamma}} + 1\right)^{-1}\right] \left(1 - e^{-e^{\boldsymbol{x}_i' \boldsymbol{\alpha}}}\right) \\ wz2_i(\beta, \boldsymbol{\gamma}, \boldsymbol{\alpha}) &= \left[1 - \left(e^{-\boldsymbol{z}_i' \boldsymbol{\gamma}} + 1\right)^{-1}\right] \left(e^{-y_i^{\beta} e^{\boldsymbol{x}_i' \boldsymbol{\alpha}}} - e^{-(y_i+1)^{\beta} e^{\boldsymbol{x}_i' \boldsymbol{\alpha}}}\right)\end{aligned} \tag{5.6}$$

The above equations are not in closed form, and their system do not have an analytic solution. Therefore, a numerical method is required to find the numerical solution of the log-likelihood equation in Equation 5.2, to yield the MLEs, $\hat{\beta}_{ML}, \hat{\boldsymbol{\alpha}}_{ML}$ and $\hat{\boldsymbol{\gamma}}_{ML}$.

5.2.2 Simulation study

There are two objectives in the following simulation study. First, it shows that the excessive zero data might exhibit a different pattern in regard to data dispersion. In addition, it evaluates the performance of the MLEs for the ZIDW regression model.

Different sample sizes $n_1 = 70$, $n_2 = 150$ and $n_3 = 500$ and $n_4 = 1000$ are considered. A single regression with one predictor is examined. As mentioned before, this study assumes that a similar set of covariates affect both q and π, that is, $\boldsymbol{X}$ is equivalent to $\boldsymbol{Z}$. All the results are based on an average over 1000 repetitions. In each iteration, MLEs and the asymptotic two-sided CIs are computed using “optim” in R. The simulation follows these steps:

- **Step 1:** Generate random samples with size n to present the covariate from the uniform distribution with parameters $(0, 1.5)$.
- **Step 2:** The true values of the parameters are chosen to be:
  - The regression parameters are assumed to be as follows:

$$\boldsymbol{\alpha} = (\alpha_0, \alpha_1) = (-2, -1.7)$$

$$\boldsymbol{\gamma} = (\gamma_0, \gamma_1) = (1.5, -0.9)$$

  - The shape parameter β of the DW is supposed to be $\beta = 2.2$.
  - Then, q and π can be calculated for each $\boldsymbol{X}$ respectively, as in Equation 3.2 and Equation 1.16.
- **Step 3:** Use the following as initial values:

- ◇ $(-\alpha)$ from the Poisson regression fitting, for (α)
- ◇ $(-\alpha)$ from the Poisson regression fitting, for (γ)
- ◇ The MLE for (β) from fitting Y using the unconditional DW distribution, for (β)

Then, for each sample size n, the simulations are conducted, in which for each iteration $(1 : 1000)$ the following is done:

- – generate a random sample from the population whose pmf is given by Equation 5.1, as follows:
  - ∗ Simulate a random number U from uniform distribution $U(0, 1)$. Then,
  - ∗ If $U \leq \pi_i$,
    set $y_i = 0$, otherwise
    generate y_i from DW with parameters q_i and β.
- – Fit this data by ZIP and ZINB using the *"pscl"* package in R (Jackman (2008)), and find their AICs.
- – Using the initial values discussed above, the MLE$_{itr}$ of the parameters $\boldsymbol{\alpha}$, $\boldsymbol{\gamma}$ and β, denoted as $\hat{\theta}_{itr}$, is computed by maximizing the log-likelihood function in Equation 5.2, using *"optim"* in R.
- – In addition to the MLEs_{itr}, the lower-limit LL_{itr} and upper-limit UL_{itr} of the 95% CIs for each MLE are determined.

- **Step 4:** The above three steps are repeated 1000 times.

To investigate the possibility of the data experiencing different levels of dispersion, a sample from one of the iterations for $n = 500$ is considered. Then, Figure 5.1 shows that for a simulated zero-inflated sample that the overall pattern for the relation between the observed mean and variance can be different than the pattern for its subgroup. In other words, the left figure shows the overall dispersion, that is, when all cases are considered, seems to be over-dispersed relative to the Poisson. From the right plot, it can be seen that in the cases with at least one response, that is, where the zeros are not included, then the dispersion structure would be different.

Subsequently, 1000 values of the MLEs and their CIs bounds are found. The average of these values is computed to obtain the MLEs of the unknown parameters. Furthermore, the average of each lower and upper bound for the CIs are calculated. Consequently, the lengths of these asymptotic CIs are found. Additionally, the average of the AICs for each model is found. Then, the estimators are evaluated using the bias and MSE, respectively. These measurements show a good behavior for the MLEs for the ZIDW. In addition, the best fitting among the ZIP, ZINB and ZIDW models can be regarded as the one with the minimum AIC. These results are displayed in Table 5.1.

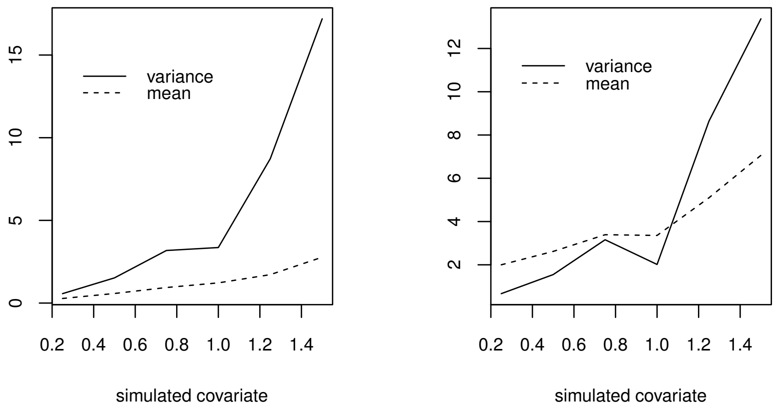


Figure 5.1: The overall pattern for the relation between the observed mean and variance for a sample simulated by ZIDW regression model (on the left), and the pattern for this sample's subgroup (on the right).

Table 5.1: MLEs based on the simulation study for the ZIDW regression model with true parameters $\boldsymbol{\alpha} = (-2, -1.7), \boldsymbol{\gamma} = (1.5, -0.9)$ *and* $\beta = 2.2$.

n	parameter	MLE	Bias	MSE	Length
70	α_0	-2.2017	0.2017	0.715	2.9911
	α_1	-1.8957	0.1957	0.5432	2.4727
	γ_0	1.1159	0.3841	0.2081	2.4272
	γ_1	-0.6466	-0.2534	0.1096	2.5337
	β	2.4428	-0.2428	0.3275	1.8732
ZIP AIC=194.1823		ZINB AIC=195.099		ZIDW AIC=194.1542	
150	α_0	-2.0587	0.0587	0.3479	2.1623
	α_1	-1.8339	0.1339	0.2418	1.7536
	γ_0	1.6062	-0.1062	0.0317	1.673
	γ_1	-1.0257	0.1257	0.0303	1.7246
	β	2.3176	-0.1176	0.1371	1.2905
ZIP AIC=372.9154		ZINB AIC=372.8397		ZIDW AIC=371.8752	
500	α_0	-2.0341	0.0341	0.1014	1.1975
	α_1	-1.739	0.039	0.0612	0.9662
	γ_0	1.6323	-0.1323	0.0234	0.9199
	γ_1	-0.9653	0.0653	0.0088	0.9666
	β	2.244	-0.044	0.0346	0.7056
ZIP AIC=1177.069		ZINB AIC=1174.391		ZIDW AIC=1172.021	
1000	α_0	-2.0208	0.0208	0.0396	0.7827
	α_1	-1.7173	0.0173	0.0254	0.6288
	γ_0	1.4701	0.0299	0.0037	0.6095
	γ_1	-1.001	0.1014	0.0125	0.6568
	β	2.2231	-0.0231	0.0148	0.4693
ZIP AIC=2528.186		ZINB AIC=2520.54		ZIDW AIC=2514.935	

5.3 Hurdle discrete Weibull

From Equation 1.19 with $f_p(y)$ is the DW, and the probability of Y being zero is $f_p(0) = 1 - q$, then, the zero truncated DW can be defined as $\dfrac{q^{y^\beta} - q^{(y+1)^\beta}}{q}$. Thus, from Equation 1.20, the response variable Y in the HDW has the following likelihood:

$$L = \prod_{i=1}^{n} [\pi_i]^{\delta_{z_i}} [1 - \pi_i]^{1-\delta_{z_i}} \prod_{i=1}^{n} \left[\frac{q_i^{y_i^\beta} - q_i^{(y_i+1)^\beta}}{q_i} \right]^{1-\delta_{z_i}} \tag{5.7}$$

where, $q_i \equiv q(\boldsymbol{x}_i)$ and $\pi_i \equiv \pi(\boldsymbol{z}_i)$ are related to some covariates, and these sets may or may not be the same. For this study, we make the same assumptions as for the ZIDW, where, $\boldsymbol{X} = \boldsymbol{Z}$ with a different regression parameter, as in

Equation 3.2 and Equation 1.16.

Then, the log-likelihood can be found as follows:

$$\ell = \ell_1(\gamma) + \ell_2(\beta, \alpha) \tag{5.8}$$

where

$$\ell_1(\gamma) = \sum_{i=1}^{n} (1 - \delta_{z_i}) \log \left[1 - \left(e^{-\boldsymbol{z}_i' \boldsymbol{\gamma}} + 1 \right)^{-1} \right] - \sum_{i=1}^{n} \delta_{z_i} \log \left(e^{-\boldsymbol{z}_i' \boldsymbol{\gamma}} + 1 \right)$$

$$\ell_2(\beta, \alpha) = \sum_{i=1}^{n} (1 - \delta_{z_i}) \log \left[e^{-y_i^\beta e^{\boldsymbol{x}_i' \boldsymbol{\alpha}}} - e^{-(y_i+1)^\beta e^{\boldsymbol{x}_i' \boldsymbol{\alpha}}} \right] + \sum_{i=1}^{n} (1 - \delta_{z_i})\, e^{\boldsymbol{x}_i' \boldsymbol{\alpha}}$$

Thus, it can be seen that the log-likelihood for the binary process, $\ell_1(\gamma)$, can be specified independently of the log-likelihood for the truncated count model, $\ell_2(\beta, \alpha)$.

5.3.1 Maximum likelihood estimation

The estimation of the parameters $(\beta, \boldsymbol{\alpha}, \boldsymbol{\gamma})$ can be obtained using the maximum likelihood method. Thus, the partial derivatives of the log-likelihood, following Equation 5.8, are obtained for each parameter. The first partial derivative of ℓ with respect to parameter β is:

$$\frac{\partial \ell}{\partial \beta} = \sum_{i=1}^{n} \frac{(\delta_{z_i} - 1)\, e^{\boldsymbol{x}_i' \boldsymbol{\alpha}}}{wh2(\beta, \boldsymbol{\alpha})} \left[y_i^\beta e^{-y_i^\beta e^{\boldsymbol{x}_i' \boldsymbol{\alpha}}} \log(y_i) - (y_i + 1)^\beta\, e^{-(y_i+1)^\beta e^{\boldsymbol{x}_i' \boldsymbol{\alpha}}} \log(y_i + 1) \right]$$

The first partial derivative of ℓ with respect to parameter $\boldsymbol{\alpha}$ is:

$$\frac{\partial \ell}{\partial \alpha_p} = \sum_{i=1}^{n} \frac{(\delta_{z_i} - 1)\, x_{ip} e^{x_{ip} \alpha_p}}{wh2(\beta, \alpha_p)} \left[y_i^\beta e^{-y_i^\beta e^{x_{ip} \alpha_p}} - (y_i + 1)^\beta\, e^{-(y_i+1)^\beta e^{x_{ip} \alpha_p}} \right] + \sum_{i=1}^{n} (1 - \delta_{z_i})\, x_{ip} e^{x_{ip} \alpha_p}$$

The first partial derivative of ℓ with respect to parameter $\boldsymbol{\gamma}$ is:

$$\frac{\partial \ell}{\partial \gamma_l} = \sum_{i=1}^{n} \frac{(\delta_{z_i} - 1)\, z_{il} e^{-z_{il} \gamma_l} \left(wh1(\gamma_l) \right)^{-2}}{1 - \left(wh1(\gamma_l) \right)^{-1}} + \sum_{i=1}^{n} \frac{\delta_{z_i} z_{il} e^{-z_{il} \gamma_l}}{wh1(\gamma_l)}$$

where

$$\begin{aligned} wh1_i(\gamma) &= e^{-\boldsymbol{z}_i'\boldsymbol{\gamma}} + 1 \\ wh2_i(\beta, \alpha) &= e^{-y_i^{\beta} e^{\boldsymbol{x}_i'\boldsymbol{\alpha}}} - e^{-(y_i+1)^{\beta} e^{\boldsymbol{x}_i'\boldsymbol{\alpha}}} \end{aligned} \tag{5.9}$$

These derivatives cannot be obtained analytically, and a numerical tool is required to obtain the MLEs of the unknown parameters. Optimizing the log-likelihood in Equation 5.8 can be considered. Then, the joint log-likelihood, ℓ, can be optimized by maximizing its two parts separately. Hence, using any of the standard optimization tools, the MLEs for the unknown parameters $\boldsymbol{\gamma}$ can be obtained by numerically maximizing the log-likelihood $\ell_1(\gamma)$. Additionally, by numerically maximizing the log-likelihood $\ell_2(\beta, \alpha)$, the MLEs for $\boldsymbol{\alpha}$ and β can be obtained.

5.3.2 Simulation study

A simulation study, is conducted to evaluate the performance of the MLEs of the HDW regression model. The same assumptions as in the ZIDW simulation study are considered here for HDW. Different sample sizes and a single regression with one predictor are considered. In addition, $\boldsymbol{X}$ and $\boldsymbol{Z}$ are the same. All the results are based on an average over 1000 repetitions. In each iteration, MLEs and the asymptotic two-sided CIs are computed using "optim" in R. Following McDowell et al. (2003), the simulation is applied as follows:

- **Step 1:** Generate a random samples with size n to present the covariate from the uniform distribution with parameters $(0, 1.5)$.

- **Step 2:** The true values of the parameters are assumed to be as mentioned earlier for the ZIDW. Then, q and π can be calculated for each $\boldsymbol{X}$ respectively, as in Equation 3.2 and Equation 1.16.

- **Step 3:** Using the true values in **Step 2** and for each sample size n, the simulation is conducted, and for each iteration $(1 : 1000)$,

  – generate random samples from the population whose pmf is given by Equation 5.7, as follows:

    - $*$ Simulate a sample from the zero-truncated DW, called *trunc*, in such a way that a sample from DW, with the parameters (q_i, β), is generated. Then, if a zero is simulated, drop it and re-sample again until a non-zero sample is generated.
    - $*$ Generate a random sample, called *bern*, from a Bernoulli, with the parameter $(1 - \pi)$.
    - $*$ Then, the sample from HDW, Y, can be repopulated, in which $y = 0$ if $bern = 0$; otherwise, $y = trunc$.
  - Fit this data by HP using the *"pscl"* package in R (Jackman (2008)).
  - Minus the regression coefficients from the HP model are assumed to be the initial values for $\boldsymbol{\alpha}$, $\boldsymbol{\gamma}$, and the MLE for β from the unconditional DW distribution as initial value for β. Then, the MLEs_{itr} of the parameters $\boldsymbol{\alpha}$, $\boldsymbol{\gamma}$ and β, which are denoted as $\hat{\theta}_{itr}$, are computed by maximizing the two log-likelihood functions in Equation 5.8 separately, using "optim" in R.
  - In addition to the MLEs_{itr}, the lower-limit LL_{itr} and upper-limit UL_{itr} of the 95% CIs for each MLE are determined.

- **Step 4:** The three steps above are repeated 1000 times.

Subsequently, 1000 values of the MLEs and their CIs bounds are found. The average of these values is computed to obtain the MLEs of the unknown parameter. Furthermore, the average of each lower and upper bound for the CI is calculated. Consequently, the lengths of these asymptotic CIs are found.

Then, the estimators are evaluated using the bias and MSE, respectively. Additionally, the average of the AICs for HP, HNB and HNB are found. The best fitting model is the one with the minimum AIC. The results are displayed in Table 5.2, showing a good performance for the MLEs.

Table 5.2: MLEs based on the simulation study for the HDW regression model with true parameters $\boldsymbol{\alpha} = (-2, -1.7), \boldsymbol{\gamma} = (1.5, -0.9)$ *and* $\beta = 2.2$.

n	parameter	MLE	Bias	MSE	Length
70	α_0	-2.2086	0.2086	0.8831	3.2912
	α_1	-1.9114	0.2114	0.6963	2.7456
	γ_0	1.5634	-0.0634	0.4455	2.4837
	γ_1	-0.9573	0.0573	0.4871	2.6001
	β	2.4654	-0.2654	0.3861	1.9701
HP AIC=184.2689		HNB AIC=185.2339		HDW AIC=184.3379	
150	α_0	-2.11	0.11	0.3138	2.0883
	α_1	-1.8074	0.1074	0.2032	1.6854
	γ_0	1.5256	-0.0256	0.1681	1.5455
	γ_1	-0.9199	0.0199	0.1896	1.6403
	β	2.3333	-0.1333	0.1335	1.282
HP AIC=381.9166		HNB AIC=382.0269		HDW AIC=380.7658	
500	α_0	-2.0387	0.0387	0.0856	1.118
	α_1	-1.7318	0.0318	0.0571	0.8975
	γ_0	1.4993	0.0007	0.046	0.8403
	γ_1	-0.8981	-0.0019	0.0524	0.909
	β	2.2424	-0.0424	0.0348	0.6768
HP AIC=1254.709		HNB AIC=1251.69		HDW AIC=1248.902	
1000	α_0	-2.0098	0.0098	0.0405	0.783
	α_1	-1.7128	0.0128	0.0259	0.6272
	γ_0	1.5004	-0.0004	0.021	0.5792
	γ_1	-0.8983	-0.0017	0.0258	0.6366
	β	2.2151	-0.0151	0.016	0.4786
HP AIC=2471.745		HNB AIC=2464.401		HDW AIC=2458.939	

5.4 Numerical examples

In this section, a data with a too many zeros response are analyzed using the zero-inflated and hurdle models. For the Poisson and NB, their zero-inflated and hurdle models are applied using the *pscl* package.

5.4.1 Fish data

This dataset was collected by state wildlife biologists, who were interested in studying the number of fish caught by fisherman in a particular park. The data is available at *http://www.ats.ucla.edu/stat/r/dae/zipoisson.htm* and have been analyzed in some articles, including Saffari and Adnan (2011) and Saffari et al. (2012). In this study, 250 visitors were asked whether or not they brought a camper on the visit (*camper=1* or *camper=0*), how many persons took part in the visit (*persons*), how many children took part in the visit (*child*) and how many fish were caught (*count*).

The effect of the three predictors, *camper*, *persons*, and *child*, on the response variable *count* is investigated using ZIP, ZINB, ZIDW, HP, HNB and HDW. This is due to the excessive zeros in the count response variable since few visitors caught any fish. The MLEs for the parameters are shown in Table 5.3 and Table 5.4.

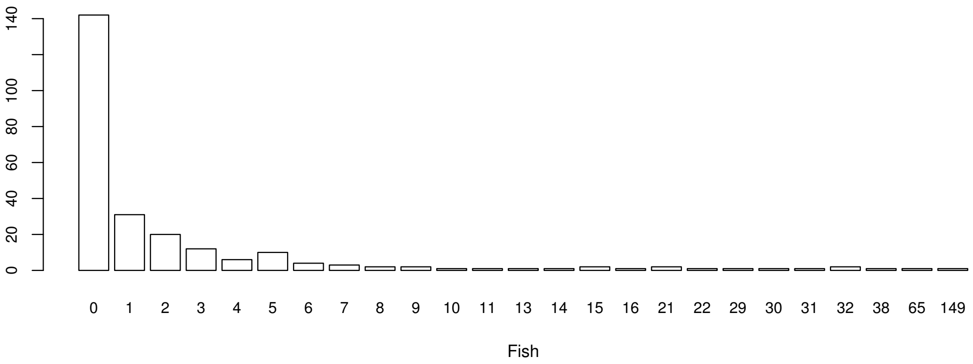


Figure 5.2: Histogram for the observed frequencies for the fish dataset.

Table 5.3: MLEs, SEs (in parentheses) and AIC from the zero-inflated: Poisson, NB and DW regression models for the fish dataset.

	intercept	camper	persons	child	other	AIC
logit-ZIDW						
count model	0.5400 (0.2709)	-0.1426 (0.2158)	-0.6828 (0.1029)	0.8736 (0.1975)	$\hat{\beta}$=0.8097 (0.0780)	814.2158
zero model	1.9025 (0.9813)	-2.4910 (1.1028)	-1.4571 (0.5662)	2.9433 (0.8568)	-	
logit-ZIP						
count model	-0.7983 (0.1708)	0.7243 (0.0931)	0.8290 (0.0440)	-1.1367 (0.0930)	-	1521.463
zero model	1.6636 (0.5155)	-0.8336 (0.3527)	-0.9228 (0.1992)	1.9046 (0.3261)	-	
logit-ZINB						
count model	-1.6177 (0.3202)	0.3856 (0.2461)	1.0901 (0.1117)	-1.2613 (0.2473)	$\log(\hat{k})$=-0.5929 (0.1580)	809.0788
zero model	-11.9920 (64.4408)	-10.7704 (64.3725)	0.2902 (0.7314)	10.9517 (64.3569)	-	

Table 5.5: Observed and expected frequencies for the standard, zero-inflated and hurdle: Poisson, NB and DW regression models for the fish dataset.

model	observed	DW	Poisson	NB	ZIDW	ZIP	ZINB	HDW	HP	HNB
0	142	144.8	94.6	139.5	146.0	142.1	141.6	142.0	142.0	142.0
1	31	33.4	56.1	38.1	26.5	16.2	33.9	34.6	16.8	33.2
2	20	17.2	28.7	18.6	16.2	15.9	18.2	17.8	16.0	17.7
3	12	10.7	15.9	11.1	11.0	13.0	11.2	11.0	12.8	11.1
4	6	7.3	9.9	7.3	8.0	9.9	7.6	7.5	9.7	7.7
5	10	5.3	6.7	5.2	6.0	7.5	5.4	5.5	7.4	5.6
6	4	4.0	4.9	3.9	4.7	5.9	4.1	4.1	5.7	4.3
7	3	3.2	3.9	3.0	3.8	4.7	3.2	3.3	4.7	3.4
8	2	2.6	3.3	2.4	3.1	4.0	2.6	2.6	3.9	2.7
9	2	2.1	2.9	2.0	2.6	3.5	2.1	2.2	3.5	2.2
10	1	1.8	2.6	1.7	2.2	3.1	1.8	1.8	3.1	1.9
11	1	1.5	2.2	1.4	1.9	2.8	1.5	1.6	2.8	1.6
13	1	1.1	1.3	1.1	1.4	2.0	1.2	1.2	2.0	1.2
14	1	1.0	0.9	0.9	1.2	1.6	1.0	1.0	1.6	1.1
15	2	0.9	0.6	0.8	1.1	1.2	0.9	0.9	1.3	0.9
16	1	0.8	0.5	0.7	1.0	1.0	0.8	0.8	1.0	0.8
21	2	0.5	0.5	0.5	0.6	0.8	0.5	0.5	0.8	0.5
22	1	0.4	0.6	0.4	0.5	0.8	0.5	0.4	0.8	0.5
29	1	0.3	0.9	0.2	0.3	0.7	0.3	0.3	0.8	0.3
30	1	0.2	0.8	0.2	0.3	0.6	0.3	0.2	0.6	0.2
31	1	0.2	0.8	0.2	0.3	0.5	0.2	0.2	0.5	0.2
32	2	0.2	0.6	0.2	0.2	0.4	0.2	0.2	0.4	0.2
38	1	0.1	0.1	0.1	0.2	0.1	0.2	0.1	0.1	0.2
65	1	0.0	0.0	0.0	0.0	0.0	0.0	0.0	0.0	0.0
149	1	0.0	0.0	0.0	0.0	0.0	0.0	0.0	0.0	0.0
AIC	-	805.6505	1682.145	820.444	814.2158	1521.463	809.0788	803.9418	1519.236	808.318

Table 5.4: MLEs, SEs (in parentheses) and AIC from the hurdle: Poisson, NB and DW regression models for the fish dataset.

	intercept	camper	persons	child	other	AIC
hurdle models						
binomial-logit model	2.3087 (0.4612)	-1.0179 (0.3246)	-1.1104 (0.1911)	2.1380 (0.3107)	-	-
HDW count model	1.1552 (0.3503)	-0.3201 (0.2100)	-0.6081 (0.1050)	0.7158 (0.2039)	$\hat{\beta}$=0.6193 (0.1053)	803.9418
HP count model	-0.8262 (0.1723)	0.7336 (0.0934)	0.8348 (0.0441)	-1.1390 (0.0929)	-	1519.236
HNB count model	-1.6215 (0.5960)	0.3746 (0.3360)	1.0029 (0.1551)	-1.0945 (0.3198)	$\log(\hat{k})$=-1.0530 (0.4974)	808.318

5.4.2 The case of a mixed level of dispersion: bids data

As mentioned earlier, the expected frequencies based on the DW model for the data in subsection 3.9.4 were not perfectly acurate. Thus, this data is analyzed again here using the zero-inflated and hurdle models.

Table 5.6: Observed and expected frequencies for the zero-inflated and hurdle: Poisson, NB and DW regression models for the bids dataset.

model	observed	ZIDW	ZIP	ZINB	HDW	HP	HNB
0	9	20.78	23.4	24.4	9.0	9.0	9.0
1	63	40.8	38.3	38.1	64.1	55.9	64.7
2	31	33.7	32.3	31.5	28.4	35.2	28.0
3	12	18.3	18.8	18.4	12.8	16.3	12.5
4	6	7.6	8.5	8.5	5.9	6.3	5.8
5	1	2.7	3.2	3.4	2.8	2.2	2.8
6	2	1.0	1.1	1.2	1.4	0.7	1.4
7	1	0.4	0.3	0.4	0.7	0.2	0.7
8	0	0.2	0.1	0.1	0.4	0.1	0.4
9	0	0.2	0.0	0.0	0.2	0.0	0.2
10	1	0.1	0.0	0.0	0.1	0.0	0.1
AIC	-	403.1906	410.2602	411.9481	373.4775	385.3196	373.9456

Hence, it can be seen from Table 5.6, that the HDW and HNB provide the closest expected frequencies to those that are observed. This could be as a result of the structure for these models that depends on mixture components.

5.5 Discrete Weibull excessive zero with right censored count models

In this section, CZIDW and CHDW are investigated to cope with the censored response with too many zero counts as follows:

5.5.1 Censored zero-inflated discrete Weibull model

From Equation 1.15, Equation 1.10 and Equation 2.2, the likelihood of the ZIDW regression model with a right censored count data can be formed as follows:

$$L = \prod_{i=1}^{n} \left\{ \left[\pi_i + (1-\pi_i)(1-q_i)\right]^{\delta_{z_i}} \left[(1-\pi_i)\left(q_i^{y_i^{\beta}} - q_i^{(y_i+1)^{\beta}}\right)\right]^{1-\delta_{z_i}} \right\}^{1-\delta_{c_i}} \left\{1 - \left[\pi_i + (1-\pi_i)\left(1 - q_i^{C^{\beta}}\right)\right]\right\}^{\delta_{c_i}} \tag{5.10}$$

Then, using π in Equation 1.16 and q in Equation 3.2, the log-likelihood can be found as:

$$\begin{aligned}\ell = &\sum_{i=1}^{n} \left(1-\delta_{c_i}\right)\delta_{z_i} \log\left\{\left[\left(e^{-\boldsymbol{z}_i'\boldsymbol{\gamma}}+1\right)^{-1}\right] + \left[1-\left(e^{-\boldsymbol{z}_i'\boldsymbol{\gamma}}+1\right)^{-1}\left(1-e^{-e^{\boldsymbol{x}_i'\boldsymbol{\alpha}}}\right)\right]\right\} + \\ &\sum_{i=1}^{n} \left(1-\delta_{c_i}\right)\left(1-\delta_{z_i}\right) \log\left[1-\left(e^{-\boldsymbol{z}_i'\boldsymbol{\gamma}}+1\right)^{-1}\right] + \\ &\sum_{i=1}^{n} \left(1-\delta_{c_i}\right)\left(1-\delta_{z_i}\right) \log\left[e^{-y_i^{\beta} e^{\boldsymbol{x}_i'\boldsymbol{\alpha}}} - e^{-(y_i+1)^{\beta} e^{\boldsymbol{x}_i'\boldsymbol{\alpha}}}\right] + \\ &\sum_{i=1}^{n} \delta_{c_i} \log\left\{1-\left[\left(e^{-\boldsymbol{z}_i'\boldsymbol{\gamma}}+1\right)^{-1} + \left[1-\left(e^{-\boldsymbol{z}_i'\boldsymbol{\gamma}}+1\right)^{-1}\right]\left(1-e^{-C^{\beta} e^{\boldsymbol{x}_i'\boldsymbol{\alpha}}}\right)\right]\right\}\end{aligned} \tag{5.11}$$

where δ_c and δ_z are defined in Equation 1.9 and Equation 1.14, respectively.

5.5.1.1 Maximum likelihood estimation

In order to estimate the unknown parameters $(\beta, \boldsymbol{\alpha}, \boldsymbol{\gamma})$, the partial derivatives of log-likelihood in Equation 5.11, for $p, l = 1, 2, \ldots, P$, are found as follows:

$$\begin{aligned}\frac{\partial \ell}{\partial \beta} = &\sum_{i=1}^{n} \frac{\left(\delta_{c_i}-1\right)\left(1-\delta_{z_i}\right) e^{\boldsymbol{x}_i'\boldsymbol{\alpha}}}{wh2(\beta, \boldsymbol{\alpha})} \left\{y_i^{\beta} e^{-y_i^{\beta} e^{\boldsymbol{x}_i'\boldsymbol{\alpha}}} \log(y_i) - (y_i+1)^{\beta} e^{-(y_i+1)^{\beta} e^{\boldsymbol{x}_i'\boldsymbol{\alpha}}} \log(y_i+1)\right\} + \\ &\sum_{i=1}^{n} \frac{-\delta_{c_i}}{wcz4(\beta, \boldsymbol{\alpha}, \boldsymbol{\gamma})} C^{\beta} e^{-C^{\beta} e^{\boldsymbol{x}_i'\boldsymbol{\alpha}} + \boldsymbol{x}_i'\boldsymbol{\alpha}} \log(C) \left[1-\left(e^{-\boldsymbol{z}_i'\boldsymbol{\gamma}}+1\right)^{-1}\right]\end{aligned}$$

$$\begin{aligned}\frac{\partial \ell}{\partial \alpha_p} = &\sum_{i=1}^{n} \frac{\left(\delta_{c_i}-1\right)\delta_{z_i}}{wz1(\alpha_p, \boldsymbol{\gamma})} \left[x_{ip} e^{-e^{x_{ip}\alpha_p} + x_{ip}\alpha_p} \left(e^{-\boldsymbol{z}_i'\boldsymbol{\gamma}}+1\right)^{-1}\right] + \\ &\sum_{i=1}^{n} \frac{\left(\delta_{c_i}-1\right)\left(1-\delta_{z_i}\right) x_{ip} e^{x_{ip}\alpha_p}}{wh2(\beta, \alpha_p)} \left[y_i^{\beta} e^{-y_i^{\beta} e^{x_{ip}\alpha_p}} - (y_i+1)^{\beta} e^{-(y_i+1)^{\beta} e^{x_{ip}\alpha_p}}\right] + \\ &\sum_{i=1}^{n} \frac{-\delta_{c_i}}{wcz4(\beta, \alpha_p, \boldsymbol{\gamma})} x_{ip} C^{\beta} e^{-C^{\beta} e^{x_{ip}\alpha_p} + x_{ip}\alpha_p} \left[1-\left(e^{-\boldsymbol{z}_i'\boldsymbol{\gamma}}+1\right)^{-1}\right]\end{aligned}$$

$$
\begin{aligned}
\frac{\partial \ell}{\partial \gamma_l} = & \sum_{i=1}^{n} \frac{(1-\delta_{c_i})\,\delta_{z_i}}{wz1(\boldsymbol{\alpha},\gamma_l)} z_{il} \left(e^{-z_{il}\gamma_l}+1\right)^{-2} e^{-e^{\boldsymbol{x}_i'\boldsymbol{\alpha}}-z_{il}\gamma_l} + \\
& \sum_{i=1}^{n} \frac{(\delta_{c_i}-1)\,(1-\delta_{z_i})}{1-(wh1(\gamma_l))^{-1}} z_{il} e^{-z_i\gamma_l} \left(e^{-z_{il}\gamma_l}+1\right)^{-2} + \\
& \sum_{i=1}^{n} \frac{-\delta_{c_i}}{wcz4(\beta,\boldsymbol{\alpha},\gamma_l)} z_{il} \left(e^{-z_{il}\gamma_l}+1\right)^{-2} e^{-C^{\beta}e^{\boldsymbol{x}_i'\boldsymbol{\alpha}}-z_{il}\gamma_l}
\end{aligned}
$$

where

$$
\begin{aligned}
wcz4(\beta,\boldsymbol{\alpha},\boldsymbol{\gamma}) = & 1- \\
& \left[\left(e^{-\boldsymbol{z}_i'\boldsymbol{\gamma}}+1\right)^{-1} + \left[1-\left(e^{-\boldsymbol{z}_i'\boldsymbol{\gamma}}+1\right)^{-1}\right]\left(1-e^{-c^{\beta}e^{\boldsymbol{x}_i'\boldsymbol{\alpha}}}\right)\right]
\end{aligned} \tag{5.12}
$$

$wz1(\alpha,\gamma)$, $wh1(\gamma)$, $wh2(\beta,\alpha)$ can be found, respectively in, Equation 5.6 and Equation 5.9. The above equations cannot be found in a closed form system, which requires a numerical solution to find the MLEs of the unknown parameter $(\beta,\boldsymbol{\alpha},\boldsymbol{\gamma})$.

5.5.1.2 Simulation study

A simulation study to evaluate the CZIDW regression model has been conducted. Different sample sizes are considered, $n_1 = 370, n_2 = 500, n_3 = 1200$, with different censoring points. This study depends on the same algorithm as in subsection 5.2.2, under the censoring scheme consideration.

The data are generated from a ZIDW, then a censored point will be considered to cut this sample from the right, and this censored data is fitted by Equation 5.10. That is, Equation 5.11 is optimized using the same initial values as in subsection 5.2.2 and the results are shown in Table 5.7.

Table 5.7: MLEs based on the simulation study for the CZIDW regression model with true parameters $\boldsymbol{\alpha} = (-2, -1.7), \boldsymbol{\gamma} = (1.5, -0.9)$ *and* $\beta = 2.2$.

n	parameter	MLE	Bias	MSE	Length
370	α_0	-2.0189	0.0189	0.2068	1.7711
	α_1	-1.7551	0.0551	0.0981	1.1482
	γ_0	1.4485	0.0515	0.0131	1.0385
	γ_1	-0.827	-0.073	0.0124	1.0936
	β	2.2434	-0.0434	0.0849	1.1224
		C=6 , (7.5959)%			
500	α_0	-2.0068	0.0068	0.1712	1.5887
	α_1	-1.7401	0.0401	0.0701	1.0331
	γ_0	1.6242	-0.1242	0.0227	0.9356
	γ_1	-0.96	0.06	0.0088	0.9743
	β	2.2278	-0.0278	0.0682	0.9817
		C=6 , (7.539)%			
1200	α_0	-2.0117	0.0117	0.0583	0.9482
	α_1	-1.7176	0.0176	0.0253	0.6139
	γ_0	1.3456	0.1544	0.0268	0.5618
	γ_1	-0.7398	-0.1602	0.0278	0.603
	β	2.2185	-0.0185	0.0246	0.605
		C=6 , (7.4863)%			

5.5.2 Censored hurdle discrete Weibull model

The HDW regression model with right censoring can be formulated with likelihood as follows:

$$L = \prod_{i=1}^{n} \left\{ [\pi_i]^{\delta_{z_i}} [1-\pi_i]^{1-\delta_{z_i}} \left[\frac{q_i^{y_i^\beta} - q_i^{(y_i+1)^\beta}}{q_i} \right]^{1-\delta_{z_i}} \right\}^{1-\delta_{c_i}} \left\{ 1 - \left[\pi_i + (1-\pi_i) \left(1 - q_i^{C^\beta - 1} \right) \right] \right\}^{\delta_{c_i}} \tag{5.13}$$

Then, using π in Equation 1.16 and q in Equation 3.2, the log-likelihood can be found as follows:

$$\begin{aligned}\ell = &\sum_{i=1}^{n} (\delta_{c_i} - 1)\, \delta_{z_i} \log\left(e^{-\boldsymbol{z}_i'\boldsymbol{\gamma}} + 1\right) + \sum_{i=1}^{n} (1-\delta_{c_i})\,(1-\delta_{z_i}) \log\left[1 - \left(e^{-\boldsymbol{z}_i'\boldsymbol{\gamma}} + 1\right)^{-1}\right] + \\ &\sum_{i=1}^{n} (1-\delta_{c_i})\,(1-\delta_{z_i}) \log\left[e^{-y_i^{\beta} e^{\boldsymbol{x}_i'\boldsymbol{\alpha}}} - e^{-(y_i+1)^{\beta} e^{\boldsymbol{x}_i'\boldsymbol{\alpha}}}\right] + \sum_{i=1}^{n} (1-\delta_{c_i})\,(1-\delta_{z_i})\, e^{\boldsymbol{x}_i'\boldsymbol{\alpha}} + \\ &\sum_{i=1}^{n} \delta_{c_i} \log\left\{1 - \left[\left(e^{-\boldsymbol{z}_i'\boldsymbol{\gamma}} + 1\right)^{-1} + \left[1 - \left(e^{-\boldsymbol{z}_i'\boldsymbol{\gamma}} + 1\right)^{-1}\right]\left(1 - e^{-(C^{\beta}-1)e^{\boldsymbol{x}_i'\boldsymbol{\alpha}}}\right)\right]\right\}\end{aligned} \tag{5.14}$$

where, δ_c and δ_z are the indicator variables in Equation 1.9 and Equation 1.14, respectively.

5.5.2.1 Maximum likelihood estimation

The MLEs of the unknown parameters $(\beta, \boldsymbol{\alpha}, \boldsymbol{\gamma})$ can be found by equating the partial derivatives of the log-likelihood in Equation 5.14 to zeros. Hence, for $p, l = 1, 2, \ldots, P$, we have

The partial derivative of ℓ with respect to β is as follows:

$$\begin{aligned}\frac{\partial \ell}{\partial \beta} = &\sum_{i=1}^{n} \frac{(\delta_{c_i} - 1)\,(1-\delta_{z_i})\, e^{\boldsymbol{x}_i'\boldsymbol{\alpha}}}{wh2(\beta, \boldsymbol{\alpha})} \left\{y_i^{\beta} e^{-y_i^{\beta} e^{\boldsymbol{x}_i'\boldsymbol{\alpha}}} \log(y_i) - (y_i+1)^{\beta} e^{-(y_i+1)^{\beta} e^{\boldsymbol{x}_i'\boldsymbol{\alpha}}} \log(y_i+1)\right\} + \\ &\sum_{i=1}^{n} \frac{-\delta_{c_i}}{wch4(\beta, \boldsymbol{\alpha}, \boldsymbol{\gamma})} \left\{\left[1 - \left(e^{-\boldsymbol{z}_i'\boldsymbol{\gamma}} + 1\right)^{-1}\right]\left(C^{\beta} e^{-(C^{\beta}-1)e^{\boldsymbol{x}_i'\boldsymbol{\alpha}} + \boldsymbol{x}_i'\boldsymbol{\alpha}}\right) \log(C)\right\}\end{aligned}$$

The partial derivative of ℓ with respect to $\boldsymbol{\alpha}$ is as follows:

$$\begin{aligned}\frac{\partial \ell}{\partial \alpha_p} = &\sum_{i=1}^{n} \frac{(\delta_{c_i} - 1)\,(1-\delta_{z_i})\, x_{ip} e^{x_{ip}\alpha}}{wh2(\beta, \alpha_p)} \left\{y_i^{\beta} e^{-y_i^{\beta} e^{x_{ip}\alpha_p}} - (y_i+1)^{\beta} e^{-(y_i+1)^{\beta} e^{x_{ip}\alpha_p}}\right\} + \\ &\sum_{i=1}^{n} (1-\delta_{c_i})\,(1-\delta_{z_i})\, x_{ip} e^{x_{ip}\alpha_p} + \\ &\sum_{i=1}^{n} \frac{-\delta_{c_i}}{wch4(\beta, \alpha_p, \boldsymbol{\gamma})} \left\{\left[1 - \left(e^{-\boldsymbol{z}_i'\boldsymbol{\gamma}} + 1\right)^{-1}\right]\left(x_{ip}(C^{\beta} - 1) e^{-(C^{\beta}-1)e^{x_{ip}\alpha_p} + x_{ip}\alpha_p}\right)\right\}\end{aligned}$$

The partial derivative of ℓ with respect to $\boldsymbol{\gamma}$ is as follows:

$$\frac{\partial \ell}{\partial \gamma_l} = \sum_{i=1}^{n} \frac{(1-\delta_{c_i})\,\delta_{z_i}}{wh1(\gamma_l)} z_{il} e^{-z_{il}\gamma_l} + \sum_{i=1}^{n} \frac{(\delta_{c_i}-1)\,(1-\delta_{z_i})}{1-(wh1(\gamma_l))^{-1}} \left[z_{il} e^{-z_{il}\gamma_l} \left(e^{-z_{il}\gamma_l}+1\right)^{-2} \right] + \\ \sum_{i=1}^{n} \frac{-\delta_{c_i}}{wch4(\beta,\boldsymbol{\alpha},\gamma_l)} z_{il} \left(e^{-z_{il}\gamma_l}+1\right)^{-2} e^{-(C^{\beta}-1)e^{\boldsymbol{x}_i'\boldsymbol{\alpha}}-z_{il}\gamma_l}$$

where

$$wch4(\beta,\boldsymbol{\alpha},\boldsymbol{\gamma}) = 1- \\ \left[\left(e^{-\boldsymbol{z}_i'\boldsymbol{\gamma}}+1\right)^{-1} + \left[1-\left(e^{-\boldsymbol{z}_i'\boldsymbol{\gamma}}+1\right)^{-1}\right] \left(1-e^{-(c^{\beta}-1)e^{\boldsymbol{x}_i'\boldsymbol{\alpha}}}\right) \right] \tag{5.15}$$

$wh1(\gamma)$ and $wh2(\beta,\alpha)$ can be found in Equation 5.9.

5.5.2.2 Simulation study

To examine the behavior of the MLEs under the CHDW model, a simulation study is performed. This study considers some censoring schemes and follows the same algorithm as in subsection 5.3.2.

In other words, data from HDW is generated, subsequently censored at some point C and fitted using Equation 5.13. That is, the log-likelihood in Equation 5.14 is maximized, using the same initial points as in subsection 5.3.2. These results are reported in Table 5.8.

Table 5.8: MLEs based on the simulation study for the CHDW regression model with true parameters $\boldsymbol{\alpha} = (-2, -1.7), \boldsymbol{\gamma} = (1.5, -0.9)$ and $\beta = 2.2$.

n	Parameter	MLE	Bias	MSE	Length
370	α_0	-2.0117	0.0117	0.2141	1.713
	α_1	-1.7335	0.0335	0.084	1.1056
	γ_0	1.512	-0.012	0.0631	1.713
	γ_1	-0.912	0.012	0.0741	1.1056
	β	2.2333	-0.0333	0.0892	0.9618
		C=6 , (7.8984)%			
500	α_0	-2.0262	0.0262	0.1455	1.4783
	α_1	-1.7363	0.0363	0.0645	0.9569
	γ_0	1.4993	0.0007	0.046	1.4783
	γ_1	-0.898	-0.002	0.0523	0.9569
	β	2.2379	-0.0379	0.065	0.8403
		C=6 , (8.022)%			
1200	α_0	-2.0154	0.0154	0.061	0.9465
	α_1	-1.7122	0.0122	0.0243	0.612
	γ_0	1.507	-0.007	0.0188	0.9465
	γ_1	-0.9058	0.0058	0.0235	0.612
	β	2.2177	-0.0177	0.0247	0.5337
		C=6 , (7.7508)%			

5.6 Numerical example: unwanted pursuit behavior perpetrations data

In this section, a data with a too many zeros response are analyzed using the zero-inflated, hurdle, and their corresponding from censored models. Regarding the Poisson and NB, their censored zero-inflated and censored hurdle are fitted by optimizing the logarithm of their likelihood in Equation 1.11, Equation 1.24, Equation 1.27 and Equation 1.28.

A dataset from Loeys et al. (2012), assessing the extent of UPB committed after couples have broken up, is considered. In this study, to explain the perpetration, 28 questions representing UPBs (ranging from “leaving unwanted gifts” to “threatening to hurt yourself”), each measured using a five-point Likert scale (from 0=never to 4=over five times) were applied. Then, the higher scores pointing out higher levels of perpetrations.

A zero count is occurs for the individuals who answer “never” to all 28 questions. Additionally, for those who choose “over five times” to “leaving unwanted gifts” and “never” to all other questions, for example, will have an UPB count equal to 4, and so on.

As in Loeys et al. (2012), two covariates are considered in this study to examine their effect on the UPB: education and anxious attachment levels. The education level is a binary predictor, where 0 = lower than bachelor degree, or 1 = at least bachelor degree. While the anxious attachment level is represented by a continuous variable. For more details on this experiment, see Loeys et al. (2012). This dataset contains 387 observations with 246 zero counts; thus, a zero-inflated case is considered. Therefore, ZIP, ZINB, ZIDW, HP, HNB and HDW can be applied. In addition, the observed frequencies are shown in Figure 5.3 and as can be seen from this histogram, around 4.6512% of the count variable are greater than or equal to 12. Thus, $C = 12$ is considered as a cut point in this example, to see the effect of censoring on the zero-inflation dataset. The results of the MLEs and their SEs are shown in Table 5.9 for the zero-inflated and their censored models, while Table 5.10 shows the results for the hurdle and their censored models.

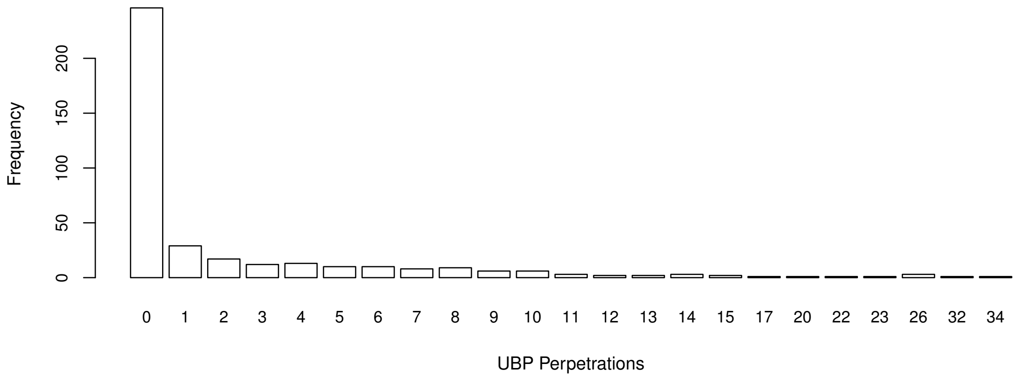


Figure 5.3: Histogram for the observed frequencies for the UPB data.

Table 5.9: MLEs, SEs (in parentheses) and AIC from the zero-inflated and censored zero-inflted: Poisson, NB and DW regression models for the UPB dataset.

	intercept	education	anxiety	other	AIC
logit-ZIDW					
count model	-1.4857 (0.3010)	0.3744 (0.1761)	-0.1931 (0.0910)	$\hat{\beta}$=0.8783 (0.0996)	1266.334
zero model	0.3124 (0.2026)	-0.4241 (0.3001)	-0.4652 (0.1455)	-	
logit-ZIP					
count model	1.9208 (0.0445)	-0.3502 (0.0713)	0.1334 (0.0345)	-	1616.901
zero model	0.6729 (0.1419)	-0.2321 (0.2219)	-0.4831 (0.1112)	-	
logit-ZINB					
count model	1.7234 (0.1495)	-0.4897 (0.2062)	0.2048 (0.1078)	log($\hat{k}$)=-0.1975 (0.2752)	1266.282
zero model	0.3398 (0.2101)	-0.4589 (0.2969)	-0.5200 (0.1467)	-	
logit-CZIDW					
count model	-1.7381 (0.3604)	0.3532 (0.1857)	-0.2704 (0.1018)	$\hat{\beta}$=0.9988 (0.1417)	1153.82
zero model	0.3881 (0.1917)	-0.3968 (0.2815)	-0.4665 (0.1368)	-	
logit-CZIP					
count model	1.6885 (0.0511)	-0.1867 (0.0763)	0.1850 (0.0384)	-	1298.707
zero model	0.6664 (0.1421)	-0.2312 (0.2223)	-0.4778 (0.1115)	-	
logit-CZINB					
count model	1.6437 (0.1411)	-0.3661 (0.2062)	0.2872 (0.1115)	$\hat{k}$=0.9974 (0.3116)	1154.096
zero model	0.3880 (0.1979)	-0.3862 (0.2783)	-0.4747 (0.1370)	-	

Table 5.10: MLEs, SEs (in parentheses) and AIC from the hurdle and censored hurdle: Poisson, NB and DW regression models for the UPB dataset.

	intercept	education	anxiety	other	AIC
hurdle models					
binomial-logit model	0.6751	-0.2203	-0.4863	-	
	(0.1418)	(0.2211)	(0.1109)		
HDW count model	-1.6219	0.4129	-0.1897	$\hat{\beta}$=0.9175	
	(0.3132)	(0.1780)	(0.0911)	(0.1045)	1266.209
HP count model	1.9209	-0.3501	0.1331	-	
	(0.0445)	(0.0713)	(0.0345)		1616.921
HNB count model	1.7252	-0.4871	0.2070	log($\hat{k}$)=-0.1871	
	(0.1484)	(0.2055)	(0.1066)	(0.2727)	1266.526
censored hurdle models					
logit-CHDW					
count model	-1.8294	0.4029	-0.2695	$\hat{\beta}$=1.0239	
	(0.3573)	(0.1862)	(0.1011)	(0.1407)	1154.129
zero model	0.6814	-0.2266	-0.4891	-	
	(0.1420)	(0.2213)	(0.1110)		
logit-CHP					
count model	1.6889	-0.1868	0.1847	-	
	(0.0511)	(0.0763)	(0.0383)		1298.756
zero model	0.6755	-0.2203	-0.4867	-	
	(0.1418)	(0.2211)	(0.1109)		
logit-CHNB					
count model	1.6470	-0.3713	0.2876	$\hat{k}$=1.006	
	(0.1404)	(0.2057)	(0.1102)	(0.3117)	1154.257
zero model	0.6747	-0.2196	-0.4861	-	
	(0.1418)	(0.2211)	(0.1109)		

The results from Table 5.3, Table 5.4, Table 5.9 and Table 5.10 show that the DW regression models are only marginally superior to the NB models, but both the DW and NB models fit the data much better than the Poisson regression models. Additionally, Table 5.5 and Table 5.11 confirm that the DW and NB models work alternatively, better than Poisson, and yield expected frequencies that are close to those that are observed.

5.7 Concluding remarks

This chapter is concerned with types of data that experience a lot of zero counts, a case known as zero inflation. Some popular models for this condition of excessive zeros are ZIP, ZINB, HP and HNB. However, these models are not the best to apply when the data show under-dispersion within subgroups. This is

Table 5.11: Observed and expected frequencies for the standard, zero-inflated, hurdle, censored zero-inflated and censored hurdle: Poisson, NB and DW regression models for the UPB data.

model	observed	DW	Poisson	NB	ZIDW	ZIP	ZINB	HDW	HP	HNB	CZIDW	CZIP	CZINB	CHDW	CHP	CHNB
0	246	239.6	57.6	241.3	248.4	246.0	245.9	246.0	246.0	246.0	246.0	246.0	245.0	246.3	246.0	246.0
1	29	43.6	94.6	42.2	26.9	2.8	26.1	26.0	2.8	26.0	25.1	4.8	24.9	24.4	4.8	24.9
2	17	23.6	88.0	22.7	20.2	7.2	20.0	20.1	7.2	20.0	20.3	11.1	20.2	20.0	11.1	20.2
3	12	15.3	62.9	15.1	15.8	12.8	15.9	16.0	12.8	15.9	16.4	17.7	16.4	16.4	17.7	16.4
4	13	10.8	38.9	10.9	12.5	17.6	12.8	12.9	17.6	12.8	13.4	21.9	13.4	13.4	21.9	13.4
5	10	8.1	22.1	8.3	10.1	20.1	10.5	10.5	20.1	10.5	10.9	22.4	11.0	11.0	22.4	11.0
6	10	6.3	11.8	6.6	8.2	19.9	8.6	8.6	19.9	8.6	9.0	19.8	9.0	9.1	19.8	9.0
7	8	5.0	6.0	5.3	6.8	17.6	7.1	7.1	17.6	7.2	7.4	15.5	7.4	7.5	15.5	7.4
8	9	4.1	2.9	4.4	5.6	14.2	6.0	5.9	14.2	6.0	6.1	11.0	6.1	6.2	11.0	6.1
9	6	3.4	1.3	3.6	4.7	10.5	5.0	4.9	10.5	5.0	5.1	7.2	5.1	5.1	7.2	5.1
10	6	2.8	0.6	3.1	3.9	7.3	4.2	4.1	7.3	4.2	4.2	4.4	4.2	4.3	4.4	4.2
11	3	2.4	0.3	2.6	3.3	4.7	3.5	3.5	4.7	3.5	3.5	2.5	3.5	3.6	2.5	3.5
12	2	2.1	0.1	2.2	2.8	2.9	3.0	3.0	2.9	3.0	19.7	2.6	19.6	19.8	2.6	19.6
13	2	1.8	0.0	1.9	2.4	1.6	2.5	2.5	1.6	2.5	-	-	-	-	-	-
14	3	1.6	0.0	1.7	2.0	0.9	2.2	2.1	0.9	2.2	-	-	-	-	-	-
15	2	1.4	0.0	1.5	1.7	0.5	1.8	1.8	0.5	1.8	-	-	-	-	-	-
17	1	1.1	0.0	1.1	1.3	0.1	1.4	1.3	0.1	1.4	-	-	-	-	-	-
20	1	0.8	0.0	0.8	0.8	0.0	0.9	0.9	0.0	0.9	-	-	-	-	-	-
22	1	0.6	0.0	0.6	0.6	0.0	0.6	0.6	0.0	0.6	-	-	-	-	-	-
23	1	0.6	0.0	0.6	0.5	0.0	0.6	0.6	0.0	0.6	-	-	-	-	-	-
26	3	0.4	0.0	0.4	0.4	0.0	0.4	0.4	0.0	0.4	-	-	-	-	-	-
32	1	0.3	0.0	0.2	0.2	0.0	0.2	0.2	0.0	0.2	-	-	-	-	-	-
34	1	0.2	0.0	0.2	0.1	0.0	0.1	0.1	0.0	0.1	-	-	-	-	-	-
AIC	-	1285.213	2782.39	1285.919	1265.964	1616.901	1266.282	1266.209	1616.921	1266.526	1153.825	1298.707	1154.094	1154.129	1298.756	1154.257

because the Poisson and NB models can cope with equi- and over-dispersion count data. Therefore, it would be better to apply a model with the flexibility to handle different levels of dispersion. Here, the DW regression model is modified for these situations with excessive zero counts. Two modifications of the DW regression model for the case of zero inflation have been developed, namely, ZIDW and HDW. These models can handle data with excessive zero counts with different infra-dispersion.

On the other hand, the censoring mechanism has been applied to such excessive zeros data. Some simulation studies and numerical examples have been discussed in this chapter. The simulation studies show good behavior for the MLEs for the ZIDW, HDW, CZIDW and CHDW. Thus, it could be noted from these studies that the measurements of accuracy, bias and MSE, as well as the length of the CI are close to zero and generally decrease as the sample size n increases, showing the consistency of these MLEs. Additionally, the applications show that the modifications of the DW regression models work well in comparison to their corresponding from Poisson and NB regression models.

Chapter 6

Median Discrete Weibull Regression Model

6.1 Introduction

The DW regression model introduced so far relies on the parameter q, that is, it is on a different scale than the common models for discrete response variables, where the regression is introduced through the mean, such as GLMs, including the Poisson and NB regression models. Thus, to achieve an interpretation equivalent to the regression models from GLMs, where the effect of the covariates is investigated on a central tendency measurement, specific approach might be employed. First, as mentioned in chapter 3, the interpretation approach in Equation 3.25, based on the median, can be applied. Alternatively, this chapter develops a new regression structure for DW through its median. Another reason for considering the median rather than the mean is the common skewed nature of count data, hence the median is more representative than the mean as a measurement for location in most discrete data analyses (see for example, Steinberg (2010) and Sellers and Shmueli (2010)).

6.2 Median discrete Weibull regression model

The form of the median in Equation 3.24, that is given by

$$M+1=\left(-\frac{\log(2)}{\log(\hat{q}(x))}\right)^{\frac{1}{\beta}}$$

will be considered in this chapter. Then, the MDW regression model is developed by first re-parameterizing the DW in Equation 2.2 in terms of the median of Y and then introducing a regression-based functional form. Therefore, q can be obtained as follows:

$$q=exp\left(\frac{-\log(2)}{(M+1)^{\beta}}\right)$$

where $0<q<1$. Hence, it follows from Equation 2.2 that

$$f(y)=e^{-\log(2)\left(\dfrac{y}{M+1}\right)^{\beta}}-e^{-\log(2)\left(\dfrac{y+1}{M+1}\right)^{\beta}}$$

where, $M+1>0$ and $\beta>0$.
Consequently, the regression structure for DW could be started by assuming the median is directly related to a set of predictors, as follows:

$$g(M_i)=\boldsymbol{x}_i'\boldsymbol{\alpha} \qquad , \qquad \boldsymbol{x}_i'\boldsymbol{\alpha}=\alpha_0+x_{i1}\alpha_1+\ldots+x_{iP}\alpha_P$$

For a link function g and regression coefficient $\alpha_0,\alpha_1\ldots,\alpha_P$. This link function g can take a number of possible choices; however, in the context of DW, it is convenient to assume that $g(M)=\log(M+1)$. Then, the MDW regression model is introduced with the link function, which defines the median of y_i as:

$$M+1=e^{\boldsymbol{x}'\boldsymbol{\alpha}} \tag{6.1}$$

Then, substituting Equation 6.1, q can be obtained as:

$$q=exp\left(\frac{-\log(2)}{e^{\beta\boldsymbol{x}'\boldsymbol{\alpha}}}\right) \tag{6.2}$$

For a sample of n independent observations (x_i, y_i), $i = 1, \ldots, n$, the likelihood for the MDW can be defined as:

$$L = \prod_{i=1}^{n} \left\{ e^{-\log(2)\left(\frac{y_i}{e^{\boldsymbol{x}_i'\boldsymbol{\alpha}}}\right)^{\beta}} - e^{-\log(2)\left(\frac{y_i+1}{e^{\boldsymbol{x}_i'\boldsymbol{\alpha}}}\right)^{\beta}} \right\} \tag{6.3}$$

6.3 Maximum likelihood estimation

To obtain the MLEs for the unknown parameters, the logarithm of the likelihood function in Equation 6.3 is required:

$$\ell = \sum_{i=1}^{n} \log\left\{wm_i(\beta, \boldsymbol{\alpha})\right\} \tag{6.4}$$

where

$$wm_i(\beta, \boldsymbol{\alpha}) = e^{-\log(2)\left(\frac{y}{e^{\boldsymbol{x}_i'\boldsymbol{\alpha}}}\right)^{\beta}} - e^{-\log(2)\left(\frac{y+1}{e^{\boldsymbol{x}_i'\boldsymbol{\alpha}}}\right)^{\beta}}$$

Consequently, the partial derivatives with respect to each parameter are obtained as follows:

$$\frac{\partial \ell}{\partial \beta} = \sum_{i=1}^{n} \frac{-\log(2)}{wm_i(\beta, \boldsymbol{\alpha})} \left\{ \left(\frac{y_i}{e^{\boldsymbol{x}_i'\boldsymbol{\alpha}}}\right)^{\beta} e^{-\log(2)\left(\frac{y_i}{e^{\boldsymbol{x}_i'\boldsymbol{\alpha}}}\right)^{\beta}} \left[\log(y_i) - \boldsymbol{x}_i'\boldsymbol{\alpha}\right] - \left(\frac{y_i+1}{e^{\boldsymbol{x}_i'\boldsymbol{\alpha}}}\right)^{\beta} e^{-\log(2)\left(\frac{y_i+1}{e^{\boldsymbol{x}_i'\boldsymbol{\alpha}}}\right)^{\beta}} \left[\log(y_i+1) - \boldsymbol{x}_i'\boldsymbol{\alpha}\right] \right\}$$

$$\frac{\partial \ell}{\partial \alpha_p} = \sum_{i=1}^{n} \frac{\log(2)\beta x_{ip} e^{-x_{ip}\alpha_p}}{wm_i(\beta, \alpha_p)} \left\{ y_i \left(\frac{y_i}{e^{x_{ip}\alpha_p}}\right)^{\beta-1} e^{-\log(2)\left(\frac{y_i}{e^{x_{ip}\alpha_p}}\right)^{\beta}} - (y_i+1)\left(\frac{y_i+1}{e^{x_{ip}\alpha_p}}\right)^{\beta-1} e^{-\log(2)\left(\frac{y_i+1}{e^{x_{ip}\alpha_p}}\right)^{\beta}} \right\}$$

The common maximum likelihood problem is experienced again here, for which there is no closed form solution for the above non-linear equations. Therefore, a numerical optimization method is required to directly maximize the log-likelihood function in Equation 6.4 and find the MLEs.

6.4 Simulation study

A simulation study is conducted in order to evaluate the MDW regression model described in this chapter. A multiple regression with three covariates is considered. The parameters in this simulation includes the set of regression coefficients $(\alpha_0 = 1.5, \alpha_1 = 0.4, \alpha_2 = -0.2, \alpha_3 = 0.8)$ and $\beta = 1.6$.
Then, the covariates $\boldsymbol{X}_1$, $\boldsymbol{X}_2$ and $\boldsymbol{X}_3$ are chosen to be generated from $N(0, 1)$, $unif(0, 10)$ and $binomial(1, 0.6)$, respectively. In this simulation study, samples of size $n_1 = 50$, $n_2 = 100$, $n_3 = 250$, $n_4 = 500$ and $n_5 = 1000$ are generated from the DW regression model with the parameters β and q defined in Equation 6.2. The simulation study is carried out over 1000 iterations, following the same method as in section 3.8, and the results are shown in Table 6.1. However, the initial values for the regression coefficient $\boldsymbol{\alpha}$ are selected using the Poisson regression fit without the minus sign. This is due to the regression structure of the MDW, which links some covariates to the median with the log link function, similar to the Poisson model.

Table 6.1: MLEs based on the simulation study for the MDW regression model with true parameters $\boldsymbol{\alpha} = (1.5, 0.4, -0.2, 0.8)$ and $\beta = 1.6$.

n	parameter	MLE	Bias	MSE	Length
50	α_0	1.5051	0.0051	0.0471	0.7947
	α_1	0.4068	0.0068	0.0117	0.3907
	α_2	-0.2011	-0.0011	0.0011	0.1223
	α_3	0.7954	-0.0046	0.0371	0.6955
	β	1.7385	0.1385	0.078	0.6955
100	α_0	1.5063	0.0063	0.0216	0.5731
	α_1	0.4038	0.0038	0.0049	0.267
	α_2	-0.2013	-0.0013	0.0006	0.0896
	α_3	0.7996	-0.0004	0.0187	0.502
	β	1.6715	0.0715	0.0287	0.502
250	α_0	1.498	-0.002	0.0087	0.372
	α_1	0.3995	-0.0005	0.002	0.1691
	α_2	-0.2003	-0.0003	0.0002	0.0552
	α_3	0.8075	0.0075	0.0068	0.3298
	β	1.6313	0.0313	0.0101	0.3298
500	α_0	1.4988	-0.0012	0.0046	0.2713
	α_1	0.3993	-0.0007	0.0009	0.1186
	α_2	-0.2	0	0.0001	0.0405
	α_3	0.8024	0.0024	0.0032	0.227
	β	1.617	0.017	0.0049	0.227
1000	α_0	1.5016	0.0016	0.0023	0.1871
	α_1	0.3991	-0.0009	0.0004	0.0841
	α_2	-0.2002	-0.0002	0.0001	0.0273
	α_3	0.7996	-0.0004	0.0017	0.1635
	β	1.6086	0.0086	0.002	0.1635

6.5 Numerical examples

In this section some of the examples in chapter 3 are applied to investigate the MDW regression model and compare its results with the results obtained previously.

6.5.1 The case of under-dispersion: inhaler use data

The data in subsection 3.9.1 is applied to represent the under-dispersion case relative to the Poisson. The data is fitted by the MDW regression model, and the results are summarized in Table 6.2.

Table 6.2: MLEs, SEs (in parentheses) and AIC from MDW fitted to the inhaler use data.

intercept	humidity	pressure	temperature	particles	other	AIC
-1.8612 (0.9226)	-0.0991 (0.0474)	3.9766 (1.4658)	-0.1597 (0.0731)	0.0125 (0.0073)	$\hat{\beta}$=2.1405 (0.0260)	13486.14

6.5.2 The case of over-dispersion: strikes data

To show the case of over-dispersed data relative to the Poisson, the data in subsection 3.9.2 is applied. This data is fitted using the MDW regression model, and the results are summarized in Table 6.3.

Table 6.3: MLEs, SEs (in parentheses) and AIC from MDW fitted to the strikes data.

intercept	economic activity	other	AIC
1.6362 (0.0694)	3.2028 (1.1192)	$\hat{\beta}$=1.6525 (0.1302)	564.157

6.5.3 The case of excessive zeros: doctor visits from the German health survey data

The example in subsection 3.9.3 is fitted in Table 6.4 using the MDW regression model, representing a case of excessive zero count.

Table 6.4: MLEs, SEs (in parentheses) and AIC from MDW fitted to the doctor visits from the German health dataset.

intercept	bad health	age	other	AIC
0.2875 (0.1086)	0.9770 (0.1023)	0.0058 (0.0028)	$\hat{\beta}$=0.9887 (0.0265)	4474.974

From the above examples, it can be seen that the results for the regression coefficients from the MDW fitting are similar to the corresponding results form the Poisson and NB provided in chapter 3.

6.6 Concluding remarks

This chapter develops a different methodology for structuring the regression based on DW distribution. In other words, the DW distribution is re-parameterized in term of the median. Thus, compared to the GLMs, including the Poisson and NB regression models, both models are investigating the central tendency of the data explained by some covariates, through the log link function; however, the MDW is more appropriate for count data due to the common skewed nature of count data. Additionally, the DW model has the ability to handle different types of data dispersion.

The simulation study shows good behavior for the MLEs for the MDW, where the measurements of accuracy generally are close to zero and decrease as the sample size n increases, showing the consistency of the MLEs. In addition, the numerical examples for different types of data exhibit similar results to those from the DW model in chapter 3, which introduces the regression via the parameter q but interprets it through the median. Additionally, these estimators are similar to those in the Poisson and NB regression models.

Chapter 7

Conclusions and Future Research

Count data can be found in several disciplines, representing the number of times an event occurs. The type of dispersion of the data is central to the modeling of count data and plays an essential role in their analysis. Hence, they have been attracting great interest, and it has become a challenge for practitioners to select a proper model that takes into account the varying levels of dispersion that typically occur in count data sets. It would be highly desirable to have a unified model that can automatically adapt to the underlying dispersion and be easily implemented in practice.

Count data regression is widely performed by models such as the Poisson, NB and zero-inflated regression models. This thesis focuses on introducing the DW as a simple regression model for count data and shows how this model can capture different levels of dispersion adaptively. A summary of this thesis and some future research topics are discussed below.

7.1 Summary

DW distribution is investigated in chapter 2 as a unified model for capturing different levels of dispersion in count data, namely, under-dispersion and over-dispersion relative to the Poisson, in addition to the common case of excessive zeros. This is an attractive feature of DW, in addition to its simplicity with a closed form pmf with two parameters.

Then, the DW regression model is introduced in chapter 3, by generalizing the

DW distribution and allowing its parameter to be related to a set of covariates. Unlike the GLMs, in which the conditional mean is central to the interpretation, the DW regression model proposed in this chapter has the advantage of modeling the whole conditional density, including all the conditional quantiles that can be easily extracted from the fitted model. This is particularly useful as most count data have a highly skewed distribution. To assess the performance of the DW regression model, the different levels of dispersion have been explained through simulation studies and real data applications for each level of dispersion. The simulation studies show a good behavior for the MLEs under the fitting of the DW regression model. Additionally, the goodness of fit for this model shows a very good performance compared to the related models from the Poisson and NB that could be applied for the same situations.

In chapter 4, the DW regression model is modified to model the right censored data, that commonly arise in count data. In other words, the CDW regression model is developed to analyze the dependent variable, which is available for a limited range although the covariate values are always observed. These right censored count data are modeled in simulations and numerical examples for all the types of dispersion using two different fittings. The first is the truncated case, where the censoring has been ignored and the data are considered to be complete and modeled using the standard DW regression model presented in chapter 3. The second is the censored case, where the data are analyzed using the proposed CDW regression model. A comparison of the results shows that if the experiment is based on censored data, then ignoring this censoring and applying the standard DW regression model will yield misleading results. Hence, for this case the CDW is recommended for censored data.

Although the DW regression model shows good performance for the count data with too many zero responses, there are some experiments where there is an interest in distinguishing between the generating processes for the zero and the non-zero counts, which can be explained using the mixture models. Hence, chapter 5 introduces two modifications for the case of excessive zero counts: namely, the ZIDW and HDW regression models. Some simulation studies and real data applications have been carried out to evaluate the behavior of these proposed models and compare them to the corresponding zero-inflated and hurdle models from the Poisson and NB. The ZIDW and HDW show a well fitting for these situations.

Altogether, the DW regression models introduced so far are based on investigating the effect of covariates through one of its parameter. Thus, this model is on a different scale compared to the GLMs, which study the effect on the mean. Therefore, to achieve an equivalent scale, chapter 6 suggests structuring the regression model through the median. The median is considered rather than the mean due to the common skewness nature of count data. Hence, the median is more representative than the mean as a measure for location for most discrete data analyses. The result for the simulation and numerical examples from MDW are promising compared to the Poisson and NB regression models.

7.2 Recommendations for future research

Although this thesis has covered many important and interesting aspects of the DW model, there are points worthy of further study. Some of the ideas that deserve further attention are listed below.

- The maximum likelihood approach has been applied for the inferences in this thesis. However, the *Expectation-Maximization (EM)* algorithm, which is commonly applied for mixture models, can be considered for estimating the parameters of the DW models. This is due to the similarity between the mixture and DW models both of which are based on a summation of the log in their log-likelihood (Equation 2.11).

- It might be useful and more general to consider both parameters q and β of the DW as functions of the covariates.

- The data in subsection 3.9.4 motivates us to try a mixture of DW components.

- Different sets of covariate might be considered to affect q and π in chapter 5, that is, $\boldsymbol{X} \neq \boldsymbol{Z}$.

- Although chapter 6 concerns with the MDW regression based on the standard DW likelihood, censored and zero-inflation likelihood might be considered for the median regression.

- Further extensions of this study, which has applied with most of the common regression models, might be considered. For example, the DW regression models could be considered for bivariate counts, where an experiment results in two joint responses. Another topic for consideration might be the variable selection. Also, the DW model could be considered for the time series of counts.

Bibliography

1. Aa, M. A. and Naing, N. N. Analysis death rate of age model with excess zeros using zero inflated negative binomial and negative binomial death rate: Mortality aids co-infection patients, kelantan malaysia. *Procedia Economics and Finance*, 2:275–283, 2012.

2. Abdel-Aty, M. A. and Radwan, A. E. Modeling traffic accident occurrence and involvement. *Accident Analysis & Prevention*, 32(5):633–642, 2000.

3. Allison, P. D. *Logistic regression using SAS: Theory and application*. SAS Institute, 2012.

4. Atkinson, A. C. *Plots, transformations, and regression: an introduction to graphical methods of diagnostic regression analysis*. Clarendon Press Oxford, 1985.

5. Bailey, B. A model for function word counts. *Applied statistics*, pages 107–114, 1990.

6. Barbiero, A. *DiscreteWeibull: Discrete Weibull Distributions (Type 1 and 3)*, 2015. URL http://CRAN.R-project.org/package=DiscreteWeibull. R package version 1.0.1.

7. Bilgic, A. and Florkowski, W. J. Application of a hurdle negative binomial count data model to demand for bass fishing in the southeastern united states. *Journal of environmental management*, 83(4):478–490, 2007.

8. Bracquemond, C. and Gaudoin, O. A survey on discrete lifetime distributions. *International Journal of Reliability, Quality and Safety Engineering*, 10(01): 69–98, 2003.

9. Brännäs, K. Limited dependent poisson regression. *The Statistician*, pages 413–423, 1992.

10. Byers, A. L., Allore, H., Gill, T. M., and Peduzzi, P. N. Application of negative binomial modeling for discrete outcomes: a case study in aging research. *Journal of clinical epidemiology*, 56(6):559–564, 2003.

11. Cameron, A. C. and Johansson, P. Count data regression using series expansions: with applications. *Journal of Applied Econometrics*, 12:203–223, 1997.

12. Cameron, A. C. and Trivedi, P. K. Econometric models based on count data. comparisons and applications of some estimators and tests. *Journal of applied econometrics*, 1(1):29–53, 1986.

13. Cameron, A. C. and Trivedi, P. K. *Regression analysis of count data*. Cambridge university press, 2013.

14. Canale, A. and Dunson, D. B. A bayesian nonparametric model for count functional data. In *46TH SCIENTIFIC MEETING OF THE ITALIAN STATISTICAL SOCIETY*, 2012.

15. Caudill, S. B. and Mixon Jr, F. G. Modeling household fertility decisions: Estimation and testing of censored regression models for count data. *Empirical Economics*, 20(2):183–196, 1995.

16. Cheol Jung, B., Jhun, M., and Heun Song, S. Testing for overdispersion in a censored poisson regression model. *Statistics*, 40(6):533–543, 2006.

17. Chipeta, M. G., Ngwira, B. M., Simoonga, C., and Kazembe, L. N. Zero adjusted models with applications to analysing helminths count data. *BMC research notes*, 7(1):856, 2014.

18. Cragg, J. G. Some statistical models for limited dependent variables with application to the demand for durable goods. *Econometrica: Journal of the Econometric Society*, pages 829–844, 1971.

19. Croissant, Y. *Ecdat: Data Sets for Econometrics*, 2015. URL http://CRAN.R-project.org/package=Ecdat. R package version 0.2-9.

20. Da Silva, M. F., Ferrari, S. L., and Cribari-Neto, F. Improved likelihood inference for the shape parameter in Weibull regression. *Journal of Statistical Computation and Simulation*, 78(9):789–811, 2008.

21. Dayton, C. M. Model comparisons using information measures. *Journal of Modern Applied Statistical Methods*, 2(2):2, 2003.

22. Dunn, P. K. and Smyth, G. K. Randomized quantile residuals. *Journal of Computational and Graphical Statistics*, 5(3):236–244, 1996.

23. Efron, B. Double exponential families and their use in generalized linear regression. *Journal of the American Statistical Association*, 81(395):709–721, 1986.

24. Englehardt, J., Swartout, J., and Loewenstine, C. A new theoretical discrete growth distribution with verification for microbial counts in water. *Risk Analysis*, 29(6):841–856, 2009.

25. Englehardt, J. D. and Li, R. The discrete Weibull distribution: an alternative for correlated counts with confirmation for microbial counts in water. *Risk Analysis*, 31(3):370–381, 2011.

26. Englehardt, J. D., Ashbolt, N. J., Loewenstine, C., Gadzinski, E. R., and Ayenu-Prah, A. Y. Methods for assessing long-term mean pathogen count in drinking water and risk management implications. *Journal of water and health*, 10(2):197–208, 2012.

27. Famoye, F. Restricted generalized Poisson regression model. *Communications in Statistics-Theory and Methods*, 22(5):1335–1354, 1993.

28. Famoye, F. and Wang, W. Censored generalized poisson regression model. *Computational statistics & data analysis*, 46(3):547–560, 2004.

29. Ferrari, S. and Cribari-Neto, F. Beta regression for modelling rates and proportions. *Journal of Applied Statistics*, 31(7):799–815, 2004.

30. Garay, A. M., Hashimoto, E. M., Ortega, E. M., and Lachos, V. H. On estimation and influence diagnostics for zero-inflated negative binomial regression models. *Computational Statistics & Data Analysis*, 55(3):1304–1318, 2011.

31. Gardner, W., Mulvey, E. P., and Shaw, E. C. Regression analyses of counts and rates: Poisson, overdispersed poisson, and negative binomial models. *Psychological bulletin*, 118(3):392, 1995.

32. Greene, W. H. *Econometric analysis.* Pearson Education India, 2003.

33. Grunwald, G. K., Bruce, S. L., Jiang, L., Strand, M., and Rabinovitch, N. A statistical model for under-or overdispersed clustered and longitudinal count data. *Biometrical Journal*, 53(4):578–594, 2011.

34. Hand, D. J., Daly, F., McConway, K., Lunn, D., and Ostrowski, E. *A handbook of small data sets*, volume 1. CRC Press, 1993.

35. Hilbe, J. *Negative binomial regression*. Cambridge University Press, 2011.

36. Hilbe, J. M. *COUNT: Functions, data and code for count data.*, 2014. URL http://CRAN.R-project.org/package=COUNT. R package version 1.3.2.

37. Hilbe, J. M. *Modeling Count Data*. Cambridge University Press, 2014.

38. Hosmer Jr, D. W. and Lemeshow, S. *Applied logistic regression*. John Wiley & Sons, 2004.

39. Hu, M.-C., Pavlicova, M., and Nunes, E. V. Zero-inflated and hurdle models of count data with extra zeros: examples from an hiv-risk reduction intervention trial. *The American journal of drug and alcohol abuse*, 37(5):367–375, 2011.

40. Hutchinson, M. K. and Holtman, M. C. Analysis of count data using poisson regression. *Research in nursing & health*, 28(5):408–418, 2005.

41. Jackman, S. pscl: Classes and methods for r developed in the political science computational laboratory, 2008.

42. Johnson, N. L., Kemp, A. W., and Kotz, S. *Univariate discrete distributions*, volume 444. John Wiley & Sons, 2005.

43. Khan, M. A., Khalique, A., and Abouammoh, A. On estimating parameters in a discrete Weibull distribution. *IEEE Transactions on Reliability*, 38(3): 348–350, 1989.

44. Kong, M., Xu, S., Levy, S. M., and Datta, S. Gee type inference for clustered zero-inflated negative binomial regression with application to dental caries. *Computational Statistics & Data Analysis*, 85:54–66, 2015.

45. Kuha, J. Aic and bic comparisons of assumptions and performance. *Sociological Methods & Research*, 33(2):188–229, 2004.

46. Kulasekera, K. Approximate MLE's of the parameters of a discrete Weibull distribution with type i censored data. *Microelectronics Reliability*, 34(7): 1185–1188, 1994.

47. Lam, K., Xue, H., and Bun Cheung, Y. Semiparametric analysis of zero-inflated count data. *Biometrics*, 62(4):996–1003, 2006.

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

49. Lawless, J. F. Negative binomial and mixed Poisson regression. *Canadian Journal of Statistics*, 15(3):209–225, 1987.

50. Lee, E. T. and Wang, J. *Statistical methods for survival data analysis*, volume 476. John Wiley & Sons, 2003.

51. Loeys, T., Moerkerke, B., De Smet, O., and Buysse, A. The analysis of zero-inflated count data: Beyond zero-inflated Poisson regression. *British Journal of Mathematical and Statistical Psychology*, 65(1):163–180, 2012.

52. Mahmoud, M. and Alderiny, M. On estimating parameters of censored generalized poisson regression model. *Applied Mathematical Sciences*, 4(13): 623–635, 2010.

53. McDowell, A. et al. From the help desk: hurdle models. *The Stata Journal*, 3(2):178–184, 2003.

54. Montgomery, D. C. and Peck, E. A. *Introduction to linear regression analysis*. John Wiley and Sons, New York; Chichester, 1982.

55. Mullahy, J. Specification and testing of some modified count data models. *Journal of econometrics*, 33(3):341–365, 1986.

56. Nakagawa, T. and Osaki, S. The discrete Weibull distribution. *IEEE Transactions on Reliability*, 24(5):300–301, 1975.

57. Nelder, J. A. and Wedderburn, R. W. Generalized linear models. *Journal of the Royal Statistical Society. Series A*, pages 370–384, 1972.

58. Ospina, R. and Ferrari, S. L. A general class of zero-or-one inflated beta regression models. *Computational Statistics & Data Analysis*, 56(6):1609–1623, 2012.

59. Padgett, W. and Spurrier, J. D. On discrete failure models. *Reliability, IEEE Transactions on*, 34(3):253–256, 1985.

60. Posada, D. and Buckley, T. R. Model selection and model averaging in phylogenetics: advantages of akaike information criterion and bayesian approaches over likelihood ratio tests. *Systematic biology*, 53(5):793–808, 2004.

61. R Core Team. *R: A Language and Environment for Statistical Computing*. R Foundation for Statistical Computing, Vienna, Austria, 2014. URL http://www.R-project.org/.

62. Raciborski, R. et al. Right-censored poisson regression model. *Stata Journal*, 11(1):95, 2011.

63. Rigby, R. and Stasinopoulos, D. Generalized additive models for location, scale and shape. *Journal of the Royal Statistical Society: Series C (Applied Statistics)*, 54(3):507–554, 2005.

64. Rinne, H. *The Weibull distribution: a handbook*. CRC Press, 2008.

65. Roy, D. Discrete Rayleigh distribution. *IEEE Transactions on Reliability*, 53 (2):255–260, 2004.

66. Sáez-Castillo, A. and Conde-Sánchez, A. A hyper-Poisson regression model for overdispersed and underdispersed count data. *Computational Statistics & Data Analysis*, 61:148–157, 2013.

67. Sáez-Castillo, A. J. and Conde-Sánchez, A. Detecting over-and under-dispersion in zero inflated data with the hyper-poisson regression model. *Statistical Papers*, pages 1–15, 2015.

68. SAFFAR, S. E., ADNAN, R., and GREENE, W. Parameter estimation on hurdle poisson regression model with censored data. *Jurnal Teknologi*, 57(1), 2012.

69. Saffari, S. E. and Adnan, R. Zero-inflated poisson regression models with right censored count data. *Matematika*, 27:21–29, 2011.

70. Saffari, S. E. and Adnan, R. Zero-inflated negative binomial regression model with right censoring count data. *Journal of Materials Science and Engineering B*, 1(4):551–554, 2011.

71. Saffari, S. E., Adnan, R., and Greene, W. Hurdle negative binomial regression model with right censored count data. *SORT. 2012, vol. 36, núm. 2*, 2012.

72. Saffari, S. E., Adnan, R., Greene, W., and Ahmad, M. H. A poisson regression model for analysis of censored count data with excess zeroes. *Jurnal Teknologi*, 63(2), 2013.

73. Sato, H., Ikota, M., Sugimoto, A., and Masuda, H. A new defect distribution metrology with a consistent discrete exponential formula and its applications. *IEEE Transactions on Semiconductor Manufacturing*, 12(4):409–418, 1999.

74. Schmidt, M. and Hurling, R. A spatially-explicit count data regression for modeling the density of forest cockchafer (melolontha hippocastani) larvae in the hessian ried (germany). *Forest Ecosystems*, 1(1):1, 2014.

75. Sellers, K. F. and Shmueli, G. A flexible regression model for count data. *Annals of Applied Statistics*, 4(2):943–961, 2010.

76. Sellers, K. F. and Shmueli, G. Predicting censored count data with com-poisson regression. *Robert H. Smith School Research Paper No. RHS-06-129*, 2010.

77. Serfling, R. J. *Approximation theorems of mathematical statistics*. John Wiley & Sons, 1980.

78. Silva, J. S. and Covas, F. A modified hurdle model for completed fertility. *Journal of Population Economics*, 13(2):173–188, 2000.

79. Spyroglou, I. I., Chatzimichail, E. A., Spanou, E., Paraskakis, E., and Rigas, A. G. Ridge regression and bootstrapping in asthma prediction. In *New Developments in Pure and Applied Mathematics INASE Conference proceedings (MMSSE 15), Vienna, Austria*, pages 44–48, 2015.

80. Staub, K. E. and Winkelmann, R. Consistent estimation of zero-inflated count models. *Health economics*, 22(6):673–686, 2013.

81. Steenberger, M. Maximum likelihood programming in r. *University of North Carolina, Chapel Hill*, 2006.

82. Stein, W. E. and Dattero, R. A new discrete Weibull distribution. *IEEE transactions on reliability*, 33(2):196–197, 1984.

83. Steinberg, W. J. *Statistics alive!* Sage Publications, 2010.

84. Terza, J. V. A tobit-type estimator for the censored poisson regression model. *Economics Letters*, 18(4):361–365, 1985.

85. Tin, A. Modeling zero-inflated count data with underdispersion and overdispersion. In *SAS Global Forum Proceedings*, 2008.

86. Vanegas, L. H., Rondón, L. M., and Cordeiro, G. M. Diagnostic tools in generalized weibull linear regression models. *Journal of Statistical Computation and Simulation*, 83(12):2315–2338, 2013.

87. Xie, H., Tao, J., McHugo, G. J., and Drake, R. E. Comparing statistical methods for analyzing skewed longitudinal count data with many zeros: An example of smoking cessation. *Journal of substance abuse treatment*, 45(1): 99–108, 2013.

88. Zeileis, A., Kleiber, C., and Jackman, S. Regression models for count data in r. *Journal of Statistical Software*, 27, 2008.

89. Zeileis, A. and Hothorn, T. Diagnostic checking in regression relationships. *R News*, 2(3):7–10, 2002. URL http://CRAN.R-project.org/doc/Rnews/.

90. Zeileis, A., Kleiber, C., and Jackman, S. Regression models for count data in r. 2007.

91. Zorn, C. J. Evaluating zero-inflated and hurdle Poisson specifications. *Midwest Political Science Association*, 18(20):1–16, 1996.

Appendix

In this appendix we obtain the following,
$\dfrac{\partial^2 \ell(q_i, \beta)}{\partial q_i^2}$, $\dfrac{\partial^2 \ell(q_i, \beta)}{\partial q_i \partial \beta}$ and $\dfrac{\partial^2 \ell(\boldsymbol{\alpha}, \beta)}{\partial \beta^2}$

- From Equation 3.13,

$$
\begin{aligned}
\frac{\partial^2 \ell(q_i, \beta)}{\partial q_i^2} &= \frac{\partial}{\partial q_i}\left[\frac{\partial \ell(q_i, \beta))}{\partial q_i}\right] \\
&= \frac{y_i^\beta (y_i^\beta - 1) q_i^{y_i^\beta - 2} - (y_i + 1)^\beta \left((y_i + 1)^\beta - 1\right) q_i^{(y_i+1)^\beta - 2}}{q_i^{y_i^\beta} - q_i^{(y_i+1)^\beta}} \\
&- \frac{\left(y_i^\beta q_i^{y_i^\beta - 1} - (y_i + 1)^\beta q_i^{(y_i+1)^\beta - 1}\right)\left[y_i^\beta q_i^{y_i^\beta - 1} - (y_i + 1)^\beta q_i^{(y_i+1)^\beta - 1}\right]}{\left(q_i^{y_i^\beta} - q_i^{(y_i+1)^\beta}\right)^2}
\end{aligned}
$$

- From Equation 3.13,

$$
\begin{aligned}
\frac{\partial^2 \ell(q_i,\beta)}{\partial q_i \partial \beta} &= \frac{\partial}{\partial \beta}\left[\frac{\partial \ell(q_i,\beta))}{\partial q_i}\right] \\
&= \frac{\log(y_i)\left[y_i^{\beta} q_i^{y_i^{\beta}-1} + y_i^{2\beta} q_i^{y_i^{\beta}-1}\log(q_i)\right] - \log(y_i+1)\left[(y_i+1)^{\beta} q_i^{(y_i+1)^{\beta}-1} + (y_i+1)^{2\beta} q_i^{(y_i+1)^{\beta}-1}\log(q)\right]}{q_i^{y_i^{\beta}} - q_i^{(y_i+1)^{\beta}}} \\
&- \frac{\left(y_i^{\beta} q_i^{y_i^{\beta}}\log(y_i)\log(q_i) - (y_i+1)^{\beta} q_i^{(y_i+1)^{\beta}}\log(y_i+1)\log(q_i)\right)\left[y_i^{\beta} q_i^{y_i^{\beta}-1} - (y_i+1)^{\beta} q_i^{(y_i+1)^{\beta}-1}\right]}{\left(q_i^{y_i^{\beta}} - q_i^{(y_i+1)^{\beta}}\right)^2}
\end{aligned}
$$

- From Equation 3.15,

$$
\begin{aligned}
\frac{\partial^2 \ell(\boldsymbol{\alpha},\beta)}{\partial \beta^2} &= \frac{\partial}{\partial \beta}\left[\frac{\partial \ell(\boldsymbol{\alpha},\beta))}{\partial \beta}\right] \\
&= \frac{(\log(y_i))^2\left[y_i^{2\beta} q_i^{y_i^{\beta}}(\log(q_i))^2 + y_i^{\beta} q_i^{y_i^{\beta}}\log(q_i)\right] - (\log(y_i+1))^2\left[(y_i+1)^{2\beta} q_i^{(y_i+1)^{\beta}}(\log(q_i))^2 - (y_i+1)^{\beta} q_i^{(y_i+1)^{\beta}}\log(q)\right]}{q_i^{y_i^{\beta}} - q_i^{(y_i+1)^{\beta}}} \\
&- \frac{\left(y_i^{\beta} q_i^{y_i^{\beta}}\log(y_i)\log(q_i) - (y_i+1)^{\beta} q_i^{(y_i+1)^{\beta}}\log(y_i+1)\log(q_i)\right)\left[y_i^{\beta} q_i^{y_i^{\beta}}\log(y_i)\log(q_i) - (y_i+1)^{\beta} q_i^{(y_i+1)^{\beta}}\log(y_i+1)\log(q_i)\right]}{\left(q_i^{y_i^{\beta}} - q_i^{(y_i+1)^{\beta}}\right)^2}
\end{aligned}
$$