2. Review of Literature

More studies have been made on urbanisation across the world, the review of such studies provides in insight for the research in the field of urbanisation. An attempt has been made in this chapter to list out the systematic development of studies on urbanisation, methodologies adopted in the study and results and summarised.

S. Gurmu (1971) has derived scores test for extra Poisson variation in the positive or truncated at zero Poisson regression (Positive Poisson regression) model against truncated at zero negative binomial family alternatives. He has developed the size corrected tests for over dispersion. He suggested that the Monte Carlo experiments indicate the size correction, based on the asymptotic expansions of the score function is effective in improving the accuracy of the size and power of the tests in small samples. He extended the previous work in specification tests for count models by developing the test for overdispersion in the Poisson regression model that is truncated at below zero. He suggested that the truncated at zero negative binomial family is useful for analysis of data too heterogeneous to be fitted by the positive Poisson model. He has studied the specification and estimation of the Poisson and negative binomial regression models that are truncated at the left at zero. He has taken strike data and industrial output to study the relationship between them. He has applied positive Poisson and truncated zero negative binomial model. He has used strike data to the empirical analysis and find that the positive Poisson model is not adequate for overdispersion.

Davis and Swanson (1972) observed that the growth of cities may be viewed as stochastic process and the distribution of city sizes as the outcome of that process. He showed a special theory of regional growth to lead to particular stochastic process
in the growth rates of population centres. He find that the theory indicates a definite
time path for the size of a city’s labour force, which, when conjoined with a
specification of the distribution of efficiency progress, permits the characterization of
city growth rate as a stochastic process. Each city is assumed to be constrained in its
accumulation plan by the amount of savings generated within that city, and each city
saves a constant fraction of its real product. Net investment it identically the time
derive of the capital stock capital goods are indestructible. He suggested that the
distribution of proportional growth rates among centres has been log normally
distributed.

Edward L. Frome, Michael H. Kutner and John J. Beauchamp (1973), used the
principle of maximum likelihood is to obtain estimates of the parameters in the
Poisson regression model. They find that the maximum likelihood estimates are
equivalent to those obtained by minimization of a quadratic form which reduces to a
modified chi square under the Poisson assumption. Also they find, both of these
estimation procedures are equivalent to a weighted least squares analysis. They
proposed the approximate tests of the assumed Poisson variation and goodness of fit.
They discussed the applications of the estimation procedure to linear and nonlinear
regression models.

Richardson H.W (1973) suggested the new theories capable of generating the
lognormal, Pareto and rank size distribution’s. He analysed the theory of hierarchy,
stochastic, economic and allometric growth models to represent city size distribution.
He focussed to the Markov chain models of interurban migration and capital flows.
He suggested that the Pareto may also be preferred if the upper tail is highly skewed.
It is feasible to devise a model based on the hypothesis that the size distribution of
cities is a function of the age distribution of cities.
Kenneth T. Rosen and Mitchel Rensick (1980) examined the Pareto and primacy of the size distribution of cities. He confirmed that the Pareto exponents are a better reflection of the overall city size distribution that are primacy measures. They are two measures of primacy: 1. the ratio of the largest city to the sum of the top five cities (primacy I) and 2. The ratio of the largest to the sum of the top 50 cities (primacy II). Primacy II should provide a better reflection of how the top city compares to the distribution as a whole, rather than just to rest of the upper tail. He has examined 44 countries with total population, area, GNP per capita and railway mileage to study the validity of the rank size rule. He find that countries which rank significantly higher using the primacy II have relatively small intermediate sized cities and Pareto exponents are less that the mean. He suggested that the single Pareto exponent exposes a feature which both primacy measures together were needed to detect. From their results they suggested that the growth rates of cites are positively correlated to size in the countries. It may be reflected scale economics. He observed that heavily populated countries, intermediate sized cities are likely to develop since the major cities may reach a size where negative externalities discourage growth in the cities. He find that the value of the Pareto exponent is quite sensitive to the definition of the city. The significance of non-linear terms in variants of the Pareto distribution also indicate that the rank size rule is only a first approximation to a complete characterization of the low correlation between primacy and Pareto measures. He suggested that the large cities are moving faster than small cities in most of the countries.

E.L. Frome (1983), studied the IRLS algorithm is equivalent to using the method of scoring to obtain maximum likelihood estimates in the Poisson regression model. The approach has been considered to concentrate on describing the relation
between the dependent variable and the predictor variables through the regression model. He used deviance statistics as a measure of residual variation to develop an ANOVA for the Poisson regression model to study the lung cancer death rates.

Gershon Alperovich (1984, 1989) proposed joint test and new testing procedure for testing the validity of the rank size rule for city size distribution. He used the population of cities have 100000 and more inhabitants in different countries to study the city size distribution under Pareto distribution. From his empirical analysis he confirmed that the rank size distribution when the Pareto exponent is equal to unit. The constant of the distribution is equal to the average product of a city’s rank and its population. He explained if excluding cities form the lower size range of the distribution constitutes misspecification of the estimated equation which leads to biased estimates. This will be the case if small cities are consistently smaller or larger than implied by the rank size rule. He observed that the Pareto exponent turned out to be significantly different a unit, the joint test showed that the rank size distribution is not confirmed by the data. His findings support the validity of other studies which have accepted or rejected the rank size distribution relying on whether the Pareto exponent is equal to or is different from a unit. He also studied city size distribution using rank size distribution under standard statistical model and improved model and improved model \( \ln \left( 1 + \frac{1}{m} \right) = \ln \left( \frac{B_{m+1}}{B_m} \right) + \alpha \left( \frac{P_{m+1}}{P_m} \right) \). Where \( R_m \) and \( P_m \) are the rank and population of the \( m \)th level (\( m=1,2,...,M \)) respectively. \( A \) is a parameter and \( \alpha \) is the Pareto exponent. He has used SMSAs data population for the year 1970 and 1980.

He concluded that the rank size distribution is confirmed by the improved model, it is not supported by the standard model and standard model is highly sensitive to exclusion of SMSAs and the estimates and test statistics of the improved model are
relatively robust to exclusion of SMSAs from the lower tail of the distribution. He showed from his empirical analysis that the Pareto distribution provides a very accurate description of the distribution of population sizes of urban areas. From his findings the standard model of rank size distribution is not an accurate representation of the size distribution of urban areas and the new model is preferred to describe the city size distribution.

J.F. Lawless (1987) has studied negative binomial regression models and examined efficiency and robustness properties of inference procedures based on them. He suggested that the maximum likelihood, quasi-likelihood or weighted least squares method for dealing with extra Poisson variation or overdispersion. He focussed over dispersion or extra Poisson variation. He has used ship damage incidents data to the empirical study.

C. Dean and J.F. Lawless (1989) have developed tests for detecting extra-Poisson variation in Poisson regression models. They have obtained the tests as score tests against arbitrary mixed Poisson alternatives, and provided approximations for computing the significance level and power of the tests against the negative binomial alternatives. The model assumed that the random effects had finite the first and second moments. If the random effects follow a gamma distribution, then the random variable has a negative binomial distribution. They have obtained test for over dispersion and other generalised model using $t$-test.

Norman Breslow (1990) has developed the test statistics for evaluating the significance of added regression equation. He estimated the regression co-efficient by quasi-likelihood method. He indicated three tests under likelihood and quasi likelihood theory. He has explained the Wald test based on the comparison of
estimated coefficients with their standard errors, the likelihood ratio test based on comparison of deviances under full and reduced model. He has developed a score test to overdispersed quasi-likelihood models. He concluded that the quasi-likelihood analysis with the correct specification of the variance function yielded reasonably unbiased estimates of the regression coefficients with their standard errors in moderate sized samples.

David F. Signorini (1991) obtained the Fisher's information matrix based on the moment generating function of the distribution of covariates for the Poisson regression model. They were used the asymptotic variance of the maximum likelihood estimate of the parameters to calculate the sample size. They have derived the methods for calculating sample size for various distributions of a single covariate, and for a family of multivariate exponential-type distributions of multiple covariates.

Bradley Efron (1992), discussed the asymmetric maximum likelihood (AML) for estimating and describing overdispersion in generalized linear regression models. They studied the asymmetric maximum likelihood is a method of fitting regressions for the conditional percentiles of the response variable as a function of the predictors (e.g., the conditional 90th percentile of $y$ given $x$). They proposed the distances between the various regression percentiles give a direct assessment of overdispersion. They suggested that the AML estimates are easy to compute and relate nicely to the usual maximum likelihood estimates for generalized linear regression.

C.B. Dean (1992) has generalized the idea to the over dispersed exponential family. He has derived a method for obtaining tests for over dispersion with respect to a natural exponential family. He obtained various test for extra Poisson and extra binomial model. He suggested that the derivation of all the tests for extra Poisson
variation assumes that the regression specification is correct. He found that the correlated binomial also stems from a logical explanation of the extra binomial variation. He has claimed that the tests are powerful against arbitrary alternative mixture models where only the first and second moments of the mixed distribution are specified.

Felix Famoye (1993), formulated and studied a restricted generalized Poisson regression (RGPR) model for count data that are affected by a number of known predictor variables. He discussed the method of maximum likelihood, least square method and method of moments are to estimate the RGPR model parameter. He applied goodness of fit statistics to examine the adequacy of the RGPR model. He suggested that the RGPR model provide a better fit than the Poisson and compound Poisson regression models.

Eduardo S. Schwartz and Walter N. Torous (1993) used the Poisson regression model to estimate the parameters of a proportional hazards model for prepayment and default decisions. They suggested that the Poisson regression for grouped survival data has several advantages over partial likelihood methods. First, while dealing with time dependent covariates. It is considerably more efficient in terms of computations. Second, it is possible to estimate full hazard models which include, for example, functions of time as well as multiple time scales (i.e., age of the loan and calendar time). In a straightforward manner than partial likelihood methods for ungrouped data. Third, Poisson regression can be used to estimate non proportional hazards models such as additive excess risk specifications.

Brad McNeney and John Petkau (1994), applied overdispersed Poisson regression models to study the relationships between ambient levels of air pollution and human health. They studied the estimation of the dispersion parameter,
estimation of regression coefficients and their standard errors, and the performance of model selection tests. From the preliminary work, they concluded that the estimation of the dispersion parameter should be based on Pearson's chi-square statistic rather than the Poisson deviance.

S.P Miaou (1994) has studied the performance of Poisson and negative binomial (NB) regression models in establishing the relationship between truck accidents and geometric design of road sections. He used Maximum likelihood method to estimate the unknown parameters of these models. He suggested that the Poisson regression model be used as an initial model for developing the relationship. He found that the unsatisfactory property of linear regression models has led to the investigation of the Poisson and negative binomial (NB) regression models in recent studies. He suggested that the Zero inflated Poisson (ZIP) regression is more flexible than the Poisson regression, the interpretation of the ZIP regression can be difficult.

Rainer Winkelmann and Klaus F. Zimmermann (1994), deals with the estimation of single equation models in which the counts are regressed on a set of observed individual characteristics such as age, gender, or nationality. They proposed a generalized event count model to simultaneously allow for a wide class of count data models and account for over- and underdispersion. They pointed that the application of the simple Poisson model may lead to seriously biased asymptotic t-ratios, where the direction and the size of the bias depend on the nature of the deviation from the mean – variance equality. They suggested that the generalized event count model in the case of deviation from the mean - variance equality.

Winkelmann and Zimmermann (1995) deals with statistical methods for modelling individual behaviour when the endogenous variable is a nonnegative
integer. Several approaches such as Poisson, robust Poisson, negative Binomial, hurdle Poisson, truncated at zero Poisson are discussed with a focus on specification, estimation and testing. Semi parametric approaches have been proposed.

Weiren Wang, Felix Famoye (1997), studied the generalized Poisson regression model for household fertility decisions. They discussed the approximate tests for the dispersion and goodness-of-fit measures for comparing alternative models. From the empirical results they showed that the generalized Poisson regression model is a useful general framework and is more appropriate than the standard Poisson regression model for the present fertility data.

Claudia Czado and Aleksey Min (2000), studied regression models associated with the generalized Poisson distribution. They proved consistency and asymptotic normality of the maximum likelihood estimators in zero-inflated generalized Poisson regression models. They compared the score test for detecting zero-inflation or zero-deflation in a zero-modified Poisson (ZMP) model to the Wald test based on the asymptotic results in a simulation study. They showed that for small and moderate sample sizes, the Wald test for detecting zero-inflation or zero-deflation in a ZMP regression model is considerable more powerful than the score test.

D. Wayne Osgood (2000) has introduced the use of regression models based on the Poisson distribution as a tool for resolving common problems in analyzing aggregate crime rates. He found that the negative binomial variant of Poisson regression effectively resolved difficulties that arise in ordinary least-squares analyses. Poisson-based regression models of counts of offenses are preferable because they are built on assumptions about error distributions that are consistent with the nature of event counts. He has proved that the Poisson-based regression models will have broad applicability for the study of crime at the aggregate level.
K.P. Singh, J.T. Wulu and A.A. Bartolucci (2001), explored the assumptions and properties of GPR from a public health point. Poisson regression and generalised Poisson regression (GPR) models are used to identify relationships between covariates and the response variable. The Wald t test is used for testing the significance of each regression parameter. The GPR model performs better than the GPR model based on the dispersion parameter and the goodness of fit measure for each data. They described nonlinear regression techniques appropriate for the analysis of household trip frequencies data and showed that the GPR Model gives better fits than standard Poisson regression model when estimating determinants of household trip frequencies.

Daniel B. Hall and Kenneth S. Berenhaut (2002), has described zero-inflated Poisson and binomial regression models that include random effects to account for excess zeros and additional sources of heterogeneity in the data. They proposed a general score test for the null hypothesis that variance components associated with these random effects are zero in a ZIP-mixed or ZIB-mixed regression model. They studied the new test that reduces to an alternative to the overdispersion test for a zero-inflated Poisson model with random intercept. They also examined the general test in the special case of the zero-inflated binomial model with random intercept and proposed an overdispersion test in that context which is based on a beta-binomial alternative.

J.T. Wulu, Jr., K.P. Singh, F. Famoye, T.N. Thomas and G. McGwin (2002), used Poisson Regression (PR), Negative Binomial Regression (NBR), and Generalized Poisson Regression (GPR) for assessing the effects of risk factors on agricultural injuries from farm injury data. They find that the dispersion parameter
estimates and their standard errors for GPR models were consistently smaller than that of NBR models. Also find that the estimated dispersion parameters in the NBR and GPR models were positive and significantly different from zero. The Estimated goodness-of-fit measures showed that GPR models outperformed the NBR and PR models. They were used the standard error, Wald t-statistics, dispersion parameters, Pearson's and generalized chi-square, deviance, and log-likelihood estimates for their study.

Amy L. Byers, Heather Allore, Thomas M. Gill, Peter N. Peduzzi (2003), applied Negative Binomial regression to the clinical trial data. They have find that the Negative binomial regression is the best fit than the Poisson regression model when the overdispersion is present. They concluded that the Negative binomial model is an alternative approach for the discrete data where the overdispersion is a problem, provided that the model is correctly specified and adequately fits the data.

Harald Heinzl, Martina Mittlbock (2003), studied R-squared measures for over and underdispersion. They generalized the adjusted R-squared measure for over and underdispersed Poisson regression model in an attempt to adjust for bias in the presence of small samples or many covariates. They investigated the performance of R-squared measures for Poisson regression model under over or underdispersion when ignoring the fact of over or underdispersion and compare the results with the performance of the proposed adjustment R-square measures. They were used deviance statistic and Pearson chi-square statistic to the overdispersed Poisson regression model.
Felix Famoye, John T. Wulu, Jr. and Karan P. Singh (2004) studied the generalized Poisson regression (GPR) model for identifying the relationship between the number of accidents and some covariates. They selected a random sample of drivers aged sixty-five years or older. The data in the sample has information on many variables including the number of accidents, demographic information, driving habits, and medication. They find that the GPR model performs as good as or better than the other regression models based on the test for the dispersion parameter and the goodness-of-fit measure for the accident data. They applied Negative Binomial Regression model to the data and found the results to be similar to that of Generalised Poisson Regression model.

Jansakul. N and Hinde. J.P (2004), discussed the possible use of the Newton-Raphson algorithm to obtain maximum likelihood estimates of the linear mean-variance negative binomial (NB1) regression model and of the overdispersion parameter. They proposed the use of a half-normal plot with a simulated envelope for checking the adequacy of fitted NB1 models. They applied these procedures to analyze data of a number of embryos from an orange tissue culture experiment. From the empirical analysis they showed that the NB1 regression model with a cubic response function over the dose levels is consistent with the data.

Dianliang Deng and Sudhir R. Paul (2005) studied the goodness of fit tests of the generalized linear model that deals separately against zero-inflation and against over-dispersion. They deals with the class of zero-inflated overdispersed generalized linear models and proposed procedures based on score tests for selecting a model that fits such data. They considered a general over-dispersion model and specific overdispersion models for overdispersion. They obtained score tests (i) for zero-inflation in presence of overdispersion, (ii) for over-dispersion in presence of zero-
inflation, and (iii) simultaneously for zero-inflation and over-dispersion. They showed that for Poisson data the score test statistics, for testing over-dispersion in the presence of zero-inflation and that for testing simultaneously for zero-inflation and over-dispersion, using the zero-inflated over-dispersed generalized linear model with log-link is identical to the corresponding score test statistics using the zero-inflated negative binomial model. Also, for binomial data the score test statistics, for testing over-dispersion in the presence of zero-inflation and that for testing simultaneously for zero-inflation and over-dispersion, using the zero-inflated over-dispersed generalized linear model with logit link is identical to the corresponding score test statistics using the zero-inflated beta-binomial model. However, for testing zero-inflation in presence of over-dispersion the score test statistics obtained as special cases do not coincide with those obtained from the specific over-dispersed zero-inflated models. They conducted a small scale simulation experiment to study the small sample behaviours of the likelihood ratio statistic and the Wald test statistic in testing simultaneously for no zero-inflation and no over-dispersion.

Noriszura Ismail and Abdul Aziz Jemain (2005), studied the Poisson and Generalized Poisson regression models which are fitted, tested and compared on two types of Malaysian motor insurance claims count data as Own Damage (OD) and Third Party Bodily Injury (TPBI). The Poisson regression model for OD claims gives large values for Pearson chi-squares and deviance, indicating possible existence of overdispersion. They found that the Generalized Poisson was superior to the Poisson based on the results of goodness-of-fit tests. They concluded, the small deviance for Poisson regression model in TPBI claims implies that the model is adequate. Based on the likelihood ratio test, the likelihood ratio is insignificant, implying that the Poisson is adequate.
Felix Famoye and Karan P. Singh (2006), studied the generalized Poisson regression to dispersed count data. Zero-inflated Poisson and zero-inflated negative binomial regression models have been proposed for the situations where the data generating process results into too many zeros. They proposed a zero-inflated generalized Poisson (ZIGP) regression model to model domestic violence data with too many zeros. They provided maximum likelihood estimation method to estimated the model parameters. A score test is proposed to test whether the number of zeros is too large for the generalized Poisson model to adequately fit the domestic violence data. He finds the ZIP regression model is not an appropriate model for the domestic violence data. The ZIGP regression model fits the data better than the ZIP model with almost one fold increase in the value of the likelihood.

Claudia Czado, Vinzenz Erhardt, Aleksey Min and Stefan Wagner (2007), focused an extension of zero-inflated generalized Poisson (ZIGP) regression models for patent outsourcing data. They discussed the generalized Poisson (GP) models where dispersion is modelled by an additional model parameter. They were discussed the zero-inflated models, in which the overdispersion is assumed to be caused by an excessive number of zeros. They proposed tools for an exploratory data analysis on the dispersion and zero-inflation. They estimated the standard errors using the asymptotic normality of the estimates. They find that the extended ZIGP regression model will prove to be superior over GP and zero-inflated Poisson (ZIP) models, and even over ZIGP models.

Ilknur Ozmen and Felix Famoye (2007), presented an evaluation framework to the suitability of applying the Poisson, NB, GP, ZIP and ZIGP to zoological data set where the count data may exhibit evidence of many zeros and over-dispersion. They
provided estimation of the model parameters using the method of maximum likelihood (ML). They find that the GP regression model performs better than other count regression models based on the score test and the goodness of fit measure for zoological data. They suggested that the NB, GP, ZINB and ZIGP regression models can be appropriate for the over-dispersion situation. They applied the score test to check whether the ZIP regression model is a significant improvement over the Poisson regression model and finds that the ZIGP regression model provides an adequate (but not better than GP) fit to the data. They suggested the NB and GP regression models seem to perform better than the Poisson and ZIP regression models.

Liming Xiang, Andy H. Lee, Kelvin K. W. Yau and Geoffrey J. McLachlan (2007), proposed a score test for testing the ZIP mixed regression model against the zero-inflated negative binomial alternative. They have developed extensions of the ZIP regression model in which random effects are incorporated within the Poisson and binary components of the ZIP model to handle the clustered heterogeneous counts. Marginal models utilizing generalized estimating equations have also been suggested as alternatives to the inclusion of random effects. They studied score test for overdispersion. They suggested that in the case of overdispersion, the alternative ZINB mixed regression model may considered for fitting the heterogeneous and correlated data.

Joseph M. Hilbe and William H. Greene (2008) have studied Poisson regression model. They found that the Poisson models are inappropriate for many count modelling situations under the assumption of the distributional mean and variance. They suggested that the negative binomial regression model to the
overdispersed Poisson regression model. They proposed zero inflated model alternatives for overdispersed Poisson and negative binomial models. They proposed likelihood based test, Wald and lagrange multiplier test for the Poisson regression model.

Santosh C. Sutradhara, Nagaraj K. Neerchal, Jorge G. Morel (2008) focussed the goodness of fit statistic for evaluating model adequacy of overdispersion model. They find that the goodness of fit is a direct analogue of the Pearson chi-square statistic. They suggested that the goodness of fit test is applicable even when all the clusters are not all of the same sizes. They find that the goodness of fit test is performed well in the overdispersion model.

Byung-Jung Park, Dominique Lord (2009), have proposed an alternative formulation, that could capture the heterogeneity in crash count models through the use of finite mixture model. They suggested that the finite mixtures of Poisson or Negative binomial regression models are especially useful where count data were drawn from heterogeneous populations. They suggested that the finite mixture model to the overdispersed data.

Zhang Yang, James W. Hardin, Cheryl L. Addy (2009), have proposed a score test for overdispersion based on the Generalised Poisson regression model and compared the power of the test with the Likelihood ration test and Wald tests. They proposed that the score test has an advantage over the Likelihood ratio test and over the Wald test in that the score test only requires that the parameter of interest to estimated under the null hypothesis.
Marie M. Mahmoud and Mahmoud M. Alderiny (2010) studied the Censored generalized Poisson regression (CGPR) model for identifying the relationship between the dependent, and the independent variables. Based on the goodness-of-fit measure for the absent days data, they find that the CGPR model perform better than the CPR model in identifying daily wage, worker's age, experience years, family size, number of rooms in worker’s house, motivations system, department, industry trips and the reason for working in industry use associated with the number of absent days for worker. Also find the estimated dispersion parameter from the data is negative and it is significantly different from zero.

Wan Rozita WM, Rasimah A, Mazrura S, Lim KH, Thana S (2010), used maximum likelihood procedures to estimate the model parameters of a Poisson regression model. They studied the Negative Binomial distribution has been widely suggested as the alternative to the Poisson when there is overdispersion. They modelled the mortality cases as the dependent variable using Poisson and Negative Binomial regression and compare both of the models. They find that the mortality data in Poisson regression exhibit large ratio values between deviance to degree of freedom which indicate model misspecification or overdispersion. Also they find that this large ratio have been reduced in Negative Binomial regression. They suggested that the NB regression model provides a better account of the probability distribution of the data than the Poisson regression model.

Kimberly F. Sellers and Galit Shmueli (2010), proposed a regression model based on the Conway–Maxwell-Poisson (COM- Poisson) distribution to overdispersion or underdispersion. The COM-Poisson regression generalizes the well-known Poisson and logistic regression models, and is suitable for fitting count data with a wide range
of dispersion levels. It includes as special cases the Poisson, Bernoulli, and geometric distributions, as well as distributions with dispersions levels between these three well-known cases (governed by the dispersion parameter). They discussed the model estimation, inference, interpretation, and diagnostics, obtaining the fitted values, and testing for dispersion for a COM-Poisson regression over a standard Poisson regression.

Shahid Ullaha, Caroline F. Finch, Lesley Day (2010), studied the modified Poisson and Negative binomial models as a alternatives to Poisson regression for the analysis of falls count data. They have compared the finite mixture of Poisson and Negative binomial regression to the standard Negative binomial model. They have find that the Negative binomial model gives the better performance than Poisson, Zero inflated Poisson and Zero inflated Negative binomial models based on statistical fit. They have found that the Negative binomial model fit was also better than that achieved with finite mixture of Poisson and Negative binomial regression models.

Zhao Wei-hua, Feng Yu, Li Ze-an (2010), introduced the zero truncated generalized Poisson(ZTGP) regression model and its estimation method. They were obtained an algorithm for estimating parameters and two score tests were developed for testing the ZTP regression model against the ZTGP regression model, and for testing the significance of regression coefficients. They used the hospital length of stay data in this method and the power of score tests were investigated by Monte Carlo simulation.

M. Fazil baksh, Dankmar Bohning, Rattana Lerdsuwansri (2011), proposed a simple and efficient method of testing for overdispersion in zero-truncated count data. The proposed method genearlizes an overdispersion test previously suggested for
un-truncated count data and used for testing residual overdispersion in zero-inflated data. Simulations suggest that the asymptotic distribution of the test statistics is standard normal and that this approximation is also reasonable for small sample sizes. They showed that the proposed method more efficient than an existing test for overdispersion adapted for the capture-recapture setting. The proposed method test procedures have been applied to the zero-truncated and zero-inflated count data.

Andreas Linden and Samu Mantyniemi (2011), proposed a parameterization of the negative binomial distribution, which can be used to approximate overdispersed Poisson processes with a wide range of mean–variance relationships. Using bird migration data they presented hypothetical scenarios on how overdispersion can arise due to sampling, flocking behaviour or aggregation, Environmental variability or combinations of these factors. They found that the proposed framework can be a useful approximation for modelling marginal distributions of independent count data in likelihood-based analyses. In addition to the proposed scenarios, a potential area of application for the proposed parameterization of the negative binomial distribution is population dynamics, where process error in population size (or density) is likely to take a quadratic mean–variance relationship. They found that the negative binomial distribution is a viable option for describing many common types of mean–variance relationships in ecological data.

Naser A. Rashwan and Maie M. Kamel (2011), have studied generalized Poisson regression for two way contingency tables. They were used the study variable as urban population, rural population and age group. They have applied Poisson regression model and used Pearson chi-square and deviance statistic for the goodness of fit. They were found the overdispersion. From the empirical study they suggested that the GPR is better than the Poisson regression model when the occurrence of overdispersion in the Poisson regression model.
Seyed Ehsan Saffari, Robiah Adnan and William Greene (2011), introduced a Poisson regression model for truncated data. They suggested that to cut the large value of response variable to overcome the overdispersion. They have discussed the estimation of regression parameters using the maximum likelihood method and the goodness-of-fit for the regression model for the right truncated data. They have studied the effects of truncation in terms of parameters estimation and their standard errors through the fish catching study.

Background papers showed that the city size distribution is studied using rank size rule, Pareto distribution and stochastic model. These models were being used to analyse population data, area, GNP per capita and railway mileage. Poisson regression and Negative Binomial regression models were used for count data such as crime rates, road accidents and geometric design of the road sections and lung cancer death rates. Overdispersed Poisson regression is applied to the air pollution data. Truncated Poisson regression model has been used to analyse the fish catching study. Generalised Poisson regression model is used to the fertility data and two way contingency tables which include the study variable such as urban population, rural population and age group. Poisson regression and Generalised Poisson regression models are used to the zoological data and agricultural injuries data. Zero inflated generalised Poisson regression model has been used to patent outsourcing data and domestic violence data. From the past studies, it is found that the effect of age of the city has not been taken to describe the city size distribution. Poisson regression, Negative binomial regression, Generalised Poisson regression and Zero inflated Generalised Poisson regression models did not deal with the city sizes. Poisson regression and Negative binomial regression models are proposed to describe the administrative city sizes based on the factors such as age of the city, area of the city, education and transport of the city.