A POISSON REGRESSION MODEL WITH AUTOCORRELATON

A Thesis Submitted to the
Graduate School of Natural and Applied Sciences of
Dokuz Eylül University
in partial Fulfillment of the Requirements for
the Degree of Master of Science in Statistics

by
Burcu ÜÇER

July, 2003
İZMİR

Ms.Sc. THESIS EXAMINATION RESULT FORM

We certify that we have read the thesis, entitled "A POISSON REGRESSION MODEL WITH AUTOCORRELATION" completed by Burcu ÜÇER (HÜDAVERDİ) under supervision of Prof.Dr.Serdar KURT and that in our opinion it is fully adequate, in scope and in quality, as a thesis for the degree of Master of Science.

Prof.Dr.Serdar KURT

Supervisor

Yard.Doç.Dr.Ahmet KAYA

Committee Member

Yard.Doç.Dr.Özlem EGE ORUÇ

Committee Member

Approved by the

Graduate School of Natural and Applied Sciences

Prof.Dr.Cahit Helvacı

Director

ACKNOWLEDGEMENTS

I wish to express my sincere gratitude to my supervisor Prof. Dr. Serdar KURT for his close guidance and help throughout the course of this work.

I am also grateful to my husband Göksel ÜÇER for his encouragement and great assistance throughout the study.

I wish to express my gratitude to Esin FİRUZAN, Neslihan ORTABAŞ and all my friends for their assistance.

Finally, I also wish to express my gratitude to my dear family for their encouragement.

Burcu ÜÇER

ABSTRACT

In this study, time series of Poisson count model is concerned. In real situations, mean-variance equality, which is the basic property of Poisson data, cannot be provided. Generally, in such data variance exceeds mean, this is called overdispersion. When the overdisperison is detected, then there may be autocorrelation in latent process for Poisson regression model.

Correlation is assumed to result from a latent process which is added to the linear predictor in a Poisson regression model. A quasi-likelihood approach is used as a parameter estimation technique. Tests for the presence of the latent process and autocorrelation of the latent process are examined. Asymptotic properties of the regression coefficients are investigated by using a simulation study.

As an illustration, monthly number of deathes who were infected by pulmonary tuberculosis for the years 1996 to 2002 in Izmir are investigated as a parameter-driven model and the asymptotic properties of the regression coefficients are investigated, then a suitable model is constructed for forecasting.

Keywords: Quasi-Likelihood Method, Latent Process, Poisson Regression, Overdispersion, Autocorrelation.

ÖZET

Bu çalışmada, Poisson sayımlarının zaman serisi modeli incelendi. Poisson dağılan bir verinin temel özelliklerinden olan ortalama-varyans eşitliği uygulamada sağlanamaz. Genellikle, aşırı yayılım olarak adlandırılan varyans değerinin ortalamayı aştığı durum söz konusu olur. Aşırı yayılımın söz konusu olduğu durumda, Poisson regresyon modeli için gizli süreçte otokorelasyonun varlığından söz edilebilir.

Poisson regresyon modelinde doğrusal tahmin ediciye eklenen gizli süreçten kaynaklanan bir korelasyon durumunun olduğu varsayılır. Bu durumda yarı-olabilirlilik yöntemi, parametre tahmin yöntemi olarak kullanılabilir. Gizli sürecin varlığının testi ve gizli sürecin otokorelasyon yapısı bu çalışmada incelenmiştir. Aynı zamanda regresyon katsayılarının asimptotik özellikleri yapılan simulasyon çalışmasıyla araştırılmıştır.

Yapılan uygulamada, 1996'dan 2002'ye kadar İzmir'deki akciğer tüberkulozu ölümleri aylık olarak incelenmiştir. Yapılan çözümleme sonucunda katsayıların asimptotik özellikleri saptanarak uygun bir model oluşturulmaya çalışılmıştır.

Anahtar Kelimeler : Yarı-Olabilirlilik Yöntemi, Gizli Süreç, Poisson Regresyon, Aşırı Yayılım, Otokorelasyon

CONTENTS

Page

Chapter Three

AUTOCORRELATION

Chapter Four

TIME SERIES MODEL

Chapter Five

THE USE OF MODELS HAVING TRIGONEMETRIC TERMS

Chapter Six

APPLICATION

Chapter Seven

CONCLUSION

LIST OF TABLES

Page

LIST OF FIGURES

Page

CHAPTER ONE
INTRODUCTION

In the last recent years, there has been a great interest in the analysis of time series of counts. In order to get satisfactory modeling in the integer-valued characteristic of the data, time series of count analysis methods are used. The Poisson regression model is the basic member in such count data models. Although Poisson regression model has the characteristic property that the expected values and the variances are equal, generally the variance exceeds the mean in count data in real situations. This is overdispersion in the Poisson model, so with some empirical support, there may be autocorrelation in such data.

The techniques for the identification of a suitable model for the correlation structure in the noise latent process have been studied for many years. In count models with correlated latent process, there is a need for diagnostic techniques for Poisson counts. When there is autocorrelation among residuals, a method will be presented for valid statistical inference. Several regression application of time series of counts find little or no serial correlation. Then there is no need to use these mentioned models.

This study contains seven chapters. In chapter two, basic features of General Regression Model and Generalized Linear Model are given. The structure of a Poisson Regression model is examined. In order to identify the correlation structure of the latent process, a consistent estimation procedure for the regression coefficient is needed. Also, the estimation methods are studied .

In chapter three, general information about autocorrelation in a linear model is given. First-order autocorrelation structure and its properties are mentioned. Autocorrelation tests in count data model are proposed.

In chapter four, general information about the time series model is given.Time series count data models are studied. Presence of the latent process tests and the autocovariance and autocorrelation function of a latent process are determined.

In chapter five, as a consequence of the seasonality, the use of models having trigonometric terms is investigated for the time series data.

Chapter six aims to construct a model for the monthly number of deathes who were infected by pulmonary tuberculosis for the years 1996 to 2002 in İzmir. The Quasi-Likelihood estimation method is used to estimate the regression coefficients. Overdispersion is detected and then the presence of the latent process is decided. The correlation structure of the latent process is determined. Simulation study is done to investigate the asymptotic behaviour of these coefficients while there is autocorrelation in the latent process. In the last chapter, the results of this study are discussed.

CHAPTER TWO

POISSON REGRESSION MODEL

2.1 General Regression Model

A model

$$Y = X\beta + \varepsilon \tag{2.1}$$

is the form of the general linear models in which the error term ε is assumed to be $NID(0,\sigma^2)$. The general model has the following features:

1. For mutually independent random variables, $Y_i \sim N(\mu_i,\sigma^2)$ for i=1,2,...,n.
2. The explanatory variables provide a set of linear predictors $\eta_i = \beta_1 x_{i1} + \beta_2 x_{i2} + ... + \beta_p x_{ip}$ for $i = 1,2,...,n$.
3. The mean of the dependent variable for any observation is the linear predictor formed from that observation's values on the explanatory variables. This means that the relationship between 1 and 2 is that $\mu_i = \eta_i$.

The regression function is related with the expected value of dependent variable to the independent variables and the parameters, and the aim is to estimate the unknown parameters under given experimental conditions and data. The maximum likelihood or the least squares estimation methods are most commonly used methods in estimation. And these estimates are identical when the y_{ij}'s are independent and normally distributed, then they are obtained using linear regression analysis.

New methods are developed for linear models with the help of statistical theory in the following situations:

1. The distribution of the response variables is different from the Normal distribution. (The response variable may be categorical data, not continuous.)
2. The relationship between the response variable and explanatory variable need not be of the simple linear form in equation (2.1).

Many of the properties of Normal distribution are shared by a wider class of distributions called The Exponential Family of Distributions. (Dobson,1990)

2.2 Genaralized Linear Model

A generalized linear model is defined in terms of a set of independent random variables $Y_1, Y_2,..., Y_n$. Each of $Y_1, Y_2,..., Y_n$ has the distribution from the exponential family with the following properties:

1. Distribution function for each Y_i depends on a single parameter θ_i (the θ_i's do not all have to be the same) and each has canonical form
$$f(y_i;\theta_i) = \exp[y_i b_i(\theta_i) + c_i(\theta_i) + d_i(y_i)] \quad (2.2)$$
2. In order that the subscripts on b,c and d are not necessary, all distribution function of Y_i have the same form (all Poisson or all Binomial), so the joint pdf of $Y_1, Y_2,..., Y_n$ is
$$f(y_1,...,y_n;\theta_1,...,\theta_n) = \exp\left[\sum_{i=1}^{n} y_i b(\theta_i) + \sum_{i=1}^{n} c(\theta_i) + \sum_{i=1}^{n} d(y_i)\right] \quad (2.3)$$

A smaller set of parameters $\beta_1, \beta_2,..., \beta_p$ ($p<n$) is considered for the generalized linear model. Linear combination of β's are equal to the same function of the expected value μ_i of Y_i. (Dobson,1990)

Generalized linear model has three components:

1. Response variables $Y_1, Y_2, ..., Y_n$ are assumed to belong to the same distribution from the exponential family.
2. A set of parameters β and explanatory variables,
   $X = [x_1^T \cdots x_2^T]^T$
3. A monotone link function g such that
   $g(\mu_i) = x_i^T \beta$
   where
   $\mu_i = E(Y_i)$.

The general linear model characterizes the generalized linear model with two extensions:

1. It is applicable to any member of the exponential family of distributions.
2. There is a link function when connecting the linear predictor η to the mean μ of Y. (Wojtek&Krazanovski,1998)

2.2.1 Log-Linear Model

Association patterns among categorical variables are described by log-linear models. Log-linear models are used to analyze the data in multi-dimensional table or contingency table. The cell counts are modelled in a contingency table in terms of associations among the variables. Log-linear models are special case of generalized linear models family.

Let N = IJ be the cells of an I x J contingency table. The probabilities $\{\pi_{ij}\}$ form the joint distribution of two categorical responses. And these responses are statistically independent when $\pi_{ij} = \pi_{i.}\pi_{.j}$, i=1,2,...,I, j=1,2,...,J. For the expected frequencies $\{\mu_{ij} = n\pi_{ij}\}$ is $\mu_{ij} = n\pi_{i.}\pi_{.j}$ for all i and j. In order to apply for the

Poisson sampling model for N cell counts with expectation $\{\mu_{ij}\}$, log-linear models are constructed by using $\{\mu_{ij}\}$ rather than $\{\pi_{ij}\}$. (Agresti,1990)

In Poisson models, the most commonly used link function is the log, which is also the canonical link

$$\log \mu_i = \eta = x_i^T \beta \tag{2.4}$$

Using a log-link ensures that the fitted values of μ_i will remain in the parameter space $[0,\infty)$. A Poisson model with a log-link is sometimes called a log-linear model.

i. Independent Model

The log expected frequency for cell (i,j) is an additive function of an i^{th} row effect and a j^{th} column effect. The model

$$\log \mu_{ij} = \mu + \alpha_i + \beta_j \tag{2.5}$$

where

μ : overall mean

α_i: i^{th} level of first variable (row effect)

β_j: j^{th} level of second variable (column effect)

is called the log-linear model of independence in a two way contingency table. (Agresti, 1990)

ii. Saturated Model

If there is dependency between the variables, with all $\mu_{ij} > 0$, then the model will be

$$\log \mu_{ij} = \mu + \alpha_i + \beta_j + (\alpha\beta)_{ij} \qquad (2.6)$$

Here $(\alpha\beta)_{ij}$ shows the interaction between the variables. This model which is called *saturated model* describes perfectly any set of positive expected frequencies.

There is no error term in (2.6), because columns and rows are the repetation of a set of variable levels rather than construction of different part. In this case, based on the assumption that there is no real interaction between the columns and rows, $(\alpha\beta)_{ij}$ can be assumed as error term of the model. (Özmen,1998)

2.3 Poisson Distribution

In the Poisson distribution, the trails must be very large while the probability of occurences of the outcome under observation must be small, in addition to this independence of trials and consistency of probability from trial to trial properties are required.

The probability function for the discrete random variable Y is

$$f(y;\mu) = \frac{\mu^y e^{-\mu}}{y!}, \qquad y = 0,1,2... \qquad (2.7)$$

or

$$f(y;\mu) = \exp[y \log \mu - \mu - \log y!] \qquad (2.8)$$

which is in the canonical form with $\log \mu$ as the natural parameter.

Characteristics of the Poisson distribution are

1. The variance is equal to the mean
2. Poisson distribution with larger means tend to be well-approximated by a Normal distribution.

2.4 Overdispersion

Since observed count data often exhibit overdispersion, the Poisson assumption of mean-variance equality is no longer valid. Many authors have studied the effects of overdispersion on inference made on a Poisson model. The quasi-likelihood models suggest that the presence of overdispersion may result in less efficient regression coefficient estimates, but when the true regression model is specified, these estimates are true consistent estimates. (Wang et al., 1996) In fact, overdispersion is a real aspect of the data while modelling. Although such excess variation has little effect on estimation of the regression coefficients of primary interest, standard errors, tests and confidence intervals may be seriously in error unless it is approximately taken into account. (Breslow, 1990) Overdispersion may result in either seriously biased parameter or standard error or both.

The generalized Poisson distribution is used when the sample mean and the sample variance of this distribution are not equal. The generalized Poisson regression (GPR) model is useful in predicting a response variable affected by one or more covariates in the case of over-dispersed and under-dispersed count data. This regression model is suitable for both types of dispersions or no dispersion at all. This generalized Poisson distribution (GPD) is very useful model in many fields of study such as genetics, queueing, insurance, labor absenteeism and marketing research, etc.

Score tests have been developed to diagnose whether observed data is overdispersed respect to Poisson Regression by Fisher, Willimas (1982), Cameron and Trivedi (1986,1990), Dean and Lawless (1989) and Dean (1992).

The score test statistics are obtained by fitting the linear regression of squared residuals on the explanatory variables. Here residulas are deviance residuals for the double exponential family formulation and Pearson residulas for the pseudo-likelihood formulation. (Ganio & Schafer, 1992)

Three score tests proposed by Dean (1992) is used to determine whether the data is overdispersed. The hypothesis is that there is no overdispersion against alternatives representing different forms of overdispersion.

The test statistics are

$$P_a = \frac{\sum\left((y_i - \hat{\mu}_i)^2 - \hat{\mu}_i\right)}{\sqrt{2\sum \hat{\mu}_i^2}} \tag{2.9}$$

$$P_b = \frac{\sum\left((y_i - \hat{\mu}_i)^2 - y_i\right)}{\sqrt{2\sum \hat{\mu}_i^2}} \tag{2.10}$$

and

$$P_c = \frac{1}{\sqrt{2n}} \sum \frac{\left((y_i - \hat{\mu}_i)^2 - y_i\right)}{\hat{\mu}_i} \tag{2.11}$$

corresponding to the following specification of overdispersion:

(a) $E(y_i) \approx \mu_i$, $Var(y_i) = \mu_i(1 + \tau\mu_i)$ for τ small.
(b) $E(y_i) = \mu_i$, $Var(y_i) = \mu_i(1 + \tau\mu_i)$.
(c) $E(y_i) = \mu_i$, $Var(y_i) = \mu_i(1 + \tau)$.

Here $\hat{\mu}_i$ is the estimated value for the independent identical observations based on Poisson regression, and $H_0 : \tau = 0$ is tested and each asymptotically follows a Standard Normal distribution. It should be noted that (a) has the approximate forms for the first two moments, and (b) has the exact ones.

In cases where the dispersion parameter is not known, an estimate can be used to obtain an approximation to the scaled deviance and Pearson's chi-square statistic. Fitting a model that contains a sufficient number of parameters so that all systematic

variation is removed is one of the strategy in estimating the dispersion parameter ϕ from this model. An estimate of ϕ based on Pearson's chi-square is $\phi = \frac{\chi^2}{n-p}$ where $\chi^2 = \sum \frac{(y_i - \hat{\mu}_i)^2}{\hat{\mu}_i}$ and n is the number of observations and p is the number of parameters. Similarly, since the limiting chi-square distribution of the scaled deviance $D^* = \frac{D}{\phi}$ has (n-p) degrees of freedom, where equating D^* to its mean and solving for ϕ yields $\hat{\phi} = \frac{D}{n-p}$. If this dispersion parameter ϕ is significantly larger than 1, then overdispersion is indicated.

Unmeasured effects, clustering of events, or other concomitant influences combine to produce more variation in the responses than is predicted by the Poisson model. When this extra Poisson variation is present, there are three conseqeunces in Poisson log-linear regression;

1. Parameter estimates are still unbiased.
2. Standard errors are small.
3. Tests give smaller p-values than they are truly warrented by the data.

Extra-Poisson variation in log-linear model is expected

1. When important explanatory variables are not available.
2. When individuals with the same level of explanatory variables may behave differently.
3. When the events making up the count are clustered or spaced systematically through time or space.

2.5 Poisson Regression Analysis

The Poisson model describes a process when the succesive events occur at the same time independently. Poisson regression analysis is proposed instead of known linear regression analysis because of the fact that the normality assumption is not provided since the dependent variable is discrete. In log-linear models which do not have restriction on independent variables, there is no difference whether these variables are discrete or continuous.

Poisson regression is used when the outcome is count with large size and being rare events. The regression model may represent the number of failures of a piece of equipment per unit time or the number of purchases of a particular commodity, or the number of the bacteria per unit volume of suspension. (Frome et al., 1973)

In Poisson regression model, the normal link function is natural logarithm. The predicted counts are permitted to have a Poisson distributed error distribution to allow for discrepancies between the observed counts and those predicted by the model.

Poisson regression analysis is the method for analyzing the relationship between the explanatory variables and the dependent variable which is distributed as Poisson.

The expected value of Poisson distributed Y random variable is

$$E(Y_{ij}) = \mu_{ij} = n_{ij}\lambda_{ij} = n_{ij}\exp(\mu + \alpha_i + \beta_j) \tag{2.12}$$

or it may be used as

$$E(Y_{ij}) = \mu_{ij} = \exp(\mu + \alpha_i + \beta_j) \tag{2.13}$$

And also it can be rewritten as

$$E(Y_{ij}) = \mu_{ij} = n_{ij}\mu(X_i,\beta) = n_{ij}[\exp(X'\beta)] \quad (2.14)$$

or

$$E(Y_{ij}) = \mu_{ij} = \mu(X_i,\beta) = e^{X'\beta} \quad (2.15)$$

$\mu(X_i,\beta)$ is the expected rate for i^{th} sub-group. This rate function can be thought as the regression function that relates the mean response to X_i, the values of the predictor variables for case i, and β, the values of regression coefficients.

Y_{ij} : the number of interested cases for i^{th} level of first variable and j^{th} level of second variable

n_{ij} : the total number of individuals for i^{th} level of first variable and j^{th} level of second variable

μ_{ij} : risk function

If regression function can be assumed to have log-linear form with the number of parameters p = 1+I+J, then the regression function will be

$$\mu(X_i,\beta) = \exp(X_i'\beta) = \exp\left[\sum_{j=0}^{p} X_{ij}\beta_j\right] \qquad i = 1,2,..,n \quad (2.16)$$

$X_i' = [X_{i0} \cdots X_{ip}]$: i^{th} row vector of n x p input matrix of dummy variables

$\beta' = [\mu, \alpha_1, \alpha_2, \cdots, \alpha_i, \beta_1, \beta_2, \cdots, \beta_j]$: p x 1 dimensional vector of unknown parameters

Log-linear models which are used in Poisson regression analysis are represented by μ_{ij} regression function. Since the independent variables are categorical and qualitative, they must be taken into the model as dummy variables. Corner-point constraint is used in coding the dummy variables. (Dobson.1990) Corner-point constraint provides that the other parameters of corresponding (interested) variables

are estimated by giving "0" to the parameter value of one of the level of the variable. Under this constraint, the log-linear models which will be used in Poisson regression analysis are given below:

Model I :

This is the model that the first and the second variables are neglected as

$$\ln \mu_{ij} = \mu \tag{2.17}$$

Model II :

This is the model that the second variable is neglected as

$$\ln \mu_{ij} = \mu + \sum_{i=2}^{I} \alpha_i U_i \tag{2.18}$$

$\alpha_1 = 0$ according to the corner-point constraint.

Model III:

This is the model that the first variable is neglected as

$$\ln \mu_{ij} = \mu + \sum_{j=2}^{J} \beta_j E_j \tag{2.19}$$

$\beta_1 = 0$, according to the corner-point constraint.

Model IV :

This is the additive model as

$$\ln \mu_{ij} = \mu + \sum_{i=2}^{I} \alpha_i U_i + \sum_{j=2}^{J} \beta_j E_j \tag{2.20}$$

$\alpha_1 = \beta_1 = 0$, according to the corner-point constraint.

Model V:

This is the saturated model as

$$\ln \mu_{ij} = \mu + \sum_{i=2}^{I} \alpha_i U_i + \sum_{j=2}^{J} \beta_j E_j + \sum_{i=2}^{I} \sum_{j=2}^{J} (\alpha\beta)_{ij} (U_i E_j) \tag{2.21}$$

$\alpha_1 = \beta_1 = 0$ and $(\alpha\beta)_{1i} = (\alpha\beta)_{j1} = 0$, according to the corner-point constraint.

U_i and E_j are used as dummy variables.

Since the function $\mu(X_i,\beta)$ is exponential, β parameter estimation is found with the help of iterative procedures. Number of expected conditions are

$$E(Y_i) = \mu_i = n_i\mu(X_i,\beta) \qquad i=1,2,...,n \tag{2.22}$$

then the maximum likelihood function relating to the Poisson distribution is

$$L(\beta;y) = \prod_{i=1}^{n} f(y_i;\beta) = \prod_i \frac{\mu_i^{y_i} e^{-\mu_i}}{y_i!} = \prod_i \left\{ \frac{[n_i\mu(X_i,\beta)]^{y_i} e^{-n_i\mu(X_i;\beta)}}{y_i!} \right\}$$

$$= \frac{\left\{ \prod_i [n_i\mu(X_i,\beta)]^{y_i} \right\} \exp\left[-\sum_i n_i\mu_i(X_i,\beta) \right]}{\prod y_i!} \tag{2.23}$$

2.6. Generalized Poisson Regression Model

Generalized and restricted generalized Poisson regression model may be used instead of Poisson Regression when there is overdispersion or underdispersion.

The mean of the Y_i is given by

$$E(Y_i|x_i) = \mu_{x_i}. \tag{2.24}$$

The variance of the Y_i is given by

$$Var(Y_i|x_i) = \phi\mu_{x_i}. \tag{2.25}$$

Here, μ_{x_i} is given in log-linear form, and ϕ represents dispersion parameter.

2.7 Estimation Methods for Poisson Regression Analysis

There are three common types of statistical estimation methods:

i) Method of Maximum Likelihood

ii) Nonlinear Least Squares Method

iii) Quasi-likelihood Approach

The method of maximum likelihood is used for generalized linear models. By an iterative procedure which is related to weighted least squares estimation, the estimates have to be obtained numerically.

2.7.1 Maximum Likelihood Methods

Maximum likelihood estimators are the values of β parameters which correspond to the maximum value of the likelihood function. Generally, working with log-likelihood function is easier than the likelihood function itself. If the Poisson log-likelihood function is taken, then the function will be

$$\ln L(\beta; y) = \sum_{i=1}^{n} y_i \ln[n_i \mu(X_i, \beta)] - \sum_{i=1}^{n} n_i \mu(X_i, \beta) - \sum_{i=1}^{n} \ln(y_i!) \qquad (2.26)$$

Let Ω denote the set of all possible value of parameter vector β . $\hat{\beta}$, which is the maximum likelihood estimator of β , maximizes the likelihood function that is

$$L(\hat{\beta}; y) \geq L(\beta; y) \text{ for all } \beta \text{ in } \Omega$$

and also $\hat{\beta}$ maximizes the log-likelihood function $l(\hat{\beta}; y) = \log L(\hat{\beta}; y)$. (Since the logarithmic function is monotonic.) Thus,

$$l(\hat{\beta}; y) \geq l(\beta; y) \text{ for all } \beta \text{ in } \Omega .$$

Likelihood equations are obtained first by differentiating the log-likelihood function with respect to each element β_j of β, second solving the equations

$$\frac{\partial \ln L(\beta; y)}{\partial \beta_j} = 0 \qquad \text{for } j = 1,2,...,p$$

Then, it is needed to prove that the matrix of second derivatives

$$\frac{\partial^2 l(\beta; y)}{\partial \beta_j \partial \beta_k}$$

evaluated at $\beta = \hat{\beta}$ is negative definite, because checking that the solutions corresponds to maxima of $l(\beta;)$. (Dobson, 1990) All local maxima of $l(\beta; y)$ must be identified at the edges of Ω parameter space then the value $\hat{\beta}$ corresponding to the largest one is the maximum likelihood estimator.

Since $\mu(X_i, \beta)$ regression function is exponential, likelihood function is not linear. So, Newton-Raphson Iterative procedure is used to find the β maximum likelihood estimator. First, initial values are appointed to β's during iterative procedure. Genarally, zero may be taken as initial value, but also trial value may be taken as initial value. (Dobson, 1990)

2.7.2 Nonlinear Least Squares Method

Estimator of β holds consistency even though there is autocorrelation, by using the estimation methods NLS and Poisson MLE. But in fact we have difficulty in finding the consistent estimator of the variance matrix of these estimators. It is assumed that autocorrelation is present at lag k, then

$$w_{tj} = E[(y_t - \mu_t)(y_{t-j} - \mu_{t-j}) | x_1, ..., x_n] \qquad j=0,1,...,k \qquad (2.27)$$

is defined where $\mu_t = \exp(x_t'\beta)$.

NLS estimator of $\hat{\beta}_{NLS}$ minimizes $\sum_{t=1}^{n}(y_t - \exp(x_t'\beta))^2$. Cameron and Trivedi (1998) suggested that $\hat{\beta}_{NLS}$ is asymptotically normal with mean β and variance matrix

$$V[\hat{\beta}_{NLS}] = \left(\sum_{t=1}^{n}\mu_t^2 x_t x_t'\right)^{-1} B_{NLS}\left(\sum_{t=1}^{n}\mu_t^2 x_t x_t'\right)^{-1} \tag{2.28}$$

where

$$B_{NLS} = \sum_{t=1}^{n} w_{t0}\mu_t^2 x_t x_t' + \sum_{j=1}^{k}\sum_{t=k}^{n} w_{tj}\mu_t\mu_{t-j}\left(x_t x_{t-j}' + x_{t-j}x_t'\right) \tag{2.29}$$

If $w_{tj} = 0$ for $j \neq 0$, there is no autocorrelation at all in y_t.

Or with Poisson MLE we get the same result, with variance matrix

$$V[\hat{\beta}_P] = \left(\sum_{t=1}^{n}\mu_t x_t x_t'\right)^{-1} B_P\left(\sum_{t=1}^{n}\mu_t x_t x_t'\right)^{-1} \tag{2.30}$$

where

$$B_P = \sum_{t=1}^{n} w_{t0} x_t x_t' + \sum_{j=1}^{k}\sum_{t=k}^{n} w_{tj}\left(x_t x_{t-j}' + x_{t-j}x_t'\right) \tag{2.31}$$

Generally, estimation by using Poisson MLE is more efficient than NLS, because the Poisson MLE uses a working matrix that allows for heteroscedasticity (Cameron &Trivedi, 1998).

If lagged dependent variables are regressors and there is serial correlation in y_t after controlling regressors, then the results do not apply, so the NLS and Poisson MLE estimators are inconsistent.

2.7.3 Quasi-Likelihood Approach

If the likelihood function is in complex form, with the help of mean-variance relations, quasi-likelihood approach that is based on only the first and the second moments of the distributions is suggested. The only assumptions on the distribution of the data are first and second moments and some additional regularity conditions relating to the regression equation $E(Y)=\mu=\mu(\beta)$.

Quasi-likelihood considered as a function of mean μ_i and its variance $V(\mu_i)$ where $Var(Y_i)\approx V(\mu_i)$ and Y_i (i=1,2,...,n) are independent observations. μ_i are assumed to be known function of β_i parameters. Then for each observation, quasi-likelihood function

$$Q(y_i;\mu_i)=\int_{y_i}^{\mu_i}\frac{y_i-\mu_i'}{V(\mu_i')}d\mu_i'+f(y_i) \qquad (2.32)$$

or equivalently

$$\frac{\partial Q(y_i,\mu_i)}{\partial \mu_i}=\frac{y_i-\mu_i}{V(\mu_i)}$$

If Y_i's come from a one-parameter exponential family, it is found that Q is the log-likelihood function of the distribution. (Wedderburn, 1974) Log-likelihood function and log quasi-likelihood function have the similar properties and β parameters are asymptotically normal. (McCullagh, 1983). For one-parameter exponential family distributions, log-likelihood function and log quasi-likelihood function are the same, so it is valid for one-parameter distribution Poisson.

Maximum quasi-likelihood estimates may be estimated from the expected second derivatives of Q as in the case of maximum-likelihood estimates from the log-likelihood.

Let z be the vector whose components are $\frac{\partial Q}{\partial \beta_i}$. According to the Quasi-Likelihood function properties (Wedderburn,1974), z has 0 mean and dispersion matrix with elements

$$-E\left(\frac{\partial^2 Q}{\partial \beta_i \partial \beta_j}\right)$$

Let $K=\frac{\partial^2\left(\sum Q_i\right)}{\partial \beta_i \partial \beta_j}$, if the observations are independent, then it is considered that $\sum z_i$ has 0 mean and dispersion D=-E(K). Let $\hat{\beta}$ be the maximum quasi-likelihood estimates of β, then we have $\sum z_i \cong K\left(\beta-\hat{\beta}\right)$. From here we get

$$\beta-\hat{\beta} \approx K^{-1}\left(\sum z_i\right) \tag{2.33}$$

When we use expectation of K, instead of K, then we get

$$\hat{\beta} \approx \beta+D^{-1}\left(\sum z_i\right) \tag{2.34}$$

and from here, $D^{-1}\left(\sum \mu_i\right)$ has dispersion D^{-1}.

When the mean-variance relation is not known completely, but the variance is known, $\hat{\beta}$ can be calculated; for example when the variance-mean equality assumption is violated in Poisson model, generally variance exceeds the mean, so it causes overdispersion. Although this excess variation has little effect on parameter estimates, standard errors, tests and confidence intervals may be wrong unless it is

appropriately taken into account. (Breslow,1990) The variance in this case is considered as

$$\mathrm{Var}(Y_i) = \phi V(\mu_i) = \phi \mu_i \tag{2.35}$$

where $V(\mu_i)$ is a known function and ϕ is an unknown parameter. Here although it is unkown, we can calculate $\hat{\beta}$ assuming $\phi = 1$ because, maximum quasi-likelihood estimate of β is not affected by the value ϕ. But only, ϕ is needed to estimate in getting error estimates. Since it is considered that μ is approximately linear in β, we have

$$E\left(\sum\left(\frac{(y_i - \mu_i)^2}{V(\mu_i)}\right)\right) \cong n - p \tag{2.36}$$

then

$$\hat{\phi} = \frac{1}{n-p}\sum_{i=1}^{n}\frac{(y_i - \mu_i)^2}{V(\mu_i)} \tag{2.37}$$

For normal linear models, this gives the usual estimate of variance. (Wedderburn, 1974)

In the quasi-likelihood approach, variance ($\phi\mu_i$) of the Poisson Regression model is equal to the variance of the Generalized Poisson Regression Model, so the log quasi-likelihood functions of the Poisson Regression and the Generalized Poisson Regression are the same, then the quasi-likelihood estimates for both function will be the same. However, the log-likelihood functions of the Poisson Regression and the Generalized Poisson Regression are different, so the maximum likelihood estimates are also different.

Quasi-likelihood leads quite generally to consistent estimates of the regression coefficients even if the variance function is misspecified. (Breslow, 1990) In application, testing significancy of added variables in a regression model is a considerable problem. In this case, three tests are available under likelihood and quasi-likelihood theory. These are Wald tests, which is based on comparison of estimated coefficients with their standard errors; quasi-likelihood ratio test which is based on comparison of deviances under full and reduced models; and the third is the quasi-likelihood score test using the estimating equations themselves for inference.

2.8 Goodness of Fit Criterions in Poisson Regression Analysis

For the goodness of fit tests, nonparametric criterions like G^2, χ^2 ve T^2 are used since normality assumption is not provided in Poisson regression model. There are many alternative regression models for the given set of data; but the most suitable one should be chosen. To reach this decision, the deviances D are compared for the regression model where

$$D = -2 \ln\left(\frac{ELM}{SLM}\right)$$

ELM : Estimated Likelihood Model

SLM : Saturated Likelihood Model

$$D = -2\left(\frac{\ln L(\hat{\beta}; y)}{\ln L(\hat{\mu}; y)}\right) \tag{2.38}$$

or more easily

$$D = 2\sum_{i=1}^{n}\left[(y_i \ln\left(\frac{y_i}{\hat{\mu}_i}\right) - (y_i - \hat{\mu}_i)\right] \tag{2.39}$$

D is distributed χ^2 with $n-p$ degrees of freedom, where p is the total number of estimated parameters. If D is too large, the regression function spesification should be rejected.

The regression model which has the minumum deviance among all possible regression models is the best model for the given data set.

One of the other methods is Pearson test statistic defined as

$$\chi^2 = \sum_i^n \frac{(y_i - \hat{\mu}_i)^2}{\hat{\mu}_i} \tag{2.40}$$

The other method used in goodness of fit test is pseudo-R^2 value defined as

$$R^2_{pse} = \frac{D_0 - D}{D_0} \tag{2.41}$$

where D_0 is the deviance relating to the minimal model and D is the deviance of estimated model. If this criterion which is between 0 and 1 approaches to 1, then this chosen model is the best.

In Poisson regression anlaysis, the test of the variables in the model and the test of the contribution of the interaction term is applied according to the deviance difference after providing the degrees of freedom $n-p \geq 1$ restriction. Comparing this deviance difference with the χ^2 table value, the contribution of the tested variable is significant or not.

Model construction is thought as a process. The simplest model with only one parameter which is representing the mean value generally gives little insight into the processes behind the data. The most complicated model (saturated model) which has

one parameter for each observation is also of little use, because it replaces the observations themselves with an equal number of regression parameters.

For a given set of data, there will be larger discrepancies for the simpler model, and smaller discrepancies for the complicated model.

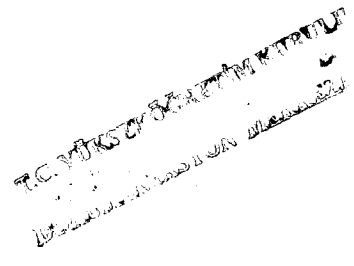
The deviance is formed from the logarithm of the ratio of two likelihoods. One of them is the likelihood of the current model or estimated model; the other is the likelihood of the saturated model.

2.9 Other Model Selection Methods

Two widely used model selection criteria are Akaike's Information Criterion (AIC) and the Bayesian Information Criterion (BIC) (or it is called Schwarz Information Criterion). These criterions are derived from the factor of Error Sum of Squares. While ESS are decreasing, number of selected factors are increasing. So the criterion sets the relationship between goodness of fit and complex model. Which criterion of the model is less, then it is decided that model is acceptable.

The AIC arises from maximizing an estimated expected log-likelihood. It penalizes the log-likelihood by the number of parameters fit to the data to avoid overfitting. The basic of the BIC is a prior distribution on the parameter space including all dimensions and models considered. (Wang et al., 1996)

When AIC selected the wrong model, it always choose a model with too many components. BIC always choose the correct model, suggesting that BIC may not overpenalize the number of parameters.



CHAPTER THREE
AUTOCORRELATION

When we are analyzing time series data, *serial independence* assumption is generally violated. Error terms for time periods may be correlated. This property is known as *autocorrelation* or *serial correlation.* There are various number of factors that cause the errors like, ommiting variables, measurement errors, ignoring nonlinearities, determining wrong model, random unpredictible effects. The first three effects can also lead to a serially correlated errors.

3.1 General Information about Autocorrelation in a Linear Model

The model

$$y_t = x_t^T \beta + \varepsilon_t \tag{3.1}$$

is a linear model with time series errors while y_t's are continuous variable. While working with such series, first of all autocovariance structure of the time series errors ε_t is determined. $(\beta_1,...,\beta_n)$ parameters are estimated by regressing the data vector $(y_1,...,y_n)'$ onto the $(x_1,...,x_n)$ using Ordinary Least Squares (OLS) Method. These estimates ignore the dependence structure of the ε_t. In this condition, these OLS estimate has the same asymptotic efficiency as the maximum likelihood estimate. The asymptotic covariance matrix of the OLS and ML estimate depends on the covariance structure of the ε_t. If β is the consistent estimator, then Autocovariance Function (ACVF) of ε_t can be consistently estimated from the sample ACVF of the residuals which is defined as $\hat{\varepsilon}_t = y_t - x_t'\hat{\beta}$.

3.1.1 First-Order Autocorrelation

This is the simplest case of autocorrelation. If serial correlation is present, the error for the period t is correlated with the error for the period k, such as $\mathrm{Cov}(\varepsilon_t, \varepsilon_k) \neq 0$ for $t \neq k$.

Assumption1

$$y_t = \beta_0 + \beta_1 x_t + \varepsilon_t \tag{3.2}$$

$$\varepsilon_t = \rho \varepsilon_{t-1} + \xi_t \qquad -1 < \rho < 1 \tag{3.3}$$

The error ε_t is related to the previous period's error ε_{t-1}. $|\rho| < 1$ for stationarity condition. Because ρ is the coefficient of the error term lagged one period, then it is called *first-order autocorrelation coefficient.* Equation 3.3 describes the *first-order autoregressive process AR(1).*

Assumption2

ξ_t, errors are independently and identically distributed with zero mean and constant variance so that $E(\xi_t) = 0$, $E(\xi_t^2) = \sigma_\varepsilon^2$, and $E(\xi_t \xi_{t-k}) = 0$ for $k \neq 0$

In the time series literature, a series in Assumption 2 is known as *white noise series* with zero mean. Since ε_t depends on ε_{t-1}, it is expected that they are correlated. And it is seen that ε_t does not directly depend on ε_{t-2}; but ε_t is indirectly correlated with all past errors. If the covariance is positive, then there is positive autocorrelation; if the covariance is negative, then there is negative autocorrelation.

3.1.2 Consequences of Autocorrelation

Under the Least Squares assumptions, the OLS estimates are unbiased and consistent. Consequently, when the error terms are serially correlated, OLS estimates are unbiased and consistent. The problem is with the efficiency of the estimates. While we are minimizing the variance of the linear combination $\sum a_t \varepsilon_t$,

$$\mathrm{Var}\left(\sum a_t \varepsilon_t\right) = \sum a_t^2 \sigma^2 + \sum_{t \neq k} \sum a_t a_k \mathrm{Cov}(\varepsilon_t, \varepsilon_k) \qquad (3.4)$$

for efficiency.

If $\mathrm{Cov}(\varepsilon_t, \varepsilon_k) \neq 0$, the second term on the right-hand side will not vanish. So this will not be equal with minimizing normal OLS equations, $\sum a_t^2 \sigma^2$, then OLS estimates are not BLUE any more and hence not efficient. Note that if the lagged dependent variable such as Y_{t-1} are included as explanatory variables, then autocorrelation causes inconsistent OLS estimates. The estimated variance of regression coefficient will be unbiased and inconsistent, so tests of hypotheses are invalid.

And also, if the serial correlation in ε_t is positive and X_t is getting larger over time, then the estimated residual variance $\hat{\sigma}^2$ will be an underestimate, so R^2 will be an overestimate. Also, t-statistics in such a case will tend to appear more significant than they actually are.

3.1.3 Properties of Error When it is AR(1)

$$\varepsilon_t = \rho \varepsilon_{t-1} + \xi_t = \xi_t + \rho(\xi_{t-1} + \rho \xi_{t-2}) = \xi_t + \rho \xi_{t-1} + \rho^2 \xi_{t-2} + \ldots \qquad (3.5)$$

Since $E(\xi_t) = 0$, then we have $E(\varepsilon_t) = 0$, and also by the independence assumption of the ξ's,

$$\sigma_\varepsilon^2 = \mathrm{Var}(\varepsilon_t) = \mathrm{Var}(\xi_t) + \rho^2\mathrm{Var}(\xi_{t-1}) + \rho^4\mathrm{Var}(\xi_{t-2}) + \ldots$$

$$= \sigma_\xi^2(1 + \rho^2 + \rho^4 + \ldots) = \frac{\sigma_\xi^2}{1-\rho^2} \tag{3.6}$$

The infinite series will sum to a finite value only if $|\rho| < 1$. When the first-order autocorrelation is less than 1 in absolute value, $|\rho| < 1$, it means that the necessary condition for the stationarity is provided. If $\rho = 1$, then the error process becomes $\varepsilon_t = \varepsilon_{t-1} + \xi_t$. The value of error term at time t is equal to its past value in the previous period plus a random effect. This process is called *random walk model.*

The covariance between ε_t and ε_{t-k}, for $k \neq 0$, is given by

$$E(\varepsilon_t\varepsilon_{t-k}) = E\left[(\xi_t + \rho\xi_{t-1} + \rho^2\xi_{t-2} + \ldots)(\xi_{t-k} + \rho\xi_{t-k-1} + \rho^2\xi_{t-k-2} + \ldots)\right] \tag{3.7}$$

All the cross-product terms of type $\xi_t\,\xi_{t-k}$, $\xi_{t-1}\xi_{t-k}\ldots$ have zero expectations because on assumptions of indepenedency ξ_t and ξ_{t-k}. Only the square terms remain. Therefore

$$\mathrm{Cov}(\varepsilon_t, \varepsilon_k) = E(\varepsilon_t\varepsilon_{t-k}) = E(\rho^k\xi_{t-k}^2 + \rho^{k+2}\xi_{t-k-1}^2 + \rho^{k+4}\xi_{t-k-2}^2 + \ldots)$$

$$= \rho^k\sigma_\xi^2(1 + \rho^2 + \rho^4 + \ldots)$$

$$= \frac{\rho^k\sigma_\xi^2}{1-\rho^2} = \rho^k\sigma_\varepsilon^2 \tag{3.8}$$

The correlation coefficient denoted r(k) between ε_t and ε_{t-k} is autocorrelation function

$$r(k) = \frac{\mathrm{Cov}(\varepsilon_t, \varepsilon_{t-k})}{\mathrm{Var}(\varepsilon_t)} = \rho^k \tag{3.9}$$

where for all lags $k = 0, \pm1, \pm2, \ldots$

Since $|\rho| < 1$, as k increases, the autocorrelation function decreases in absolute value. The autocorrelation coefficient for a lag k = 0 is $\rho(0) = 1$ which is the largest value. The smallest possible value is -1. Generally, the autocorrelation coefficients drop down to zero for any substantial lag for noisy data. (Brown,1962) Pure random noise would have zero correlation between samples which are not identically equal to each other.

When y_t is positive, y_{t+k} is also positive, or when y_t is negative then y_{t+k} is also negative, so the autocovariance will be large and positive. Similarly, when y_t is positive and y_{t+k} is negative or vice versa, the autocovariance will be large and negative. In both situation, one can be used to forecast the other. When the autocovariance is close to zero, the information about one observation doesn't help much in forecasting a late observation in the same sequence. (Brown, 1962)

3.2 Testing for Autocorrelation in Count Data Model

The aim is to detect the autocorrelation for the analysis; but for this, firstly overdispersion should be determined since if there is no overdispersion, this means there is no autocorrelation and there is no need for correction in the covariance matrix estimator of the Poisson maximum likelihood estimator nor for quasi-likelihood estimator. When overdispersion is detected, then possible correction of the Poisson maximum likelihood covariance matrix will be our interest, because we need to see whether this overdispersion is generated by ε_t which has an autocorrelation.

When the time series correction is necessary, it is helpful to use residual based test from Poisson Regression. Residual based test from dynamic regression makes clear that after inclusion of lagged variables, whether there is still autocorrelation. By using any of the estimators, autocorelation coefficients for ε_t can be estimated from the residulas (Cameron&Trivedi, 1998).

Let $z_t = y_t - \hat{\mu}_t$ be the residual from Poisson regression model. The standard measure of time series correlation is the autocorrelation at lag k.

$$\rho_k = \frac{E(z_t z_{t-k})}{\sqrt{E(z_t^2)E(z_{t-k}^2)}} \tag{3.10}$$

These residuals are nonstationary, with nonconstant variance because the variance equals to the nonconstant mean. Thus, before using the standard tests of serial correlation used in linear time series modeling, these raw residuals from Poisson Regression need to be standardized.

One of the standardized residual is Pearson Residual

$$z^*_t = \frac{y_t - \hat{\mu}_t}{\sqrt{\hat{w}_t}} \tag{3.11}$$

where $\hat{\mu}_t = \mu(x_t, \hat{\beta})$

$$\hat{w}_t = V[y_t | x_t] = w(\mu_t, \alpha)$$

where α are variance function parameters.

Box-Jenkins modeling in the continuous case can be applied using the autcorrelation function if z^*_t is standardized at least asymptotically constant variance where

$$\hat{\rho}_k = \frac{\sum_{t=k+1}^{n} z_t^* z_{t-k}^*}{\sum_{t=1}^{n} (z_t^*)^2} \tag{3.12}$$

In the case of $\hat{\rho}_k \cong 0, k \neq 0$, then we can say that there is no autocorrelation. Box-Pierce portmanteau statistic for serial correlation where K is the maximum lag is

$$Q_{BP} = n\sum_{k=1}^{K}\rho_k^2 \qquad \text{(Box and Pierce (1970))} \qquad (3.13)$$

$$Q_{LB} = n(n+2)\sum_{k=1}^{K}\frac{\rho_k^2}{n-k} \qquad \text{(Ljung and Box (1978))} \qquad (3.14)$$

where ρ_k is the estimated autocorrelation at lag k.

Assuming z_t^* is normalized to have constant variance, under the null hypothesis that no autocorrelation, it is tested with $\chi^2(K)$. As in the continuous case, for the count data, the degree of serial correlation can depend on whether first differences or levels are modeled.

Davis, Dunsmur and Wang (1998) suggested another test statistic, since there is a problem with the correlated Poisson model. This problem is that the variance and covariances have different forms of dependence on the mean function μ_t and there is no single normalization of residuals. These normalization eliminate the dependence from the variance and from the covariance terms required to construct autocorrelations.

$$H^2 = \sum_{k=1}^{K}\left[\frac{\hat{\gamma}_\varepsilon(k)}{s.e.(\hat{\gamma}_\varepsilon(k))}\right]^2 \qquad (3.15)$$

where K is the maximum lag and $\hat{\rho}_\varepsilon(k)$ is the autocorrelation function of ε, is proposed for testing for serial correlation in the mean of the observed count time series. It is analogous to the Box-Jenkin's statistics. Under the hypothesis, of independence H^2 will have an approximate χ^2 distribution with K degrees of freedom.

CHAPTER FOUR

TIME SERIES MODEL

4.1 Linear Model

Time series models for a continuous dependent variable where the only explanatory variables are lagged values of dependent variable is the standard class of linear models. Autoregressive moving average, ARMA(p,q), model is the sample of these linear models where p is the order of the autoregressive part and q is the order of the moving average part.

$$y_t = \rho_1 y_{t-1} + \ldots + \rho_p y_{t-p} + \xi_t + \gamma_1 \xi_{t-1} + \ldots + \gamma_q \xi_{t-q}, \quad t = p+1,\ldots n \tag{4.1}$$

is the ARMA(p,q) model where ξ_t is $N(0,\sigma^2)$.

Linear time series regression model (autoregressive or dynamic) has explanatory variables and lagged dependent variables as regressors such as

$$y_t = \rho y_{t-1} + x_t'\beta + \xi_t \tag{4.2}$$

where the error term ξ_t is $N(0,\sigma^2)$.

The model is called a *distributed lag model* if there are only x_t and lags of x_t in the model but if there is x_t alone in the model, this model is called a *static* model. One of the alternative time series regression model is *serially correlated error model* with a static regression function such as

$$y_t = x_t'\beta + \varepsilon_t \tag{4.3}$$

where the error term ε_t is serially correlated. As an example of the simplest form of the autoregressive error can be written as

$$\varepsilon_t = \rho\varepsilon_{t-1} + \xi_t \tag{4.4}$$

The coefficient of the model which autoregressive and serial correlation model is combined is estimated by the method of least squares or by the method of maximum likelihood if a distribution is specified for ξ_t.

When the ρ takes the value $\rho=1$, then y_t is nonstationary, due to a nonstationary stochastic trend, so asymptotic normal theory of estimators no longer valid. Nonstationary stochastic trends have not been studied for count regression (Cameron&Trivedi, 1998).

4.2 Count Models

There are many possible time series models for count data. Different models of the dependency of y_t on past y, and current and past x, and the latent process or error process ε_t causes different time series models.

Various Count Models:

1. **Integer-valued ARMA (INARMA) Models :** y_t is the sum of an integer whose value is determined by past y_t and independent innovation. Assumptions of Poisson or Negative Binomial distributions lead to a count marginal distribution of y_t. This is a generalization of the autoregressive model in equation (4.2).

2. **Autoregressive Models (Markov Models) :** Conditional distributions of y_t are count distributions such as Poisson or Negative Binomial. The mean parameter of this conditional distributions of y_t is a function of lagged values of y_t. Different from INARMA model in which marginal distributions of y_t is specified, here the conditional distributions of y_t is specified. This model is an extension of equation (4.2).

3. **Serially Correlated Error Models or Latent Variable Models :** y_t depends on a static component and a serially correlated latent variable. This is an extension of the serially correlated error model in equation (4.3) and equation (4.4).

4. **State-Space Models or Time-Varying Parameter Models :** y_t is a count distribution such as Poisson or Negative Binomial. In these distributions, conditional mean or parameters of conditional mean depend on their values in previous periods.

$$y_t = x_t'\beta_t + \varepsilon_t$$

$$\beta_t - \overline{\beta} = \phi(\beta_{t-1} - \overline{\beta}) + \xi_t \qquad (4.5)$$

where ε_t is $N(0,\sigma^2)$, ρ is a k x k matrix, and ξ_t is a k x 1, $N(0,\Sigma)$ error vector. This model is also widely used in Bayesian analysis of time series and it is called *dynamic linear.*

5. **Hidden Markov Models or Regime Shift Models :** y_t is a count distribution such as Poisson or Negative Binomial. In these distributions, paramaters vary according to which of a finite number of regimes is currently in effect.

6. **Discrete ARMA (DARMA) Models :** This model introduce time dependency through a mixture process (Cameron & Trivedi,1998).

4.2.1 Time Series Count Data Regression

The outcomes $\{Y_t : t = 1,...,n\}$ are time series of counts. Log-linear models can be used to describe $\mu_t = E(Y_t)$ as a function of a p x 1 vector of covarites x_t with independent observations. Mean function is specified by a linear predictor modified by a 'latent' process. If Y_t is Poisson, likelihood methods can be used to estimate β in the case $var(Y_t) = \mu_t$. Quasi-likelihood methods which allow a variety of variance-mean relations are appropriate in the case of $var(Y_t) > \mu_t$. The two common assumptions are

i) $var(Y_t) = \mu_t \phi$

ii) $var(Y_t) = \mu_t + \mu_t^2 \sigma^2$

where ϕ and σ^2 are unknown scale parameters. (Zeger, 1988)

In time series data, neighbouring observations are dependent. For non-Gaussian time series as count data following a Poisson distribution are divided two groups by Cox (1981) : observation-driven models and parameter-driven models:

i) Observation-driven : The conditional distribution of Y_t is specified as a function of past observations, $Y_{t-1},...,Y_1$. Autoregressive models for Gaussian series or Markov chains for discrete data is an example for this type of model. Assume that $Y_t|\mu_t$ is Poisson(μ_t), then the model,

$$\log \mu_t = x_t'\beta + \alpha_t, \qquad (4.6)$$

where α_t is a function of past observations Y_s, $s < t$ i.e. $\alpha_t = \gamma_1 Y_{t-1} + ... + \gamma_p Y_{t-p}$ is called as onservation-driven model.

ii) Parameter-driven : A latent process is thought as generating autocorrelation. Let $\theta_t = \log \mu_t$ be the canonical parameter for the log-linear model. Here, θ_t is assumed to depend on an unobservable noise process (ε_t), so $\theta_t = \theta(\varepsilon_t, Y_{t-1}, ..., Y_1)$. If Y_t given ε_t is Poisson, then it is $E(Y_t|\varepsilon_t) = \exp(x_t'\beta)\varepsilon_t$. The latent process, ε_t, introduces both overdispersion and autocorrelation in Y_t. (Zeger, 1988) *The parameter-driven models* has the advantage of incorporating both overdisperison and autocorrelation and the model specifies an unobserved latent process.

Assume that $Y_t|\mu_t$ is Poisson(μ_t), then the model

$$\log \mu_t = x_t'\beta + \alpha_t, \tag{4.7}$$

where α_t is a stationary Gaussian AR(1) latent process, i.e. $\alpha_t = \rho\alpha_{t-1} + e_t$ where $|\rho| < 1$, e_t are i.i.d. $N(0, \sigma^2)$, is called as paramater-driven model. As a property of this kind of model,

$$E(Y_t) = \exp(x_t'\beta)E(\exp(\alpha_t)) = \exp(x_t'\beta), \text{ if } E(\exp(\alpha_t)) = 1$$

4.2.1.1 The Model

Let the Poisson probability density function is again defined as

$$P(y_t) = \frac{\mu_t^{y_t} e^{-\mu_t}}{y_t!} \quad (t=1,2,...,n) \tag{4.8}$$

where Y_t is the count or frequency variable at time t.

Conditional on ε_t is assumed as stationary latent process. The marginal moments of Y_t is determined as a function of the log-linear coefficients and the parameters of ε_t since Y_t follows a log-linear model (Zeger,1988).

To introduce both overdispersion and autocorrelation in, conditional on a latent process ε_t, Y_t is a sequence of independent counts with properties

$$E(Y_t|\varepsilon_t)=\varepsilon_t\mu_t=\exp(x_t'\beta)\varepsilon_t \qquad var(Y_t|\varepsilon_t)=\varepsilon_t\mu_t=\exp(x_t'\beta)\varepsilon_t \tag{4.9}$$

Suppose that the ε_t is a stationary process with $E(\varepsilon_t)=1$ and $Cov(\varepsilon_t,\varepsilon_{t+k})=\sigma^2\rho_\varepsilon(k)$ where σ^2 is the variance and $\rho_\varepsilon(k)$ the autocorrelation function at lag k of the ε_t process (Brannas and Johansson, 1994). Then,

$$E(Y_t)=\mu_t=\exp(x_t'\beta)\ , \quad Var(Y_t)=\mu_t+\sigma^2\mu_t^2 \tag{4.10}$$

and

$$\rho_y(t,k)=corr(Y_t,Y_{t+k})=\frac{\rho_\varepsilon(k)}{\left[\left\{1+(\sigma^2\mu_t)^{-1}\right\}\left\{1+(\sigma^2\mu_{t+k})^{-1}\right\}\right]^{1/2}} \qquad (k\neq 0) \tag{4.11}$$

The latent process, ε_t, introduces both overdispersion and autocorrelation into y_t; this autocorrelation function $\rho_y(t,k)$ varies with t and k. The autocorrelation in y_t must be less than or equal to ε_t. The degree of autocorrelation in y_t relative to ε_t decreases as μ_t and σ^2 decrease.

While using MLE method, we need to have both density for $Y_t|\varepsilon_t$, and a multivariate density for $(\varepsilon_1,\varepsilon_2,...,\varepsilon_t)'$. No closed-form solution is possible, except in

trivial cases such as ε_t independently and identical distributed gamma if $Y_t|\varepsilon_t$ independently and identical distributed Poisson. (Cameron&Trivedi, 1998)

4.2.1.2 Estimation of β

Poisson maximum likelihood estimator is consistent even if the autocorrelation is not accounted for. In such cases maximum quasi-likelihood estimator is used. The conventional covariance matrix of this maximum likelihood estimator is inconsistent (Brannas and Johansson, 1994).

$$E(Y_t|x_t) = \exp(x_t'\beta) \tag{4.12}$$

model is considered with exponential conditional mean where y_t is dependent variable and x_t is static regressor. Estimator of β holds consistency even though there is autocorrelation, by using the estimation methods Quasi Likelihood and Poisson MLE. But in fact we have difficulty in finding the consistent estimator of the variance matrix of these estimators. It is assumed that autocorrelation is present at lag k, then

$$w_{tj} = E[(y_t - \mu_t)(y_{t-j} - \mu_{t-j})|x_1,...,x_n] \qquad j = 0,...,k \tag{4.13}$$

is defined where $\mu_t = \exp(x_t'\beta)$.

If lagged dependent variables are regressors and there is a serial correlation in Y_t after controlling regressors, then the results do not apply.

4.3 About the Latent Process

Before finding again the parameter estimates, it is needed to test for the existence of a latent process. Once a latent process is detected, then autocorrelation should be

tested. While covariance estimates are often biased, the standard estimates of correlation proposed by Zeger are in use.

4.3.1 Test for a Latent Process

There are tests that are used to detect overdispersion in Poisson distribution. Brannas and Johansson (1994) used the following statistic

$$S = \frac{\sum_{i=1}^{n}\left[(y_t - \hat{\mu}_t)^2 - y_t\right]}{\sqrt{\left[2\sum_{t=1}^{n}\hat{\mu}_t^2\right]}} \tag{4.14}$$

under the hypothesis Lagrange multiplier test of the Poisson distribution against Negative Binomial or more general Katz distribution. Dean and Lawless (1989) improved this test statistic for the small samples as

$$S_a = \frac{\sum_{t=1}^{n}\left[(y_t - \hat{\mu}_t)^2 - y_t + \hat{h}_t\hat{\mu}_t\right]}{\sqrt{\left[2\sum_{t=1}^{n}\hat{\mu}_t^2\right]}} \tag{4.15}$$

where h_t is the t^{th} diagonal element of the *hat* matrix. "Hat" matrix is $H = \Lambda^{1/2}X\left(X^T\Lambda X\right)^{-1}X^T\Lambda^{1/2}$, where $\Lambda = \mathrm{diag}(\mu_1,...,\mu_n)$ and $X = (x_1,...,x_n)^T$ is the design matrix (Fahrmeir &Tutz, 1994). Both of these test statistic asymptotically distributed as N(0,1) under "there is no latent process" hypothesis.

Davis, Dunsmuir and Wang (1998) introduced an alternative test specifically designed for overdisperison in the case of latent process in Poisson process. Since this test uses higher moment properties of Poisson observation, it is more powerfull than S_a.

Under the null hypothesis that there is no latent process $(\text{i.e.}, \varepsilon_t \equiv 1)$ the Pearson residuals,

$$e_t = \frac{y_t - \hat{\mu}_t}{\sqrt{\hat{\mu}_t}} \tag{4.16}$$

have approximately zero mean and unit variance.

$$Q = \frac{\left(\frac{1}{n}\sum_{t=1}^{n} e_t^2 - 1\right)}{\hat{\sigma}_Q} \tag{4.17}$$

where

$$\hat{\sigma}_Q^2 = \frac{1}{n}\left(\frac{1}{n}\sum_{t=1}^{n}\frac{1}{\hat{\mu}_t} + 2\right) \tag{4.18}$$

may be used to test the presence of a latent process. (Davis and Dunsmur, 1998) Q statistic is distributed as N(0,1) approximately under the hypothesis that the variance of a latent process is zero.

4.3.2 Estimation of the Autocorrelation Function and Autocovariance Function of the Latent Process

In literature, there are many suggestions on various estimates of the autocovariances. Zeger (1988) obtained the estimation of nuisance paramaters with the help of moments method. In this method, σ^2 is estimated as

$$\hat{\sigma}_\varepsilon^2 = \frac{\sum_{t=1}^{n}\left\{(y_t - \hat{\mu}_t)^2 - \hat{\mu}_t\right\}}{\sum_{t=1}^{n}\hat{\mu}_t^2} \tag{4.19}$$

where $\text{var}(Y_t) = \mu_t + \sigma^2\mu_t^2$.

Similarly, the autocorrelation function of ε_t (latent process) can be estimated by

$$\hat{\rho}_\varepsilon(k) = \hat{\sigma}_\varepsilon^{-2} \sum_{t=k+1}^{n} \frac{\{(y_t - \hat{\mu}_t)(y_{t-k} - \hat{\mu}_{t-k})\}}{\sum_{t=k+1}^{n} \hat{\mu}_t \hat{\mu}_{t-k}} = \frac{\hat{\gamma}_\varepsilon(k)}{\hat{\sigma}_\varepsilon^2} \tag{4.20}$$

where autocovariance function of ε_t (latent process) is given by

$$\hat{\gamma}_\varepsilon(k) = \frac{\sum_{t=k+1}^{n} \{(y_t - \hat{\mu}_t)(y_{t-k} - \hat{\mu}_{t-k})\}}{\sum_{t=k+1}^{n} \hat{\mu}_t \hat{\mu}_{t-k}} \quad \text{(Zeger, 1988)} \tag{4.21}$$

The pattern of the estimated autocorrelations is useful for the identification of basic autoregressive moving average (ARMA) model.

CHAPTER FIVE
THE USE OF MODELS HAVING TRIGONOMETRIC TERMS

It is suspected that data can be represented as a linear combination of these function values

$$y_t = \hat{y}_t + \hat{\varepsilon}_t \tag{5.1}$$

where the residuals $\hat{\varepsilon}_t$ are caused by two elements. The first is the noise in the observations: the observed values may be equal to true value plus a random stochastic variation. The second is errors in the model : Because of various reasons, the model can be identified wrong, or even the correct model is determined, there will be errors in estimating the coefficients because of the noise in the observations.

Regression deals with the problem of estimating values for vector of coefficients in the model. The fitting functions must be linearly independent. It means that any function is used to express as a linear combination of other functions. Independent variables may be emprical variables.

If the process can be represented in terms of previous values of the dependent variable, then this means that it has *autoregressive* form. The *autocovariance* of the data, or the mean, variance and autocorrelation coefficients contain all the information necessary for coefficient estimation that minimize the sum of squared residuals (least squares) in an autoregressive form (Brown, 1962).

5.1 Regression with Trigonometric Terms

When the process to be forecast is periodic, it is appropriate to describe it in terms of sines and cosines. Periodic function describes the repetation of the same values all over again. The length of time before the sequence starts to repeat is called the *period* and we are interested in the number of observations in these period. Quarterly observations of an annual cycle will have 4 observations per period; 12 observations per period are quite common because of the twelve hours in a half day, and twelve months in a year or if a weekly cycle may be observed seven times.

Simple trigonometric identity can be shown as

$$y_t = \beta_0 + \beta_1 \sin\left(\frac{2\pi t}{p}\right) + \beta_2 \cos\left(\frac{2\pi t}{p}\right) \tag{5.2}$$

In 1822, Fourier showed that any reasonable periodic function of time could be represented by taking a sufficient number of terms in series

$$\begin{aligned} y_t = c + \alpha_1 \sin\left(\frac{2\pi t}{p}\right) + \alpha_2 \sin\left(\frac{4\pi t}{p}\right) + \ldots + \alpha_k \sin\left(\frac{2k\pi t}{p}\right) + \ldots \\ + \beta_1 \cos\left(\frac{2\pi t}{p}\right) + \beta_2 \cos\left(\frac{4\pi t}{p}\right) + \ldots + \beta_k \cos\left(\frac{2k\pi t}{p}\right) + \ldots \end{aligned} \tag{5.3}$$

If there is a complicated waveform in the data and if it is necessary to forecast this form accurately, sufficient number of terms should be in the model. Trigonometric model should be used only where there is a known cause that makes a periodic phenomenon.

Three groups of model have been used to describe process for which forecasts are required. First group includes simple polynomials of degree 0, 1 and 2. The second group includes simple sinusoids which is required to represent periodic process with

12 observations per period. Third group includes various combinations of linear and sinusoidal change in the process. (Brown, 1963)

i) Simple Polynomials

1. *Constant Process :*

$$y_t = \beta_0 \tag{5.4}$$

This model is used when the process does not change in any deterministic way with time.

2. *Linear Process*

$$y_t = \beta_0 + \beta_1 t \tag{5.5}$$

This model is used when the process is growing (or falling) at a steady rate in time. This rate may change by small random increments from time to time.

3. *Quadratic Process*

$$y_t = \beta_0 + \beta_1 t + \frac{1}{2}\beta_2 t^2 \tag{5.6}$$

This model is used, for example, in forecasting future positions of an automobile in which the accelaration is constant for substantial periods of time, but in the long run changes by small, random increments. (Brown, 1962)

ii) Simple Sinusoids

1. *Simple 12-point sine*

$$y_t = \beta_0 + \beta_1 \sin\left(\frac{2\pi t}{12}\right) + \beta_2 \cos\left(\frac{2\pi t}{12}\right) \tag{5.7}$$

This model is used when the first approximation to a periodic process, such as a seasonal sales pattern observed once a month.

2. *Twelve-point sine with harmonic*

$$y_t = \beta_0 + \beta_1 \sin\left(\frac{2\pi t}{12}\right) + \beta_2 \cos\left(\frac{2\pi t}{12}\right) + \beta_3 \sin\left(\frac{4\pi t}{12}\right) + \beta_4 \cos\left(\frac{4\pi t}{12}\right) \tag{5.8}$$

This model includes one harmonic of the basic waveform.

3. *Irregular 12-point Periodic Function*

$$y_t = \beta_0 + \sum_{i=1}^{4}\left[\alpha_i \sin\left(\frac{2\pi it}{12}\right) + \beta_i \cos\left(\frac{2\pi it}{12}\right)\right] \tag{5.9}$$

This model includes four harmonic frequencies and so this model can describe almost any periodic process with an irregular pattern.

iii) Linear and Sinusoidal Combinations

Linear Trend with Simple sine Wave

$$y_t = \beta_0 + \beta_1 t + \beta_2 \sin\left(\frac{2\pi t}{12}\right) + \beta_3 \cos\left(\frac{2\pi t}{12}\right) \tag{5.10}$$

This model would be used instead the first model of the Simple 12-point Sine, where the sinusoidal variation is superimposed on a steady increase (or decrease).

CHAPTER SIX
APPLICATION

6.1 The Data

As an illustration, monthly number of deathes who were infected by pulmonary tuberculosis, for the years 1996 to 2002 in İzmir are investigated as a parameter-driven model. These data are reported by Health Directorate of İzmir Administrative Province. Our interest is to determine a long-term decrease in the rate of pulmonary tuberculosis infection.

Table 6.1 Monthly number of deathes who were infected by pulmonary tuberculosis, for the years 1996 to 2002 in İzmir.

	Jan	Feb	Mar	Apr	May	June	July	Aug	Sep	Oct	Nov	Dec
1996	10	3	6	3	1	2	5	8	2	9	7	10
1997	0	10	12	3	4	3	3	0	4	5	5	3
1998	3	1	6	2	0	3	3	2	1	5	5	3
1999	5	8	2	0	6	3	4	6	2	2	0	7
2000	4	3	4	3	2	2	2	1	4	3	2	6
2001	3	8	6	1	3	2	6	3	6	1	3	2
2002	3	0	0	1	4	1	2	1	0	1	2	3

6.2 The Method

Our aim is to identify the β parameters, for the Poisson model

$$f(y|x,\beta)=\frac{e^{-\mu(x,\beta)}\mu(x,\beta)^{y}}{y!}$$

where

$$\mu(x,\beta)=E(y|x,\beta)=\exp(x',\beta)$$

The MLE of the parameter β is obtained by maximizing the loglikelihood function

$$L(\beta)=\sum_{i=1}^{n} y_t \log\mu(x_t,\beta)-\mu(x_t,\beta)-\log(y_t!)$$

On specifiying correct conditional mean function and conditional Poisson distribution of Y, the MLE is consistent, efficient and asymptotically normally distributed, with variance matrix consistently estimated with

$$Var(\hat{\beta})=\left(\frac{\sum_{i=1}^{n}\frac{\partial\mu(x,\beta)}{\partial\beta}\frac{\partial\mu(x,\beta)}{\partial\beta'}}{\mu(x,\beta)}\right)^{-1}$$

In the case of rejection of the mean-variance equality assumption, the model is misspecified. Here the Poisson estimator may also interpreted as a Quasi Maximum Likelihood Estimator (QMLE). These QMLEs are robust in the sense of producing consistent estimates of the parameters of a correctly specified conditional mean, even whether the distribution is incorrectly specified. For these QL models, only a correct spesification of $\mu(x,\beta)$ is needed for consistency.

Hovewer, the estimated standard errors won't be consistent unless the conditional distribution of Y is correctly specified. But, it is possible to get the robust standard errors in order to make valid inferences even whether the distribution is incorrectly specified by using QL standard errors. But it doesn't posess any efficiency properties.

6.3 The Model

Since there is seasonality, monthly number of infected people is regressed on a linear trend with sine and cosine pairs at the annual and semi-annual frequencies. Figure 6.1 shows the trend analysis of the response variable.

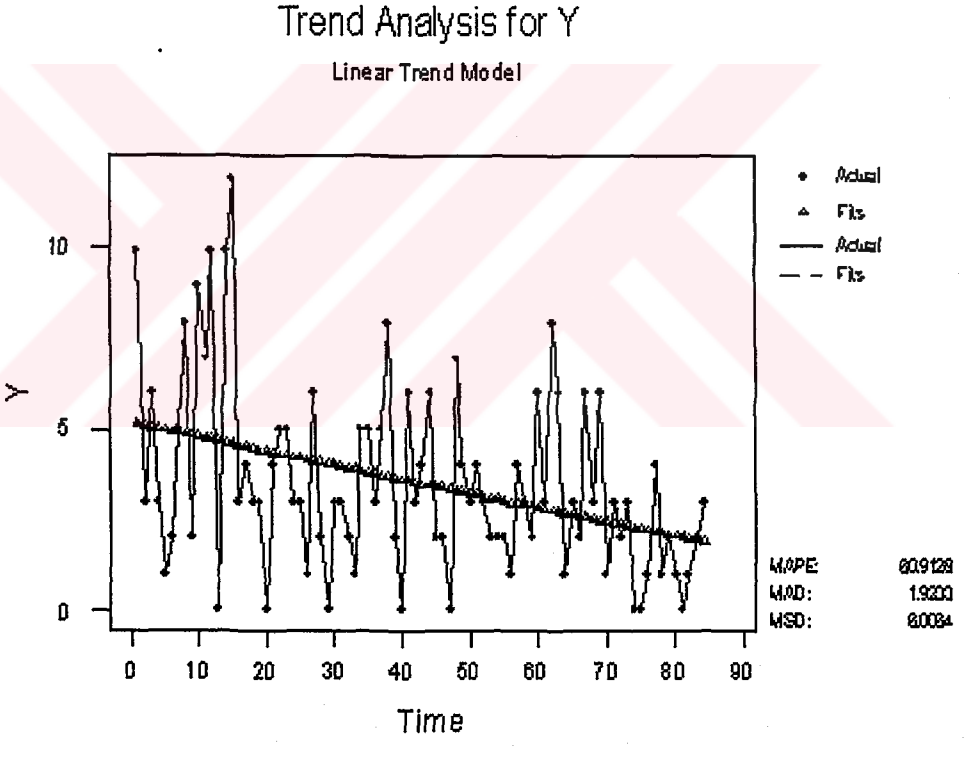


Figure 6.1 Trend Analysis for Monthly Number of Deathes

Table 6.2 Coefficients and Standard Errors of Regression Parameters

	QML		PML		Asym	Simulations	
	Coeff.	Std. Error	Coeff	Std. Error	s.e. (Coeff)	Ave (Coeff)	s.d. (Coeff)
Intercept	1.259704	0.069613	1.259704	0.059526	0.059951	1.258452	0.060632
$Trendx10^{-3}$	-11.00053	3.050726	-11.00053	2.471986	2.472688	-10.58603	2.513145
$cos(2\pi t/12)$	0.247722	0.096764	0.247722	0.082393	0.082619	0.234196	0.082685
$sin(2\pi t/12)$	-0.087644	0.103449	-0.087644	0.086066	0.086446	-0.074159	0.087014
$cos(2\pi t/6)$	0.082642	0.093502	0.082642	0.083067	0.083324	0.083165	0.086735
$sin(2\pi t/6)$	0.073830	0.10719	0.07383	0.083638	0.083871	0.072462	0.084761

An intercept term, a linear trend, and harmonics of 6 and 12 months are used as regressors. The design matrix is

$$x_t = \left(1, t'/1000, \cos\left(2\pi t'/12\right), \sin\left(2\pi t'/12\right), \cos\left(2\pi t'/6\right), \sin\left(2\pi t'/6\right)\right)$$

where $t' = t - 37$ is the intercept term at January 1999. Here, trend function has the form $x_{nt} = (1, t/n)'$. If the sample size in the linear component is omitted, then the parameter estimate would not be consistent for negative values of the slope parameter. This is due to the fact that the Poisson mean is converging to 0 rapidly. (Davis et al., 1998)

A simulation study was done to investigate these estimates over many trails, and 1000 realizations on the parameter-driven model fitted to the data were generated and this is also repeated 1000 times. On using the correct standard errors for the trend term it is concluded that the trend term is significant. Simulations columns of the Table 6.2 gives the true regression parameters. The latent process in this simulation was assumed to be a lognormal AR(1) with $\rho = 0.81$. On comparing the true value of the parameters, it can be concluded that there is no significant bias.

Non-negative assumption on ε_t provides the non-negativity of Y_t. And also $E(\varepsilon_t)=1$ is needed, because if $E(\varepsilon_t)\neq 1$, then it can be absorbed into the intercept term in the component of μ_t. In order to ensure the non-negativity constraint $\alpha_t = \ln \varepsilon_t$, then the model will be

$$y_t = \exp(x_t'\beta + \alpha_t)$$

where α_t is a stationary Gaussian process, so ε_t will be stationary lognormal process.

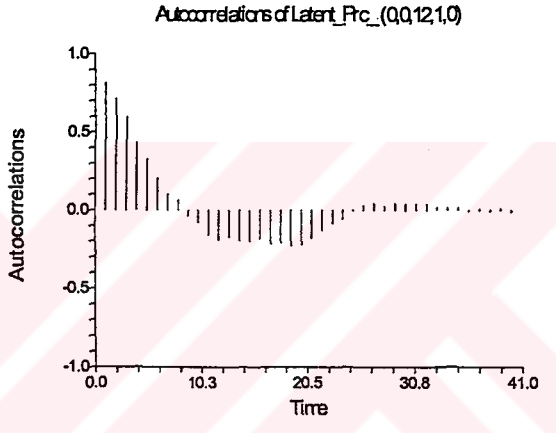


Figure 6.2 Autocorrelation Plot of Latent Process

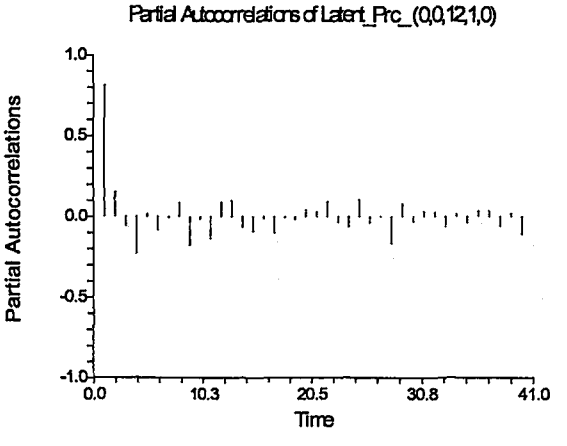


Figure 6.3 Partial Autocorrelation Plot of Latent Process

Table 6.3 Q statistics of Latent Process

Autocorrelation	Partial Correlation		AC	PAC	Q-Stat	Prob
. |****** |	. |****** |	1	0.814	0.814	57.607	0.000
. |****** |	. |*. |	2	0.714	0.154	102.52	0.000
. |***** |	.*| . |	3	0.597	-0.059	134.34	0.000
. |*** |	**| . |	4	0.431	-0.196	151.13	0.000
. |*** |	. | . |	5	0.328	0.019	160.99	0.000
. |** |	.*| . |	6	0.201	-0.085	164.75	0.000
. |*. |	. | . |	7	0.105	-0.009	165.78	0.000
. |*. |	. |*. |	8	0.066	0.091	166.20	0.000
. | . |	.*| . |	9	-0.039	-0.179	166.35	0.000
.*| . |	. | . |	10	-0.084	-0.016	167.04	0.000
.*| . |	.*| . |	11	-0.160	-0.138	169.58	0.000
**| . |	. |*. |	12	-0.191	0.088	173.23	0.000
.*| . |	. |*. |	13	-0.171	0.102	176.22	0.000
**| . |	.*| . |	14	-0.198	-0.068	180.25	0.000
**| . |	.*| . |	15	-0.201	-0.094	184.46	0.000
.*| . |	. | . |	16	-0.185	-0.017	188.08	0.000
**| . |	.*| . |	17	-0.219	-0.103	193.24	0.000
**| . |	. | . |	18	-0.210	-0.007	198.08	0.000
**| . |	. | . |	19	-0.231	-0.020	204.02	0.000
**| . |	. | . |	20	-0.220	0.040	209.50	0.000
.*| . |	. | . |	21	-0.180	0.031	213.20	0.000
.*| . |	. |*. |	22	-0.130	0.092	215.17	0.000
.*| . |	. | . |	23	-0.086	-0.038	216.04	0.000
.*| . |	.*| . |	24	-0.059	-0.069	216.46	0.000
. | . |	. |*. |	25	-0.005	0.107	216.47	0.000
. | . |	. | . |	26	0.028	-0.039	216.56	0.000
. | . |	. | . |	27	0.040	0.001	216.76	0.000
. | . |	.*| . |	28	0.025	-0.166	216.84	0.000
. | . |	. |*. |	29	0.042	0.081	217.08	0.000
. | . |	. | . |	30	0.040	-0.032	217.29	0.000
. | . |	. | . |	31	0.038	0.033	217.48	0.000
. | . |	. | . |	32	0.040	0.027	217.71	0.000
. | . |	.*| . |	33	0.017	-0.064	217.74	0.000
. | . |	. | . |	34	0.014	0.014	217.77	0.000
. | . |	. | . |	35	0.011	-0.039	217.79	0.000
. | . |	. | . |	36	-0.008	0.035	217.80	0.000

If the sample autocorrelation function of the original observations either dies down or cuts off fairly quickly, it can be assumed that the original series is stationary. For the autoregressive models, theoretical partial autocorrelation function of this model cuts off after lag k, and the theoratical autocorrelation function dies down.

If there is no autocorrelation in the latent process, the autocorrelations and the partial autocorrelations at all lags should be nearly zero and all Q-statistics should be insignificant with large p-values.

Table 6.4 Descriptive Statistics of the Simulation

	β_0	β_1	β_2	β_3	β_4	β_5
Mean	1.258452	-10.58603	0.234196	-0.074159	0.083165	0.072462
Median	1.261703	-10.66275	0.233171	-0.071803	0.086023	0.075050
Maximum	1.448640	-2.913805	0.553063	0.171008	0.351765	0.388515
Minimum	1.026540	-18.69673	-0.024499	-0.350561	-0.219456	-0.201225
Std. Dev.	0.060632	2.513145	0.082685	0.087014	0.086735	0.084761
Skewness	-0.154144	-0.121542	0.017149	-0.136640	-0.102956	-0.049945
Kurtosis	3.132887	2.986629	3.078314	3.053805	3.389450	3.192151
Jarque-Bera	4.695835	2.469523	0.304563	3.232361	8.086281	1.954178
Probability	0.095568	0.290904	0.858747	0.198656	0.017542	0.376405
Observations	1000	1000	1000	1000	1000	1000

In simulation firstly, 1000 population is generated for the pulmonary tuberculosis data, then the model is constructed 1000 times and the parameters are estimated by using both Quasi-Likelihood and Maximum Likelihood methods. These are made with the help of Eviews and Minitab code shown in Appendices.

Since ϕ (Dispersion Parameter = $\frac{\text{Pearson}\_\chi^2}{\text{df}}$) = 1.525 > 1 then we can conclude that there is overdispersion. When overdispersion is detected then we investigate the presence of latent process. With the help of Q = 2.57 statistics we decide that there is latent process.

Table 6.5 Relative Efficiency for QL Estimation to ML Estimation

Regressors	Relative Efficiency
Intercept	0.855099
$Trend x 10^{-3}$	0.810294
$cos(2\pi t/12)$	0.851484
$sin(2\pi t/12)$	0.831966
$cos(2\pi t/6)$	0.888398
$sin(2\pi t/6)$	0.780278

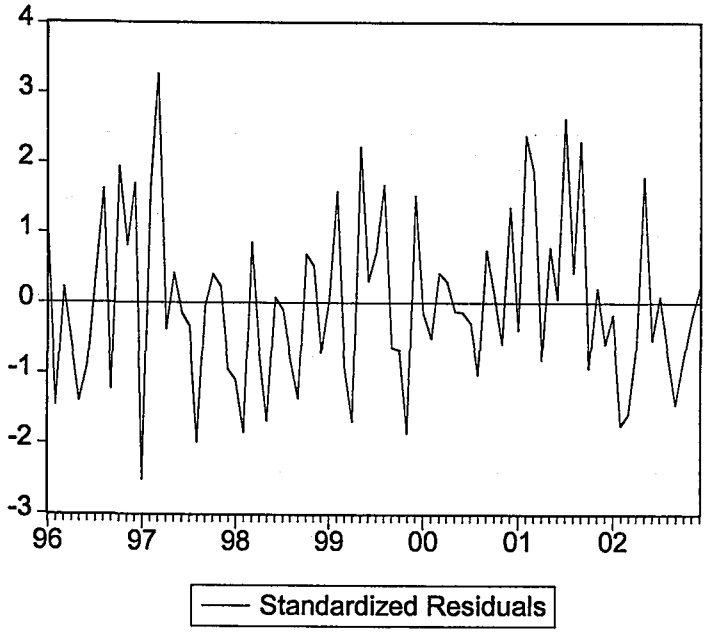


Figure 6.4 Standardized Residual Plot

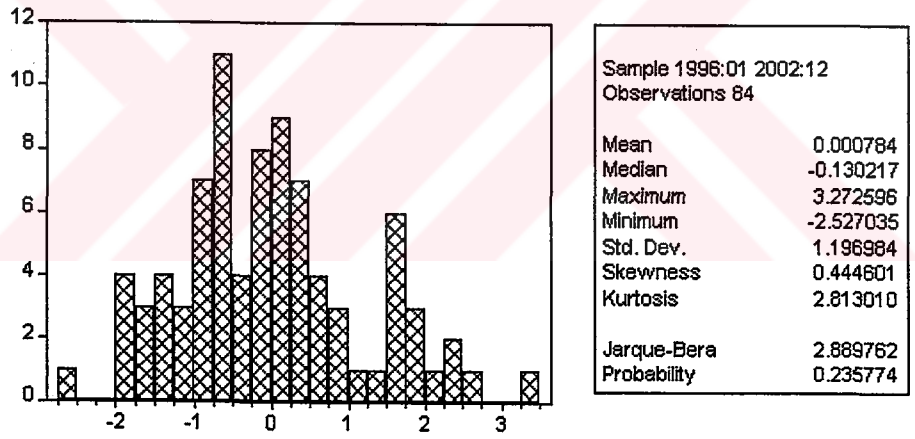


Figure 6.5 Descriptive Statistics of the Pearson Residuals

Since Pearson residuals have the same correlation structure as ε_t (Zeger, 1988), it is assumed that there is lag one autoregressive correlation structure with $E(\varepsilon_t)=1$. The latent process has a lognormal distribution as depicted Figure 6.6.

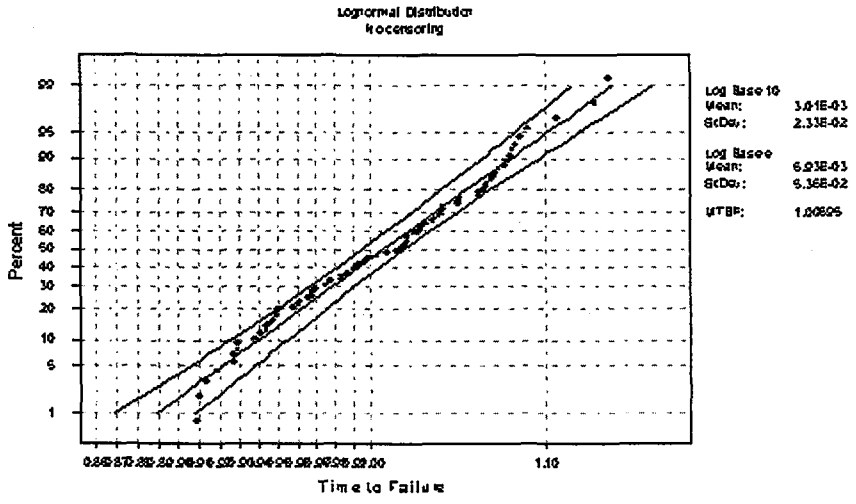


Figure 6.6 Probability Plot of the Latent Process

The maximum quasi-likelihood estimator is consistent and asymptotically normal. This is a robust approach in making consistent inferences about β only that $E(Y_t) = \mu_t$ whether or not equal to $var(Y_t)$.

Table 6.6 Model with an Intercept Term, a Linear Trend, and Harmonics of 6 and 12 Months

Variable	Coeff.	Std. Error
C	1.259704	0.069613
TREND	-11.00053	3.050726
COS12	0.247722	0.096764
SIN12	-0.087644	0.103449
COS6	0.082642	0.093502
SIN6	0.07383	0.10719
S.E. of regression		2.411857
Log likelihood		-184.4295
LR statistic (5 df)		33.61566
Probability(LR stat)		2.84E-06
Akaike criterion		4.534037
Schwarz criterion		4.707666

Table 6.7 Model with an Intercept Term, a Linear Trend, and Harmonics of 12 Months

Variable	Coeff.	Std. Error
C	1.263258	0.071392
TREND	-11.20847	3.046832
COS12	0.257259	0.102682
SIN12	-0.07705	0.100826
S.E. of regression		2.386169
Log likelihood		-185.3152
LR statistic (3 df)		31.84426
Probability(LR stat)		5.64E-07
Akaike criterion		4.5075
Schwarz criterion		4.6232

Table 6.8 Model with an Intercept Term, a Linear Trend, and Harmonics of 6 Months

Variable	Coeff.	Std. Error
C	1.276383	0.074122
TREND	-11.01905	3.184534
COS6	0.100139	0.097493
SIN6	0.069366	0.115191
S.E. of regression		2.493434
Log likelihood		-189.4683
LR statistic (3 df)		23.53815
Probability(LR stat)		3.12E-05
Akaike criterion		4.606388
Schwarz criterion		4.722141

Table 6.9 Model with an Intercept Term and a Linear Trend

Variable	Coeff.	Std. Error
C	1.281134	0.075442
TREND	-11.21862	3.206079
S.E. of regression		2.479241
Log likelihood		-190.5601
LR statistic (1 df)		21.35459
Probability(LR stat)		3.82E-06
Akaike criterion		4.5847
Schwarz criterion		4.6426

Since the parameter estimates are consistent and have robust covariance estimates, also asymptotically efficient, and other model selection criterions, we choose the second model for forecasting. Then the model will be

$$\log Y = 1.263258 - 11.20847((t-37)/1000) + 0.257259(\cos 2\Pi(t-37)/12) \\ 0.077050(\sin 2\Pi(t-37)/12)$$

This model gives very close forecasts as in the first model. The model has the significant trend function.This means that the pulmonary tuberculosis deathes are decreasing by the time.

Even there is latent process with autocorrelation for this Poisson regression model, consistent coefficient estimates and robust variance estimates can be obtained by using Quasi Likelihood method, so valid inferences can be made.

CHAPTER SEVEN
CONCLUSION

In this study, time dependent Poisson regression model is investigated. For the Poisson data, generally variance exceeds mean, this case is called overdispersion. When overdispersion presents, for this kind of time dependent Poisson data, presence of latent process should be investigated. A latent process is thought as generating autocorrelation in such models, these type of count model is called Parameter-Driven model.

Monthly number of deathes who were infected by the pulmonary tuberculosis data as a Parameter-Driven model is examined for this study. Our aim is to identify a model for forecasting. For this, firstly overdispersion is detected, then it is decided that there is a latent process and the correlation structure of this latent process is determined.

In order to identify the latent process and its correlation structure, a consistent estimation procedure for the regression coefficient is needed. For consistent estimation procedure, a quasi-likelihood method which is based on only the first and the second moments of the distribution is suggested. In the case of rejection of the mean-variance equality assumption, the model is misspecified. These Quasi maximum likelihood estimates are robust in the sense of producing consistent estimates of the parameters of the correctly specified conditional mean, even whether the distribution is incorrectly specified. For these QML models, only a correct specification of the mean function is needed for consistency.

With simulation study, QMLEs asymptotic behaviour is examined. For this monthly pulmonary tuberculosis data, parameter estimates are consistent, asymptotically normal

and asymptotically efficient. And also since the variance estimates of the parameters are robust, we can make valid inferences about this model only that $E(Y_t) = \mu_t$ whether or not equal to $Var(Y_t)$.

REFERENCES

Agresti A. (1984). Analysis of Ordinal Categorical Data. Canada: John Wiley & Sons.

Agresti A. (1990). Categorical Data Analysis. Canada: John Wiley & Sons.

Bartholomew D. & Knott M. (1999), Latent Variable Models and Factor Analysis (2nd ed.).London: Arnold.

Bowerman B. & O'Conell R. (1987). Time Series Forecasting. (2nd ed.). Canada: Duxbury Press.

Brannas K. & Johanson P. (1994). Time Series Count Data Regression. Communication Statistics- Theory and Methods, 23, 2907-2925.

Blais M. & MacGibbon B. & Roy R. (2000). Limit theorems for regression models of time series of counts. Statistics&Probability Letters, 46, 161-168.

Breslow N.E. (1984). Extra-Poisson Variation in Log-Linear Models. Appl. Statist., 33, 38-44.

Breslow N. (1990). Tests of Hypotheses in Overdispersed Poisson Regression and Other Quasi-Likelihood Models. JASA, 85, 565-571.

Brown R. (1963). Smoothing Forecasting and Prediction of Discrete Time Series. USA: Prentice Hall.

Cameron A. C. & Trivedi K. P. (1998). Regression Analysis of Count Data. USA: Cambridge University Press.

Campell M.J. (1994). Time Series Regression for Counts: An Investigation into the Relationship between Sudden Infant Death Syndrome and Enviromental Temperature. JASA, 157, 191-208.

Chan K.S. & Ledolter J. (1995). Monte Carlo EM Estimation for Time Series Models Involving Counts, JASA, 90, 242-252.

Consul P.C. & Famoye F. (1988). Maximum Likelihood Estimation for the Generalized Poisson distribution hen the sample mean larger than the sample variance. Commun. Statist.-Theory and Methods, 17, 299-309.

Consul P.C. (1988). On some models leading to the generalized Poisson distribution, Commun. Statist.-Theory and Methods, 17, 423-442.

Consul P.C. & Famoye F. (1992). Generalized Poisson Regression Model. Commun. Statist.-Theory and Methods, 21, 89-109.

Cox R. (1981). Statistical Analysis of Time Series: Some Recent Developments. Scandinavian Journal of Statistics, 8, 93-115.

Cox R. (1983). Some Remarks on Overdispersion. Biometrika, 70, 269-274.

Davis R., Dunsmuir T.M. & Wang Y. (2000). On Autocorrelation In a Poisson Regression Model. Biometrics, 87, 191-208.

Dean C.B. (1992). Testing for Overdipersion in Poisson and Binomial Regression Models. JASA, 87, 451-457.

Dobson J. A. (1990). An Introduction to Generalized Linear Models. London: Chapman & Hall.

Famoye F. (1993). Restricted Generalized Poisson Model. Commun. Statist.-Theory and Methods, 22, 1335-1354.

Frome L. E., Kutner M.H. & Beauchamp J.J. (1973). Regression Analysis of Poisson- Distributed Data. JASA, 68, 935-940 .

Frome E. (1983). The Analysis of Rates Using Poisson Regression Models. Biometrics, 39,

Frome E. & Checkoway H. (1985). Use of Poisson Regression Models in estimating incidence rates and ratios. American Journal of Epidemiology, 121, 309-323.

Ganio L.M. & Schafer D.W. (1992). Diagnostics for Overdisperion. JASA, 87, 795-804

Guo J. & Li T. (2002). Poisson Regression Models with errors-in-variables: implication and treatment. Journal of Statistical Planning and Inference, 104, 391-401.

Haberman J.S. (1978). Analysis of Qualitative Data Volume 1. USA: Introductory Topics Academic Press.

Johnson N., Kotz S. & Kemp A. (1992). Univariate Discrete Distributions (2nd ed.). Canada: John Wiley & Sons.

Jorgensen B., Christensen S., Song P. & Son L. (1999). A state space model for multivariate longitudinal count data. Biometrika, 86, 169-181.

Kleinbaum, Kupper, Muller & Nizam. (1998). Applied Regression Analysis and Other Multivariable Methods (3th ed.). Canada: Duxbury Press.

Krzanowski J. W. (1998). An Introduction to Statistical Modelling, London: Arnold.

McCullagh P. (1983). Quasi-Likelihood Functions. The Annals of Statistics, 11, 59-67.

Oh M. & Li Y. (2001). Bayesian analysis of Time Series Poisson data. Journal of Applied Statistics, 28, 259-271.

Özmen İ. (1992). Poisson Regresyon Çözümlemesinde Logaritmik Doğrusal Modellerin Kullanımı ve Uygulaması. Bilim Uzmanlığı Tezi, H.Ü.Fen Fakültesi, Ankara.

Özmen İ. (1998). Poisson Regresyon Çözümleme Teknikleri. Doktora Tezi, H.Ü. Fen Fakültesi, Ankara.

Ramanathan R. (1998). Introductory Econometrics with Applications. USA : The Dryden Press

Tuenter H.J.H. (2000). On the Generalized Poisson distribution. Statistica Neerlandica, 54, 374-376.

Wang P., Puterman M.L., Cockburn I. & Le N. (1996). Mixed poisson Regression Models with Covariate Dependent Rates. Biometrics, 52, 381-400.

Wedderburn R.W.M. (1974). Quasi-Likelihood Functions, generalized linear models, and the Gauss-Newton method. Biometrika, 61, 439-447.

Zeger S.L. (1988). A Regression Model for Time Series of Counts. Biometrika, 75, 621-629.

Zeger S.L. & Qaqish B. (1988). Markov Regression Models for Time Series: A Quasi-Likelihood approach, Biometrics, 44, 1019-1031.

APPENDICES

Appendix A MINITAB code for generating 1000 Poisson Data with Latent Process for the pulmonary tuberculosis data

```
Let k1 = k1+1
Let k2 = k2+1
Let k3 = C1(k2)
Random 1000 Ck1;
        Poisson k3.
```

Appendix B Eviews Code for Generating Random Columns

```
'number of random columns
!Count = 1000

'Number of Data in Columns
!RowCount = 84

'Generating random columns...
For !Sayac = 1 To !Count
 matrix(84, 1) Rnd{!Sayac}
 For !I = 1 To !RowCount
  sym(1) Satir
  rndint (Satir,  999)
  Rnd{!Sayac}(!I, 1) = Y{!I}(Satir(1,1) + 1)
 Next
 series Ser{!Sayac}
 mtos(Rnd{!Sayac}, Ser{!Sayac})
 Delete Rnd{!Sayac}
Next
```

Appendix C Eviews Code for Estimation of Every Simulated Data and Descriptive Statistics of Parameter Estimates

```
'declare constant variable
!N =1000

'init
!top1 = 0
!top2 = 0
!top3 = 0
!top4 = 0
!top5 = 0
!top6 = 0

'a1 = standart errors
matrix(!N, 6)    a1

'a2= coefficients
matrix(!N, 6)    a2


'calc equations
for !i=1 to !N
  equation burcu.COUNT(D=P,H,R) Ser{!i} c TREND COS12 SIN12 COS6 SIN6
 FOR !j = 1 TO 6
    A1(!i, !j) = burcu.@stderrs(!j)
    A2(!i, !j) = burcu.@coefs(!J)
 Next

 !top1 = !top1 + burcu.@coefs(1)
 !top2 = !top2 + burcu.@coefs(2)
 !top3 = !top3 + burcu.@coefs(3)
 !top4 = !top4 + burcu.@coefs(4)
 !top5 = !top5 + burcu.@coefs(5)
 !top6 = !top6 + burcu.@coefs(6)
Next

'find avg
!top1 = !top1/ !N
!top2 = !top2/ !N
!top3 = !top3/ !N
!top4 = !top4/ !N
```

```
!top5 = !top5/ !N
!top6 = !top6/ !N

series toplam
toplam.fill !top1, !top2, !top3, !top4, !top5, !top6
show toplam
```