2

Poisson Models and Parametric Inference

2.1 Introduction

It is well-known that for the analysis of count data, Poisson model is perhaps the most commonly used model or assumption (Breslow, 1984, 1990; Cameron and Trivedi, 1998). Thus for the analysis of panel count data, it is helpful to first consider Poisson-based approaches for the motivation of more general inference procedures and their comparison. As mentioned before, the focus of this chapter is on parametric approaches, which also can be seen as a motivation to many semiparametric inference procedures discussed later.

To be complete, we start with the discussion on regression analysis of simple count data in Sect. 2.2. Count data can be seen as a special case of panel count data in which all study subjects are observed only once at the same time point. They occur in many fields including actuarial studies, demography, economics, political and social sciences, and reliability studies. Examples of count data include the occurrences of certain tumors and the frequency of recurrent events such as auto accidents and visits to doctor's offices or clinics. For inference about regression parameters, both the likelihood-based method and the estimating equation-based method are described. The former is developed under the Poisson assumption and the latter can be regarded as a generalization of the former. Section 2.3 considers regression analysis of panel count data with the focus on the maximum likelihood approach. For this, two situations are discussed. One is that the underlying recurrent event processes of interest are non-homogeneous Poisson processes and the other is that the recurrent event processes are mixed Poisson processes (Dean, 1991; Lawless, 1987a,b).

As mentioned above, the focus of this book is on nonparametric and semiparametric inference procedures. One reason for the discussion of parametric methods is that the development of the former is often motivated by the latter. Of course, the latter itself could be useful too when the parametric assumption used is reasonable. Sometimes one may prefer some approaches or compromises between the two types of procedures. One such type of approaches is piecewise procedures that are essentially

J. Sun and X. Zhao, *Statistical Analysis of Panel Count Data*, Statistics for Biology and Health 80, DOI 10.1007/978-1-4614-8715-9_2,

parametric methods but can be made to be close to nonparametric or semiparametric inference methods. Section 2.4 describes two piecewise approaches for regression analysis of panel count data. Similar to the two methods discussed in Sect. 2.2 for count data, one is a likelihood-based method and the other is an estimating equation-based method. In Sect. 2.5, some bibliographical notes and remarks are given. Throughout the chapter, it is assumed that the observation process is independent of the underlying recurrent event process of interest.

2.2 Regression Analysis of Count Data

Consider a recurrent event study that consists of n independent subjects and in which the observed information from each subject is only the number of the events that have occurred over some time interval. For subject i, let N_i denote the observed count and suppose that there exists a vector of covariates $\boldsymbol{Z}_i$, $i = 1, \ldots, n$. Then the observed data are $\{ (N_i, \boldsymbol{Z}_i);\, i = 1, \ldots, n \}$. It is assumed that the main goal is to make inference about the effects of covariates on the occurrence rate of the event of interest.

To describe the covariate effects, we assume that given $\boldsymbol{Z}_i$, the conditional mean of N_i has the form

$$E(N_i | \boldsymbol{Z}_i) = \exp(\boldsymbol{\beta}^T \boldsymbol{Z}_i) \,, \tag{2.1}$$

where $\boldsymbol{\beta}$ denotes the vector of regression parameters. Note that an alternative or a more natural choice is to assume $E(N_i | \boldsymbol{Z}_i) = \alpha \exp(\boldsymbol{\beta}^T \boldsymbol{Z}_i)$, a special case of the proportional mean model (1.4), where α is an unknown parameter. It is easy to see that model (2.1) includes this latter choice as a special case by setting the first component of $\boldsymbol{Z}_i$ as one. For inference about $\boldsymbol{\beta}$, in the following, we first discuss some likelihood-based procedures developed for the situation where the N_i's follow the Poisson distribution. Some estimating equation-based procedures are then presented that do not require the Poisson assumption and followed by some discussions.

2.2.1 Likelihood-Based Procedures

In this subsection, we suppose that the N_i's follow Poisson distributions with the mean given by model (2.1). Then it is easy to see that the log-likelihood function of $\boldsymbol{\beta}$ is proportional to

$$l(\boldsymbol{\beta}) = \sum_{i=1}^{n} \left\{ N_i \, \boldsymbol{\beta}^T \boldsymbol{Z}_i - \exp(\boldsymbol{\beta}^T \boldsymbol{Z}_i) \right\} .$$

It follows that the maximum likelihood estimator, denoted by $\hat{\boldsymbol{\beta}}_P$, of $\boldsymbol{\beta}$ is given by the solution to the score estimating equation

$$\sum_{i=1}^{n} \boldsymbol{Z}_i \left\{ N_i - \exp(\boldsymbol{\beta}^T \boldsymbol{Z}_i) \right\} = 0 . \tag{2.2}$$

It is easy to show that $\hat{\boldsymbol{\beta}}_P$ is consistent and its distribution can be approximated by the normal distribution with mean $\boldsymbol{\beta}_0$, the true value of $\boldsymbol{\beta}$, and the covariance matrix

$$V_{ML}(\boldsymbol{\beta}) = n^{-1} \left\{ \sum_{i=1}^{n} \boldsymbol{Z}_i \boldsymbol{Z}_i^T \exp(\boldsymbol{\beta}^T \boldsymbol{Z}_i) \right\}^{-1}$$

with $\boldsymbol{\beta}$ replaced by $\hat{\boldsymbol{\beta}}_P$.

Under the Poisson assumption, we have that $Var(N_i|\boldsymbol{Z}_i) = E(N_i|\boldsymbol{Z}_i)$ and it is well-known that this often does not hold in practice. To relax this restriction, one common approach is to assume that there exists a latent variable ν_i and given ν_i, N_i follows the Poisson distribution with the mean

$$E(N_i|\boldsymbol{Z}_i, \nu_i) = \nu_i \exp(\boldsymbol{\beta}^T \boldsymbol{Z}_i) , \tag{2.3}$$

$i = 1, \ldots, n$. That is, the N_i's follow the mixed Poisson distribution. Here it is assumed that the ν_i's are i.i.d. with $E(\nu_i) = 1$ and $Var(\nu_i) = \sigma_\nu^2$. Under the model above, it is easy to show that

$$E(N_i|\boldsymbol{Z}_i) = \mu_i = \exp(\boldsymbol{\beta}^T \boldsymbol{Z}_i) , \; Var(N_i|\boldsymbol{Z}_i) = \mu_i (1 + \sigma_\nu^2 \mu_i) .$$

That is, the variance of N_i can be equal to or larger than its mean.

Now we assume that the ν_i's follow the Gamma distribution with the density function

$$g(\nu;\gamma) = \frac{\gamma^{-1/\gamma}}{\Gamma(\gamma^{-1})} \nu^{\gamma^{-1}-1} e^{-\nu/\gamma} ,$$

where $\Gamma(a) = \int_0^\infty t^{a-1} \exp(-t)\, dt$, the Gamma function. In this case, we have $\sigma_\nu^2 = \gamma$. Then the marginal density function of N_i is given by

$$f(N_i|\boldsymbol{\beta},\gamma) = \frac{\Gamma(\gamma^{-1} + N_i)}{\Gamma(\gamma^{-1})\, \Gamma(N_i+1)} \left(\frac{\gamma^{-1}}{\gamma^{-1} + \mu_i} \right)^{\gamma^{-1}} \left(\frac{\mu_i}{\mu_i + \gamma^{-1}} \right)^{N_i} .$$

That is, N_i follows the negative binomial distribution. It follows that the log likelihood function of $\boldsymbol{\beta}$ and γ is proportional to

$$l(\boldsymbol{\beta},\gamma) = \sum_{i=1}^{n} \left\{ \sum_{j=0}^{N_i-1} \log(j+\gamma^{-1}) - (N_i + \gamma^{-1}) \log(1+\gamma\mu_i) \right.$$
$$\left. + N_i \log\gamma + N_i \boldsymbol{\beta}^T \boldsymbol{Z}_i \right\} .$$

Let $\hat{\boldsymbol{\beta}}_{NB}$ and $\hat{\gamma}_{NB}$ denote the resulting maximum likelihood estimators of $\boldsymbol{\beta}$ and γ, respectively. Then they can be obtained by solving the score equations

$$\sum_{i=1}^{n} \frac{N_i - \mu_i}{1 + \gamma \mu_i} \boldsymbol{Z}_i = 0$$

and

$$\sum_{i=1}^{n} \left\{ \frac{1}{\gamma^2} \left(\log(1 + \gamma \mu_i) - \sum_{j=0}^{N_i - 1} \frac{1}{j + \gamma^{-1}} \right) + \frac{N_i - \mu_i}{\gamma(1 + \gamma \mu_i)} \right\} = 0$$

together. One can easily show that $\hat{\boldsymbol{\beta}}_{NB}$ and $\hat{\gamma}_{NB}$ are consistent. Furthermore, their joint distribution can be asymptotically approximated by the multivariate normal distribution with mean $(\boldsymbol{\beta}_0^T, \gamma_0)^T$ and the covariance matrix determined by

$$Var(\hat{\boldsymbol{\beta}}_{NB}) = n^{-1} \left(\sum_{i=1}^{n} \frac{\hat{\mu}_i}{1 + \tilde{\mu}_i} \boldsymbol{Z}_i \boldsymbol{Z}_i^T \right)^{-1},$$

$$Var(\hat{\gamma}_{NB}) = n^{-1} \left\{ \sum_{i=1}^{n} \left(\log(1 + \tilde{\mu}_i) - \sum_{j=0}^{N_i - 1} \frac{1}{j + \hat{\gamma}_{NB}^{-1}} \right)^2 + \frac{\hat{\mu}_i}{\hat{\gamma}_{NB}^2 (1 + \tilde{\mu}_i)} \right\}^{-1}$$

and $Cov(\hat{\boldsymbol{\beta}}_{NB}, \hat{\gamma}_{NB}) = 0$. That is, $\hat{\boldsymbol{\beta}}_{NB}$ and $\hat{\gamma}_{NB}$ are asymptotically independent. In the above, again $\boldsymbol{\beta}_0$ and γ_0 denote the true values of $\boldsymbol{\beta}$ and γ, respectively, $\hat{\mu}_i = \exp(\hat{\boldsymbol{\beta}}_{NB}^T \boldsymbol{Z}_i)$ and $\tilde{\mu}_i = \hat{\gamma}_{NB} \, \hat{\mu}_i$.

2.2.2 Estimating Equation-Based Procedures

As mentioned above, under the negative binomial model, the variance of N_i does not have to be equal to its mean as under the Poisson model, but cannot be smaller than its mean. It is obvious that this may still be too restrictive in reality. In this subsection, we assume that the N_i's still satisfy model (2.1) but do not make any assumption about the distribution of the N_i's. By following the estimating equation theory, it is clear that one can still employ the estimating Eq. (2.2) and to use its solution, denoted by $\hat{\boldsymbol{\beta}}_{PP}$, to estimate $\boldsymbol{\beta}$. Again by using the estimating equation theory (White, 1982), one can show that $\hat{\boldsymbol{\beta}}_{PP}$ is consistent and its distribution can be asymptotically approximated by the normal distribution with mean $\boldsymbol{\beta}_0$ and the covariance matrix

$$Var(\hat{\boldsymbol{\beta}}_{PP}) = n^{-1} \left(\sum_{i=1}^{n} \hat{\mu}_i \, \boldsymbol{Z}_i \boldsymbol{Z}_i^T \right)^{-1} \left(\sum_{i=1}^{n} w_i \, \boldsymbol{Z}_i \boldsymbol{Z}_i^T \right) \left(\sum_{i=1}^{n} \hat{\mu}_i \, \boldsymbol{Z}_i \boldsymbol{Z}_i^T \right)^{-1}, \tag{2.4}$$

where $\hat{\mu}_i = \exp(\hat{\boldsymbol{\beta}}_{PP}^T \boldsymbol{Z}_i)$ and $w_i = Var(N_i|\boldsymbol{Z}_i)$, $i = 1, \ldots, n$. The estimator $\hat{\boldsymbol{\beta}}_{PP}$ is often referred to as the Poisson pseudo- or quasi-maximum likelihood estimator.

To use the formula (2.4), one usually needs to specify the variance function w_i's. For this, one common choice is to let

$$w_i = \phi \, \mu_i \, , \tag{2.5}$$

where ϕ is an unknown parameter. In this case, the formula reduces to

$$Var(\hat{\boldsymbol{\beta}}_{PP}) = \frac{\phi}{n} \left(\sum_{i=1}^{n} \hat{\mu}_i \, \boldsymbol{Z}_i \boldsymbol{Z}_i^T \right)^{-1}$$

and one can estimate ϕ empirically by

$$\hat{\phi} = \frac{1}{n-p} \sum_{i=1}^{n} \frac{(N_i - \hat{\mu}_i)^2}{\hat{\mu}_i} \, ,$$

where p denotes the dimension of $\boldsymbol{\beta}$. Another common choice for the w_i's is to assume that

$$w_i = \mu_i + \alpha \, \mu_i^2 \tag{2.6}$$

with α being an unknown parameter as ϕ. Under the model above, we have

$$Var(\hat{\boldsymbol{\beta}}_{PP}) = n^{-1} \left(\sum_{i=1}^{n} \hat{\mu}_i \boldsymbol{Z}_i \boldsymbol{Z}_i^T \right)^{-1} \left(\sum_{i=1}^{n} (\hat{\mu}_i + \alpha \hat{\mu}_i^2) \boldsymbol{Z}_i \boldsymbol{Z}_i^T \right) \left(\sum_{i=1}^{n} \hat{\mu}_i \boldsymbol{Z}_i \boldsymbol{Z}_i^T \right)^{-1}$$

and one can also estimate α empirically by

$$\hat{\alpha} = \frac{1}{n-p} \sum_{i=1}^{n} \frac{\{(N_i - \hat{\mu}_i)^2 - \hat{\mu}_i\}}{\hat{\mu}_i^2} \, .$$

Of course, sometimes one may not want to impose any form on the w_i's. In this case, assuming that we can treat $\{ (N_i, \boldsymbol{Z}_i) \}_{i=1}^n$ as i.i.d., one can estimate $Var(\hat{\boldsymbol{\beta}}_{PP})$ by the robust estimator

$$n^{-1} \left(\sum_{i=1}^{n} \hat{\mu}_i \boldsymbol{Z}_i \boldsymbol{Z}_i^T \right)^{-1} \left(\sum_{i=1}^{n} (N_i - \hat{\mu}_i)^2 \boldsymbol{Z}_i \boldsymbol{Z}_i^T \right) \left(\sum_{i=1}^{n} \hat{\mu}_i \boldsymbol{Z}_i \boldsymbol{Z}_i^T \right)^{-1} . \tag{2.7}$$

2.2.3 Discussion

There exists extensive literature on the analysis of count data (Cameron and Trivedi, 1998; Vermunt, 1997). For example, Cameron and Trivedi (1998) discussed the count data arising from natural and social sciences. Also they gave a relatively complete review of the analysis approaches commonly used in the field. In particular, in these books, one can find some real count data and the applications of the methods discussed above to real count data. As mentioned before, the main purpose of this section is to give some introduction of the existing literature as the count data can be regarded as a special case of panel count data. More importantly, the methods described above serve as a motivation to many inference procedures discussed below for the analysis of panel count data.

To relax the Poisson assumption used in Sect. 2.2.1, instead of using the mixed Poisson model, an alternative is to develop a latent class-based Poisson model as follows (Wedel et al., 1993). Suppose that study subjects are from K unknown classes and each subject belongs to one and only one class. Let α_k denote the unknown probability that a subject belongs to class k with $\sum_{k=1}^{K} \alpha_k = 1$. It is assumed that conditional on subject i belonging to class k, N_i follows the Poisson distribution with mean

$$\lambda_{i|k} = \exp(\boldsymbol{\beta}_k^T \boldsymbol{Z}_i)\,,$$

where $\boldsymbol{\beta}_k$ is a vector of unknown regression parameters as $\boldsymbol{\beta}$, $i = 1, \ldots, n$, $k = 1, \ldots, K$. Then the likelihood function of the $\boldsymbol{\beta}_k$'s and α_k's has the form

$$L(\boldsymbol{\beta}_k' s, \alpha_k' s) = \prod_{i=1}^{n} \sum_{k=1}^{K} \alpha_k \, \exp\left\{- \exp(\boldsymbol{\beta}_k^T \boldsymbol{Z}_i)\right\} \frac{\exp(N_i \boldsymbol{\beta}_k^T \boldsymbol{Z}_i)}{N_i!}\,.$$

It follows that one can naturally estimate the $\boldsymbol{\beta}_k$'s and α_k's by their maximum likelihood estimators.

For estimation of $\boldsymbol{\beta}$ in model (2.1) without making a distribution assumption, instead of the estimating Eq. (2.2), one can use its weighted version given by

$$\sum_{i=1}^{n} w_i \, \boldsymbol{Z}_i \left\{ N_i - \exp(\boldsymbol{\beta}^T \boldsymbol{Z}_i) \right\} = 0\,,$$

where the w_i's are some weights. One can easily show that the estimator of $\boldsymbol{\beta}$ given by the solution to the equation above is consistent. Furthermore, one can asymptotically approximate its distribution by the normal distribution with mean $\boldsymbol{\beta}_0$ and the covariance matrix

$$n^{-1} \left(\sum_{i=1}^{n} w_i \hat{\mu}_i \boldsymbol{Z}_i \boldsymbol{Z}_i^T \right)^{-1} \left(\sum_{i=1}^{n} w_i^2 (N_i - \hat{\mu}_i)^2 \boldsymbol{Z}_i \boldsymbol{Z}_i^T \right) \left(\sum_{i=1}^{n} w_i \hat{\mu}_i \boldsymbol{Z}_i \boldsymbol{Z}_i^T \right)^{-1}$$

with $\boldsymbol{\beta}$ replaced by its estimator. It is easy to see that the estimator above reduces to (2.7) if one takes $w_i = 1$ for all i.

2.3 Parametric Maximum Likelihood Estimation of Panel Count Data

Consider an event history study concerning certain recurrent events that involves n independent subjects and in which each subject gives rise to a counting process $N_i(t)$. Here $N_i(t)$ represents the total number of the occurrences of the recurrent event of interest from subject i up to time t, $i = 1, \ldots, n$. For each subject, as before, suppose that there is a p-dimensional vector $\boldsymbol{Z}_i$ of covariates whose effects on $N_i(t)$ are of main interest. Also suppose that $N_i(t)$ is observed only at finite time points $t_{i,1} < \cdots < t_{i,m_i}$, where m_i denotes the number of observation times, $i = 1, \ldots, n$. That is, we only observe panel count data given by

$$\{ (t_{i,j}, n_{i,j} = N_i(t_{i,j}), \boldsymbol{Z}_i) \, ; \, j = 1, \ldots, m_i, i = 1, \ldots, n \} \, .$$

For estimation of the effects of covariates on the $N_i(t)$'s, in the following, we first consider the situation where it is reasonable to assume that the $N_i(t)$'s are non-homogeneous Poisson processes. A more general situation where the $N_i(t)$'s are mixed Poisson processes is then discussed and followed by an illustrative example and some discussions.

2.3.1 Analysis Under Poisson Models

In this subsection, we assume that the N_i's are non-homogeneous Poisson processes with the rate function

$$E\{ dN_i(t) | \boldsymbol{Z}_i \} = r(t; \boldsymbol{\beta}, \boldsymbol{Z}_i) \, dt = r_0(t; \boldsymbol{\beta}_1) \exp(\boldsymbol{\beta}_2^T \boldsymbol{Z}_i) \, dt \, , \tag{2.8}$$

$i = 1, \ldots, n$. In the above, $\boldsymbol{\beta} = (\boldsymbol{\beta}_1^T, \boldsymbol{\beta}_2^T)^T$ denotes the unknown parameters and $r_0(t; \boldsymbol{\beta}_1)$ is a function of t known up to $\boldsymbol{\beta}_1$. Some simple and commonly used choices for $r_0(t; \boldsymbol{\beta}_1)$ include $r_0(t; \boldsymbol{\beta}_1) = \boldsymbol{\beta}_1^T \, \phi(t)$ and

$$r_0(t; \boldsymbol{\beta}_1) = \exp\left\{ \boldsymbol{\beta}_1^T \, \boldsymbol{\phi}(t) \right\} . \tag{2.9}$$

In the above, $\boldsymbol{\phi}(t)$ is a vector of known functions of t such as $\boldsymbol{\phi}(t) = (1, t, \log(t))^T$ or

$$\boldsymbol{\phi}(t) = (1, t, \ldots, t^q)^T \tag{2.10}$$

with q being a known integer.

It is apparent that under the formulation (2.9), the rate function $r(t; \boldsymbol{\beta}, \boldsymbol{Z}_i)$ can be rewritten as

$$r(t; \boldsymbol{\beta}, \boldsymbol{Z}_i) = \exp\left\{ \boldsymbol{\beta}^T \boldsymbol{Z}_i^*(t) \right\} ,$$

where $\boldsymbol{Z}_i^*(t) = (\boldsymbol{\phi}^T(t), \boldsymbol{Z}_i^T)^T$. Also the likelihood function of $\boldsymbol{\beta}$ is proportional to

$$L(\boldsymbol{\beta}) = \prod_{i=1}^{n} L_i(\boldsymbol{\beta})$$

$$= \prod_{i=1}^{n} \prod_{j=1}^{m_i} \exp\{-\Delta\mu_{i,j}(\boldsymbol{\beta})\} \{\Delta\mu_{i,j}(\boldsymbol{\beta})\}^{\Delta n_{i,j}}$$

$$= \prod_{i=1}^{n} \exp\{-\mu_i(\boldsymbol{\beta})\} \prod_{j=1}^{m_i} \{\Delta\mu_{i,j}(\boldsymbol{\beta})\}^{\Delta n_{i,j}} ,$$

where $\Delta\mu_{i,j}(\boldsymbol{\beta}) = \int_{t_{i,j-1}}^{t_{i,j}} r(t; \boldsymbol{\beta}, \boldsymbol{Z}_i)\, dt$, $\mu_i(\boldsymbol{\beta}) = \int_0^{t_{i,m_i}} r(t; \boldsymbol{\beta}, \boldsymbol{Z}_i)\, dt$ and $\Delta n_{i,j} = n_{i,j} - n_{i,j-1}$ with $t_{i,0} = 0$ and $n_{i,0} = 0$. It follows that one can obtain the maximum likelihood estimator of $\boldsymbol{\beta}$, which is consistent and asymptotically has a normal distribution. The determination of the estimator is discussed in more details in the next subsection for a more general situation.

Suppose that $t_{1,m_1} = \cdots = t_{n,m_n}$. That is, all subjects have the same last observation time point. In this case, for inference about $\boldsymbol{\beta}_2$, a conditional likelihood can actually be derived as

$$L_c(\boldsymbol{\beta}) = \exp\left(\sum_{i=1}^{n} n_{i,m_i}\, \boldsymbol{\beta}_2^T\, \boldsymbol{Z}_i \right) \left\{ \sum_{i=1}^{n} \exp(\boldsymbol{\beta}_2^T\, \boldsymbol{Z}_i) \right\}^{-\sum_{i=1}^{n} n_{i,m_i}} .$$

It is easy to see that for this simple situation, the inference about $\boldsymbol{\beta}_2$ depends only on the observed numbers of the events at the last observation time or the total numbers of occurrences of the events during the whole follow-up period. In other words, the number of observations and observation times before the last observation time do not contain relevant information about $\boldsymbol{\beta}_2$.

2.3.2 Analysis Under Mixed Poisson Models

Similar to the situation considered in Sect. 2.2, the Poisson process assumption used in the previous subsection may be questionable in practice, and one way to relax it is to consider mixed Poisson processes (Thall, 1988). More specifically, assume that there exists a latent variable ν_i and given ν_i and $\boldsymbol{Z}_i$, $N_i(t)$ is a non-homogeneous Poisson process with the rate function

$$E\{dN_i(t)|\boldsymbol{Z}_i, \nu_i\} = \nu_i\, r(t; \boldsymbol{\beta}, \boldsymbol{Z}_i)\, dt = \nu_i\, r_0(t; \boldsymbol{\beta}_1) \exp(\boldsymbol{\beta}_2^T \boldsymbol{Z}_i)\, dt . \quad (2.11)$$

Here $\boldsymbol{\beta}$ and $r_0(t; \boldsymbol{\beta}_1)$ are defined as in model (2.8). Furthermore, it is assumed that the ν_i's are i.i.d. with the density function $g(\nu; \boldsymbol{\alpha})$ known up to the unknown vector of parameters $\boldsymbol{\alpha}$. Then it is easy to see that the likelihood function of $\boldsymbol{\theta} = (\boldsymbol{\beta}^T, \boldsymbol{\alpha}^T)^T$ is proportional to

$$L(\boldsymbol{\theta}) = \prod_{i=1}^{n} L_i(\boldsymbol{\theta})$$
$$= \prod_{i=1}^{n} \int_0^\infty \prod_{j=1}^{m_i} \exp\{-\nu\, \Delta\mu_{i,j}(\boldsymbol{\beta})\}\, \{\nu\, \Delta\mu_{i,j}(\boldsymbol{\beta})\}^{\Delta n_{i,j}}\, g(\nu; \boldsymbol{\alpha})\, d\nu$$
$$= \prod_{i=1}^{n} \int_0^\infty \exp\{-\nu\, \mu_i(\boldsymbol{\beta})\} \prod_{j=1}^{m_i} \{\nu\, \Delta\mu_{i,j}(\boldsymbol{\beta})\}^{\Delta n_{i,j}}\, g(\nu; \boldsymbol{\alpha})\, d\nu\,.$$

It follows that one can estimate both $\boldsymbol{\beta}$ and $\boldsymbol{\alpha}$ by maximizing the likelihood function above.

Suppose that the baseline rate function $r_0(t; \boldsymbol{\beta}_1)$ has the form (2.9) and the latent variables ν_i's follow the gamma distribution with the density function

$$g(\nu; \alpha_1, \alpha_2) = \frac{\nu^{\alpha_1 - 1}}{\alpha_2^{\alpha_1}\, \Gamma(\alpha_1)} \exp(-\nu/\alpha_2)\,,\ \nu > 0\,.$$

That is, the $N_i(t)$'s are negative binomial processes (Lawless, 1987b). Then one can show that $L_i(\boldsymbol{\theta})$ is equivalent to

$$L_i(\boldsymbol{\theta}) = \frac{\Gamma(\alpha_1 + n_{i,m_i})}{\Gamma(\alpha_1)}\, \alpha_2^{-\alpha_1}\, (\mu_i(\boldsymbol{\beta}) + \alpha_2^{-1})^{-(n_{i,m_i}+\alpha_1)} \prod_{j=1}^{m_i} (\Delta\mu_{i,j}(\boldsymbol{\beta}))^{\Delta n_{i,j}}\,.$$

Define $\mathbf{a}_i^T = (\Delta\mu_{i,1}, \ldots, \Delta\mu_{i,m_i})$, $\mathbf{a}^T = (\mathbf{a}_1^T, \ldots, \mathbf{a}_n^T)$,

$$W = \frac{\partial \log L(\boldsymbol{\theta})}{\partial \mathbf{a}} = (W_1^T, \ldots, W_n^T)^T$$

with $W_i = \partial \log L(\boldsymbol{\theta})/\partial \mathbf{a}_i$, and

$$D = \frac{\partial \mathbf{a}}{\partial \boldsymbol{\beta}} = \begin{bmatrix} D_1 \\ \vdots \\ D_n \end{bmatrix} = \begin{bmatrix} \partial \mathbf{a}_1/\partial \boldsymbol{\beta} \\ \vdots \\ \partial \mathbf{a}_n/\partial \boldsymbol{\beta} \end{bmatrix} = \begin{bmatrix} \mathrm{diag}(\mathbf{a}_1)\, \mathbf{X}_1 \\ \vdots \\ \mathrm{diag}(\mathbf{a}_n)\, \mathbf{X}_n \end{bmatrix},$$

where $\mathbf{X}_i = (\boldsymbol{Z}_i^{*T}(t_{i,1}), \ldots, \boldsymbol{Z}_i^{*T}(t_{i,m_i}))$. Then we have

$$\log L_i(\boldsymbol{\theta}) \propto I(n_{i,m_i} > 0) \sum_{k=0}^{n_{i,m_i}-1} \log(\alpha_1 + k) - \alpha_1 \log(\alpha_2)$$
$$-\, (n_{i,m_i} + \alpha_1)\, \log\left(\mu_{i,m_i} + \alpha_2^{-1}\right) + \sum_{j=1}^{m_i} \Delta n_{i,j}\, \log(\Delta\mu_{i,j})\,,$$

$i = 1, \ldots, n$. It follows that the score function $U(\boldsymbol{\beta}) = \partial \log L(\boldsymbol{\theta})/\partial \boldsymbol{\beta}$ has the form

$$U(\boldsymbol{\beta}) \,=\, D^T\,W \,=\, \sum_{i=1}^{n} D_i^T\,W_i$$
$$= \sum_{i=1}^{n} \mathbf{X}_i^T \left(\Delta\mathbf{n}_i \,-\, \frac{n_{i,m_i} + \alpha_1}{\mu_{i,m_i} + \alpha_2^{-1}}\,\mathbf{a}_i \right)$$

since

$$W_{i,j} \,=\, \frac{\Delta n_{i,j}}{\Delta\mu_{i,j}} \,-\, \frac{n_{i,m_i} + \alpha_1}{\mu_{i,m_i} + \alpha_2^{-1}}\,,$$

where $\Delta\mathbf{n}_i \,=\, (\Delta n_{i,1}, \ldots, \Delta n_{i,m_i})^T$. The computation of the score function $U(\boldsymbol{\alpha}) \,=\, \partial \log L(\boldsymbol{\theta})/\partial\boldsymbol{\alpha}$ is straightforward.

It follows that one can obtain the maximum likelihood estimators, denoted by $\hat{\boldsymbol{\beta}}_{MPL}$ and $\hat{\boldsymbol{\alpha}}_{MPL}$, of $\boldsymbol{\beta}$ and $\boldsymbol{\alpha}$ by solving the score equations $U(\boldsymbol{\beta}) \,=\, 0$ and $U(\boldsymbol{\alpha}) \,=\, 0$ together. By the standard maximum likelihood theory, $\hat{\boldsymbol{\beta}}_{MPL}$ and $\hat{\boldsymbol{\alpha}}_{MPL}$ are consistent and have joint asymptotic normal distribution with their covariance matrix consistently estimated by the observed Fisher information matrix. Note that in general, there is no closed form for the integration involved in the likelihood function $L(\boldsymbol{\theta})$ and thus some numerical algorithms have to be used. Some discussions on this can be found in Thall (1988) among others.

2.3.3 An Illustration

To illustrate the maximum likelihood estimation procedures described above, we apply them to a set of current status data arising from a tumorigenicity experiment on multiple incidental tumors. The experiment consists of 99 female and 100 male rats. The observed data, presented in Table 2.1 and reproduced from Ii et al. (1987) and Sun and Kalbfleisch (1993), give the total numbers of the tumors that each rat had developed up to the 10-week interval within which they died. In other words, each animal is observed only once at the death and the death times are given by 10-week intervals. For the convenience, it is assumed below that the observation is at the endpoint of each 10-week interval. The number in the table denotes the number of rats which died in the ith interval and in which k tumors were found. Note that the term incidental means that the presence of such tumors has no effect on the death rate. In other words, the death or observation time is independent of the occurrences of the tumors.

To compare the tumor occurrence rates between female and male rats, let $N_i(t)$ denote the number of tumors that have occurred up to time t for the ith animal and define $Z_i \,=\, 1$ if animal i is male and 0 otherwise, $i \,=\, 1, \ldots, 199$. Suppose that the $N_i(t)$'s are mixed Poisson processes with the rate function

$$E\,\{\, dN_i(t) | Z_i\,,\, \nu_i\,\} \,=\, \nu_i\, \exp(\beta_1 \,+\, \beta_2 Z_i)\, dt\,.$$

Table 2.1. Observed number of k-tumored animals at interval i

		(a) Males								(b) Females							
		k								k							
Week	Interval i	0	1	2	3	4	5	6	7	0	1	2	3	4	5	6	7
1–10	1																
11–20	2																
21–30	3																
31–40	4	3															
41–50	5	11								9	1						
51–60	6									2	1						
61–70	7	17	1							12	2						
71–80	8	2															
81–90	9	3	1								2						
91–100	10	5	3	1						1	1	2	2				
101–110	11	5	7	1	2					5	4	1	2				
111–120	12	8	5	2	1						9	5	3	1			
121–130	13		6	1						1	4	3	3				
131–140	14	1	4	1		1				2	5	4	3			1	1
141–150	15	3		2	2	1				1	2	1	3				

In the above, the ν_i's are defined as in model (2.11) with the density function $g(\nu; \alpha_1, \alpha_2)$ given in the previous subsection. The application of the maximum likelihood estimation procedure given above yields $\hat{\beta}^*_{MPL,1} = \exp(\hat{\beta}_{MPL,1}) = 0.421$ and $\hat{\beta}_{MPL,2} = -0.601$ with the estimated standard errors of 0.066 and 0.165, respectively. This indicates that the male rats seem to have a significantly lower tumor occurrence rate than the female rates. By assuming $E(\nu_i) = 1$ for all i, we obtain $\hat{\beta}^*_{MPL,1} = 0.120$ and $\hat{\beta}_{MPL,2} = -0.601$ with the estimated standard errors being 0.011 and 0.155, which give the same conclusion.

2.3.4 Discussion

The focus of this section has been on the Poisson process and parametric analysis. It is apparent that it is straightforward to generalize the inference procedures described above to or develop similar parametric inference procedures under different parametric models. Some references on this include Albert (1991), Lawless (1987a), Thall (1988) and Thall and Lachin (1988).

It is well-known that in general, parametric models and analyses should be preferred than nonparametric and semiparametric models and analyses if there is some evidence indicating or suggesting that the parametric models are reasonable or appropriate. In addition to being more efficient, parametric analyses are usually more straightforward than nonparametric and semiparametric analyses. On the other hand, in many situations, there may not exist such evidence or appropriate parametric models, or there do not exist data or information that can be used to assess the appropriateness of an assumed parametric model. In consequence, one may want to employ or rely on nonparametric and semiparametric models and the corresponding inference procedures. One advantage is that they could avoid making assumptions on parametric models and give more reasonable and/or robust analysis results. It is apparent that these general arguments apply to the analysis of panel count data considered here.

In addition to the two types of procedures mentioned above, sometimes one may prefer a third type of models or procedures or a compromise between the two. One such procedure is described in the next section, which models the baseline rate function $r_0(t)$ in model (1.3) or $r_0(t; \boldsymbol{\beta}_1)$ in model (2.8) by using a piecewise constant function (Lawless and Zhan, 1998). It is obvious that by controlling the number of steps, one can push the resulting analysis procedure more similar to either a parametric procedure or a semiparametric procedure. As another compromise between parametric and semiparametric procedures, instead of using the piecewise step function, one can employ some smooth functions such as monotone splines (Lu et al., 2009). More discussions on this are given below. Of course, as mentioned above, nonparametric and semiparametric procedures for the analysis of panel count data are discussed in later chapters.

2.4 Regression Analysis with Piecewise Models

In this section, we consider the same problem and also the same type of inference procedures in nature as those discussed in the previous section. On the other hand, as mentioned above, the inference procedures to be described can also be regarded as compromises between parametric and semiparametric procedures. Specifically, consider a recurrent event study for which we only observe panel count data. Let the $N_i(t)$'s, $\boldsymbol{Z}_i$'s, $t_{i,j}$'s, m_i's, $n_{i,j}$'s and $\Delta n_{i,j}$'s be defined as in the previous section and suppose that one is mainly interested in estimating the effects of the covariates $\boldsymbol{Z}_i$'s on the $N_i(t)$'s as before.

To describe the effects of the covariates, we assume that there exist i.i.d. latent variables $\{\nu_i\}_{i=1}^n$ with $E(\nu_i) = 1$ and given ν_i and $\boldsymbol{Z}_i$, the rate function of $N_i(t)$ has the form

$$E\{dN_i(t)|\boldsymbol{Z}_i, \nu_i\} = \nu_i\, r_0(t)\, \exp(\boldsymbol{\beta}^T \boldsymbol{Z}_i)\, dt, \tag{2.12}$$

$i = 1, \ldots, n$. In the above, $r_0(t)$ denotes an unknown baseline rate function and $\boldsymbol{\beta}$ is a vector of regression parameters. Furthermore, it is assumed that there exists a prespecified partition $0 = s_0 < \cdots < s_k < \infty$ such that $r_0(t) = \alpha_l$ for $t \in S_l = (s_{l-1}, s_l]$, where the α_l's are unknown constants. That is, the baseline rate function $r_0(t)$ is a step function. It is apparent that the model above can be seen a special case of model (2.11) and implies the proportional rate model (1.3).

For estimation of the regression parameter $\boldsymbol{\beta}$ in model (2.12), in the following, we consider two inference procedures. First we assume that the $N_i(t)$'s are non-homogeneous Poisson processes and develop the maximum likelihood estimation procedure. A generalized estimating equation procedure is then discussed and followed by an illustration and some discussions.

2.4.1 Likelihood-Based Approach

In this subsection, we assume that the $N_i(t)$'s are non-homogeneous Poisson processes with the rate function given by model (2.12). It follows that we have

$$E\{\, N_i(t) | \boldsymbol{Z}_i, \nu_i \,\} \,=\, \nu_i \, \mu_0(t) \, \exp(\boldsymbol{\beta}^T \boldsymbol{Z}_i) \,, \tag{2.13}$$

where $\mu_0(t) = \sum_{l=1}^k \alpha_l \, u_l(t)$ with $u_l(t) = \max\{0, \min(s_l, t) - s_{l-1}\}$, representing the length of the intersection of the two intervals $(0, t]$ and S_l. For each (i, j), define $\mu_{i,j} = \mu_0(t_{i,j}) \exp(\boldsymbol{\beta}^T \boldsymbol{Z}_i)$ and

$$\Delta\mu_{i,j} \,=\, \mu_{i,j} \,-\, \mu_{i,j-1} \,=\, \mu_{0,i,j} \, \exp(\boldsymbol{\beta}^T \boldsymbol{Z}_i) \,,$$

$j = 1, \ldots, m_i$, $i = 1, \ldots, n$. Here $\mu_{0,i,j} = \sum_{l=1}^k \alpha_l \, u_l(i,j)$ and

$$u_l(i,j) \,=\, \max\{\, 0, \min(s_l, t_{i,j}) \,-\, \max(s_{l-1}, t_{i,j-1}) \,\} \,,$$

denoting the length of the intersection of the two intervals $(t_{i,j-1}, t_{i,j}]$ and S_l, $l = 1, \ldots, k$. Then under the assumption above, one can easily show that

$$E\{\, N_i(t_{i,j}) \,-\, N_i(t_{i,j-1}) | \boldsymbol{Z}_i, \nu_i \,\} \,=\, \nu_i \, \Delta\mu_{i,j} \,.$$

For the simplicity, we assume that the ν_i's follow the gamma distribution with the density function $g(\nu; \gamma)$ given in Sect. 2.2.1. That is, the ν_i's have the mean one and variance γ. It follows that the likelihood function of $\boldsymbol{\beta}$, $\boldsymbol{\alpha} = (\alpha_1, \ldots, \alpha_k)^T$ and γ is proportional to

$$L(\boldsymbol{\beta}, \boldsymbol{\alpha}, \gamma) \,=\, \prod_{i=1}^n \int_0^\infty \prod_{j=1}^{m_i} \exp(-\nu_i \Delta\mu_{i,j}) \, (\nu_i \Delta\mu_{i,j})^{\Delta n_{i,j}} \, g(\nu_i; \gamma) \, d\nu_i$$

or

$$L(\boldsymbol{\beta}, \boldsymbol{\alpha}, \gamma) = \prod_{i=1}^{n} \left(\prod_{j=1}^{m_i} \Delta\mu_{i,j}^{\Delta n_{i,j}} \right) \frac{\Gamma(n_{i,m_i} + 1/\gamma)\, \gamma^{n_{i,m_i}}}{\Gamma(1/\gamma)\, (1 + \gamma\, \mu_{i,m_i})^{n_{i,m_i} + 1/\gamma}} .$$

The resulting log likelihood function has the form

$$l(\boldsymbol{\beta}, \boldsymbol{\alpha}, \gamma) = \sum_{i=1}^{n} \left\{ \sum_{j=1}^{m_i} (\, \Delta n_{i,j} \log \Delta\mu_{i,j}\,) + n_{i,m_i} \log\gamma + \log\Gamma\left(n_{i,m_i} + \frac{1}{\gamma} \right) \right.$$
$$\left. - \log\Gamma\left(\frac{1}{\gamma}\right) - \left(n_{i,m_i} + \frac{1}{\gamma} \right) \log(1 + \gamma\mu_{i,m_i}) \right\} .$$

For the determination of the maximum likelihood estimators of $\boldsymbol{\beta}$, $\boldsymbol{\alpha}$ and γ, we need their score functions, which have the form

$$\frac{\partial l(\boldsymbol{\beta}, \boldsymbol{\alpha}, \gamma)}{\partial \boldsymbol{\beta}} = \sum_{i=1}^{n} \frac{n_{i,m_i} - \mu_{i,m_i}}{1 + \gamma\, \mu_{i,m_i}} \boldsymbol{Z}_i , \tag{2.14}$$

$$\frac{\partial l(\boldsymbol{\beta}, \boldsymbol{\alpha}, \gamma)}{\partial \alpha_l} = \sum_{i=1}^{n} \sum_{j=1}^{m_i} \frac{(\Delta n_{i,j} - \Delta\mu_{i,j}) u_l(i,j) \exp(\boldsymbol{\beta}^T \boldsymbol{Z}_i)}{\Delta\mu_{i,j}}$$
$$- \sum_{i=1}^{n} \frac{\gamma(n_{i,m_i} - \mu_{i,m_i})\, u_l(i,+)\, \exp(\boldsymbol{\beta}^T \boldsymbol{Z}_i)}{1 + \gamma\mu_{i,m_i}} , \tag{2.15}$$

and

$$\frac{\partial l(\boldsymbol{\beta}, \boldsymbol{\alpha}, \gamma)}{\partial \gamma} = \sum_{i=1}^{n} \left\{ \frac{n_{i,m_i} - \mu_{i,m_i}}{\gamma(1 + \gamma\mu_{i,m_i})} + \gamma^{-2} \log(1 + \gamma\mu_{i,m_i}) \right\}$$
$$- \gamma^{-1} \sum_{i=1}^{n} \sum_{s=1}^{n_{i,m_i}} \{1 + \gamma(s-1)\}^{-1} ,$$

respectively, where $u_l(i,+) = \sum_{j=1}^{m_i} u_j(i,j)$, $l = 1, \ldots, k$. Thus it is natural to solve the score equations

$$\frac{\partial l(\boldsymbol{\beta}, \boldsymbol{\alpha}, \gamma)}{\partial \boldsymbol{\beta}} = 0 , \quad \frac{\partial l(\boldsymbol{\beta}, \boldsymbol{\alpha}, \gamma)}{\partial \alpha_l} = 0 , \quad \frac{\partial l(\boldsymbol{\beta}, \boldsymbol{\alpha}, \gamma)}{\partial \gamma} = 0 , \ l = 1, \ldots, k$$

together by using, for example, the Newton-Raphson algorithm. As an alternative, one could apply the EM algorithm (Dempster et al., 1977) given below and developed by Lawless and Zhan (1998).

To define the pseudo-complete data, assume that one observes the ν_i's and c_{ijl}, the number of the occurrences of the recurrent event of interest within the intersection of S_l and $(t_{i,j-1}, t_{i,j}]$, $j = 1, \ldots, m_i$, $i = 1, \ldots, n$,

$l = 1, \ldots, k$. Define $u_{ijl} = \alpha_l \, u_l(i,j) \, \exp(\boldsymbol{\beta}^T \boldsymbol{Z}_i)$. Then the log likelihood function based on the pseudo-complete data ν_i's and c_{ijl}'s can be written as

$$l_{pl}(\boldsymbol{\beta}, \boldsymbol{\alpha}, \gamma) = l_{pl,1}(\gamma) + l_{pl,2}(\boldsymbol{\beta}, \boldsymbol{\alpha}),$$

where

$$l_{pl,1}(\gamma) = -n \left\{ \log \Gamma\left(\frac{1}{\gamma}\right) + \frac{\log \gamma}{\gamma} \right\} + \gamma^{-1} \sum_i (\log \nu_i - \nu_i)$$

and

$$l_{pl,2}(\boldsymbol{\beta}, \boldsymbol{\alpha}) = \sum_{i=1}^{n} \sum_{j=1}^{m_i} \sum_{l=1}^{k} c_{ijl} \, \log u_{ijl} - \sum_{i=1}^{n} \nu_i \, \mu_{i,m_i} \, .$$

Denote $\boldsymbol{\theta} = (\boldsymbol{\beta}^T, \boldsymbol{\alpha}^T, \gamma)^T$. The EM algorithm can be carried out as follows.
Step 1. Choose an initial estimator $\boldsymbol{\theta}^{(0)}$.
Step 2. E-step. At the mth iteration, compute

$$l_{pl,1}^{(m)}(\gamma \,|\, \boldsymbol{\theta}^{(m-1)}) = E\left\{ l_{pl,1}(\gamma \,|\, n'_{i,j}s, \boldsymbol{\theta}^{(m-1)}) \right\}$$
$$= -n \left\{ \log \Gamma\left(\frac{1}{\gamma}\right) + \frac{\log \gamma}{\gamma} \right\} + \gamma^{-1} \sum_i \left(\widetilde{\log \nu_i}^{(m)} - \widetilde{\nu_i}^{(m)} \right)$$

and

$$l_{pl,2}^{(m)}(\boldsymbol{\beta}, \boldsymbol{\alpha} \,|\, \boldsymbol{\theta}^{(m-1)}) = E\left\{ l_{p,2}(\boldsymbol{\beta}, \boldsymbol{\alpha} \,|\, n'_{i,j}s, \boldsymbol{\theta}^{(m-1)}) \right\}$$
$$= \sum_{i=1}^{n} \sum_{j=1}^{m_i} \sum_{l=1}^{k} \widetilde{c_{ijl}}^{(m)} \log u_{ijl} - \sum_{i=1}^{n} \widetilde{\nu_i}^{(m)} \mu_{i,m_i} \, .$$

In the above,

$$\widetilde{\log \nu_i}^{(m)} = \Phi(C_{i1}^{(m)}) - \log(C_{i2}^{(m)}),$$
$$\widetilde{\nu_i}^{(m)} = \frac{C_{i1}^{(m)}}{C_{i2}^{(m)}},$$

and

$$\widetilde{c_{ijl}}^{(m)} = \frac{\Delta n_{i,j} \, \alpha_l^{(m-1)} \, u_l(i,j)}{\sum_{b=1}^{k} \alpha_b^{(m-1)} \, u_b(i,j)},$$

where

$$C_{i1}^{(m)} = n_{i,m_i} + \frac{1}{\gamma^{(m-1)}}, \; C_{i2}^{(m)} = \mu_{i,m_i}^{(m-1)} + \frac{1}{\gamma^{(m-1)}}$$

and $\Phi(t) = d \log \Gamma(t)/dt$.
Step 3. M-step. Maximize $l_{pl,1}^{(m)}(\gamma \,|\, \boldsymbol{\theta}^{(m-1)})$ and $l_{pl,2}^{(m)}(\boldsymbol{\beta}, \boldsymbol{\alpha} \,|\, \boldsymbol{\theta}^{(m-1)})$ with respect to $\boldsymbol{\theta}$ to obtain the estimator $\boldsymbol{\theta}^{(m)}$.
Step 4. Repeat Steps 2 and 3 until the convergence.

To implement the EM algorithm above, one needs to choose an initial estimator $\boldsymbol{\theta}^{(0)}$ and a convergence criterion. For the former, a simple and natural approach is to set $\nu_i = 1$ for all i and the α_l's to be identical in (2.13) and then to employ the resulting estimators as the initial estimators of $\boldsymbol{\beta}$ and $\boldsymbol{\alpha}$. For the parameter γ, one can use the moment estimator given by

$$\frac{1}{n} \sum_{i=1}^{n} \left\{ \frac{n_{i,m_i}}{\mu_0(t) \exp(\boldsymbol{\beta}^T \boldsymbol{Z}_i)} - 1 \right\}^2$$

with replacing $\mu_0(t)$ and $\boldsymbol{\beta}$ by their initial estimators. In practice, of course, one may want to employ several different initial estimators to hope that they all result in the same final estimators. For the convergence criterion, a common one is to compare the consecutive values of the estimators $\boldsymbol{\theta}^{(m-1)}$ and $\boldsymbol{\theta}^{(m)}$ or the values of the log likelihood function $l_{pl}(\boldsymbol{\beta}, \boldsymbol{\alpha}, \gamma)$ at $\boldsymbol{\theta}^{(m-1)}$ and $\boldsymbol{\theta}^{(m)}$. More specifically, for given positive numbers ϵ_1 and ϵ_2, one can stop the iteration if

$$\max_l \mid \theta_l^{(m)} - \theta_l^{(m-1)} \mid \leq \epsilon_1$$

or

$$\mid l_{pl}(\boldsymbol{\theta}^{(m)}) - l_{pl}(\boldsymbol{\theta}^{(m-1)}) \mid \leq \epsilon_2 \,,$$

where the maximum above is over all components of $\boldsymbol{\theta}$. An alternative, suggested by Lawless and Zhan (1998), is to use

$$\max_l \frac{\mid \theta_l^{(m)} - \theta_l^{(m-1)} \mid}{\mid \theta_l^{(m-1)} \mid + 10^{-5}} \leq \epsilon_1$$

and

$$\frac{\mid l_{pl}(\boldsymbol{\theta}^{(m)}) - l_{pl}(\boldsymbol{\theta}^{(m-1)}) \mid}{\mid l_{pl}(\boldsymbol{\theta}^{(m-1)}) \mid + 10^{-5}} \leq \epsilon_2$$

together.

Let $\hat{\boldsymbol{\theta}}_L = (\hat{\boldsymbol{\beta}}_L^T, \hat{\boldsymbol{\alpha}}_L^T, \hat{\gamma}_L)^T$ denote the maximum likelihood estimator of $\boldsymbol{\theta}$ obtained above. Then it follows from the standard maximum likelihood theory that $\hat{\boldsymbol{\theta}}_L$ is consistent and asymptotically follows a multivariate normal distribution. Furthermore, its covariance matrix can be consistently estimated by the observed Fisher information matrix or the negative second derivative of the log likelihood function $l(\boldsymbol{\beta}, \boldsymbol{\alpha}, \gamma)$ calculated at the maximum likelihood estimator. For this, one can directly find the second derivative or use the EM algorithm (Louis, 1982).

2.4.2 Estimating Equation-Based Approach

As discussed in Sect. 2.2, the Poisson process or mixed Poisson process assumption may not hold in practice, and one way to address it is to employ the estimating equation or generalized estimating equation approach (McCulluagh and Nelder, 1989). The general idea behind the latter approach is to only model the mean function and the covariance matrix of the underlying response process or the recurrent event process, and the resulting estimation procedure is usually robust. Also to follow the idea discussed in Sect. 2.2, for estimation of $\boldsymbol{\beta}$ and $\boldsymbol{\alpha}$, one could directly employ the score functions given in (2.14) and (2.15) and solve the estimating equations

$$\frac{\partial l(\boldsymbol{\beta}, \boldsymbol{\alpha}, \gamma)}{\partial \boldsymbol{\beta}} = 0 \,, \; \frac{\partial l(\boldsymbol{\beta}, \boldsymbol{\alpha}, \gamma)}{\partial \alpha_l} = 0 \,, \; l = 1, \ldots, k$$

while ignoring the mixed Poisson process assumption. In the following, we describe this using the generalized estimating equation theory (McCulluagh and Nelder, 1989).

In this subsection, we use the same notation defined in the previous subsection. Also define $\boldsymbol{Y}_i = (\Delta n_{i,1}, \cdots, \Delta n_{i,m_i})^T$, $\boldsymbol{a}_i = (\Delta \mu_{i,1}, \cdots, \Delta \mu_{i,m_i})^T$ as in Sect. 2.3.2, and $\boldsymbol{b}_i = \mathrm{diag}(\mathbf{a}_i)$, $i = 1, \ldots, n$. Then it is easy to see that the covariance matrix of $\boldsymbol{Y}_i$ under the mixed Poisson model specified in the previous subsection has the form

$$\mathbf{V}_i = \mathbf{b}_i + \gamma \, \mathbf{a}_i \, \mathbf{a}_i^T \,. \tag{2.16}$$

Now assume that the recurrent event processes $N_i(t)$'s only satisfy (2.13) and (2.16), and let $\mathbf{D}_i = \partial \mathbf{a}_i / \partial(\boldsymbol{\beta}^T, \boldsymbol{\alpha}^T)$ and $\boldsymbol{S}_i = \boldsymbol{Y}_i - \mathbf{a}_i$. Then by following the generalized estimating equation theory, for estimation of $\boldsymbol{\beta}$ and $\boldsymbol{\alpha}$, we have the generalized estimating equations

$$\mathbf{U}_1(\boldsymbol{\beta}, \boldsymbol{\alpha}, \gamma) = \sum_{i=1}^{n} \mathbf{D}_i^T \, \mathbf{V}_i^{-1} \, \mathbf{S}_i = 0 \,. \tag{2.17}$$

One can easily show that

$$\mathbf{U}_1(\boldsymbol{\beta}, \boldsymbol{\alpha}, \gamma) = \begin{pmatrix} \frac{\partial l(\boldsymbol{\beta}, \boldsymbol{\alpha}, \gamma)}{\partial \boldsymbol{\beta}} \\ \frac{\partial l(\boldsymbol{\beta}, \boldsymbol{\alpha}, \gamma)}{\partial \alpha_1} \\ \cdots \\ \frac{\partial l(\boldsymbol{\beta}, \boldsymbol{\alpha}, \gamma)}{\partial \alpha_k} \end{pmatrix} .$$

That is, the equations defined in (2.17) are the same as those used in the previous subsection for estimation of $\boldsymbol{\beta}$ and $\boldsymbol{\alpha}$. Note that $\mathbf{V}_i$ given in (2.16) is a working covariance matrix, which may be correct or may not, and also one may use other forms. For the estimation of $\boldsymbol{\beta}$ and $\boldsymbol{\alpha}$, a simple approach is to adopt (2.16) and solve the Eqs. (2.17) based on a given value of γ such as $\gamma = 0$. Alternatively and more generally, one may want to develop an additional estimating equation for γ and estimate all parameters together.

One such estimating equation for γ is the simple moment equation

$$U_2(\boldsymbol{\beta}, \boldsymbol{\alpha}, \gamma) = \sum_{i=1}^{n} w_i \left\{ (n_{i,m_i} - \mu_{i,m_i})^2 - \sigma_i^2 \right\} = 0, \tag{2.18}$$

suggested by Lawless and Zhan (1998), where

$$\sigma_i^2 = \mathrm{Var}(n_{i,m_i}) = \mu_{i,m_i} + \gamma\, \mu_{i,m_i}^2$$

and the w_i's are some weights. Some simple choices for the weights include $w_i = 1$, $w_i = 1/\sigma_i^2$ and $w_i = \mu_{i,m_i}^2/\sigma_i^4$. Now one can estimate $\boldsymbol{\beta}$, $\boldsymbol{\alpha}$ and γ by iteratively solving the Eqs. (2.17) and (2.18) as follows.
Step 1. Choose an initial estimator $\boldsymbol{\theta}^{(0)}$.
Step 2. At the mth iteration, obtain the updated estimators of $\boldsymbol{\beta}$ and $\boldsymbol{\alpha}$ as

$$\begin{pmatrix} \boldsymbol{\beta}^{(m)} \\ \boldsymbol{\alpha}^{(m)} \end{pmatrix} = \left\{ \begin{pmatrix} \boldsymbol{\beta} \\ \boldsymbol{\alpha} \end{pmatrix} + \left(\sum_{i=1}^{n} \mathbf{D}_i^T \mathbf{V}_i^{-1} \mathbf{D}_i \right)^{-1} \right.$$
$$\left. \times \left(\sum_{i=1}^{n} \mathbf{D}_i^T \mathbf{V}_i^{-1} \mathbf{S}_i \right) \right\} \Big|_{\boldsymbol{\beta}=\boldsymbol{\beta}^{(m-1)}, \boldsymbol{\alpha}=\boldsymbol{\alpha}^{(m-1)}, \gamma=\gamma^{(m-1)}} .$$

Step 3. Also at the mth iteration, obtain the updated estimator of γ as

$$\gamma^{(m)} = \gamma^{(m-1)} - \left\{ \left(\frac{\partial U_2}{\partial \gamma} \right)^{-1} U_2 \right\} \Big|_{\boldsymbol{\beta}=\boldsymbol{\beta}^{(m)}, \boldsymbol{\alpha}=\boldsymbol{\alpha}^{(m)}, \gamma=\gamma^{(m-1)}} .$$

Step 4. Repeat Steps 2 and 3 until the convergence.

It is apparent that the discussion on the selection of initial estimators and the convergence criterion given in the previous subsection applies here. Let $\hat{\boldsymbol{\theta}}_E = (\hat{\boldsymbol{\beta}}_E^T, \hat{\boldsymbol{\alpha}}_E^T, \hat{\gamma}_E)^T$ denote the estimator of $\boldsymbol{\theta}$ defined above. Then it can be shown by using the estimating equation theory that under some mild conditions, $\hat{\boldsymbol{\beta}}_E$ and $\hat{\boldsymbol{\alpha}}_E$ are consistent and their joint distribution can be asymptotically approximated by a multivariate normal distribution (Lawless and Zhan, 1998; Liang and Zeger, 1986; White, 1982). These results hold no matter whether the covariance matrices V_i's specified by (2.16) are correct or not.

For estimation of the covariance matrix of $\hat{\boldsymbol{\beta}}_E$ and $\hat{\boldsymbol{\alpha}}_E$, define

$$\boldsymbol{\Sigma}_n(\boldsymbol{\beta}, \boldsymbol{\alpha}, \gamma) = \begin{pmatrix} \boldsymbol{\Sigma}_{n,11} & \boldsymbol{\Sigma}_{n,12} \\ \boldsymbol{\Sigma}_{n,21} & \boldsymbol{\Sigma}_{n,22} \end{pmatrix}$$

and

$$\boldsymbol{\Gamma}_n(\boldsymbol{\beta}, \boldsymbol{\alpha}, \gamma) = \begin{pmatrix} \boldsymbol{\Gamma}_{n,11} & \boldsymbol{\Gamma}_{n,12} \\ \boldsymbol{\Gamma}_{n,21} & \boldsymbol{\Gamma}_{n,22} \end{pmatrix}.$$

In the above,

$$\boldsymbol{\Sigma}_{n,11} = \frac{1}{n} E\left\{ - \frac{\partial \mathbf{U}_1}{\partial(\boldsymbol{\beta}^T, \boldsymbol{\alpha}^T)} \right\} = \frac{1}{n} \sum_{i=1}^{n} \mathbf{D}_i^T \mathbf{V}_i^{-1} \mathbf{D}_i$$

$$\boldsymbol{\Sigma}_{n,12} = \frac{1}{n} E\left(- \frac{\partial \mathbf{U}_1}{\partial \gamma} \right) = 0,$$

$$\boldsymbol{\Sigma}_{n,21} = \frac{1}{n} E\left\{ - \frac{\partial U_2}{\partial(\boldsymbol{\beta}^T, \boldsymbol{\alpha}^T)} \right\} = \frac{1}{n} \sum_{i=1}^{n} w_i \, (1 + 2\,\gamma\,\mu_{i,m_i}) \, \frac{\partial \mu_{i,m_i}}{\partial(\boldsymbol{\beta}^T, \boldsymbol{\alpha}^T)},$$

$$\boldsymbol{\Sigma}_{n,22} = \frac{1}{n} E\left(- \frac{\partial U_2}{\partial \gamma} \right) = \frac{1}{n} \sum_{i=1}^{n} w_i \, \mu_{i,m_i}^2,$$

$$\boldsymbol{\Gamma}_{n,11} = \frac{1}{n} \sum_{i=1}^{n} \mathbf{D}_i^T \mathbf{V}_i^{-1} \mathbf{S}_i \mathbf{S}_i^T \mathbf{V}_i^{-1} \mathbf{D}_i,$$

$$\boldsymbol{\Gamma}_{n,12} = \frac{1}{n} \sum_{i=1}^{n} w_i \left\{ (n_{i,m_i} - \mu_{i,m_i})^2 - \sigma_i^2 \right\} \mathbf{D}_i^T \mathbf{V}_i^{-1} \mathbf{S}_i,$$

$$\boldsymbol{\Gamma}_{n,22} = \frac{1}{n} \sum_{i=1}^{n} w_i^2 \left\{ (n_{i,m_i} - \mu_{i,m_i})^2 - \sigma_i^2 \right\}^2,$$

and $\boldsymbol{\Gamma}_{n,21} = \boldsymbol{\Gamma}_{n,12}^T$. Then if the covariance matrices V_i's specified in (2.16) are correct, one can consistently estimate the asymptotic covariance matrix of $\sqrt{n}\left(\hat{\boldsymbol{\theta}}_E - \boldsymbol{\theta}_0\right)$ by

$$\boldsymbol{\Sigma}_n^{-1}(\hat{\boldsymbol{\beta}}_E^T, \hat{\boldsymbol{\alpha}}_E^T, \hat{\gamma}_E)\, \boldsymbol{\Gamma}_n(\hat{\boldsymbol{\beta}}_E^T, \hat{\boldsymbol{\alpha}}_E^T, \hat{\gamma}_E)\, \boldsymbol{\Sigma}_n^{-T}(\hat{\boldsymbol{\beta}}_E^T, \hat{\boldsymbol{\alpha}}_E^T, \hat{\gamma}_E).$$

In the above, $\boldsymbol{\theta}_0$ denotes the true value of $\boldsymbol{\theta}$ and $\boldsymbol{\Sigma}_n^{-T}$ the transpose of the inverse of the matrix $\boldsymbol{\Sigma}_n$. In this case, $\hat{\gamma}_E$ is also consistent.

In general, as mentioned above, the specification given in (2.16) may not be correct. In this case, a robust estimator of the asymptotic covariance matrix of $\hat{\boldsymbol{\beta}}_E$ and $\hat{\boldsymbol{\alpha}}_E$ is given by

$$\boldsymbol{\Sigma}_{n,11}^{-1}(\hat{\boldsymbol{\beta}}_E^T, \hat{\boldsymbol{\alpha}}_E^T, \hat{\gamma}_E)\, \boldsymbol{\Gamma}_{n,11}(\hat{\boldsymbol{\beta}}_E^T, \hat{\boldsymbol{\alpha}}_E^T, \hat{\gamma}_E)\, \boldsymbol{\Sigma}_{n,11}^{-T}(\hat{\boldsymbol{\beta}}_E^T, \hat{\boldsymbol{\alpha}}_E^T, \hat{\gamma}_E).$$

To implement both the likelihood-based and estimating equation-based procedures described above, one also needs to choose the number of partitions k and the partition points s_l's. For the selection of the s_l's, a common approach is to choose them such that they divide the observed data evenly. For k, which determines the smoothness of the baseline rate function, Lawless and Zhan (1998) suggested the range of 4–10 if the main goal is estimation of regression parameters. On the other hand, it is apparent that if one wants a smoother estimator of the baseline rate function, some large k should be used.

2.4.3 An Illustration

To illustrate the two estimation procedures described above, we apply them to the bladder tumor data discussed in Sect. 1.2.3 and given in the data set II of Chap. 9. As mentioned before, this is a set of panel count data arising from 85 patients with superficial bladder tumors. The patients belong to two treatment groups, the placebo (47) and thiotepa (38) groups. In addition to the information on the observation times and the numbers of recurrences of bladder tumors, the observed data also include the information on two baseline covariates. They are the number of initial tumors and the size of the largest initial tumor.

Table 2.2. Estimated covariate effects for the bladder tumor data

	Method I		Method II	
	$\hat{\boldsymbol{\beta}}_L$ (SD)	$\hat{\boldsymbol{\beta}}_E$ (SD)	$\hat{\boldsymbol{\beta}}_L$ (SD)	$\hat{\boldsymbol{\beta}}_E$ (SD)
β_1	−1.2191 (0.399)	−1.1749 (0.317)	−1.2200 (0.403)	−1.2387 (0.326)
β_2	0.3792 (0.109)	0.3716 (0.086)	0.3786 (0.108)	0.3818 (0.088)
β_3	−0.0103 (0.140)	−0.0094 (0.104)	−0.0100 (0.141)	−0.0086 (0.105)

For the analysis, we first define the covariates $\boldsymbol{Z}_i = (Z_{i1}, Z_{i2}, Z_{i3})^T$ such that $Z_{i1} = 1$ if subject i is in the thiotepa treatment group and 0 otherwise, and Z_{i2} and Z_{i3} denote the number of initial tumors and the size of the largest initial tumor, respectively, $i = 1, \ldots, 85$. To apply the two estimation procedures described above, we need to partition the whole observation period. In the following, we consider two methods for this. One, which is referred to as Method I below, is to divide the period $(0, 53]$ into five intervals with the s_l's being 0, 5.5, 15.5, 25.5, 40.5 and 53. The other, referred to as Method II below, is to divide the period $(0, 53]$ into eight intervals with the s_l's equal to 0, 5.5, 10.5, 15.5, 20.5, 25.5, 30.5, 40.5 and 53.

Tables 2.2 and 2.3 present the estimated covariate effects and recurrence rates of bladder tumors, respectively, given by the two estimation procedures. One can see that the results seem to be quite consistent with respect to both the partition method and the estimation procedure. In particular, they suggest that the patients in the thiotepa group seem to have a lower recurrence rate of bladder tumors than the patients in the placebo group. That is, the thiotepa treatment had some significant effects in reducing the recurrence rate of bladder tumors. On the two baseline covariates, the results indicate

Table 2.3. Estimated recurrence rates of the bladder tumors

	Method I		Method II	
Interval	$\hat{\boldsymbol{\alpha}}_L$ (SD)	$\hat{\boldsymbol{\alpha}}_E$ (SD)	$\hat{\boldsymbol{\alpha}}_L$ (SD)	$\hat{\boldsymbol{\alpha}}_E$ (SD)
1	0.1329 (0.060)	0.1329 (0.057)	0.1341 (0.061)	0.1338 (0.059)
2	0.0790 (0.036)	0.0791 (0.040)	0.0722 (0.034)	0.0722 (0.038)
3	0.0991 (0.045)	0.0992 (0.047)	0.0895 (0.042)	0.0896 (0.054)
4	0.1053 (0.048)	0.1047 (0.051)	0.0657 (0.033)	0.0658 (0.037)
5	0.0426 (0.023)	0.0424 (0.029)	0.1424 (0.067)	0.1421 (0.073)
6			0.0798 (0.041)	0.0789 (0.042)
7			0.1176 (0.055)	0.1167 (0.061)
8			0.0430 (0.024)	0.0427 (0.029)

that the tumor recurrence rate seems to be positively related to the number of initial tumors, but has no significant correlation with the size of the largest initial tumor. With respect to the estimation of the parameter γ, all approaches suggest that γ is significantly away from zero. That is, the latent variables ν_i's indeed have non-zero variance. For example, the likelihood-based procedure gives $\hat{\gamma}_L = 2.3632$ and 2.3697 with the estimated standard errors of 0.465 and 0.528 with the use of Methods I and II, respectively.

2.4.4 Discussion

As mentioned above, the piecewise model approaches discussed in this section are essentially parametric procedures as those investigated in Sect. 2.3. On the other hand, they are usually more flexible than fully or typical parametric procedures as one can easily change the number of partition points and thus the smoothness of the baseline rate function. The flexibility of the former can also be seen in that it is often regarded as approximate parametric procedures in the sense that the piecewise model simply provides an approximation to the underlying baseline rate function. Among others, Lawless and Zhan (1998) provided some discussion on this. In particular, they showed through a simulation study that the approaches perform well and give stable results about the regression parameters and mean function with respect to the number of partitions or steps used.

Note that instead of the baseline rate function, one can alternatively and equivalently model the baseline mean function using the piecewise constant function. For example, one could start with model (2.13) and assume that $\mu_0(t)$ has the form

$$\mu_0(t) = \sum_{l=1}^{k} \alpha_l \, I(s_{l-1} < t \leq s_l) \, .$$

That is, it is a step function that jumps only at the time points s_l's. For estimation of regression parameters, one can develop both likelihood-based and estimating equation-based approaches similarly as above.

With respect to the comparison of the two estimation procedures discussed above, it is apparent that the likelihood-based approach should be used if the mixed Poisson process assumption is reasonable. In general, it may be difficult to assess the assumption and thus one may prefer the estimating equation-based approach. Of course, one may also question the appropriateness of another assumption behind both approaches, the piecewise model assumption for the baseline rate function. To relax it, one way is to allow the number of partitions k to change with the sample size n and develop a data-driven procedure for the selection of k. Another general method is to leave the baseline rate function $r_0(t)$ or mean function $\mu_0(t)$ arbitrary and to develop semiparametric estimation procedures, the subject in the following chapters.

2.5 Bibliography, Discussion, and Remarks

In addition to those mentioned above, other references that investigated the problems similar to the ones discussed in this chapter include Hinde (1982) and Breslow (1984), and both considered the log-linear model for the event rate. More specifically, the former developed the maximum likelihood approach when the model error follows a normal distribution, while the latter proposed an iterative reweighted least squares approach. More on these methods can be found in Cameron and Trivedi (1998). As mentioned before, the focus of the book is not about Poisson-based models or parametric inference procedures. On the other hand, it is not difficult to generalize the methods discussed here to more complicated situations. One such situation is that there exists some truncation (Hu and Lawless, 1996), and another one is that the observation process depends on covariates or is informative about the underlying recurrent event process of interest as discussed later.

The Poisson process plays a major role in the parametric inference procedures discussed in this chapter. Some authors have also investigated nonparametric or semiparametric procedures under the Poisson process. For example, Staniswalls et al. (1997) considered the situation where the $N_i(t)$'s are mixed Poisson processes and the rate function satisfies model (2.12) with the baseline rate function $r_0(t)$ completely unspecified. For inference, they employed some smoothing techniques and the generalized profile likelihood method (Severini and Wong, 1992) for estimation of the baseline rate function and regression parameters, respectively. Also one can find some discussion about the comparison of parametric and semiparametric inference procedures in Staniswalls et al. (1997). In particular, they showed through an example that as expected, the parametric approach may not fully capture some patterns of the underlying rate function. In contrast, the semiparametric approach can provide substantive insights that would not be revealed by the parametric approach. More discussions on the nonparametric and semiparametric methods developed under the Poisson process assumption for the analysis of panel count data are given in both Chaps. 3 and 5.

It is worth to emphasize again that throughout the chapter, it has been assumed that the observation process or the process generating the observation times $t_{i,j}$'s is independent of the recurrent event process $N_i(t)$ of interest. As discussed before and also again in later chapters, this may not be true sometimes and in this situation, the methods described above would give biased results. In other words, some new inference procedures are needed.

Springer

http://www.springer.com/978-1-4614-8714-2

Statistical Analysis of Panel Count Data
Sun, J.; Zhao, X.
2013, XV, 271 p. 18 illus., 11 illus. in color., Hardcover
ISBN: 978-1-4614-8714-2