CAMBRIDGE

Cambridge University Press
978-1-107-01416-9 - Regression Analysis of Count Data: Second Edition
A. Colin Cameron and Pravin K. Trivedi
Excerpt
More information

CHAPTER 1

Introduction

God made the integers, all the rest is the work of man.
– Kronecker

This book is concerned with models of event counts. An event count refers to the number of times an event occurs, for example, the number of airline accidents or earthquakes. It is the realization of a nonnegative integer-valued random variable. A *univariate* statistical model of event counts usually specifies a probability distribution of the number of occurrences of the event known up to some parameters. Estimation and inference in such models are concerned with the unknown parameters, given the probability distribution and the count data. Such a specification involves no other variables, and the number of events is assumed to be independently identically distributed (iid). Much early theoretical and applied work on event counts was carried out in the univariate framework. The main focus of this book, however, is on *regression analysis* of event counts.

The statistical analysis of counts within the framework of discrete parametric distributions for univariate iid random variables has a long and rich history (Johnson, Kemp, and Kotz, 2005). The Poisson distribution was derived as a limiting case of the binomial by Poisson (1837). Early applications include the classic study of Bortkiewicz (1898) of the annual number of deaths in the Prussian army from being kicked by mules. A standard generalization of the Poisson is the negative binomial distribution. It was derived by Greenwood and Yule (1920), as a consequence of apparent contagion due to unobserved heterogeneity, and by Eggenberger and Polya (1923) as a result of true contagion. The biostatistics literature of the 1930s and 1940s, although predominantly univariate, refined and brought to the forefront seminal issues that have since permeated regression analysis of both counts and durations. The development of the counting process approach unified the treatment of counts and durations. Much of the vast literature on iid counts, which addresses issues such as heterogeneity and overdispersion, true versus apparent contagion, and identifiability of Poisson mixtures, retains its relevance in the context of count data regressions. This leads to models such as the negative binomial regression model.

© in this web service Cambridge University Press www.cambridge.org

CAMBRIDGE

Cambridge University Press
978-1-107-01416-9 - Regression Analysis of Count Data: Second Edition
A. Colin Cameron and Pravin K. Trivedi
Excerpt
More information

Significant early developments in count models took place in actuarial science, biostatistics, and demography. In recent years these models have also been used extensively in economics, political science, and sociology. The special features of data in their respective fields of application have fueled developments that have enlarged the scope of these models. An important milestone in the development of count data models for regression was the emergence of the "generalized linear models," of which the Poisson regression is a special case, first described by Nelder and Wedderburn (1972) and detailed in McCullagh and Nelder (1983, 1989). Building on these contributions, the papers by Gourieroux, Monfort, and Trognon (1984a, b) and the work on longitudinal or panel count data models by Hausman, Hall, and Griliches (1984) have also been very influential in stimulating applied work in the econometric literature.

Regression analysis of counts is motivated by the observation that in many, if not most, real-life contexts, the iid assumption is too strong. For example, the mean rate of occurrence of an event may vary from case to case and may depend on some observable variables. The investigator's main interest therefore may lie in the role of covariates (regressors) that are thought to affect the parameters of the conditional distribution of events, given the covariates. This is usually accomplished by a regression model for event count. At the simplest level we may think of this task in the conventional regression framework in which the dependent variable, y, is restricted to be a nonnegative random variable whose conditional mean depends on some vector of regressors, $\mathbf{x}$.

At a different level of abstraction, an event may be thought of as the realization of a point process governed by some specified *rate of occurrence* of the event. The number of events may be characterized as the total number of such realizations over some unit of time. The dual of the event count is the *interarrival time*, defined as the length of the period between events. Count data regression is useful in studying the occurrence rate per unit of time conditional on some covariates. One could instead study the distribution of interarrival times conditional on covariates. This leads to regression models of *waiting times* or *durations*. The type of data available – cross-sectional, time series, or longitudinal – will affect the choice of the statistical framework.

An obvious first question is whether "special methods" are required to handle count data or whether the standard Gaussian linear regression may suffice. More common regression estimators and models, such as the ordinary least squares in the linear regression model, ignore the restricted support for the dependent variable. This leads to significant deficiencies unless the mean of the counts is high, in which case normal approximation and related regression methods may be satisfactory.

The Poisson (log-linear) regression not only is motivated by the usual considerations for regression analysis but also seeks to preserve and exploit as much as possible the nonnegative and integer-valued aspect of the outcome. At one level one might simply regard it as a special type of *nonlinear* regression that respects the discreteness of the count variable. In some analyses this specific distributional assumption may be given up, while preserving nonnegativity.

© in this web service Cambridge University Press www.cambridge.org

CAMBRIDGE

Cambridge University Press
978-1-107-01416-9 - Regression Analysis of Count Data: Second Edition
A. Colin Cameron and Pravin K. Trivedi
Excerpt
More information

In econometrics the interest in count data models is a reflection of the general interest in modeling discrete aspects of individual economic behavior. For example, Pudney (1989) characterizes a large body of microeconometrics as "econometrics of corners, kinks and holes." Count data models are specific types of discrete data regressions. Discrete and limited dependent variable models have attracted a great deal of attention in econometrics and have found a rich set of applications in microeconometrics (Maddala, 1983), especially as econometricians have attempted to develop models for the many alternative types of sample data and sampling frames. Although the Poisson regression provides a starting point for many analyses, attempts to accommodate numerous real-life conditions governing observation and data collection lead to additional elaborations and complications.

This introductory chapter presents the Poisson distribution and some of its characterizations in section 1.1, the Poisson regression model in section 1.2, and some leading examples of count data in section 1.3. An outline of the book is provided in section 1.4.

The scope of count data models is very wide. This monograph addresses issues that arise in the regression models for counts, with a particular focus on features of economic data. In many cases, however, the material covered can be easily adapted for use in other social sciences and in natural sciences, which do not always share the peculiarities of economic data.

1.1 POISSON DISTRIBUTION AND ITS CHARACTERIZATIONS

The benchmark parametric model for count data is the Poisson distribution. It is useful at the outset to review some fundamental properties and characterization results of the Poisson distribution (for derivations, see Taylor and Karlin, 1998).

If the discrete random variable Y is Poisson distributed with *intensity* or rate parameter μ, $\mu > 0$, and t is the *exposure,* defined as the length of time during which the events are recorded, then Y has density

$$\Pr[Y = y] = \frac{e^{-\mu t}(\mu t)^y}{y!}, \qquad y = 0, 1, 2, \ldots, \tag{1.1}$$

where $\mathsf{E}[Y] = \mathsf{V}[Y] = \mu t$.

If we set the length of the exposure period t equal to unity, then

$$\Pr[Y = y] = \frac{e^{-\mu}\mu^y}{y!}, \qquad y = 0, 1, 2, \ldots. \tag{1.2}$$

The probabilities satisfy the recurrence relation

$$\Pr[Y = y + 1](y + 1) = \Pr[Y = y]\mu. \tag{1.3}$$

The Poisson distribution has a single parameter μ and we refer to it as $\mathsf{P}[\mu]$. Its k^{th} raw moment, $\mathsf{E}[Y^k]$, may be derived by differentiating the moment

© in this web service Cambridge University Press www.cambridge.org

CAMBRIDGE

Cambridge University Press
978-1-107-01416-9 - Regression Analysis of Count Data: Second Edition
A. Colin Cameron and Pravin K. Trivedi
Excerpt
More information

generating function (mgf) k times

$$\mathsf{M}(t) \equiv \mathsf{E}[e^{tY}] = \exp\{\mu(e^t - 1)\},$$

with respect to t and evaluating at $t = 0$. This yields the following four raw moments:

$$\begin{aligned}
\mu_1' &= \mu \\
\mu_2' &= \mu + \mu^2 \\
\mu_3' &= \mu + 3\mu^2 + \mu^3 \\
\mu_4' &= \mu + 7\mu^2 + 6\mu^3 + \mu^4.
\end{aligned}$$

Following convention, raw moments are denoted by primes, and central moments without primes. The central moments around μ can be derived from the raw moments in the standard way. Note that the first two central moments, denoted μ_1 and μ_2, respectively, are equal to μ. The central moments satisfy the recurrence relation

$$\mu_{r+1} = r\mu\mu_{r-1} + \mu\frac{\partial\mu_r}{\partial\mu}, \qquad r = 1, 2, \ldots, \tag{1.4}$$

where $\mu_0 = 0$.

Equality of the mean and variance is referred to as the *equidispersion* property of the Poisson. This property is frequently violated in real-life data. *Overdispersion* (*underdispersion*) means that the variance exceeds (is less than) the mean.

A key property of the Poisson distribution is additivity. This is stated by the following *Countable Additivity Theorem* (for a mathematically precise statement, see Kingman, 1993).

Theorem. *If* $Y_i \sim \mathsf{P}[\mu_i]$, $i = 1, 2, \ldots$ *are independent random variables, and if* $\sum \mu_i < \infty$, *then* $S_Y = \sum Y_i \sim \mathsf{P}[\sum \mu_i]$.

The binomial and the multinomial can be derived from the Poisson by appropriate conditioning. Under the already stated conditions,

$$\begin{aligned}
&\Pr[Y_1 = y_1, Y_2 = y_2, \ldots, Y_n = y_n \mid S_Y = s] \\
&= \left[\prod_{j=1}^n \frac{e^{-\mu_j}\mu_j^{y_j}}{y_j!}\right] \Big/ \left[\frac{\left(\sum \mu_i\right)^s e^{-\sum \mu_i}}{s!}\right] \\
&= \frac{s!}{y_1! y_2! \ldots y_n!}\left(\frac{\mu_1}{\sum \mu_i}\right)^{y_1}\left(\frac{\mu_2}{\sum \mu_i}\right)^{y_2} \cdots \left(\frac{\mu_n}{\sum \mu_i}\right)^{y_n} \\
&= \frac{s!}{y_1! y_2! \ldots y_n!}\pi_1^{y_1}\pi_2^{y_2} \ldots \pi_n^{y_n},
\end{aligned} \tag{1.5}$$

where $s = \sum Y_i$ and $\pi_j = \mu_j / \sum \mu_i$. This is the multinomial distribution $\mathsf{m}[s\,; \pi_1, \ldots, \pi_n]$. The binomial is the case $n = 2$.

© in this web service Cambridge University Press www.cambridge.org

CAMBRIDGE

Cambridge University Press
978-1-107-01416-9 - Regression Analysis of Count Data: Second Edition
A. Colin Cameron and Pravin K. Trivedi
Excerpt
More information

There are many characterizations of the Poisson distribution. Here we consider four. The first, the law of rare events, is a common motivation for the Poisson. The second, the Poisson counting process, is very commonly encountered in an introduction to stochastic processes. The third is simply the dual of the second, with waiting times between events replacing the count. The fourth characterization, the Poisson-stopped binomial, treats the number of events as repetitions of a binomial outcome, with the number of repetitions taken as Poisson distributed.

1.1.1 Poisson as the Law of Rare Events

The law of rare events states that the total number of events will follow, approximately, the Poisson distribution if an event may occur in any of a large number of trials, but the probability of occurrence in any given trial is small.

More formally, let $Y_{n,\pi}$ denote the total number of successes in a large number n of independent Bernoulli trials, with the success probability π of each trial being small. Then the distribution is binomial with

$$\Pr\left[Y_{n,\pi} = k\right] = \binom{n}{k}\pi^k(1-\pi)^{n-k}, \quad k = 0, 1, \ldots, n.$$

In the limiting case where $n \to \infty$, $\pi \to 0$, and $n\pi = \mu > 0$ – that is, the average μ is held constant while $n \to \infty$ – we have

$$\lim_{n\to\infty}\left[\binom{n}{k}\left(\frac{\mu}{n}\right)^k\left(1-\frac{\mu}{n}\right)^{n-k}\right] = \frac{\mu^k e^{-\mu}}{k!},$$

the Poisson probability distribution with parameter μ, denoted as $\mathsf{P}[\mu]$.

1.1.2 Poisson Counting Process

The Poisson distribution has been described as characterizing "complete randomness" (Kingman, 1993). To elaborate this feature, the connection between the Poisson distribution and the *Poisson process* needs to be made explicit. Such an exposition begins with the definition of a *counting process*.

A stochastic process $\{N(t), t \geq 0\}$ is defined to be a counting process if $N(t)$ denotes an event count up to time t. $N(t)$ is nonnegative and integer valued and must satisfy the property that $N(s) \leq N(t)$ if $s < t$, and $N(t) - N(s)$ is the number of events in the interval $(s, t]$. If the event counts in disjoint time intervals are independent, the counting process is said to have independent increments. It is said to be stationary if the distribution of the number of events depends only on the length of the interval.

The Poisson process can be represented in one dimension as a set of points on the time axis representing a random series of events occurring at points of time. The Poisson process is based on notions of independence and the Poisson distribution in the following sense.

© in this web service Cambridge University Press www.cambridge.org

Cambridge University Press
978-1-107-01416-9 - Regression Analysis of Count Data: Second Edition
A. Colin Cameron and Pravin K. Trivedi
Excerpt
More information

Define μ to be the constant rate of occurrence of the event of interest, and $N(s, s+h)$ to be the number of occurrences of the event in the time interval $(s, s+h]$. A (pure) Poisson process of rate μ occurs if events occur independently with constant probability equal to μ times the length of the interval. The numbers of events in disjoint time intervals are independent, and the distribution of events in each interval of unit length is $\mathsf{P}[\mu]$. Formally, as the length of the interval $h \to 0$,

$$\begin{aligned} \Pr[N(s, s+h) = 0] &= 1 - \mu h + o(h) \\ \Pr[N(s, s+h) = 1] &= \mu h + o(h), \end{aligned} \tag{1.6}$$

where $o(h)$ denotes a remainder term with the property $o(h)/h \to 0$ as $h \to 0$. $N(h, s+h)$ is statistically independent of the number and position of events in $(s, s+h]$. Note that in the limit the probability of two or more events occurring is zero, whereas 0 and 1 events occur with probabilities of, respectively, $(1 - \mu h)$ and μh. For this process it can be shown (Taylor and Karlin, 1998) that the number of events occurring in the interval $(s, s+h]$, for nonlimit h, is Poisson distributed with mean μh and probability

$$\Pr[N(s, s+h) = r] = \frac{e^{-\mu h}(\mu h)^r}{r!} \qquad r = 0, 1, 2, \ldots. \tag{1.7}$$

Normalizing the length of the exposure time interval to be unity, $h = 1$, leads to the Poisson density given previously. In summary, the counting process $N(t)$ with stationary and independent increments and $N(0) = 0$, which satisfies (1.6), generates events that follow the Poisson distribution.

1.1.3 Waiting-Time Distributions

We now consider a characterization of the Poisson that is the flip side of that given in the preceding paragraph. Let W_1 denote the time of the first event, and $W_r, r \geq 1$ the time between the $(r-1)^{th}$ and r^{th} event. The nonnegative random sequence $\{W_r, r \geq 1\}$ is called the sequence of *interarrival times*, *waiting times*, *durations*, or *sojourn times*. In addition to, or instead of, analyzing the number of events occurring in the interval of length h, one can analyze the duration of time between successive occurrences of the event, or the time of occurrence of the r^{th} event, W_r. This requires the distribution of W_r, which can be determined by exploiting the duality between event counts and waiting times. This is easily done for the Poisson process.

The outcome $\{W_1 > t\}$ occurs only if no events occur in the interval $[0, t]$. That is,

$$\Pr[W_1 > t] = \Pr[N(t) = 0] = e^{-\mu t}, \tag{1.8}$$

which implies that W_1 has exponential distribution with mean $1/\mu$. The waiting time to the first event, W_1, is exponentially distributed with density

© in this web service Cambridge University Press www.cambridge.org

CAMBRIDGE

Cambridge University Press
978-1-107-01416-9 - Regression Analysis of Count Data: Second Edition
A. Colin Cameron and Pravin K. Trivedi
Excerpt
More information

$f_{W_1}(t) = \mu e^{-\mu t}$, $t \geqslant 0$. Also,

$$\begin{aligned}\Pr[W_2 > t | W_1 = s] &= \Pr[N(s, s+t) = 0 | W_1 = s] \\ &= \Pr[N(s, s+t) = 0] \\ &= e^{-\mu t},\end{aligned}$$

using the properties of independent stationary increments. This argument can be repeated for W_r to yield the result that W_r, $r = 1, 2, \ldots$, are iid exponential random variables with mean $1/\mu$. This result reflects the property that the Poisson process has no memory.

In principle, the duality between the number of occurrences and time between occurrences suggests that count and duration data should be covered in the same framework. Consider the arrival time of the r^{th} event, denoted S_r,

$$S_r = \sum_{i=1}^{r} W_i, \quad r \geq 1. \tag{1.9}$$

It can be shown using results on sums of random variables of exponential random variables that S_r has one-parameter gamma distribution with density

$$f_{S_r}(t) = \frac{\mu^r t^{r-1}}{(r-1)!} e^{-\mu t}, \quad t \geq 0. \tag{1.10}$$

This is a special case of a more general result derived in Chapter 4.10. The density (1.10) can also be derived by observing that

$$N(t) \geq r \Leftrightarrow S_r \leq t. \tag{1.11}$$

Hence

$$\begin{aligned}\Pr[N(t) \geq r] &= \Pr[S_r \leq t] \\ &= \sum_{j=r}^{\infty} e^{-\mu t} \frac{(\mu t)^j}{j!}.\end{aligned} \tag{1.12}$$

To obtain the density (1.10) of S_r, the cumulative distribution function (cdf) given in (1.12) is differentiated with respect to t. Thus, the Poisson process may be characterized in terms of the implied properties of the waiting times.

Suppose one's main interest is in the role of the covariates that determine the Poisson process rate parameter μ. For example, let $\mu = \exp(\mathbf{x}'\boldsymbol{\beta})$. Then from (1.8) the mean waiting time is given by $1/\mu = \exp(-\mathbf{x}'\boldsymbol{\beta})$, confirming the intuition that the covariates affect the mean number of events and the waiting times in opposite directions. This illustrates that, from the viewpoint of studying the role of covariates, analyzing the frequency of events is the dual complement of analyzing the waiting times between events.

The Poisson process is often too restrictive in practice. Mathematically tractable and computationally feasible common links between more general count and duration models are hard to find (see Chapter 4).

© in this web service Cambridge University Press www.cambridge.org

CAMBRIDGE
Cambridge University Press
978-1-107-01416-9 - Regression Analysis of Count Data: Second Edition
A. Colin Cameron and Pravin K. Trivedi
Excerpt
More information

In the waiting time literature, emphasis is on estimating the *hazard rate*, the conditional instantaneous probability of the event occurring given that it has not yet occurred, controlling for censoring due to not always observing occurrence of the event. Andersen, Borgan, Gill and Keiding (1993), Kalbfleisch and Prentice (2002), and Fleming and Harrington (2005) present, in great detail, models for censored duration data based on the application of martingale theory to counting processes.

We focus on counts. Even when duration is the more natural entity for analysis, it may not be observed. If only event counts are available, count regressions still provide an opportunity for studying the role of covariates in explaining the mean rate of event occurrence. However, count analysis leads in general to a loss of efficiency (Dean and Balshaw, 1997).

1.1.4 Binomial Stopped by the Poisson

Yet another characterization of the Poisson involves mixtures of the Poisson and the binomial. Let n be the actual (or true) count process taking nonnegative integer values with $\mathsf{E}[n] = \mu$, and $\mathsf{V}[n] = \sigma^2$. Let $B_1, B_2, \ldots, B_n$ be a sequence of n independent and identically distributed Bernoulli trials, where each B_i takes only two values, 1 or 0, with probabilities π and $1-\pi$, respectively. Define the count variable $Y = \sum_{i=1}^{n} B_i$. For n given, Y follows a binomial distribution with parameters n and π. Hence,

$$\begin{aligned} \mathsf{E}[Y] &= \mathsf{E}[\mathsf{E}\,[Y|n]] = \mathsf{E}[n\pi] = \pi\mathsf{E}[n] = \mu\pi \\ \mathsf{V}[Y] &= \mathsf{V}[\mathsf{E}\,[Y|n]] + \mathsf{E}[\mathsf{V}\,[Y|n]] = (\sigma^2 - \mu)\pi^2 + \mu\pi. \end{aligned} \tag{1.13}$$

The actual distribution of Y depends on the distribution of n. For Poisson-distributed n it can be found using the following lemma.

Lemma. *If π is the probability that $B_i = 1$, $i = 1, \ldots, n$, and $1-\pi$ the probability that $B_i = 0$, and $n \sim \mathsf{P}[\mu]$, then $Y \sim \mathsf{P}[\mu\pi]$.*

To derive this result, begin with the probability generating function (pgf), defined as $g(s) = \mathsf{E}[s^B]$, of the Bernoulli random variable B,

$$g(s) = (1-\pi) + \pi s, \tag{1.14}$$

for any real s. Let $f(s)$ denote the pgf of the Poisson variable n, $\mathsf{E}[s^n]$; that is,

$$f(s) = \exp(-\mu + \mu s). \tag{1.15}$$

Then the pgf of Y is obtained as

$$\begin{aligned} f(g(s)) &= \exp[-\mu + \mu g(s)] \\ &= \exp[-\mu\pi + \mu\pi s], \end{aligned} \tag{1.16}$$

which is the pgf of Poisson-distributed n with parameter $\mu\pi$. This characterization of the Poisson has been called the Poisson-*stopped* binomial. This

© in this web service Cambridge University Press www.cambridge.org

CAMBRIDGE

Cambridge University Press
978-1-107-01416-9 - Regression Analysis of Count Data: Second Edition
A. Colin Cameron and Pravin K. Trivedi
Excerpt
More information

characterization is useful if the count is generated by a random number of repetitions of a binary outcome.

1.2 POISSON REGRESSION

The approach taken to the analysis of count data, especially the choice of the regression framework, sometimes depends on how the counts are assumed to arise. There are two common formulations. In the first, counts arise from a direct observation of a point process. In the second, counts arise from discretization (ordinalization) of continuous latent data. Other less used formulations appeal, for example, to the law of rare events or the binomial stopped by Poisson.

1.2.1 Counts Derived from a Point Process

Directly observed counts arise in many situations. Examples are the number of telephone calls arriving at a central telephone exchange, the number of monthly absences at the place of work, the number of airline accidents, the number of hospital admissions, and so forth. The data may also consist of interarrival times for events. In the simplest case, the underlying process is assumed to be stationary and homogeneous, with iid arrival times for events and other properties stated in the previous section.

1.2.2 Counts Derived from Continuous Data

Count-type variables sometimes arise from categorization of a latent continuous variable as the following example indicates. Credit rating of agencies may be stated as "AAA," "AAB," "AA," "A," "BBB," "B," and so forth, where "AAA" indicates the greatest credit worthiness. Suppose we code these as $y = 0, 1, \ldots, m$. These are pseudocounts that can be analyzed using a count regression. But one may also regard this categorization as an ordinal ranking that can be modeled using a suitable latent variable model such as ordered probit. Section 3.6 provides a more detailed exposition.

1.2.3 Regression Specification

The standard model for count data is the *Poisson regression model,* which is a nonlinear regression model. This regression model is derived from the Poisson distribution by allowing the *intensity parameter* μ to depend on covariates (regressors). If the dependence is parametrically exact and involves exogenous covariates but no other source of stochastic variation, we obtain the standard Poisson regression. If the function relating μ and the covariates is stochastic, possibly because it involves unobserved random variables, then one obtains a *mixed Poisson regression*, the precise form of which depends on the assumptions about the random term. Chapter 4 deals with several examples of this type.

© in this web service Cambridge University Press www.cambridge.org

CAMBRIDGE

Cambridge University Press
978-1-107-01416-9 - Regression Analysis of Count Data: Second Edition
A. Colin Cameron and Pravin K. Trivedi
Excerpt
More information

A standard application of Poisson regression is to cross-section data. Typical cross-section data for applied work consist of n independent observations, the i^{th} of which is $(y_i, \mathbf{x}_i)$. The scalar dependent variable y_i is the number of occurrences of the event of interest, and $\mathbf{x}_i$ is the vector of linearly independent regressors that are thought to determine y_i. A regression model based on this distribution follows by conditioning the distribution of y_i on a k-dimensional vector of covariates, $\mathbf{x}_i' = [x_{1i}, \ldots, x_{ki}]$, and parameters $\boldsymbol{\beta}$, through a continuous function $\mu(\mathbf{x}_i, \boldsymbol{\beta})$, such that $\mathsf{E}[y_i|\mathbf{x}_i] = \mu(\mathbf{x}_i, \boldsymbol{\beta})$.

That is, y_i given $\mathbf{x}_i$ is Poisson distributed with density

$$f(y_i|\mathbf{x}_i) = \frac{e^{-\mu_i}\mu_i^{y_i}}{y_i!}, \qquad y_i = 0, 1, 2, \ldots \tag{1.17}$$

In the *log-linear* version of the model the mean parameter is parameterized as

$$\mu_i = \exp(\mathbf{x}_i'\boldsymbol{\beta}), \tag{1.18}$$

to ensure $\mu > 0$. Equations (1.17) and (1.18) jointly define the Poisson (log-linear) regression model. If one does not wish to impose any distributional assumptions, (1.18) by itself may be used for (nonlinear) regression analysis.

For notational economy we write $f(y_i|\mathbf{x}_i)$ in place of the more formal $f(Y_i = y_i|\mathbf{x}_i)$, which distinguishes between the random variable Y and its realization y. Throughout this book we refer to $f(\cdot)$ as a density even though more formally it is a probability mass function.

By the property of the Poisson, $\mathsf{V}[y_i|\mathbf{x}_i] = \mathsf{E}[y_i|\mathbf{x}_i]$, implying that the conditional variance is not a constant, and hence the regression is intrinsically heteroskedastic. In the log-linear version of the model the mean parameter is parameterized as (1.18), which implies that the conditional mean has a multiplicative form given by

$$\begin{aligned}\mathsf{E}[y_i|\mathbf{x}_i] &= \exp(\mathbf{x}_i'\boldsymbol{\beta}) \\ &= \exp(x_{1i}\beta_1)\exp(x_{2i}\beta_2)\cdots\exp(x_{ki}\beta_k),\end{aligned}$$

with interest often lying in changes in this conditional mean due to changes in the regressors. The additive specification, $\mathsf{E}[y_i|\mathbf{x}_i] = \mathbf{x}_i'\boldsymbol{\beta} = \sum_{j=1}^{k} x_{ji}\beta_j$, is likely to be unsatisfactory because certain combinations of $\boldsymbol{\beta}$ and $\mathbf{x}_i$ will violate the nonnegativity restriction on μ_i.

The Poisson model is closely related to the models for analyzing counted data in the form of proportions or ratios of counts sometimes obtained by grouping data. In some situations – for example when the population "at risk" is changing over time in a known way – it is helpful to reparameterize the model as follows. Let y be the observed number of events (e.g., accidents), N the known total exposure to risk (i.e., number at risk), and $\mathbf{x}$ the known set of k explanatory variables. The mean number of events μ may be expressed as the product of N and π, the probability of the occurrence of event, sometimes also called the hazard rate. That is, $\mu(\mathbf{x}) = N(\mathbf{x})\pi(\mathbf{x}, \boldsymbol{\beta})$. In this case the probability π is parameterized in terms of covariates. For example, $\pi = \exp(\mathbf{x}'\boldsymbol{\beta})$. This leads

© in this web service Cambridge University Press www.cambridge.org