Recent Developments in Models of Event Counts: A Survey

Shiferaw Gurmu

Pravin K. Trivedi

September 1994

Discussion Paper 261
ABSTRACT

The main focus of this article is on specification, estimation and inference in models related to the Poisson regression model for non-negative integer valued counts, both from a formal point of view, and from that of practical data analysis. It provides an up to date survey with emphasis on developments since 1986. Further, it develops the subject matter in a connected fashion that would enable a new entrant to this field to apply the available models and methods. Beginning with Poisson process models, a variety of departures from the Poisson model will be considered, and these will be shown to motivate less restrictive extensions of the basic Poisson regression model. Some of these extensions are related to the way in which data are generated and collected. A number of parametric and semi-parametric estimators will be examined. The paper presents an empirical example to illustrate a number of issues of econometric modelling.

Shiferaw Gurmu
Department of Economics
Rouss Hall
University of Virginia
Charlottesville, VA 22903

Pravin K. Trivedi
Department of Economics
Ballantine Hall
Indiana University
Bloomington, IN 47405
RECENT DEVELOPMENTS IN MODELS OF EVENT COUNTS: A SURVEY*

Shiferaw Gurmu
Department of Economics
Rouss Hall, University of Virginia
Charlottesville, VA 22903
Phone: (804) 924-3660
e-mail: sg2m@uva.pcmail.virginia.edu

Pravin K. Trivedi
Department of Economics
Ballantine Hall, Indiana university
Bloomington, IN 47405
Phone: (812) 855-3567
e-mail: trivedi@indiana.edu

This Version: September 1994

Contents

1 INTRODUCTION ................................................. 2
2 THE POISSON PROCESS ........................................ 4
   2.1 The stationary Poisson process ........................ 4
   2.2 The Poisson regression model ....................... 5
   2.3 The interevent waiting time distribution .......... 6
   2.4 Links between event counts and waiting time models 7
3 VARIATIONS OF THE BASIC COUNT MODEL AND THEIR PROPERTIES 8
   3.1 Some common failures of the standard Poisson regression 8
   3.2 Some useful definitions and concepts ............... 9
   3.3 Generalizing the Poisson regression ................ 10
      3.3.1 The overdispersion phenomenon .................. 10
      3.3.2 The mixing distribution ........................ 11
   3.4 Some modified and generalized Poisson models .... 12
      3.4.1 Negative binomial, Double Poisson, A.L.D.P. and gamma models 13
      3.4.2 Hurdle models .................................. 15

*We thank several readers of an earlier version, including three anonymous referees and an Associate Editor for exceptionally detailed comments on a previous version of the paper. Helpful comments were also provided by Colin Cameron, Alain Monfort, John Mullahy, and Rainer Winkelmann. Teofilo Ozuna and Irma Gomaz graciously permitted us to use their data. The authors retain responsibility for the current version. Comments on the present draft may be sent to either author.
1. INTRODUCTION

The main focus of this article is on specification, estimation and inference in models related to the Poisson regression model for non-negative integer valued counts, both from a formal viewpoint, and from that of practical data analysis. Recently these methods have been used in numerous econometric and statistical applications, with examples in agricultural, health and labor economics, insurance and finance, industrial organization, marketing, and so forth. Major impetus for this research in econometrics stemmed from the publication of the important papers by Hausman, Hall and Griliches (1984), henceforth HHG (1984), who extended the application of these methods to panel data, and the important theoretical articles by Gourieroux, Monfort and Trognon (1984a, b), henceforth GMT (1984). Econometric handling of such models has been facilitated by the inclusion of such regression procedures in easily available and widely
used mainframe and PC based packages. Some of these developments were surveyed by Cameron and Trivedi (1986) and Winkelmann (1993a).

This article has two objectives. The first is to provide an up to date survey with emphasis on developments since 1986; the second is to develop the subject matter in a connected fashion that would enable a new entrant to this field to apply the available models and methods. For this purpose it is useful to understand the origins of these models in the point process literature, the alternative approaches for analyzing data on the frequency of events or on the inter-event time duration distributions, and the parallels between the methods for handling discrete and continuous random variables, and the major features and limitations of the theoretical and empirical treatment of count data regressions. Many developments for the Poisson regression type models mirror those for parametric limited dependent variable models based on, or derived from, the parent normal regression model for continuous data. These parallels can be usefully exploited both in exposition and in motivating further research.

The foundations of the Poisson regression model lie in theory of stochastic point processes. One strand of this theory deals with life tables and renewal theory, which may be described as the study of the sequence of intervals between successive occurrences of an event of interest, such as the failure of an engine component, or exit from a spell of employment. A second strand of this theory deals with counting problems which involves the number of events of a given type in some time interval, e.g. the number of hospital episodes, or spells of unemployment. The benchmark stochastic point process is the Poisson process which provides the basic link between the study of the number of events and the inter-event distribution of waiting times. These fundamental connections are explained in Section 2. It also addresses the following key issue: If one's main interest lies in the role of covariates (or exogenous variables), is the regression analysis of random counts more, less or equally informative relative to the regression analysis of waiting time distributions?

Section 3 proceeds to the Poisson regression model, with conditionally independent observations and defined over all non-negative integers. Despite its central importance as the benchmark count model, it is often found to be too restrictive in empirical application. This can occur either from the failure of the moment restrictions implicit in the Poisson model, or because the count data may be generated or collected in a manner that requires a modification of the basic model to accommodate additional complications. A variety of departures from the Poisson model will be considered, and these will be shown to motivate less restrictive extensions of the Poisson regression model. There are essentially two broad approaches to seeking generalizations of the Poisson model. The term (fully) parametric approach will be used to describe more general alternatives which specify the full probability distribution of the events having a finite-dimensional parameters of interest. The term semi-parametric will be used to cover alternatives in which estimation and inference do not require knowledge of the full
distribution of the events. Estimation of semi-parametric models may be based only on certain low-order conditional moments of the probability distribution, usually the first two. Specification and estimation of various count models will also be discussed in Section 3. Further aspects of estimation, with emphasis on extensions of quasi-likelihood method, will be presented in Section 4.

Model evaluation, which is an important facet of practical data analysis, usually requires diagnostic checks as well as goodness of fit measures of the adequacy of the maintained models. Section 5 will present and discuss these and related issues. In section 6 we shall present an empirical example to illustrate a number of issues of econometric modelling.

2. THE POISSON PROCESS

The Poisson process is of fundamental importance in the theory of point processes because it deals with a completely random series of events. Hence a discussion of the properties of this process and of a number of closely related concepts is a good starting point. A more detailed and mathematically accessible account is available in Cox and Lewis (1966).

2.1. The stationary Poisson process

Let $\theta$ denote the mean rate of occurrence of an event over a long time interval; let $n(t, t + h)$ denote the number of events in the time interval $(t, t + h)$, when $h > 0$. The homogeneous or stationary Poisson process is defined as follows: As $h \to 0$,

$$\Pr(n(t, t + h) = 0) = 1 - \theta h + o(h)$$

$$\Pr(n(t, t + h) = 1) = \theta h + o(h)$$

and $n(t, t + h)$ is independent of the position of $t$ on time axis for all $t, h > 0$, where $o(h)$ denotes a quantity which, when divided by $h$, tends to 0 as $h \to 0$. These conditions imply that $\Pr(n(t, t + h) \geq 2) = o(h)$. The key features of the Poisson process are (a) the process is stationary so event probabilities do not vary over time, i.e. trends are excluded; (b) the process is memoryless, so the probability of an event in $(t, t + h)$ is independent of the past history of events; (c) the probability of simultaneous occurrence of events is negligible.
2.2. The Poisson regression model

The number of events occurring in an interval of length \( h \), has Poisson distribution with mean \( \theta(h) \), i.e.,

\[
Pr \{n(t,t+h) = r\} = \frac{\exp[-\theta(h)] \cdot \{\theta(h)\}^r}{r!} , \quad r \geq 0. \tag{2.1}
\]

The number events in \((0,t)\), denoted \( n(t) \), is a stochastic process with independent increments in non-overlapping time periods. To proceed to the Poisson regression model, set the time interval to unity, let \( y_i, i = 1, ..., n \) be the observations of a discrete response variable (the number of events) and let \( x_i \) be a \( q \times 1 \) covariate vector, the first element of which is unity for all \( i \). It is assumed that

\[
E(y_i \mid x_i) = \theta_i = \theta(x_i, \beta), \tag{2.2}
\]

where \( \beta \) is a \( k \)-dimensional vector of unknown parameters. The most common specification for the mean parameter \( \theta \) is exponential, which ensures non-negativity of \( \theta \); i.e.,

\[
\theta_i = \exp(x_i'\beta) \tag{2.3}
\]

for which \( q = k \).

The density function of \( y_i \) is

\[
h(y_i \mid x_i) = \frac{\exp(-\theta_i) \cdot \theta_i^{y_i}}{y_i!} , \quad y_i = 0, 1, ..., \tag{2.4}
\]

and it is easy to show that the density implies the moment restriction,

\[
E(y_i \mid x_i) = \text{var}(y_i \mid x_i) = \theta_i. \tag{2.5}
\]

The maximum likelihood equations for \( \beta \) are nonlinear, but given the log-concavity of the likelihood, the solution by standard methods is straightforward, and will not be discussed here.

In some situations, for example when the population ‘at risk’ or exposed to the probability of an event is changing over time in a known way, it is useful to reparameterize the model as follows. Specifically, \( y_i \) may be the observed number of accidents in period \( i \), and \( N_i \) could be the number at risk of accident. The mean number of events \( \theta_i \) might be expressed as the product of \( N_i \) and \( \pi_i \), the probability of the occurrence of event, sometimes also called the hazard rate. Subsequently \( \pi_i \) may be parameterized as a function of vector of covariates \( x_i \); \( \pi_i = \pi(x_i, \beta) \), where a suitable mathematical function \( \pi \) is used. This leads to the rate form of Poisson count model with the density

\[
h(y_i \mid x_i) = \frac{\exp(-N_i \pi_i) \cdot (N_i \pi_i)^{y_i}}{y_i!} , \quad y_i = 0, 1, ..., \tag{2.6}
\]
The nonhomogeneous (time-dependent) Poisson process: The previous definition can be generalized to the case where the rate of the occurrence of the event, the intensity function, depends upon the elapsed time since the start of the process, i.e., we replace \( \theta \) by \( \theta(t) \). The stochastic process is now referred to as the non-homogeneous or non-stationary Poisson process. The number of events occurring in the interval \((t, t+h)\), denoted \(n(t, t+h)\), has Poisson distribution with mean \( \theta^*(t, h) = E[n(t, t+h)] \) where
\[
\theta^*(t, h) = \int_t^{t+h} \theta(s)ds = m(t+h) - m(t),
\]
and \( m(t) = \int_0^t \theta(s)ds = E[n(0, t)] \). The probability of observing \( r \) events in the interval \((t, t+h)\) is given by
\[
\Pr\{n(t, t+h) = r\} = \frac{\exp[-\theta^*(t, h)] \{\theta^*(t, h)\}^r}{r!}, \quad r \geq 0.
\]

2.3. The interevent waiting time distribution

The Poisson process implies that the random length interval of time from the origin to the first occurrence of the event, also called the first passage time distribution, is exponential. Omitting the observation subscript \( i \), if \( T \) is the interval from origin to the first event, then the probability of no event in the interval \((0, t)\) is
\[
\Pr(T > t) = \Pr(n(t) = 0) = \exp(-\theta t)
\]
which corresponds to the distribution function \( F_T(t) = 1 - \exp(-\theta t) \) of \( T \) and the p.d.f. \( f_T(t) = \theta \exp(-\theta t), t \geq 0 \). It is readily shown that
\[
E(T) = 1/\theta; \quad \text{var}(T) = 1/\theta^2.
\]

Thus the Poisson process implies an exponential distribution of inter-event durations from any arbitrary origin. These durations also display no dependence. The parameter \( \theta \) can be parameterized in terms of exogenous variables as before, and may be estimated from data on time intervals between consecutive occurrences of events for different cases. This reflects the fundamental duality relation between the Poisson regression for event counts and the exponential regression for waiting times.

The assumption of stationarity may be too strong; the waiting times between successive occurrences of events may exhibit positive or negative dependence, implying that the number of events in a given interval may depend on the past history. The distribution of times between successive occurrences of events is no longer either independently or identically distributed, but will be conditional on \( t \). A commonly used functional form for \( \theta(t) \) are the Weibull process \( \theta(t) = \mu t^{\gamma-1} \). In this case the expected number of events in a given interval \((t, t+h)\) is given by \( E[n(t, t+h)] = m(t+h) - m(t) = \mu(t+h)^\gamma - \mu t^\gamma \).
It is readily verified that the expected number of events will be an increasing or decreasing function of $t$, as $\gamma > 1$ or $\gamma < 1$, which correspond to, respectively, decreasing and increasing duration dependence. Thus the inclusion of a trend in the conditional mean function could be justified.

An empirical example of a nonhomogeneous Poisson model in the form given in equation (2.6) is Schwartz and Torous (1993) who model the mortgage prepayment and default behavior of homeowners. Let $y_i$ denote the number of pre-payments, $N_i$ the number of outstanding mortgages and $\pi_i$, the probability of prepayment. They write $\pi_i = \pi_1(t, \beta_1) \pi_2(x_i, \beta_2)$ where the first component $\pi_1(\cdot)$ depends only on the age, denoted $t$, of the mortgage and hence is a deterministic trend, and the second factor $\pi_2(\cdot)$ depends only on individual specific covariates.

2.4. Links between event counts and waiting time models

Cox and Lewis (1966; Chapter 4) have provided an excellent discussion of this topic. Our discussion will highlight parts of this discussion that have a bearing on the question: are event count regressions more or less informative compared with waiting time models? The discussion is for the stationary point process.

Consider the process of intervals between events from an arbitrary time origin. Let $W$ denote the time to the first occurrence of the event ("forward recurrence time") and $L_i (i = 1, 2, ..)$ the intervals between subsequent events. Let $n(t)$ denote the cumulative number of events in an interval following an arbitrary origin. Then the counting process $n(t)$ and the random variables $W, L_1, L_2, ...$ are connected by the following fundamental relationships:

$$\Pr(n(t) = 0) = \Pr(W > t)$$

$$\Pr(n(t) < n) = \Pr(W + L_1 + L_2 + ... + L_{n-1}) > t); \ n = 2, 3, ..$$

If one begins with an arbitrary event, and $\{Z_i\}$ denotes the sequence of intervals between events with a common marginal distribution $F_T(z)$, then the following relationship exists between the number of events in the interval $(0, t)$, denoted $n^f(t)$, and $\{Z_i\}$:

$$\Pr(n^f(t) < r) = \Pr(Z_1 + Z_2 + ... + Z_r > t).$$

In certain cases, e.g. the renewal process, the joint distribution of the $\{Z_i\}$ and $\{L_i\}$ are identical. These relationships indicate that given the distribution of counts it is theoretically possible to deduce the joint or the marginal distribution of $W$ and the $L_i$. In principle, therefore, there is an analysis of event counts that is equivalent to the analysis of waiting times. A practical difficulty is that the formulae relating the two would in general be quite complex (Cox and Lewis 1966; Chapter 4.3), being solutions of certain convolution integrals that are not always expressible in a closed form (Gourieroux and Visser 1993). The equivalence of the count data and waiting time analysis exists.
only through their complete distributions, and is potentially exploitable only in a fully parametric analysis. Winkelmann (1993b) has developed the count regression model that is an exact counterpart of a gamma waiting time model; see Section (3.4.1) for further detail. If, as will be discussed later in the paper, the data analysis is carried out by specifying only the low order moments rather than the full distribution, then both count and waiting time regressions are informative about the role of the covariates, but the two analyses are not in general equivalent.

3. VARIATIONS OF THE BASIC COUNT MODEL AND THEIR PROPERTIES

3.1. Some common failures of the standard Poisson regression

The Poisson regression model is frequently unsatisfactory or inappropriate in application. The following complications, which call for more general models, are quite common:

1. The failure of the moment restriction (2.5). Frequently the conditional variance of data exceeds the mean, which is usually referred to as extra-Poisson variation or overdispersion relative to the Poisson model. If the conditional variance is less than the mean, we have a phenomenon of underdispersion. This calls for a more general model.

2. The “excess zeros” or “zero inflation” problem. The observed data may show a higher relative frequency of zeros, or some other integer, than is consistent with the Poisson model (Mullahy 1986; Lambert 1992).

3. Truncation, censoring and sample selection. The observed counts may be left truncated (zero truncation is quite common) leading to small counts being excluded, or right censored, by having counts exceeding some value being aggregated; the included observations may be subject to a sample selection rule.

4. The failure of the conditional independence assumption. Event counts, especially if they are a time series, may be dependent.

5. The mean rate of event occurrence, the intensity function, may have a trend or some other deterministic form of time dependence.

In the remainder of this section we shall consider the parametric approach to accommodate these complications. We also consider the semiparametric approach briefly.
3.2. Some useful definitions and concepts

It is well known that if the p.d.f., denoted \( f(y, \theta) \), is specified parametrically, then under regularity conditions the (true) maximum likelihood estimator (MLE) of \( \theta \) is consistent, efficient and asymptotically normal.

If the probability specification is incorrect, the application of the MLE results in pseudo maximum likelihood estimates (PMLE). If the conditional mean is correctly specified, the PMLE will be consistent and asymptotically normal, but will not in general be efficient (White 1982; GMT 1984a); i.e. asymptotically, \( \sqrt{n} (\hat{\theta} - \theta^*) \xrightarrow{d} N(0, A^{-1}BA^{-1}) \) where \( \theta^* \) denotes the true parameter vector and \( A = E_\theta\left[ \frac{1}{\partial \theta} \frac{1}{\partial \theta} \frac{2 \ell(\theta)}{2 \ell(\theta)} \right] \), where \( \ell(\theta) \) denotes the (pseudo) likelihood function and \( E_\theta \) denotes expectations evaluated at \( \theta^* \).

Though the term quasi-MLE (QMLE) is sometimes used interchangeably with PMLE, in this article it is reserved for generalized nonlinear least squares type methods based on (usually first and second order) moment specifications rather than the full specification of the probability structure of the data. Let \( f(y, \theta_1, \theta_2) \) be the p.d.f. Suppose the observations \( y_1, \ldots, y_n \) are independent, and \( E(y \mid x) = \theta_1(x, \beta) \) and \( \text{var}(y \mid x) = v(\theta_1, \theta_2) \). Then a QMLE of \( \beta \) is defined by the solution of equations:

\[
\left[ \frac{\partial \theta_1}{\partial \beta} \right] V^{-1}(y - \theta_1(x, \beta)) = 0
\]

(3.1)

where \( V = \text{diag}(v_1(\theta_1, \theta_2)) \) is the diagonal variance matrix of observations. Another variant of QMLE of \( \beta \) is obtained by maximizing a ‘likelihood’ function based on \( f(y, \theta_1, \theta_2) \) where \( \theta_2 \) is a consistent estimator of \( \theta_2 \) (GMT 1984a). GMT (1984a) introduced quasi-generalized pseudo MLE (QGPMLE) estimators based on a family of densities, viz., the linear exponential family.

To make the preceding discussion concrete, we consider the well known and important parametric family of distributions of which the Poisson is a member, viz., the linear exponential family (LEF) with p.d.f of the form \( f(y, \Theta) = \exp[y \Theta - a(\Theta) + b(y)] \), indexed by \( \Theta \), where \( a(.) \) and \( b(.) \) are known functions. Well known properties of the LEF include \( E(y) = a'(\Theta) \), \( \text{var}(y) = a''(\Theta) \). In the Poisson case \( \Theta = \log(\theta) \) and \( E(y) = \theta = \text{var}(y) \). The variance function uniquely determines the member of the LEF family. Greater generality and flexibility is achieved by passing to an exponential dispersion family (EDF) obtained by introducing an additional dispersion parameter, \( \phi \in \Theta \). Thus \( f(y; \Theta, \phi) = \exp\{(y \Theta - a(\Theta) + b(y, \phi))\phi^{-1}\} \) permits the variance of \( y \) to depend upon both \( \Theta \) and \( \phi \). Now \( \Theta \) corresponds to \( \theta_1 \), and \( \phi \) to \( \theta_2 \) in the preceding paragraph. For this family \( E(y) = a'(\Theta) \) and \( \text{var}(y) = \phi a''(\Theta) \). For the special case of LEF a number of important results are available regarding the properties of PMLE and QMLE and have been reviewed in Cameron and Trivedi (1986). In the
context of Poisson regression type models the parameter $\phi$ has a special role since it has the interpretation of overdispersion parameter which will be considered next.

3.3. Generalizing the Poisson regression

3.3.1. The overdispersion phenomenon

The first obvious way to remove the (2.5) restriction is to pick a less restrictive p.d.f., perhaps a member of the EDF with parameter $\phi$. How might one interpret $\phi$? One very common interpretation is that overdispersion results from neglect of unobserved heterogeneity in the phenomenon being modelled, which justifies treating the Poisson parameter as a random variable. Let a random disturbance term be added to the intercept $\beta_0$, which is equivalent to introducing a multiplicative disturbance in the conditional mean function. Replace $\theta_i$ in (2.3) by

$$\theta_i^* = E(y_i \mid x_i, \nu_i) = \exp(x_i'\beta + \varepsilon_i) = \exp(x_i'\beta) \cdot \nu_i$$

(3.2)

where the unobserved heterogeneity term $\nu_i = \exp(\varepsilon_i)$ could reflect a specification error such as unobserved omitted exogenous variables. However, randomness in the heterogeneity term $\nu_i$ is distinguished from the intrinsic randomness in the endogenous count variate $y_i$. It is usually assumed that $\nu_i$'s are identically and independently distributed (iid), possibly with a known parametric distribution, and that they are independent of the $x$'s.

For example, assume that $\nu_i$ is iid with $E(\nu_i) = 1$ and $\text{var}(\nu_i) = \sigma^2_\nu$. The assumption that $E(\nu_i) = 1$ is made for identification purpose and will only affect the intercept term. Also assume that $E(y_i \mid x_i, \nu_i) = \text{var}(y_i \mid x_i, \nu_i) = \exp(x_i'\beta)$. In this setup, $\nu_i$ leads to overdispersed $y_i$ without affecting $E(y_i \mid x_i)$. The moments of $y_i$, conditional on covariates, can be derived as

$$E(y_i \mid x_i) = \exp(x_i'\beta),$$

(3.3)

$$\text{var}(y_i \mid x_i) = \exp(x_i'\beta) \left[ \exp(x_i'\beta) + \sigma^2_\nu \cdot \exp(x_i'\beta) \right] > E(y_i \mid x_i);$$

(3.4)

see, for example GMT (1984b). The advantage of this setup is that it is possible to obtain consistent parameter estimates for all distributions $g(\nu_i)$ having second order moments without arbitrarily choosing the mixing distribution.

A fully parametric method is based on full specification of the density functions of $(y_i \mid x_i, \nu_i)$ and $\nu_i$. Specifically, let $h(y_i \mid x_i, \nu_i)$ be the probability function obtained by replacing $\theta_i$ in (2.3) by $\theta_i^*$, and let $g(\nu_i)$ denote the probability density function of $\nu_i$.

---

1 These moments also provide the basis of sequential quasi-likelihood estimation (McCullagh 1983; GMT 1984b; Cameron and Trivedi 1986) and moment estimation (Moore 1986) in count models.
The mixed marginal density of \( (y_i \mid x_i) \) is then derived by integrating with respect to \( 
abla_i \):

\[
  h(y_i \mid x_i) = \int h(y_i \mid x_i, 
abla_i) g(\nabla_i) d\nabla_i.
\]

(3.5)

The precise form of this mixed Poisson distribution depends upon the specific choice of \( g(\nabla_i) \). The general property of overdispersion does not depend on \( g(\nabla_i) \) and \( h(.) \) are conjugate families, the resulting compound model will be expressible in a closed form. The most familiar example of this type arises when \( \nabla_i \) is gamma distributed which results in a negative binomial marginal distribution (Johnson and Kotz 1969; GMT 1984a,b; HHG 1984; Cameron and Trivedi 1986; Lawless 1987b). The discrete lognormal compound distribution obtained as a Poisson-lognormal mixture (Shaban 1988) does not have a closed form representation available. Dean et al (1989) consider Poisson-Inverse Gaussian mixture, and Hinde (1982) has considered a Poisson-normal mixture.

In all previous examples the distribution of the unobserved heterogeneity term has infinite points of support. If the continuous mixing distribution \( g(\nabla_i) \) can be approximated by a discrete distribution, denoted by \( \pi_j \) \( (j = 1, \ldots, q) \) with a finite number, \( q \), of support points (Laird 1978; Lindsay 1983; Heckman and Singer 1984), then the marginal distribution is

\[
  h(y_i \mid x_i) = \sum_{j=1}^{q} h(y_i \mid x_i, \nabla_j) \pi_j(\nabla_j),
\]

(3.6)

where \( \nabla_j \) is an estimated support point and \( \pi_j \) is the associated probability. This semiparametric representation of unobserved heterogeneity was examined by Heckman and Singer (1984) in duration modeling and has been recently used in count data models by Brännäs and Rosenqvist (1992). Closely related work is that of Wedel et al (1993), in which the model is generalized further to include random effects in the slope parameters of the conditional mean function, and the model is interpreted in terms of a latent class model. Issues relevant to the choice of mixing distributions are discussed further in the next section.

3.3.2. The mixing distribution

Major considerations in the choice of the mixing distribution are identifiability, flexibility in specification, tractability in computation and robustness to misspecification. Identifiability logically precedes estimation. In the context of mixture models it refers to the identification of the parameters of the conditional and the mixing distributions given a priori restrictions on these distributions.
The probability generating function (p.g.f.) of the mixture model, denoted \( P(z) \), can be expressed as the convolution integral

\[
P(z) = \int_0^\infty \exp(\theta(z - 1))dF(\theta),
\]

where \( \exp(\theta(z - 1)) \) is the p.g.f. of the Poisson distribution and \( dF(\theta) \) is the assumed distribution for \( \theta \). The mixture models are akin to 'reduced form' models and subject to one difficulty, an identification problem, which is that the same p.g.f. could be obtained from a different mixture. For example, it can be shown that the negative binomial mixture obtained as above could also be generated by taking a random sum of independent random variables in which the number of terms in the sum has a Poisson distribution. If each term is discrete and has a logarithmic distribution and if the number of terms has a Poisson distribution, the mixture would be a negative binomial (Daley and Vere-Jones 1988). Identification may be secured by restricting the conditional event distribution to be Poisson. This follows from the uniqueness property of exponential mixtures. A practical consideration is that in applied work, especially that based on small samples, it may be difficult to distinguish between alternative mixing distributions, and the choice may be largely based on the ease of computation.\(^2\)

Flexibility in specification refers to the potential for modeling over- and under-dispersion without prior restriction. While more flexible count distributions are usually derived by mixing, it may some times be appropriate to specify directly flexible functional forms for counts, without the intermediate step of introducing a distribution of unobserved heterogeneity, e.g. in aggregative time series applications. However, in microeconometric applications mixing seems natural.\(^3\)

3.4. Some modified and generalized Poisson models

The statistics literature contains many examples of mixed, compound and generalized count models. The negative binomial, derived as a Poisson-gamma mixture, is one of the oldest and most popular in applied work. Other mixtures which have been used also include Poisson-inverse Gaussian mixture (Dean et al 1988), discrete lognormal (Shaban 1988), generalized Poisson (Consul 1989; Consul and Jain 1973), and Gauss-Poisson (Johnson and Kotz 1969). Several authors generalize the basic Poisson model

\(^2\)Most of the issues are analogous to those which have been discussed extensively in the duration literature; Lancaster (1990, Chapter 7) provides an excellent discussion of the identification conditions for the proportional hazard models and gives an example of a nonidentifiable model. In the examples available in the duration literature finiteness of the mean of the mixing distribution is required for identifiability of the mixture.

\(^3\)In the duration literature, where the shape of the hazard function is of central interest, there has been an extensive discussion of how misspecified unobserved heterogeneity can lead to inconsistent estimates of the hazard function; see Heckman and Singer (1984) and Lancaster (1990, pp. 294-305) for a summary.
by replacing the underlying Poisson model by another which introduces an additional unknown dispersion parameter. Such a model can simultaneously accommodate both over- or under-dispersion (King 1989b; Winkelmann and Zimmermann 1991; Consul and Famoye 1992). Consul and Famoye (1992) do so using a generalized Poisson distribution. To date these models have not been extended to accommodate unobserved heterogeneity.

How can one choose between many available alternative specifications? One possible criterion is the fit of the distribution to the data as measured by the log-likelihood of the model, or the difference in log-likelihood for choosing between nested models. If, however, the models are non-nested, the validity of the criterion and the statistical significance of observed differences is open to question. Another factor may be the goodness of fit in some specific dimension; for example, the ability of the model to fit the tail of the distribution if this is regarded as important. A third consideration may be robustness to misspecification of the probability structure. If all specifications are essentially approximations to the true but unknown data generating process, and hence any estimator is a pseudo maximum likelihood estimator in the sense of White (1982), then an appropriate criterion is the robustness of a given pseudo MLE estimator. Finally, the ease of interpretation in terms of an underlying stochastic point process may be a consideration.

In the following subsection we consider four generalizations of the Poisson to illustrate these alternative approaches.

3.4.1. Negative binomial, Double Poisson, A.L.D.P. and gamma models

The negative binomial (Negbin) distribution can be parameterized as

\[
h(y_i) = \frac{\Gamma(y_i + \psi_i)}{\Gamma(\psi_i) \Gamma(y_i + 1)} \left( \frac{\psi_i}{\theta_i + \psi_i} \right)^{\psi_i} \left( \frac{\theta_i}{\theta_i + \psi_i} \right)^{y_i}, \quad y_i = 0, 1, \ldots, \quad (3.8)
\]

where the parameter \( \theta_i \) is the mean and \( \psi_i^{-1} \) is the precision parameter. In the context of regression, a wide range of mean-variance relationship can be obtained by setting \( \theta_i = \exp(x'_i \beta) \) as in (2.3) and \( \psi_i = (1/\alpha) \theta_i^k \), where \( \alpha > 0 \) is a dispersion parameter and \( k \) is an arbitrary constant. The negative binomial model has \( E(y_i | x_i) = \theta_i \) and \( \text{var}(y_i | x_i) = \theta_i + \alpha \theta_i^{2-k} \). The most common models are Negbin1 obtained by setting \( k = 1 \) and Negbin2 obtained by letting \( k = 0 \). Negbin1 implies a linear-in-\( \theta \) variance function and Negbin2 implies a quadratic variance function. The geometric model is a special case of Negbin2 and can be obtained by setting \( \alpha = 1 \) and \( k = 0 \). In another specification \( k \) is left unrestricted, see King (1989b) and Winkelmann and Zimmermann (1991).

The negative binomial regression model has been used widely as an alternative to the basic Poisson model and is quite useful in accommodating overdispersion in data. As a pseudo likelihood specification it was preferred to the normal and gamma
models in a Monte Carlo set-up (Bourlage-Doz 1988). It is easier to estimate by maximum likelihood than the Poisson-inverse Gaussian mixture. Maximum likelihood equations involve potentially troublesome digamma functions but these can be handled using convenient computational devices discussed in Lawless (1987b). Currently it is popular in applied work. It has been used in modelling health utilization (Cameron et al 1988), in modelling industrial injuries (Ruser 1991a), in marketing (Morrison and Schmittlein 1988). Numerous authors have used this model in the estimation of count data models and in specification tests; the familiar example of the latter being the test for overdispersion in the Poisson model, (Engel 1984; GMT 1984b; HHG 1984; Collings and Margolin 1985; Cameron and Trivedi 1986; Lee 1986; Lawless 1987b; Harvey and Fernandes 1989). A historical account of this distribution can be found in Johnson and Kotz (1969, ch. 5).

Another quite appealing Poisson mixture is the double Poisson distribution proposed by Efron (1986) within the context of the double exponential family. This distribution is obtained as an exponential combination of two Poisson distributions using a dispersion parameter for weighting. The double Poisson density is

\[ h(y_i; \theta_i, \phi) = K(\theta_i, \phi) \cdot \phi^{1/2} \exp(-\phi \theta_i) \cdot \exp(-y_i \theta_i) \cdot \left( \frac{\theta_i}{y_i} \right)^{y_i} \phi^{y_i}, \]  

(3.9)

where \( K(\theta_i, \phi) \sim 1 + \frac{1-\phi}{12\phi \theta_i} \left( 1 + \frac{1}{\theta_i \phi} \right) \) ensures \( h(.) \) integrates to unity. This distribution has mean value approximately \( \theta_i \) and variance approximately \( \theta_i / \phi \) (Efron 1986, p. 715). The parameter \( \theta_i \) is similar to the Poisson mean parameter and \( \phi \) is a dispersion parameter. Efron (1986) shows that the constant \( K(\theta_i, \phi) \) in (3.9) nearly equals 1. The Poisson model is nested in the double Poisson model for \( \phi = 1 \). The double Poisson model allows for overdispersion (\( \phi < 1 \)) as well as underdispersion (\( \phi > 1 \)). Another advantage of the double Poisson regression model is that both the mean (\( \theta_i \)) and the dispersion (\( \phi \)) parameters can be allowed to depend on observed explanatory variables. Thus, it is possible to model the mean and dispersion structure separately as is done in a heteroskedastic linear regression model.

One way of generating new count models is via a flexibly parameterized recursion for probability ratio as in the case of the Katz family of distributions. The Katz family is defined by the probability recursion \( p(y | x) = \frac{\alpha + y \beta}{y} \), \( \alpha > 0, \beta < 1 \). In a similar spirit Gourieroux and Monfort (1989) have proposed the Additive Log-Differenced Probability (ALDP) model in which they use the more general recursion

\[ p(y | x) / p(y - 1 | x) = f(y, \theta) = g(y) h(\theta) = g(y) \exp(x' \beta), \quad y = 1, 2, ..., K \]  

(3.10)

The conditional probability of \( y \) given \( x \) is given by

\[ p(y | x) = \frac{\exp(\sum_{y=0}^{K} g^{*}(y) + yx' \beta)}{\sum_{y=0}^{K} \exp(\sum_{y=0}^{y} g^{*}(y) + yx' \beta)} \quad y = 1, 2, ..., K, \]  

(3.11)
where $K$ can be finite or infinite. It is easy to show that the LEF has the ALDP structure. Hence Poisson and geometric could be embedded in it. But the model is not itself nested in a more general model like the Negbin or the double Poisson, and one does not have a point process interpretation for the model. The model has been used in empirical work (Winkelmann 1993a; Winkelmann and Zimmermann 1992).

As discussed in Section 2, in principle an event count model can be derived from a specified distribution of waiting times. This involves a convolution integral which may be sometimes solvable by Laplace transform methods. Winkelmann (1993b) begins with a gamma distributed waiting times and derives the implied probability distribution for counts (also see Gourieroux and Visser 1993), which involves an incomplete gamma function. While this precludes a closed form expression for the moments, estimation can be handled by numerical methods. Maximum likelihood estimation is straight-forward. The gamma model has a parameter that can capture negative or positive duration dependence, if it exists in the data. Winkelmann shows that if the waiting times have negative duration dependence (decreasing hazard rate), this leads to overdispersion while positive duration dependence leads to underdispersion.

While closed form mixed models are computationally convenient, advances in numerical integration techniques have now made it feasible to estimate models that may not have explicit mathematical form.

At present not a great deal is known when certain probability models out perform others. We next consider a class of models which are motivated by specific limitations of the Poisson regression model - differences in the process generating the zeros and the nonzeros and the problem of ‘zero inflation’.

### 3.4.2. Hurdle models

One limitation of standard count models is that the zeros and the nonzeros (positives) are assumed to come from the same data generating process. Mullahy (1986) suggested modified count models in which these two processes are not constrained to be the same.\(^4\) The basic idea is that a binomial probability governs the binary outcome of whether a count variate has a zero or a positive realization. If the realization is positive, the hurdle is crossed, and the conditional distribution of the positives is governed by a truncated-at-zero count data model. Mullahy (1986) provides the general form of hurdle count models, together with applications to daily consumption of various beverages. The hurdle model is the dual of the split-population survival time model (Schmidt and Witte 1989) where the probability of an eventual “death” and the timing of “death” depend separately on individual characteristics. An interesting generalization is to consider hurdle models

\(^4\)The basic idea for a hurdle model was developed by Cragg (1971) as a modification of the basic Tobit model. Cragg's specification of this modified Tobit model was employed as a basis of test of the Tobit specification by Lin and Schmidt (1984).
with arbitrary rather than zero threshold point. Wilson (1991) has proposed such a
generalized hurdle models where the data are allowed to determine the hurdle point.

As an illustration consider the hurdle version of the Negbin2 model. Let \( \theta_{1i} = \theta_1(x_i, \beta_1) \), for example, \( \theta_{1i} = \exp(x_i' \beta_1) \), be the Negbin2 mean parameter for the case of
zero counts. Similarly, let \( \theta_{2i} = \theta_2(x_i, \beta_2) \) for the positives and \( \omega = \{1, 2, \ldots \} \). Further
define the indicator function \( I_i = 1 \) if \( y_i > 0 \) and \( I_i = 0 \) if \( y_i = 0 \). From the Negbin
distribution given in (3.8) with \( k = 0 \), the following probabilities can be obtained:

\[
P(y_i = 0 | x_i) = (1 + \alpha_1 \theta_{1i})^{-1/\alpha_1}; \tag{3.12}
\]

\[
1 - P(y_i = 0 | x_i) = \sum_{y \in \omega} h(y_i | x_i) = 1 - (1 + \alpha_1 \theta_{1i})^{-1/\alpha_1}; \tag{3.13}
\]

\[
P(y_i | x_i, y_i > 0) = \frac{\Gamma \left( y_i + \alpha_2^{-1} \right)}{\Gamma(\alpha_2^{-1}) \Gamma(y_i + 1)} \left( \frac{1}{(1 + \alpha_2 \theta_{2i})^{1/\alpha_2} - 1} \right) \left( \frac{\theta_{2i}}{\theta_{2i} + \alpha_2^{-1}} \right)^{y_i}, \text{ } y_i \in \omega. \tag{3.14}
\]

The equation in (3.12) gives the probability of zero counts while (3.13) is the proba-
bility that the threshold is crossed. Equation (3.14) is the truncated-at-zero Negbin2
distribution. The density function for the observations is

\[
[P(y_i = 0 | x_i)]^{1-I_i} \cdot [(1 - P(y_i = 0 | x_i)) P(y_i | x_i, y_i > 0)]^{I_i}. \tag{3.15}
\]

The Poisson hurdle and the geometric hurdle models examined in Mullahy (1986)
can be obtained from (3.12)~(3.15) by setting \( \alpha = 0 \) and \( \alpha = 1 \), respectively.

In two recent papers (Mullahy 1986; Lambert 1992), the authors consider count
models with zeros (WZ), or zero inflated Poisson (ZIP), in which there is an excess
of zeros relative to the parent distribution with the same mean. A zero inflated count
model takes the form

\[
\begin{align*}
\zeta + (1 - \zeta)h(y_i, \delta), & \quad y_i = 0 \\
(1 - \zeta)h(y_i, \delta), & \quad y_i = 1, 2, \ldots;
\end{align*} \tag{3.16}
\]

where \( h(y_i, \delta), y_i = 0, 1, \ldots \), is the parent distribution and \( 0 < \zeta < 1 \). It is possible
to allow for decreasing proportion of zeros if \[ -h(0, \delta) \cdot (1 - h(0, \delta)) < \zeta < 1 \]. The set
up in (3.16) can be used to accommodate excess of ones for an integer-valued random
variable. In estimation, the parameter \( \zeta \) may be further parameterized, for example as
a logit function of some covariates, as suggested by Lambert (1992).
3.4.3. Finite mixtures

A related model involves a finite mixture of densities. For example, if the sample is a probabilistic mixture from two populations with p.d.f. \( h(y_1 \mid x_1, \theta_1) \) and \( h(y_2 \mid x_2, \theta_2) \), then \( p h_1(.) + (1 - p) h_2(.) \), where \( 1 \geq p \geq 0 \), defines a finite mixture. That is, observations are draws from \( h_1 \) and \( h_2 \), with probabilities \( p \) and \( 1 - p \) respectively. The parameters to be estimated are \((p, \theta_1, \theta_2)\). The parameter \( p \) may be further parameterized using, for example, the logit function. Thus \( p = \frac{\exp(\lambda)}{1+\exp(\lambda)} \) and \( \lambda \) in turn may be parameterized in terms of further observable covariates. For Poisson models without covariates, this specification has been used in Everitt and Hand (1981). Generalization to more general additive mixtures is in principle straightforward. Finite mixtures of some standard count models are discussed in Titterington et al (1985). The finite mixture model is somewhat different from the heterogeneous Poisson model because it changes the conditional mean specification of the model, not just the variance function for a given mean. Its relevance in the present context arises from practical difficulties of distinguishing between alternative mixtures. If, however, the observed distribution is strongly multi-modal, there may be a good case for a finite mixture model.

3.5. Models with truncation, censoring or sample selection

Consider count regression models in which the response variable is further constrained due to sample truncation, censoring or sample selection. The models considered here are analogous to truncated and sample selection models with continuous variables, particularly the Tobit Models, that have been used extensively in the economics literature. Amemiya (1984), Maddala (1983) and Pudney (1989) contain useful surveys and references on truncated and censored econometric models; see also Heckman (1976, 1979) and Blundell (1987).

Subsection 3.5.1 deals with truncated and censored count models where censoring is done with respect to the endogenous variable of interest. We then consider the case where censoring is defined through the value of another variable.

3.5.1. Standard truncated and censored Models

The model is truncated if observations in some range are totally lost and censored if one can observe values of explanatory variables corresponding to the missing observations on the count dependent variable. There have been recent developments in truncated and censored count models. First consider truncated models. In some studies involving count data, inclusion in the sample requires that sampled individuals have been engaged in the activity of interest or, as Johnson and Kotz (1969) put it, “the observational apparatus become active only when a specified number of events (usually one) occurs”. Examples of truncated counts include the number of bus trips made per week (in surveys taken on
buses), the number of shopping trips made by individuals sampled at a mall, and the number of unemployment spells among a pool of unemployed. These are examples of left truncation or truncation from below. Right truncation, or truncation from above, may result when high counts are not observed. A distribution is doubly truncated if it is truncated both from above and below.

For simplicity, only truncation from below will be considered. The following general framework for truncated count models will be used. Let \( H(y_i, \Lambda) = P(Y_i \leq y_i) \) denote the cumulative distribution function (cdf) of the discrete random variable with probability density function (pdf) \( h(y_i, \Lambda) \), where \( \Lambda \) is a parameter vector. If realizations of \( y \) less than a positive integer \( r \) are omitted, the ensuing distribution is called left truncated (or truncated from below). The left truncated count distribution is given by

\[
f(y_i, \Lambda | y_i \geq r) = \frac{h(y_i, \Lambda)}{1 - H(r - 1, \Lambda)}, \quad y_i = r, r + 1, \ldots
\]

(3.17)

The mean of the left-truncated model can be expressed as: \( E(y_i | y_i \geq r) = E(y_i) + \delta_i \), where \( E(y_i) \) is the mean of the untruncated parent distribution and \( \delta_i > 0 \) is an adjustment factor. The adjustment factor plays a useful role, analogous to the Mill's ratio in continuous models, in estimation and testing of count models. The most common form of truncation in count models is the left truncation-at-zero; \( r = 1 \).

In maximum likelihood estimation of truncated models a misspecification of the underlying distribution leads to inconsistency due to the presence of the adjustment factor. This result is analogous to that for the truncated normal model. We will consider the maximum likelihood (ML) estimation of left truncated Poisson models. Using (2.2), (2.4) and (3.17), the log-likelihood \( \ell(\beta) \) based on \( n \) independent observations is:

\[
\ell(\beta) = \sum_{i=1}^{n} \left[ y_i \log(\theta_i) - \theta_i - \log \left( 1 - \exp(-\theta_i) \sum_{j=0}^{r-1} \frac{\theta_i^j}{j!} \right) \right] - \log(y_i!)
\]

(3.18)

The MLE of \( \beta \) is the solution of the following equation:

\[
\sum_{i=1}^{n} [y_i - \theta_i - \delta_i] \theta_i^{-1} \frac{\partial \theta_i}{\partial \beta} = 0,
\]

(3.19)

where \( \delta_i = \theta_i \cdot h(r, \theta_i)/[1 - H(r - 1, \theta_i)] \) and \( h(.) \) and \( H(.) \) are the pdf and cdf of the Poisson model. Equation (3.19) and its interpretation parallels that for the normal truncated regression. For the exponential specification \( \theta_i = \exp(x_i' \beta), \frac{\partial \theta_i}{\partial \beta} = ... \)

---


18
\( \theta; z_i \): So (3.19) reduces to an orthogonality condition between the \( z_i \) and the weighted (generalized) residuals. The information matrix is

\[
\mathfrak{I}(\beta) = -E \left[ \frac{\partial^2 \ell(\beta)}{\partial \beta \partial \beta'} \right] = \sum_{i=1}^{n} \left[ \theta_i - \delta_i (\theta_i + \delta_i - r) \right] \theta_i^{-1} \frac{\partial \theta_i}{\partial \beta} \frac{\partial \theta_i}{\partial \beta'}.
\]

(3.20)

The MLE \( \hat{\beta} \) is asymptotically normal with mean \( \beta \) and covariance matrix \( \mathfrak{I}(\beta) \).

Now, consider the case of censored samples. Censoring of count observations may arise from aggregation or may be imposed by survey design; see, for example, Terza (1985). Alternatively, censored samples may result when high counts are not observed\(^6\). Consider count models which are censored from above at point \( c \). An implicit regression model for a latent count variable \( y_i^* \) is

\[
y_i^* = \theta(x_i, \beta) + \mu_i,
\]

(3.21)

where \( \mu_i \) is a disturbance term with \( E(\mu_i) = 0 \). For a right censored count model, the latent endogenous variable \( y_i^* \) is related to the observed dependent variable \( y_i \) as follows: \( y_i = y_i^* \) if \( y_i^* < c \), \( y_i^* = c \) if \( y_i^* = c \), and \( y_i^* \geq c \). Define the indicator function \( I_i \) as follows: \( I_i = 1 \) if \( y_i^* < c \), and \( I_i = 0 \) if \( y_i^* \geq c \). Maximum likelihood estimation of censored count models raises issues similar to those in Tobit models (Terza 1985; Gurmu 1992). It is assumed that \( \{x_i\} \) are observed for all \( i \) and that the censoring mechanism and the data generation process for the count variable are independent. The log-likelihood function for \( N \) independent observations from model (3.21) is

\[
\ell(\Lambda) = \sum_{i=1}^{n} \left[ I_i \cdot \log(h(y_i, \Lambda)) + (1 - I_i) \cdot \log(1 - H(c - 1, \Lambda)) \right],
\]

(3.22)

where \( h(y_i; \Lambda) \) is the pdf of \( y_i \) and \( H(y_i; c - 1, \Lambda) = P(Y_i \leq c - 1) \), respectively.

For the right censored Poisson model the maximum likelihood estimating equation is:

\[
\sum_{i=1}^{n} \left[ I_i (y_i - \theta_i) + (1 - I_i) \delta_i \right] \theta_i^{-1} \frac{\partial \theta_i}{\partial \beta} = 0,
\]

where now \( \delta_i = \theta_i \cdot h(c - 1, \theta_i) / [1 - H(c - 1, \theta_i)] \) is the adjustment factor associated with the left truncated Poisson model. Since \( (y_i - \theta_i) \) is the error for the uncensored Poisson model and \( E(y_i - \theta_i | y_i \geq c) = \delta_i \), the expression: \( I_i (y_i - \theta_i) + (1 - I_i) \delta_i \) given above is simply the generalized error (Gourieroux et al 1987) for the right censored Poisson model. Consequently, the above likelihood equation implies that the vector of

\(^6\)Applications of censored count models include provision of hospital beds for emergency admissions (Newell 1965) and number of shopping trips (Terza 1985; Okorwu, Terza and Nourse 1988).
generalized residuals is orthogonal to the vector of exogenous variables, a result which is analogous to the Tobit likelihood equations.

The semiparametric approach outlined in Section 3.3.1 (see equation (3.6)) can be utilized for estimation of truncated and censored count regression models based on Poisson models with unobserved random heterogeneity. In this approach, in addition to the regression parameters of interest, the mixing distribution (of \( \nu \)) is estimated consistently with a finite number, \( q \), of mass points. The approach can be illustrated for the most common case of zero-truncated sample\(^7\).

Using the framework of equation (3.6), let \( (y_i, \pi_i, \nu_j) \) have a Poisson distribution with mean \( (\theta_i\nu_j) \), where \( \theta_i = \exp(x_i'\beta) \). One needs to estimate \( \beta, \nu_j \), and the associated probability \( \pi_j \) for \( j = 1, 2, \ldots q \). The ensuing semiparametric log-likelihood function for the zero-truncated sample can be written as

\[
\ell_{sp} = \sum_{i=1}^{n} \left[ y_i \log(\theta_i) - \log(y_i!) - \log \left( 1 - \sum_{j=1}^{q} \exp(-\theta_i\nu_j)\pi_j \right) \right] + \log \left( \sum_{j=1}^{q} \exp(-\theta_i\nu_j)\nu_j^y \pi_j \right). \tag{3.23}
\]

The probabilities may be constrained to lie between 0 and 1 by using the specification \( \pi_j = \left[ 1 + \exp(-\varphi_j) \right]^{-1} \) and maximizing with respect to \( \varphi_j, j = 1, 2, \ldots q - 1 \) and obtain the \( q \)-th probability estimate as \( \pi_q = 1 - \sum_{j=1}^{q-1} \pi_j \). In practice, the estimated number of mass points \( q \) is quite low.

3.5.2. General count models with sample selection

Sample selection problems arise when a non-random sample is either not available or because, to improve the precision of probability estimates of a relatively infrequently occurring event, the sample is overweighted with individuals belonging to the class which is especially relevant for the event of interest. Greene (1994) gives an example on consumer loan behavior, default on credit card loans, and examines the number of major derogatory reports to a credit reporting agency for a group of credit card applicants. Since applicants whose applications for credit card were accepted are unlikely to have any major derogatory reports, this screened subpopulation fits nicely into the framework of sample selection models.

In two recent papers (Alessie, et al 1990; Van Praag and Vermeulen 1993), the authors consider models in which the sample design imparts sample selection effects. The authors are specifically concerned with the problem of underrecording of both counts which occurs as follows. Consumers make repeated purchases of some good such as bread or tobacco. The amount spent by the \( i \)-th individual on the \( j \)-th shopping trip is denoted by \( y_{ij}, i = 1, 2, \ldots n, j = 1, 2, \ldots M_i \) and the shopping trip is only recorded

\(^7\)Brännäls and Rosenqvist (1992) provide details of the estimation algorithm and an empirical illustration using count data.
if the amount spent on any particular occasion exceeds some cut-off point, say $y_{\text{min}}$. Since "small" purchases are ignored, there is underrecording of both counts (number of trips) and the amount spent. The recorded distribution of $y_{ij}$ is truncated at the cut-off point which in turn determines the extent of under-recording of counts. Ignoring the effect of $y_{\text{min}}$ on the recorded distribution of counts will lead to sample selection bias. Van Praag and Vermeulen propose a model in which the frequency distribution of counts $(k_i)$ is $p(k_i, X_i; \theta_1 | X_i, \text{SSR})$ and the distribution of amounts $(y_{i1}, ..., y_{iM_i})$ is $f((y_{i1}, ..., y_{iM_i}, X_i; \theta_2 | X_i, \text{SSR})$, where SSR is the sample selection rule determined by $y_{\text{min}}$. They choose $f(\cdot)$ to be (truncated) lognormal and $p(\cdot)$ to be negative binomial, and the parameters $\theta_1$ and $\theta_2$ are estimated jointly by maximum likelihood. Different SSR's will lead to different amounts of data loss and hence impact on the estimates differently, depending upon the true underlying distributions of the data. Sensitivity of results to alternative truncation rules is investigated by the authors, which provides some information about the adequacy of the specification of the parent model.

In a similar spirit, Greene (1994) and Terza (1994) develop models for sample selection for count models which is analogous to Heckman (1979) treatment of sample selection for continuous choice models. An interesting issue here is whether a two-step Heckman-type procedure analogous to that for the normal linear model can be devised and whether the "correction" term enters linearly in the conditional mean function, $E[y_i | x_i, z_i]$, where $z_i$ is the endogenous sample selection variable. Estimation is either by full information maximum likelihood or by two-step non-linear least squares methods. The former requires the specification of the joint distribution of both discrete random variables; that is, $z_i$ and $(y_i | \text{sample selection rule})$.

Sample selection models are relatively new in the count regression literature, and scope exists for further research. The choice of a satisfactory "correction" term in the conditional mean function, analogous to the Mill’s ratio, needs more work. The development of semiparametric estimators and inference procedures are issues for future research.

3.6. Multivariate count models

Multivariate count models are of interest when the joint distribution of counts is required or when counts are jointly determined with other non-count variables. A model of the frequency of entry and exit of firms to an industry is an example of a bivariate count process. Cameron et al (1988) model health-care utilization of individuals measured hospital admissions, days spent in hospitals, prescribed and nonprescribed medicines taken by individuals. Such variables can be expected to be jointly dependent.

---

*Multivariate count models are also useful in sociology and political science. For example, Good and Pirog-Good (1989) consider several bivariate count models for teenage delinquency and paternity;*
specification of models of dependent counts.

A well-established technique for deriving multivariate (especially bivariate) count distributions is the method of mixtures and convolutions. The oldest and the most studied special case is the bivariate Poisson model which can be generated by sums of independent random counts with common components in the sums. Suppose Poisson variables \( y_1 \) and \( y_2 \) are defined as \( y_1 = u + v, \ y_2 = v + w \), and \( u, v \) and \( w \) are independently distributed as Poisson variables with parameters \( \theta_1, \theta_2 \) and \( \theta_3 \), respectively. Then \( y_1 \sim \text{Poisson}(\theta_1 + \theta_2), \ y_2 \sim \text{Poisson}(\theta_2 + \theta_3), \ \text{cov}(y_1, y_2) = \text{cov}(u + v, v + w) = \text{var}(v) = \theta_2 > 0 \), and \( \rho^2 = \frac{\theta_2}{(\theta_1 + \theta_2)(\theta_2 + \theta_3)} \); see Johnson and Kotz (1969), GMT (1984). In this case correlation coefficient fully characterizes dependence. In this case also, the marginal distributions for \( y_1 \) and \( y_2 \) are both Poisson. Hence, with a correctly specified conditional mean function the marginal models can be estimated consistently, but not efficiently, by maximum likelihood. Note also that this model is restrictive in that it only permits positive correlation between counts. King (1989a) calls this a model of ‘seemingly unrelated counts’ by analogy with the well-known seemingly unrelated least squares model; he also gives an application. Bivariate Poisson is a special case of the generalized exponential family (Jupp and Mardia 1980; Kocherlakota and Kocherlakota 1993);

\[
g(y_1, y_2; \theta_1, \theta_2, \rho) = \exp \left\{ \theta_1 v(y_1) + v(y_1)^\prime \rho w(y_2) + \theta_2 w(y_2) \right\} - c(\theta_1, \theta_2, \rho) + d_1(y_1) + d_2(y_2), \tag{3.24}
\]

where \( v(\cdot), w(\cdot), d_1(\cdot) \) and \( d_2(\cdot) \) are some functions, and \( c(\theta_1, \theta_2, 0) = c_1(\theta_1) \cdot c_2(\theta_2) \). Under independence \( \rho = 0 \) in which case the right hand side is a product of two exponential families. In this family \( \rho = 0 \) is a necessary and sufficient condition for independence.

The special bivariate Poisson model given above is a well-known example of a convolution family. In general, multivariate distributions can be generated from mixtures and convolutions of product families (see Marshall and Olkin 1990) in a manner analogous to equation (3.5) which leads to compound marginal distributions. Consider the bivariate distribution \( h(y_1, y_2) \) where

\[
h(y_1, y_2 \mid x_1, x_2) = \int_0^\infty h_1(y_1 \mid x_1, \nu) \cdot h_2(y_2 \mid x_2, \nu) \cdot g(\nu) d\nu. \tag{3.25}
\]

where \( h_1, h_2, \) and \( g \) are univariate densities. Multivariate distributions generated in this way will have univariate marginals in the same family (Kocherlakota and Kocherlakota 1993). This approach suggests a way of specifying or justifying overdispersed and correlated count models, based on a suitable choice of \( g(\cdot) \), more general than in the example given above. Marshall and Olkin (1990) generate a bivariate negative binomial

---

King (1989a) presented a bivariate model of U.S. presidential vetoes of social welfare and defense policy legislation.
distribution beginning with \( h(y_1) \) and \( h(y_2) \) which are Poisson with parameters \( \alpha \theta \) and \( \gamma \theta \), respectively, and \( \theta \) has gamma distribution with parameter \( r \). That is,

\[
h(y_1, y_2 | r) = \int_0^\infty \left[ (\alpha \theta)^{y_1} \exp(-\alpha \theta)/y_1! \right] \left[ (\gamma \theta)^{y_2} \exp(-\gamma \theta)/y_2! \right] \left[ \theta^{r-1} \exp(-\theta)/\Gamma(r) \right] d\theta
\]

\[
= \frac{\Gamma(y_1+y_2+r)}{y_1!y_2!} \left[ \frac{\alpha}{\alpha+\gamma+1} \right]^{y_1} \left[ \frac{\gamma}{\alpha+\gamma+1} \right]^{y_2} \left[ \frac{1}{\alpha+\gamma+1} \right]^r.
\]

(3.26)

The marginals are again univariate negative binomial and the correlation is positive.

More flexible bivariate and multivariate parametric count data models can be constructed by introducing correlated unobserved heterogeneity components in models in which the distribution, conditional on heterogeneity, has a specified functional form, e.g. Poisson, and the heterogeneity component has a specified distribution. For example, suppose \( y_1 \) and \( y_2 \) are, respectively, Poisson(\( \mu_1 \mid \xi_1 \)) and Poisson(\( \mu_2 \mid \xi_2 \)) where \( \mu_1 \mid \xi_1 = \exp(\beta_0 + \xi_1 + \beta_1 x) \) and \( \mu_2 \mid \xi_2 = \exp(\beta_0 + \xi_2 + \beta_1 x) \), where \( \xi_1 \) and \( \xi_2 \) represent unobserved heterogeneity; their presence induces overdispersion in the marginal distributions of \( y_1 \) and \( y_2 \). Dependence between \( y_1 \) and \( y_2 \) is induced if \( \xi_1 \) and \( \xi_2 \) are correlated. In these cases the marginal distributions for \( y_1 \) and \( y_2 \) will exhibit both overdispersion and dependence, but in general in such cases neither the joint nor the marginal distributions will have closed form expressions. Maximum likelihood estimation of such models will require numerically intensive methods, such as Monte Carlo integration. In such cases semiparametric methods have greater appeal, e.g. multivariate semiparametric generalized least squares methods as in Delgado (1990).

It is convenient to mention here some econometric work which includes jointly dependent continuous and count variables. An example of the former is a labor supply model with the number of children as an explanatory variable, but if fertility is treated as an endogenous variable then one may need to specify an equation for the number of children; see Browning (1992, pp. 1464-1465). Another example involves frequency/amount models of consumption in which the quantity and frequency of purchase are jointly modeled; see Mighir and Robin (1992). Terza and Wilson (1990) is an example of a joint model of counts and discrete choice dummy variables; they combine the Poisson hurdle model with a multinomial model to jointly predict household choices among types of trips and frequency of trips. Given the difficulties of writing down joint likelihood functions in such cases, single equation instrumental variable method seems the obvious choice of estimator here, with counts being instrumented. As the specification of the counts equation is usually nonlinear, the choice of instruments raises delicate issues; see Browning for an illustration. In count models with unobserved heterogeneity it has been usual to assume that the heterogeneity term is uncorrelated with the regressors, but this is not always a plausible assumption (Mullahy 1993). Again an instrumental variable type approach seems appropriate, but further work is required on the properties of such estimators.
Up to this point in this section, the attention has been on issues important in cross-section analysis. We now turn to models of dependent counts which is a relatively under-developed area.

3.7. Time series models of counts

The standard count model assumes conditionally independent realizations. Dependent sequences are to be expected when dealing with time series of counts. However, applied time series analysis of non-Gaussian data in general, and count data in particular, is relatively underdeveloped, with much attention focusing on derivation of time series models with prespecified marginals. An example of an economic applications is Rose (1990). Time series analysis of independent observations which are realizations of a time homogeneous stochastic processes does not result in any new issues. However, one is also interested in count data analogs of autoregressive ("Markov") models in which \( y_t \) may depend upon \( (z_t, y_{t-1}, y_{t-2}, \ldots, y_{t-p}) \) since these are natural analogs of the ARMAX models for continuous data. However, to date most of the available work has been on "pure" autoregressive or ARMA time series models in which the realization of the random variable is a nonnegative integer. Al-Osh and Alzaid (1987) and Alzaid and Al-Osh (1990) deal with estimation of first order integer valued autoregressive (INAR(1)) and with p-th order integer valued autoregressive (INAR(p)) models, respectively; McKenzie (1988) with ARMA processes, Zeger (1988) and Zeger and Qaqish (1988) with a model with covariates.

There are obvious problems in writing an INAR model in the standard form \( y_t = \alpha y_{t-1} + \epsilon_t \) since this does not ensure integer valued realizations. This difficulty is overcome by replacing the product \( \alpha y_{t-1} \) by \( \alpha^0 y_{t-1} \) which introduces the key idea of the binomial thinning due to Steutel and Van Harn (1979). Let \( y \) be a non-negative integer-valued random variable; then for any \( \alpha \in [0, 1] \) the binomial thinning operator ‘\( \circ \)’ is defined by

\[
\alpha \circ y = \sum_{i=1}^{y} B_i,
\]

where \( B_i \) is a sequence of i.i.d. random variables independent of \( y \) such that

\[
\Pr(B_i = 1) = 1 - \Pr(B_i = 0) = \alpha.
\]

A discrete valued random variable \( y \) may be viewed as \( y \) distinct units. Considering each unit independently, suppose that each is retained with probability \( \alpha \), or removed with probability \( 1 - \alpha \). As in a regular AR model, larger is \( \alpha \) the greater the persistence of \( y \). The resulting variable \( \alpha \circ y \) is a discrete random variable, viewed as having been obtained by binomial thinning. The definition of the operator ‘\( \circ \)’ implies (i) \( 0 \circ y = 0 \), (ii) \( 1 \circ y = y \), (iii) \( E[\alpha \circ y] = \alpha E(y) \), (iv) \( \text{var}[\alpha \circ y] = \alpha^2 \text{var}(y) + \alpha(1 - \alpha)E(y) \), and (v)
for any $\beta \in [0, 1]$, $\beta \circ \alpha \circ y = \text{def} (\beta \alpha) \circ y$. The INAR(1) process \( \{y_t, t = 0, \pm 1, \pm 2, \ldots \} \) is defined by

\[
y_t = \alpha \circ y_{t-1} + \epsilon_t.
\] (3.27)

Thus, the value of $y$ at time $t$ is modeled as a sum of two terms. The first is obtained by applying the binomial thinning operator to $y_{t-1}$. The second term, independent of $y_{t-1}$, is the innovation generated by a discrete probability model, e.g., the Poisson or Negbin. Both terms will be integer valued. In the Poisson case where $E(\epsilon_t) = \theta$, $E(y) = \text{var}(y) = \frac{\theta}{1-\sigma}$. The time series $y_t$ has properties similar to that of the regular AR1 model but its autocorrelation function is always positive. In the stationary case, the distribution of $y_t$ is uniquely determined by the distribution of $\epsilon_t$; $y_t$ has a Poisson marginal distribution iff $\epsilon_t$ has Poisson distribution. Extensions to other marginals and to higher order INAR processes are possible.

Brännäs (1994) has given a thorough treatment of estimation and inference in this model, including results from the literature. The conditional probability is defined by

\[
\Pr(y_t \mid y_{t-1}) = \exp(-\theta) \sum_{i=0}^{\min(y_{t-1}, y_t)} \frac{\theta^{y_t-i}}{y_t-i} \binom{y_t-i}{i} \alpha^i (1-\alpha)^{y_t-1-i}, \ t = 2, \ldots, T
\] (3.28)

and the exact log-likelihood by the equation

\[
\ell_T(\alpha, \theta) = \ell_{T-1}(\alpha, \theta) + \ell_1(\alpha, \theta)
\] (3.29)

where the first term on the right hand side is based on 3.28 and the second term is

\[
\Pr[y_1] = \left[ \frac{\theta}{1-\alpha} \right] y_1 \exp \left[ -\frac{\theta}{1-\alpha} \right].
\]

Equation (3.29) can be maximized numerically.

How can covariates be introduced into (3.27)? There are at least two possibilities, one to parameterize $\alpha$ and second to parameterize $\theta$ in terms of covariates. But to date few results are available to deal with these models. A more manageable problem is a time series count regression without lagged dependent variable $y_{t-1}$ but with serially correlated error term. An important modeling issue concerns the manner in which dependence or serially correlated counts arise in practice. A particularly straightforward method is to modify the random intercept Poisson regression model. Rewrite (3.2) replacing individual subscript $i$ by time subscript $t$, $t = 1, \ldots, T$:

\[
\theta_t^* = E(y_t \mid x_t, \nu_t) = \exp(x_t' \beta + \epsilon_t) = \exp(x_t' \beta) \cdot \nu_t.
\] (3.30)

The presence of a time independent $\epsilon_t$ will lead to overdispersion. If, in addition, $\epsilon_t$ is serially correlated then the model generates both overdispersion and serial correlation in the counts in a manner that depends upon how the serial correlation process of $\epsilon_t$ is parameterized. This if further explored in Section 4.2.
3.8. Poisson process regression models

We consider regression methods for nonhomogeneous Poisson process models based on extensions of (2.7) and (2.8) to allow for covariates and unobserved heterogeneity. Assume that we observe \( i = 1, 2, \ldots, N \) individuals over the period \((0, T_i)\) and that \( y_i \) events occur at times \( t_{i0} = 0 < t_{i1} < t_{i2} < \cdots < t_{iy_i} < T_i \). In this set-up, count data for individuals with observed covariates \( x_i \) are observed at different points in time or over different intervals, so one might also wish to take explicit account of temporal variations in the rate of the process. An example of such a data set is provided by Cooil (1992) who considered Poisson process models to study medical malpractice claims filed against individual physicians during 1975-1987. In this case, \( t_{ij} \)'s \((j = 1, 2, \ldots, y_i)\) denote the incident times of the \( y_i \) claims filed against physician \( i \). Lawless (1987a) provides an excellent discussion of regression methods for Poisson process models. The basic reference for Poisson models with no covariates is Cox and Lewis (1966). Key references on development of regression methods for the closely related Cox proportional intensity model are Cox (1972, 1975) and Andersen and Gill (1982).

Consider the so-called proportional intensity Poisson process model with intensity function

\[
\theta^\dagger(t) = \theta_0(t) \cdot \exp(x_i^\dagger \beta),
\]

where \( \theta_0(t) \) is the baseline intensity assumed to be constant across individuals and the subscripts on time \( t \) have been suppressed for the moment. This form is analogous to Cox’s proportional hazard duration model. The cumulative or integrated intensity function is given by

\[
\Theta^\dagger(t) = \int_0^t \theta^\dagger(s)ds = \Theta_0(t) \cdot \exp(x_i^\dagger \beta),
\]

where \( \Theta_0(t) = \int_0^t \theta_0(s)ds \) is the cumulative baseline intensity function. In estimation, \( \theta_0(t) \) is specified up to a vector of parameters \( \gamma \) (fully parametric), e.g. the Weibull specification of Section 2.3, or is left arbitrary (semiparametric).

Lawless (1987a) has shown that under the Poisson model, the likelihood function takes the form

\[
L(\gamma, \beta; t, y) = L_1(\gamma; t) \cdot L_2(\gamma, \beta; y),
\]

where

\[
L_1(\gamma; t) = \prod_{i=1}^N \prod_{j=1}^{y_i} \frac{\theta_0(t_{ij}, \gamma)}{\Theta_0(T_i, \gamma)}
\]

and

\[
L_2(\gamma, \beta; y) = \prod_{i=1}^N \exp \left[ -\Theta_0(T_i, \gamma) \exp(x_i^\dagger \beta) \right] \left[ \Theta_0(T_i, \gamma) \exp(x_i^\dagger \beta) \right]^{y_i} / y_i!.
\]
The likelihood kernel $L_1(.)$ arises from the conditional distribution of the event times, given the counts and the kernel $L_2(.)$ arises from the Poisson distribution of the counts $y_1, y_2, ..., y_N$. The effects of covariates appear only through the counts. Thus, the counts are informative about the betas. Further, the decomposition of the likelihood function simplifies the estimation of the parameters. If the process is observed over the time period $(S_i, T_i)$, rather than $(0, T_i)$, we simply replace $\Theta_0(T_i, \gamma)$ by $[\Theta_0(T_i, \gamma) - \Theta_0(S_i, \gamma)]$ in (3.33) and elsewhere in this subsection. When $T_i$'s are equal to $T$, $L_2(.)$ can be decomposed as $L_2(\gamma, \beta; y) = L_3(\beta; y)L_4(\gamma, \beta; t, y)$ (Lawless 1987a). The fully parametric approach can be implemented using the specification $\theta_0(t_{ij}, \gamma) = \gamma t_{ij}^{\gamma - 1}$ with $\Theta_0(T_i, \gamma) = T_i^\gamma$, where now $\gamma$ is a scalar parameter.

The approach can be extended to allow for unmeasured heterogeneity by specifying the intensity function as $\theta_i^k(t) = \theta_0(t) \cdot \exp(x_i^k \beta) \cdot \nu_i$, where $\nu_i$ is the individual random effects. In the case where $\nu_i$ is gamma distributed, the likelihood function associated with the negative binomial regression model is of the form

$$L(\gamma, \alpha, \beta; t, y) = L_1(\gamma; t) \cdot L_2^*(\gamma, \alpha, \beta; y),$$

where $L_1(.)$ is the same as in (3.33) and

$$L_2^*(\gamma, \alpha, \beta; y) = \prod_{i=1}^N \frac{\Gamma(y_i + \alpha^{-1})}{\Gamma(\alpha^{-1}) \Gamma(y_i + 1)} \frac{[\Theta_0(T_i, \gamma) \exp(x_i^\beta)]^{\nu_i}}{[1 + \Theta_0(T_i, \gamma) \exp(x_i^\beta)]^{\nu_i + \alpha^{-1}}}.$$

Here $L_2^*(.)$ is the likelihood kernel based on the negative binomial model with mean $[\Theta_0(T_i, \gamma) \exp(x_i^\beta)]$ and dispersion parameter $\alpha$. In general, if $\nu_i$ are iid random variables with pdf $g(\nu_i)$, then the likelihood function is of the same form as in (3.34) with

$$L_2^*(.) = \prod_{i=1}^N \int_0^\infty \text{exp}[-\Theta_0(T_i, \gamma) \exp(x_i^\beta)\nu_i] \left[\Theta_0(T_i, \gamma) \exp(x_i^\beta)\nu_i\right]^{\nu_i - 1} \frac{1}{\nu_i!} g(\nu_i) d\nu_i.$$

We again note that for all Poisson process models with proportional intensity assumption, inference about covariates depends on the distribution of the counts. Semiparametric methods that do not require the full specification of the distribution of $\nu_i$ is a topic for future research.

Semiparametric methods that do not require parametric assumptions about $\theta_0(t)$ are available for the Poisson process model with proportional intensity assumption but with no random effects as in (3.31). This is based on Cox (1972) semiparametric approach (see, Andersen and Gill 1982; Lawless 1987a). Assume that the constant term have been incorporated in the baseline intensity function so that $x$ does not include the intercept term. Define $R_t = \{t \mid T_i \geq t_{ij}\}$ as the 'risk set', a set of individuals who are still under observation at time $t_{ij}$ so that $\Psi_t = \sum_{i \in R_t} \text{exp}(x_{ij}^\beta)$ is the sum of $[\exp(x_{ij}^\beta)]$ for the
individuals in $R_t$. Then the partial likelihood function for estimating $\beta$ alone is

$$L(\beta) = \prod_{i=1}^{N} \prod_{j=1}^{y_i} \frac{\exp(x'_i \beta)}{\Psi_t}.$$  (3.35)

The advantage of this procedure is that $\beta$ may be estimated consistently without making functional form assumptions about the intensity function. The details of the possibility of combining the partial likelihood approach with an allowance for unobserved heterogeneity remain to be investigated.

3.9. Panel data models

The most familiar example of count panel data analysis is the patents model extensively studied by HHG (1984), Hall, Griliches and Hausman (1986) and others; see references in the latter. Other applications of panel count data models include distribution of occupational injuries in manufacturing (Ruser 1991a,b). There are also related applications where a given count model accounts for both frequency and choice among different types of events. Some interesting recent applications include trip frequencies taken by households and type of trips identified by destination of the trip (Terza and Wilson 1990) and number of trips taken by the elderly and disabled people where trips are classified by modes of transportation (Stern 1991).

Consider a univariate response variable $y_{it}$ and a $k \times 1$ vector of explanatory variables (including the constant term) $x_{it}$, observed at times $t = 1, 2, \ldots, T$ for individuals $i = 1, 2, \ldots, n$. In panel data we frequently have $n$ large and $T$ small. Using the framework outlined in Section 3.3, individual effects (fixed or random effects) can be incorporated as follows: $E(y_{it} \mid x_{it}, \epsilon_i) = \exp(x'_i \beta + \epsilon_i) \equiv \theta_i^*$, where $\nu_i = \exp(\epsilon_i)$ captures the individual effects.

Fixed effects models

Standard fixed effects count models can be summarized using the framework provided by HHG (1984). In this specification, $\epsilon_i$ (or $\nu_i$) represents unobserved individual specific effects that are assumed to be constant over time. For the typical case where there are only a few time series observations per individual, separate estimation of $\nu_i$ is not feasible since this leads to inconsistent MLE for $\beta$. This is the incidental parameter problem discussed in the literature (Hsiao 1986; Maddala 1987). Accordingly, fixed effects parametric count models are usually estimated using the conditional maximum likelihood approach. The approach is also used in other models with a similar problem; for example fixed effects logit and probit models. The idea is to obtain a likelihood function conditional on a sufficient statistic for $\nu_i$. Such a sufficient statistic is available in the case of Poisson and negative binomial models. The required sufficient statistic for $[\nu_i \cdot \sum_t \exp(x'_i \beta)]$ is $\sum_t y_{it}$. 

28
In order to implement the conditional ML approach, two further steps are necessary. First, a closed form for the joint distribution of \((y_{i1}, y_{i2}, \ldots, y_{iT})\), conditional on \(\sum_t y_{it}\), must be obtained using the expression:

\[
f(y_{i1}, \ldots, y_{iT} \mid \sum_{i=1}^T y_{it}) = \frac{f \left( y_{i1}, \ldots, y_{iT-1}, \sum_{t=1}^T y_{it} - \sum_{t=1}^{T-1} y_{it} \right)}{\text{Pr} \left( \sum_{t=1}^T y_{it} \right)}, \tag{3.36}
\]

where \(f(.)\) denotes the relevant probability distribution in each case. The second step involves choosing the appropriate conditional mean function and estimating the unknown parameters from the resulting regression model, in which the joint distribution is conditioned on \(\sum_t y_{it}\) and \(x_{it}\)'s.

For a Poisson based variable the conditional joint distribution in (3.36) results in a multinomial distribution and, given the error specification, one obtains the multinomial logit specification. In this case the density function for observation \(i\) takes the form

\[
f(y_{i1}, \ldots, y_{iT} \mid \sum_{i=1}^T y_{it}) = \frac{[(\sum_t y_{it})!] [\prod_t P_{it}^{y_{it}}]}{\prod_t (y_{it}!)} \tag{3.37},
\]

where \(P_{it} = \exp(x_{it}'\beta)/\sum_t \exp(x_{it}'\beta)\). The fixed effects Poisson model also has the restriction that the mean and variance are equal. Similarly, in the case of the negative binomial distribution, such a conditioning results in a multivariate hypergeometric distribution.

The conditional ML approach assumes that the distribution of \(y\) is fully specified. Chamberlain (1992) and Wooldridge (1990b) considered distribution-free estimation of panel data models in which consistent estimates of parameters can be obtained within the Generalized Method-of-Moments (GMM) framework. This procedure can be applied to count models with multiplicative fixed effects specified above. Let \(y_i = (y_{i1}, \ldots, y_{iT})'\), \(x_i = (x_{i1}', \ldots, x_{iT}')\), \(R(x_i, \beta)\) be a vector whose \(t\)-th row is \(\exp(x_{it}'\beta)\), and \(\Phi(x_i) = E \{ \text{var}(y_{i1} \mid x_i, \nu_i) \}\) be a positive definite matrix. The objective is to estimate the unknown true regression parameter \(\beta_0\) and the unconditional mean \(\gamma_0 = E(\nu_i)\) within a semiparametric framework. The semiparametric fixed effects estimator of \((\beta, \nu_1, \nu_2, \ldots, \nu_n)\) will be based on minimizing the following generalized least squares function:

\[
\sum_{i=1}^n \left[ y_{i} - R(x_i, \beta)\nu_i \right]' \Phi^{-1}(x_i) \left[ y_{i} - R(x_i, \beta)\nu_i \right]. \tag{3.38}
\]

For given \(\beta\), the minimizing value of \(\nu_i\) is given by

\[
\hat{\nu}_i(\beta) = \left[ R'(x_i, \beta)\Phi^{-1}(x_i)R(x_i, \beta) \right]^{-1} \cdot R'(x_i, \beta)\Phi^{-1}(x_i)y_{i}. \tag{3.39}
\]
Substituting this into (3.38) gives the concentrated least squares problem. A consistent estimator, say \( \hat{\beta} \), of \( \beta \) is then obtained by minimizing the following function:

\[
\sum_{i=1}^{n} y_i \Phi^{-1}(x_i, \beta) y_i,
\]

(3.40)

where \( \Phi^{-1}(x_i, \beta) = \Phi^{-1}(x_i) - \Phi^{-1}(x_i) R(x_i, \beta) \left[ R'(x_i, \beta) \Phi^{-1}(x_i) R(x_i, \beta) \right] R'(x_i, \beta) \Phi^{-1}(x_i) \).

This and (3.39) gives \( \hat{\nu}_i(\hat{\beta}) \). We then obtain \( \hat{\gamma} \) from the average of the \( \hat{\nu}_i \)'s:

\[
\hat{\gamma} = \frac{1}{n} \sum_{i=1}^{n} \hat{\nu}_i(\hat{\beta}).
\]

**Random effects models**

In the random effects count models \( \nu_i = \exp(\epsilon_i) \) is treated as a random variable with \( E(\nu_i) = 1 \) and \( \text{var}(\nu_i) = \sigma^2 \). Assume that \( \epsilon_i \)'s are independently and randomly distributed across cross-sectional units independent of the \( x_{it} \)'s and the error term of the model. Estimation of random effects count models is usually done using ML methods (HHG 1984; Hall et al 1986; Ruser 1991a). The random effects Poisson model can be obtained from the Poisson-Gamma mixture; where \( (y_{it} \mid x_{it}, \epsilon_i) \) and \( \nu_i = \exp(\epsilon_i) \) are, respectively, Poisson and gamma distributed. The resulting negative binomial type model for panel data can then be estimated using the maximum likelihood method. For the random effects negative binomial model, HHG (1984) used the negative binomial and Beta distribution mixture. An alternative formulation of panel count data models was proposed by GMT (1984b). They propose an error component model with both individual and time specific effects. However, this will require a very long panel data for consistent estimation of the parameters.

Independence between \( \epsilon_i \) and \( x_{it} \) is a crucial assumption for the random effects models but not required for the fixed effects count models. If the assumption of orthogonality is violated, the estimator of \( \beta \) obtained from the random effects models is inconsistent while the estimator from fixed effects model is not affected. Accordingly, Hausman's (1978) test can be employed to test whether individual effects are correlated with the \( x_{it} \)'s.

Further aspects of estimation of random effects panels are dealt with in Section 4.3.

**4. FURTHER ASPECTS OF ESTIMATION**

Weighted nonlinear least squares type estimators are known to be consistent and asymptotically normal under regularity conditions (Amemiya 1985). Maximum likelihood involves strong distributional assumptions and implicitly more moment restrictions on the data (in general). Some of these may be invalid, some times leading to inconsistency. The alternative quasi-likelihood approach based on low order moment specification will be preferred on grounds of robustness. If the first two moments are correctly specified, the quasi-likelihood estimator (QLE) is asymptotically fully efficient in the special case
where the data generation process is a LEF, but in general has a lower efficiency (GMT 1984a). In this section we consider estimation of count data models using low order moment specification. The standard approach uses the first two moments. We also survey the extended QL (EQL) approach. This involves the joint estimation of mean and variance parameters such as those that appear in the extended LEF (Nelder and Pregibon 1987). The QL approach can be regarded as a special case of EQL.

4.1. Extensions of the quasi-likelihood estimation method

As a preliminary example of QL estimation, consider the estimation of the parameters \((\beta, \phi)\) given the following modified Poisson moment conditions assuming independent observations: \(E(y_i \mid x_i) = \theta(x_i, \beta) = \theta_i; \) \(\text{var}(y_i \mid x_i) = \phi \theta_i\), where the variance depends upon the nuisance parameter \(\phi\). The quasi-likelihood estimator for \(\beta\) is defined by the orthogonality conditions

\[
\sum_{i=1}^{n} \frac{\partial \theta_i}{\partial \beta_j} \frac{(y_i - \theta_i)}{\theta_i} = 0; \quad j = 1, \ldots, k. \tag{4.1}
\]

Notice that the solution does not depend upon \(\phi\). Given the variance specification, a consistent estimator of \(\phi\) is given by

\[
\hat{\phi} = \frac{1}{n} \sum_{i=1}^{n} \frac{(y_i - \hat{\theta})^2}{\hat{\theta}_i}, \tag{4.2}
\]

where \(\hat{\theta}_i = \theta(x_i, \hat{\beta})\). The estimating equation (4.1) is a score type equation even though it is not derived from the likelihood; the estimating equation (4.2) is a moment type equation. Jointly (4.1) and (4.2) define a quasi-likelihood estimator for a modified Poisson model. We could have applied the same idea using a different variance formulation. A closely related estimator is defined by (3.3) and (3.4). In the econometric literature the properties of this type of sequential estimator have been discussed by Newey (1984).

Crowder (1987), Firth (1987) and Godambe and Thompson (1989) have proposed a refinement of the quasi-likelihood approach based on the estimating functions or equations approach. In this approach the central focus of estimation is on an estimating equation, defined in terms of data and parameters, with zero mean at the true parameter, whose solution defines an estimator, rather than on an objective function that is being optimized. These are analogous to the score equations in maximum likelihood theory or the orthogonality conditions in GMM estimation. For the general approach, assume independent observations \(\{y_i\}\) with means \(E(y_i \mid x_i) = \theta_i\) and variances \(\text{var}(y_i \mid x_i) = \sigma_i^2\) depending on a vector parameter \(\rho\). Crowder (1987) has proposed the following
general equations:

\[ g_n(\rho) = \sum_{i=1}^{n} \left[ a_{i\rho}(\rho) (y_i - \theta_i) + b_{i\rho}(\rho) ((y_i - \theta_i)^2 - \sigma_i^2) \right] = 0, \]  

(4.3)

where \( a_{i\rho}(\cdot) \) and \( b_{i\rho}(\cdot) \) are some nonstochastic functions of \( \rho \), the unknown parameters to be estimated, which will include, for example, the parameters \( \beta \) in the mean function and the additional parameter vector \( \alpha \) that may appear in the variance function. If \( \rho' = (\beta', \alpha') \) is a \( q \)-dimension vector then \( a_{i\rho}(\cdot) \) and \( b_{i\rho}(\cdot) \) are \( q \)-dimensional vectors which yield a \( q \)-dimensional estimating equation. This class includes (a) unweighted least squares where \( a_{i\rho}(\cdot) = \theta_i \partial / \partial \rho, \ b_{i\rho}(\cdot) = 0 \), and (b) QL estimation in which case \( a_{i\rho}(\cdot) = \theta_i \partial / \partial \rho \cdot 1 / \sigma_i^2, \ b_{i\rho}(\cdot) = 0 \). Setting \( b_{i\rho}(\cdot) \neq 0 \) will yield quadratic estimating equations (QEE), and a potential refinement to the QL approach. QL estimation is an appropriate approach when the variance specification is doubtful, whereas the quadratic approach is better if the variance specification is more certain and \( b_{i\rho}(\cdot) \neq 0 \) also. Cubic and higher order terms may be added when there is more information about higher moments but the practical usefulness of such extensions is uncertain.

If \( \theta_i \) and \( \sigma_i^2 \) are correctly specified, and \( \hat{\rho} \) is the solution to the QEE, then from the results of Crowder (1987) it is known that the estimator is consistent, asymptotically normal and with variance matrix \( A_n^{-1} B_n A_n^{-1}' \), where

\[ A_n = -\sum_{i=1}^{n} \left[ a_i \cdot \left( \frac{\partial \theta_i}{\partial \rho} \right)' + 2 \sigma_i b_i \cdot \left( \frac{\partial \sigma_i}{\partial \rho} \right)' \right], \]  

(4.4)

\[ B_n = \sum_{i=1}^{n} \sigma_i^2 \left[ a_i a_i' + \gamma_1 \left( a_i b_i' + b_i a_i' \right) + \sigma_i^2 (\gamma_2 + 2) \left( b_i b_i' \right) \right] \]  

(4.5)

where \( a_i \) and \( b_i \) are \((q \times 1)\) vectors, primes denote row vectors and \( \gamma_1 \) and \( \gamma_2 \) are skewness and kurthosis coefficients and all functions are evaluated at the solution of the estimating equations. For unweighted least squares and QL estimation in which \( b_i = 0 \), the asymptotic covariance matrix does not depend upon skewness or kurthosis parameters.

An alternative derivation of the same estimating equations functions comes by specifying an objective function \( G(\rho) = \sum_{i=1}^{n} [w_1^2(\rho) P_{1i}(y_i, x_i, \rho \mid x_i) + w_2^2(\rho) P_{2i}(y_i, x_i, \rho \mid x_i)] \), where \( P_{1i}(\cdot) \) and \( P_{2i}(\cdot) \) are first and second order orthogonal polynomials with respect to some specified probability density, and \( w_1 \) and \( w_2 \) are weights attached to the two terms. Suppose one selects the weights inversely proportional to the variance of the respective polynomials. Then the first term in \( G(\rho) \) is the squared first order orthonormal polynomial and the second term is the squared second order orthonormal polynomial.
A general expression for \( G(\rho) \) is then given by

\[
G(\rho) = \sum_{i=1}^{n} \left[ \frac{(y_i - \theta(x_i, \beta))^2}{\sigma_i^2} + \frac{(y_i - \theta(x_i, \beta))^2 - \mu_3 (y_i - \theta(x_i, \beta)) - \gamma_2}{(\gamma_2i + 2 - \gamma_1i) \sigma_i^2} \right],
\]

(4.6)

where \( \theta \) and \( \sigma^2 \) are the first and second moments of the distribution, \( \mu_3 \) is the third moment, and \( \gamma_1 \) and \( \gamma_2 \) are the skewness and excess kurtosis coefficients, respectively, for standardized errors. For generality all moments are allowed to vary with \( i \). The estimating equation for \( \rho \) is given by \( \partial G(.) \partial \rho = g(\rho) = 0 \). The asymptotic properties of the resulting estimator can be established by applying Amemiya’s (1985) general results on extremum estimators. Note that the first term in the objective function is the sum of weighted squared errors. Hence its minimization would correspond to QL estimation. The second term will be identically zero in some cases such as the LEF in which the variance function fully characterizes the distribution, making redundant the specification of higher order moments. However, in some cases efficiency gains will result from the inclusion of a correctly specified second term. An example of this follows.

Recently Dean and Lawless (1989) and Dean (1991), following Firth (1987), Crowder (1987), and Godambe and Thompson (1989), have discussed the estimation of the mixed Poisson model using EQL approach which employs the following specification of the first four moments of a mixed Poisson model, in contrast to the standard QL approach which employs only the first two:

\[
E(y_i \mid x_i) = \theta(x_i, \beta) = \theta_i, \quad \beta \in \mathbb{R}^k,
\]

(4.7)

\[
\text{var}(y_i \mid x_i) = \sigma_i^2 = \theta_i(1 + \alpha \theta_i), \quad \alpha \geq 0,
\]

(4.8)

\[
\gamma_1i = E \left[ \frac{(y_i - \theta_i)\sigma_i^{-1}}{\sigma_i} \right]^3 = \frac{1 + 2\alpha \theta_i}{\sigma_i},
\]

(4.9)

\[
\gamma_2i = E \left[ \frac{(y_i - \theta_i)\sigma_i^{-1}}{\sigma_i} \right]^4 - 3 = \frac{1}{\sigma_i^2} + 6\alpha,
\]

(4.10)

where \( \gamma_1 \) and \( \gamma_2 \) are skewness and kurtosis coefficients. Suppose the forms of the above equations is known, but one did not wish to use the negative binomial distribution, estimation could be based on the following optimal quadratic estimating equations:

\[
\sum_{i=1}^{n} \left( \frac{y_i - \theta_i}{\sigma_i^2} \right) \frac{\partial \theta_i}{\partial \beta} = 0,
\]

(4.11)

\[
\sum_{i=1}^{n} \left[ \frac{(y_i - \theta_i)^2}{(1 + \alpha \theta_i)^2} - \frac{(y_i - \theta_i)(1 + 2\alpha \theta_i)}{(1 + \alpha \theta_i)^2} \right] = 0.
\]

(4.12)
Given only the specification of the first four moments, these equations yield the most efficient estimator for \((\beta, \alpha)\). Given \(\alpha\), the solution of the first equation yields the QL estimate of \(\beta\). Other estimators for \(\alpha\), given \(\beta\) have been suggested in the literature; for example, Breslow’s (1984) estimator based on: \(\sum_i \left( y_i - \theta_i \right)^2 / \sigma_i^2 = 0 \). This latter estimator assumes \(\gamma_{1i} = \gamma_{2i} = 0\) and hence is less efficient than the previous one if the moment assumptions given earlier are correct. Dean and Lawless (1989) have evaluated the resulting loss of efficiency in a simulation context. Dean (1991) shows that the asymptotic variance of \(\beta\) is unaffected by the choice of the estimating equation for \(\alpha\). We are not currently aware of econometric applications of the QEE approach but a statistical application to the mixed Poisson-IG regression is in Dean, Lawless and Willmot (1989).

4.2. Estimation of time series models

Consider the time series count regression given in (3.30). If the interest is in parameter estimates, and not in predicting probabilities of counts, quasi-likelihood approach is the most direct alternative as it does not require the specification of the marginal distribution of \(y_t\), given its past history. This is the approach taken in Zeger and Qaqish (1988) and Zeger (1988). Zeger’s (1988) parameter-driven model assumes that \(\nu_t\) in (3.30) is an unobserved stationary process with \(E(\nu_t) = 1\), \(\text{var}(\nu_t) = \sigma_\nu^2\) and \(\text{cov}(\nu_t, \nu_{t+r}) = \sigma_\nu^2 \cdot \rho_\nu(r)\). The first two marginal moments are as given in (3.3)~(3.4), with subscript \(i\) replaced by \(t\), and the correlation at lag \(\tau\), \(\rho_y(t, \tau) \equiv \text{corr}(y_t, y_{t+r})\), is:

\[
\rho_y(t, \tau) = \frac{\rho_\nu(\tau)}{\left[ 1 + (\sigma_\nu^2 \theta_t)^{-1} \right] \left[ 1 + (\sigma_\nu^2 \theta_{t+r})^{-1} \right]^{1/2}}. \tag{4.13}
\]

Thus, the autocorrelated latent term \(\nu_t\) introduces both overdispersion and autocorrelation into a count variate \(y_t\). Zeger then generalizes the quasi-likelihood estimation method to time series of counts. In this procedure, only the first two moments of \(y_t\) are required to obtain consistent estimates.

The application of quasi-likelihood estimation is feasible but in general computationally awkward. Let \(V\) denote a non-diagonal \((n \times n)\) variance matrix of terms which depend on parameters (including \(\beta\), dispersion and correlation parameters), and \(D\) the \((n \times k)\) matrix with the \(ij\)-th element \(\partial \theta_i / \beta_j\) \((t = 1, \ldots, n; j = 1, \ldots, k)\). Then an obvious estimating equation, analogous to the estimating equation for a nonlinear generalized least squares problem, is

\[
D'V(\hat{\beta})^{-1} \left( y - \theta(x, \hat{\beta}) \right) = 0. \tag{4.14}
\]

If \(n\) is large then the inversion of \(V\) is computationally troublesome, unless a special structure is imposed on \(V\). Zeger (1988), for example, only uses information on the
first two autocorrelation coefficients of \( y \), which results in a tri-diagonal \( V \) which can be inverted using efficient numerical algorithms. Barron (1992) has used the estimator in a study of the founding rate of national labor unions. More work is needed on estimation of time series models with general variance functions.

Another advantage of this formulation is that \( y_{it} \) can be overdispersed, but not necessarily serially correlated. In an alternative approach, considered by Harvey and Fernandes (1989) and Brännäs and Johansson (1992) overdispersion and serial correlation occur simultaneously rather than independently.

4.3. Pseudo-MLE for random effects models

For random effects count models one can use pseudo maximum likelihood estimation methods which requires weaker distributional assumptions than maximum likelihood methods. The pseudo maximum likelihood (PML) and the quasi-generalized pseudo maximum likelihood (QGPML) methods of GMT (1984a) can be used for estimation of random effects count models (see Section 3.2 above). Assume that the true distribution of \( y_{it} \) is unknown but the first two moments of \( y_{it} \), conditional on \( x_{it} \), are given as follows:

\[
E(y_{it} \mid x_{it}) = \exp(x'_{it}\beta),
\]

\[
\text{var}(y_{it} \mid x_{it}) = \exp(x'_{it}\beta) \left[ 1 + \sigma_{\nu}^2 \exp(x'_{it}\beta) \right].
\]

Further, assume that the distribution for \( y_{it} \) belongs to the LEF class presented in Section 3.2:

\[
f(y_{it}) = \exp \left\{ [y_{it}\Theta_{it} - a(\Theta_{it}) + b(y_{it}, \phi)] \phi^{-1} \right\};
\]

Assume that observations on the response variable corresponding to individual \( i \) are uncorrelated over time. The pseudo likelihood equations are

\[
\sum_{i=1}^{n} \sum_{t=1}^{T} v_{it}^{-1} [y_{it} - \exp(x'_{it}\beta)] d_{it} = 0,
\]

where \( v_{it} = a''(\Theta_{it}) \) and \( d_{it} = \exp(x'_{it}\beta)x_{it} \) is a \( k \times 1 \) vector. The first order conditions can be conveniently expressed in matrix notation as follows: \( \sum_i d_i^t v_i^{-1} \varepsilon_t^* = 0 \). Here \( d_i \) is a \( T \times k \) matrix having \( d_{it} \) as its \( t \)-th row; \( v_i = \text{diag}(v_{it}) \) is a \( T \times T \) matrix; and \( \varepsilon_t^* \) is a \( T \times 1 \) vector with \( y_{it} - \exp(x'_{it}\beta) \) as its typical row. Further let \( \Omega_i = \text{diag}(\Omega_{it}) \). Note that (4.18) shows that the PML estimators differ only in their choice of weights. For some distributions which belong to (4.17), the weights \( (v_{it}) \) are given in Table 3.1.
<table>
<thead>
<tr>
<th>Family</th>
<th>$\phi$</th>
<th>Weights $v_{it}$</th>
</tr>
</thead>
<tbody>
<tr>
<td>Poisson $\sim \exp(x_{it}^t\beta)$</td>
<td>1</td>
<td>$\exp(x_{it}^t\beta)$</td>
</tr>
<tr>
<td>Normal $\sim \exp(x_{it}^t\beta), \sigma^2$</td>
<td>$\sigma^2$</td>
<td>1</td>
</tr>
<tr>
<td>Gamma $\sim (a, a^{-1}\exp(x_{it}^t\beta))$</td>
<td>$a^{-1}$</td>
<td>$\exp(2x_{it}^t\beta)$</td>
</tr>
<tr>
<td>Inv. Gaussian $\sim \exp(x_{it}^t\beta), \sigma^2$</td>
<td>$\sigma^2$</td>
<td>$\exp(3x_{it}^t\beta)$</td>
</tr>
</tbody>
</table>

Table 4.1: Examples of the LEF with Multiplicative Dispersion Parameter

Let $\hat{\beta}_p$ be the PMLE of $\beta$. It can be shown that the asymptotic covariance matrix of $n^{1/2}(\hat{\beta}_p - \beta)$ is

$$
\left[ \frac{1}{n} \sum_i d_i v_i^{-1} d_i \right]^{-1} \left[ \frac{1}{n} \sum_i d_i v_i^{-1} \Omega_i v_i^{-1} d_i \right] \left[ \frac{1}{n} \sum_i d_i v_i^{-1} d_i \right]^{-1}.
$$

Equations (4.18) and (4.19) do not involve $\phi$. This means that, even if $\phi$ is unknown, the estimation of $\phi$ is unnecessary for PMLE estimation of $\beta$ and the covariance matrix of $\hat{\beta}_p$. Of course, this holds true if the pseudo family is only mean adapted. The asymptotic covariance matrix in (4.19) may be estimated consistently by replacing $\beta$ and $\sigma^2_{\phi}$ by $\hat{\beta}_p$ and $\hat{\sigma}^2_{\phi}$, respectively. A consistent estimator of $\sigma^2_{\phi}$, $\hat{\sigma}^2_{\phi}$, based on (4.15) - (4.16) can be obtained as a regression coefficient from: $(y_{it} - \hat{\theta}_t)^2 - \hat{\theta}_t = \sigma^2_{\phi} \cdot \hat{\theta}_t^2 + \text{error}$, where $\hat{\theta}_t = \exp(x_{it}^t\hat{\beta})$. Alternatively, a consistent estimate of $\sigma^2_{\phi}$ may be obtained using the method of moments.

The QGPMLM approach assumes a pseudo distribution with a nuisance parameter, where this parameter can be related to the mean and variance of $y_{it}$. This estimator is consistent if both the mean and the variance are correctly specified. The QGPMLM estimator $\hat{\beta}_{sp}$ of $\beta$ based on (4.17) is the solution of the following equations:

$$
\sum_{i=1}^n \sum_{t=1}^T w_{it}^{-1} v_{it}^{-1} [y_{it} - \exp(x_{it}^t\beta)] d_{it} = 0,
$$

where $w_{it} = v_{it}^{-1}(\hat{\beta}) \cdot \exp(x_{it}^t\hat{\beta}) [1 + \hat{\sigma}^2_{\phi} \exp(x_{it}^t\hat{\beta})]$ is the (estimated) weight based on consistent estimates of $\beta$ and $\sigma^2_{\phi}$. The QGPMLM based on the normal was used by Hall et al (1986). This particular QGPMLM amounts to running non-linear least squares on: $y_{it} = \exp(x_{it}^t\beta) + \text{error}$, with GLS weights $w_{it} = \exp(x_{it}^t\hat{\beta}) [1 + \hat{\sigma}^2_{\phi} \exp(x_{it}^t\hat{\beta})]$. The QGPMLM based on the negative binomial model (with $\sigma^2_{\phi} \equiv \alpha$) is based on the following first order equations:

$$
\sum_{i=1}^n \sum_{t=1}^T \frac{[y_{it} - \exp(x_{it}^t\beta)] x_{it}}{[1 + \hat{\sigma}^2_{\phi} \exp(x_{it}^t\hat{\beta})]}.
$$

(4.21)
All the QGPMLE estimators (with multiplicative dispersion parameter $\phi$) based on (4.17) and negative binomial 2 distribution have the same asymptotic covariance matrix. The asymptotic covariance matrix of $n^{1/2} \left( \hat{\beta}_g - \beta \right)$ is

$$
\left[ \frac{1}{n} \sum_i d_i' \Omega_i^{-1} d_i \right]^{-1}.
$$

(4.22)

This covariance matrix is smaller than that of PMLE given in (4.19). This follows by extending the proof of GMT (1984a) to the case of panel data.

The above framework can be extended to situations where observations for individual $i$ are correlated over time. Liang and Zeger (1986) have investigate extensions of correlated random effects models to GLM family. The application of quasi-likelihood estimation is feasible but in general computationally demanding. Again assume that the first two moments of $y_{it}$ are as given in (4.15)–(4.16). Let $\Omega_i^*(\beta, \sigma^2, \rho) \equiv \Omega_i^*$ denote a non-diagonal $(T \times T)$ covariance matrix of $y_i$, where $\rho$ is a vector of correlation parameters. An obvious estimating equation is

$$
\sum_i d_i' (\Omega_i^*)^{-1} \epsilon_i^* = 0.
$$

(4.23)

Let $(\hat{\sigma}_i^2, \hat{\rho})$ be a consistent estimator of $(\sigma^2, \rho)$. Consequently the estimating equation for $\beta$ takes the form $\sum_i d_i' [\Omega_i^*(\beta, \hat{\sigma}_i^2, \hat{\rho})]^{-1} \epsilon_i^* = 0$. To simplify computation, Liang and Zeger (1986) suggest replacing $\Omega_i^*$ by $D_i^{1/2} R(\rho^*) D_i^{1/2}$, where $D_i = \text{diag}(\text{a"}{(\Theta_{ii})})$ and $R(\rho^*)$ is a 'working' correlation matrix and $\rho^*$ is a vector of parameters which characterizes it. They discuss various choices for $R(\rho^*)$, including the case where it remains unspecified. In this case $R(\rho^*)$ may be estimated by $\frac{1}{\phi n} \sum_i D_i^{-1/2} \epsilon_i^* \epsilon_i^{*\prime} D_i^{-1/2}$.

5. MODEL EVALUATION PROCEDURES

Once a model has been estimated, it might be evaluated in terms of the properties of its residuals, the goodness-of-fit, and shortcomings in specific directions. Analogously with the normal linear regression model, we consider residual analysis, $R^2$ type fit measures and specification tests for model deficiency in specific directions. The intrinsically non-linear and heteroskedastic nature of the model is responsible for an absence of a unique measure that is an exact counterpart of linear least squares residuals.

5.1. Residual analysis and goodness-of-fit measures

Since the Poisson regression model is a LEF, the apparatus developed for the GLM (McCullagh and Nelder 1989) can be exploited. An important concept is the deviance
measure based on the logarithm of the ratio of likelihoods of the 'null model' and the 'full model'; the former attributes all variation in the data to the random term, where the latter has \( n \) parameters, one for each data point and serves as a benchmark for model evaluation. For the Poisson family, the deviance measure is defined as \( 2 \sum \{ y_i \log (\hat{\theta}_i) - (y_i - \hat{\theta}_i) \} \); when the regression specification includes an intercept, the \( \sum (y_i - \hat{\theta}_i) \) term is zero. This measure is then the same as the \( G^2 \) measure used in the general linear models literature. Writing \( G^2 = 2 \sum d_i \) where \( d_i \equiv \{ y_i \log (\hat{\theta}_i) - (y_i - \hat{\theta}_i) \} \) suggests that we interpret \( d_i \) as the \( i \)th deviance residual. For Poisson regression with an intercept \( G^2 = 2 \sum y_i \log (\hat{\theta}_i) \), and the zero value implies a perfect fit.

Scaling the 'raw' residual \( (y_i - \hat{\theta}_i) \) by its standard deviation yields the so-called Pearson residual, denoted \( r_P \), which in the Poisson regression will be \( r_P = \frac{(y_i - \hat{\theta}_i)}{\sqrt{\hat{\theta}_i}} \). Note that \( \sum r_{i,p}^2 \equiv X^2 \) is the Pearson chi-square goodness of fit statistic\(^9\). If the Poisson moment restriction is valid, then \( \sum \frac{(y_i - \hat{\theta}_i)^2}{\hat{\theta}_i} \rightarrow \theta_i \); hence the Pearson goodness of fit measure will tend to \( n \).

The asymptotic distribution of deviance and the Pearson measures are poorly approximated by the chi-squared distribution; the difference in the deviance, nevertheless, provides a satisfactory basis for testing the significance of the difference between nested pairs of Poisson models (McCullagh and Nelder 1989).

A second goodness-of-fit measure is the \( R^2 \) measure, analogous to that used in the standard linear regression. For the Poisson regression such a measure might be constructed using one of a number of available definitions of a residual, e.g., the Pearson residuals for the Poisson model, denoted \( R^2_P \). Cameron and Windmeijer (1993) have suggested the measure \( R^2_P = 1 - \frac{\sum (y_i - \hat{\theta}_i)^2/\hat{\theta}_i}{\sum (y_i - \bar{y})^2/\bar{y}} \), \( R^2_P \leq 1 \), which is based on the ratio of the standardized fitted variance and the standardized raw variance. This measure has the undesirable feature that, in small samples, it may be negative and decrease as explanatory variables are added. They also suggest a measure based on deviance residuals, viz., \( R^2_D = \frac{\sum (y_i \log (\hat{\theta}_i) - (\hat{\theta}_i - \bar{\theta}))}{\sum y_i \log (\frac{y_i}{\hat{\theta}})} \), \( 0 < R^2_D \leq 1 \). The deviance based \( R^2_D \) has an information-theoretic (Kullback-Leibler distance) interpretation and the property that it increases as regressors are added to the model.

It would be desirable to generalize the residual and goodness of fit measures to cover count models which are not LEF. The negative binomial family is a LEF only when the overdispersion parameter is given, in which case a deviance residual can be defined. For this case, Cameron and Windmeijer (1993) define \( R^2 \) measures analogous to those for

---

\(^9\)Some of the commonly used PC packages, such as LIMDEP, provide the \( G^2 \) and \( X^2 \) statistics as part of the output for Poisson regression models.
the Poisson, but they show that when an estimate of the overdispersion parameter is inserted in the definition, it does not have the property that its value must increase as regressors are added. For truncated and censored models, the raw residuals are replaced by generalized residuals (Gourieroux et al 1987). For example, for the zero truncated Poisson model, the generalized residual of the form \( y_i - E[y_i | y_i > 0] \) was given in Section 3.5. While these quantities play an important role in estimation and inference, measures of goodness of fit based on them seem not currently available.

5.2. Specification tests

In modeling empirical data it is desirable to supplement estimation with specification tests to determine whether the fitted model is adequate. There are a number of relevant general issues. First, what kinds of model restrictions are commonly most appropriate to test and what is the most convenient framework for doing so? Second, how do we computationally implement these tests? We shall not discuss standard variable addition and variable deletion tests since the likelihood ratio and Wald tests can be readily applied to these cases. As previously emphasized, departures from 'Poissonness', in the form of overdispersion and excess zeros, are important features to test for. The functional form of the conditional mean function, and possible presence of covariates in the conditional variance function are also important issues for specification tests. In the context of multivariate count models, tests of stochastic independence of counts are of interest.

A useful framework for carrying out most of these tests is the conditional moment (CM) (Newey 1985; Tauchen 1985; Pagan and Vella 1989) approach since it embraces both the likelihood based approach suitable for fully parametric models, and the moments based models. We initially focus on the case where \( y_i \) is a scalar response variable and assume that \( \{y_i\} \) are independently distributed. The true data generation process (dgp) for \( y_i \), given \( x_i \), a vector of explanatory variables, is unknown. Statistical inference is based on an assumed parameterized density \( h(y | \gamma) \), where \( \gamma \) is a \( p \times 1 \) vector of parameters with population value \( \gamma_0 \), or an assumed parameterization of moments as in generalized method of moments (GMM) estimation. CM tests are tests of the validity of moment conditions implied by these assumed parameterizations. Under the null hypothesis, a CM test satisfies the moment condition:

\[
E[m(y_i, x_i, \gamma_0)] = 0,
\]

or more generally \( E[\varphi(x_i)m(y_i, x_i, \gamma_0) = 0] \), where \( m(.) \) is a \( q \)-vector function.

The CM test is based on the magnitude of the test statistic

\[
\hat{\tau} = n^{-1} \sum_i [m(y_i, x_i, \hat{\gamma})],
\]
where \( \hat{\gamma} \) is the estimator of \( \gamma \). This estimator satisfies the condition that \( n^{-1} \sum g_i(\hat{\gamma}) = 0 \). When \( \hat{\gamma} \) is the maximum likelihood estimator, \( g_i(\gamma) \) is the score of the log-likelihood function. The appropriate test statistic is of a general form: \( T = \hat{\tau}'V^{-1}\hat{\tau} \), which under the null hypothesis has asymptotic \( \chi^2(q) \) distribution\(^\text{10}\). A simple regression-based test statistic can be obtained as: \( nR_i^2 \), where \( R_i^2 \) is the uncentered coefficient of determination from the artificial regression of unity on \( m_i(y_i, z_i; \hat{\gamma}) = \tilde{m}_i \) and \( g_i(\hat{\gamma}) = \tilde{g}_i \). This test based on the outer product forms of the gradient and the moment conditions, henceforth referred to as OPGM test, will also be \( \chi^2(q) \). A major disadvantage of the OPGM test is its poor small sample properties. This is due to the poor correspondence between the nominal and the actual size of the test when the sample size is small; i.e., the test statistic based on the OPGM procedure converges to \( \chi^2(q) \) very slowly. Various tests discussed in subsequent sub-sections differ by the choice of moment conditions used in (5.1) and the methods used in estimating \( \gamma \) and \( V \). Subsequent subsections consider a number of diagnostic tests that are employed in modeling count data.

5.3. Tests of the conditional mean

The starting point in conditional mean testing is to specify a parametric model for the conditional mean so that the null hypothesis of interest is \( H_0 : E(y_i \mid x_i) = \theta(x_i; \beta), \beta \in \mathbb{R}^k \). The most commonly employed specification for the mean is \( \theta(x_i; \beta) = \exp(x_i'\beta) \). We initially assume the existence of an alternative general mean model. A test for the adequacy of the exponential specification can then be based on the Lagrange multiplier (LM) or score test. For concreteness, consider the test of the adequacy of the exponential specification \( \theta_i = \exp(x_i'\beta) \) against the generalized mean function based on the Box-Cox transformation:

\[
\theta_{\lambda i} = (1 + \lambda x_i'\beta)^{1/\lambda}, \tag{5.3}
\]

where \( \lambda \) denotes a shape parameter measuring the sensitivity of the mean to variation in the linear predictor \( x_i'\beta \). This specification reduces to \( \theta_i = \exp(x_i'\beta) \) when \( \lambda = 0 \).

Now, consider the score test for the adequacy of the Poisson mean specification. Let \( \hat{\theta}_i = \exp(x_i'\hat{\beta}) \) and \( \hat{\rho} = (\hat{\beta}' 0)' \), where \( \hat{\beta} \) is the MLE of \( \beta \). The score test statistic for the adequacy of a Poisson mean specification (2.3) against (5.3) is

\[
T_1 = \left[ \sum_i (y_i - \hat{\theta}_i)^2 \hat{z}_i \right]^2 \cdot I_{\lambda i}^{-1}(\hat{\rho}) \sim \chi^2(1), \tag{5.4}
\]

where \( \hat{z}_i = \frac{1}{\lambda} \cdot \log^2(\hat{\theta}_i) \), and \( I_{\lambda i}(\hat{\rho}) = \left[ \sum_i \tilde{\theta}_i \tilde{z}_i \tilde{z}_i' \right] - \left[ \sum_i \tilde{\theta}_i \tilde{z}_i x_i \right] \left[ \sum_i \tilde{\theta}_i x_i x_i' \right]^{-1} \left[ \sum_i \tilde{\theta}_i x_i \tilde{z}_i' \right] \).

In terms of (5.3), this test is based on the statistic \( \hat{\tau} = n^{-1} \sum_i (y_i - \hat{\theta}_i)\hat{z}_i \). An equivalent

\(^{10}\)Issues related to estimation of the covariance matrix of \( \hat{\tau}, V \), are discussed in Pierce (1983), Newey (1985), Tauchen (1983), and Pagan and Vella (1989).
regression-based test statistic for the mean test in the Poisson model can be obtained as $nR^2_U$, where $R^2_U$ is the R-squared from the artificial regression of 1 on $\hat{m}_i = \hat{\epsilon}_i z_i$ and $\hat{\gamma}_i = \hat{\epsilon}_i x_i$, where $\hat{\epsilon}_i = y_i - \hat{\theta}_i$.

The above approach can be extended to the Negbin2 and double Poisson models, which are empirically more interesting. For the Negbin2 model based on (3.8), (5.2) becomes $\hat{\tau} = n^{-1} \sum_i \left(1 + \tilde{\alpha} \hat{\theta}_i\right)^{-1} \tilde{z}_i \hat{\epsilon}_i$. In this case, the test can be implemented as $nR^2_U$, based on the regression of unity on $\hat{m}_i = \left(1 + \tilde{\alpha} \hat{\theta}_i\right)^{-1} \tilde{\epsilon}_i \tilde{z}_i$ and $\hat{\gamma}_i = \left(1 + \tilde{\alpha} \hat{\theta}_i\right)^{-1} \tilde{\epsilon}_i x_i$.

The presence of an additional parameter $\alpha$ does not affect the asymptotic distribution of the resulting test statistic since the information matrix between the mean parameter $\beta$ and the dispersion parameter $\alpha$ in Negbin2 is block-diagonal. Similarly, the conditional mean test proposed for the regular Poisson model can be extended to cases where the count variable is truncated or censored by suitably modifying (5.4). For the left truncated Poisson model, the appropriate mean test is based on the quantity: $\hat{\tau} = n^{-1} \sum_i \tilde{z}_i (y_i - \hat{\theta}_i - \delta_i)$, where $\delta_i$ is the adjustment factor, with $\hat{m}_i = (y_i - \hat{\theta}_i - \delta_i) \tilde{z}_i$ and $\hat{\gamma}_i = (y_i - \hat{\theta}_i - \delta_i) x_i$.

One problem with the above LM test is that it is not robust to conditional variance misspecification. Wooldridge (1991a,b) considered robust specification tests for models estimated by quasi-likelihood methods using density in the linear exponential family. Wooldridge (1991a) illustrates this 'robust' testing procedure for the Poisson model based on the general mean function (5.3). Let $\hat{\theta}_i = \exp(\tilde{x}_i \hat{\beta})$ and $\hat{\epsilon}_i = y_i - \hat{\theta}_i$, where now $\hat{\beta}$ is the Poisson Quasi-MLE. The robust conditional mean test for the Poisson regression model can be implemented by running the following artificial regressions: (i) run the regression $\sqrt{\hat{\theta}_i} \cdot \log(\hat{\theta}_i)$ on $\sqrt{\hat{\theta}_i} x_i$ and save the residuals, say $\tilde{r}_i$; (ii) run the regression of 1 on $\hat{\epsilon}_i \tilde{r}_i / \sqrt{\hat{\theta}_i}$ and compute $R^2_U$. The resulting test statistic is $nR^2_U$ which is asymptotically $\chi^2(1)$.

These robust tests are not feasible in situations where observations are truncated or censored. The problem is that the conditional mean in truncated or censored models depends on the entire distribution of $y_i$, through the presence of the adjustment factor $\delta_i$.

Diagnostic tests are also available in situations where one does not explicitly specify an alternative mean model. Wooldridge (1991a) has suggested Hausman test for conditional mean specification. He gives count model examples based on comparisons of estimators: (i) Poisson Quasi-MLE versus non-linear least squares estimator in a panel data setting; and (ii) Poisson MLE versus Geometric MLE in a cross section data setting. However, it is not obvious that these tests can always be interpreted as conditional mean tests. A related test, which may interpreted as a test of overidentifying restriction, can be based on comparing the Poisson Quasi-MLE estimator with the Instrumental Variable (GMM) estimator proposed by Mullany(1993). Hausman tests and
GMM J-tests are computed for an empirical application in Mullahy (1993).

5.4. Tests for misspecification of the variance function

Unlike that of the normal regression model, the variance of the Poisson and other count models depends on the mean $\theta(x_i, \beta)$ and hence is not constant. These models are intrinsically heteroskedastic and, as argued by Pagan and Pak (1991), what needs to be tested in these types of models is not the presence of heteroskedasticity per se but whether it departs from that featured in the maintained model. This section examines two types of tests for misspecification of the variance function in some widely used count models. The first variant is a test of overdispersion in Poisson models. In these models, the variance depends only on the mean parameter $\beta$. The second variant arises when the variance of $y_i$ of the maintained model depends additionally on a dispersion parameter as in the double Poisson or negative binomial models.

5.4.1. Tests for overdispersion in Poisson models

Overdispersion (conditional variance greater than the mean) in Poisson models may arise due unobserved heterogeneity or with clustering of samples. It may also occur when the interval length of the Poisson process is random. Numerous tests for detecting overdispersion in the regular as well as the truncated or censored Poisson models have been proposed; the most common being LM-based tests obtained by embedding the Poisson distribution in more general models such as the negative binomial family, the Katz family of distributions, and local alternatives to the Poisson\(^{11}\). In terms of the setup in (5.2), a test for overdispersion in the Poisson regression models is usually based on the following quantity:

$$\hat{\tau} = n^{-1} \sum_i \hat{\omega}_i[(y_i - \hat{\theta}_i)^2 - y_i],$$

(5.5)

where normally $\hat{\theta}_i = \exp(x'_i \hat{\beta})$, $\hat{\omega}_i = w(x_i, \hat{\beta})$ a weighting function and $\hat{\beta}$ is a consistent estimate of $\beta$.

The most common score test statistic for detecting overdispersion in the Poisson model based on more general overdispersed alternative models can be summarized as:

$$T_2 = \frac{\sum_i \hat{\omega}_i[(y_i - \hat{\theta}_i)^2 - y_i]}{\left[2 \sum_i \hat{\omega}_i^2 \sigma^2_i\right]^{1/2}} \sim N(0, 1)$$

(5.6)

\(^{11}\)Tests proposed for the regular Poisson regression model include Cox (1983), Collings and Margolin (1985), Cameron and Trivedi (1986), Lee (1986), Dean and Lawless (1989), and Cameron and Trivedi (1990a). These have also been extended to testing for overdispersion in truncated and censored Poisson regression models (Gurmu 1991; Gurmu and Trivedi 1992; Gurmu 1992).
under $H_0$. In this case, the choice of $w_i$ depends on the model used under the alternative hypothesis; for example, $\hat{w}_i = 1$ for Negbin2 and Cox's local approximation to the Poisson and $\hat{w}_i = \hat{\theta}_i^{-1}$ for Negbin1 alternative. The advantage of the score test is that the model need only be estimated under the null hypothesis. The Wald and likelihood ratio tests have been used less frequently mainly due to their inapplicability when testing on the boundary of the parameter space; see, for example Lawless (1987b). The information matrix (IM) test has also been used as a valuable diagnostic tool in the analysis of the Poisson count data models (Lee 1986; Mullahy 1986; Cameron and Trivedi 1990a). In particular, the IM test based on the intercept term of the Poisson regression is also a test for overdispersion or underdispersion in the Poisson model.

Cameron and Trivedi (1990a) have proposed various regression-based tests for overdispersion in the Poisson model that require specification of only the mean-variance relationship under the alternative, rather than the complete distribution whose choice is usually arbitrary. The testing framework is based on $H_0$: \( \text{var}(y_i \mid x_i) = \theta_i \) versus $H_1$: \( \text{var}(y_i \mid x_i) = \theta_i + \alpha g(\theta_i) \), where $g(\theta_i)$ is some specified function of $\theta_i$; they suggest using $g(\theta_i) = \theta_i$ or $\theta_i^2$. Their optimal test can be implemented as a $t$-test from the OLS regression of \( [(y_i - 1/\hat{\theta}_i)^2 - \bar{y}] / \sqrt{2\hat{\theta}_i} \) on $g(\hat{\theta}_i)/\sqrt{2\hat{\theta}_i}$, where $\hat{\theta}_i = \theta(x_i, \hat{\beta})$ is evaluated at a consistent estimator $\hat{\beta}$ of $\beta$.

Other tests for overdispersion have also been proposed from a different perspective. One such possibility is a test based on hurdle count models where overdispersion or underdispersion are viewed as arising from a misspecification of the maintained parent dgp in which the relative probabilities of zero and positive realizations implied by the parent distribution are not supported by the data. Mullahy (1992) suggested a goodness-of-fit test for overdispersion in Poisson regression models. The strategy of this test is based on comparisons of the observed zero outcomes and the proportion expected in the zero cell under the null model. This strategy in turn relies on the fact that the actual proportion of zero outcomes in an arbitrarily (overdispersed) mixed Poisson model tend to exceed the proportion expected under a Poisson null. Letting $1(.)$ be the 0-1 indicator function, Mullahy suggested basing a goodness-of-fit test on the statistic: $\hat{\tau} = n^{-1} \sum_i \left[ 1(y_i = 0) - \exp(-\hat{\theta}_i) \right]$. Simulation results reported by him indicate that the proposed goodness-of-fit test tend to have better size properties when the mean of the dependent variable is small.

The LM approach can be extended to testing for overdispersion in truncated and censored Poisson regression models. In this case, overdispersion can be viewed as a phenomenon whereby the actual variance of $y_i$ exceeds the nominal variance of truncated or censored Poisson\textsuperscript{12}. Consider, for example, the left truncated Poisson model with truncation point at $r$ based on (3.17) and (2.4). Using the notation developed in Section

\textsuperscript{12}See McCullagh and Nelder (1989, P. 124) for a general definition of overdispersion.
3.5. Tests for this model can be based on the statistic

\[ \hat{\tau} = n^{-1} \sum \hat{w}_i [(y_i - \hat{\theta}_i)^2 - y_i - \hat{\delta}_i (r - 1 - \hat{\theta}_i)], \]  

(5.7)

where all arguments have been evaluated at, say, MLE \( \hat{\beta} \) and, for the score test based on truncated Negbin1 and Negbin2 alternatives, \( \hat{w}_i = \hat{\theta}_i^{-1} \) and 1, respectively. The expression \( \hat{\delta}_i (r - 1 - \hat{\theta}_i) \) constitutes an adjustment to be made to (5.5) due to sample truncation. The test for the widely used positive Poisson model is obtained when \( r = 1 \). The OPGM version of the test can be implemented by computing \( n R^2_{1i} \), where \( R^2 \) is the uncentered coefficient of determination from the regression of unity on \( \hat{m}_i = \hat{w}_i [(y_i - \hat{\theta}_i)^2 - y_i - \hat{\delta}_i (r - 1 - \hat{\theta}_i)] \) and \( g_i(\hat{\beta}) = (y_i - \hat{\theta}_i - \hat{\delta}_i) x_i^T \). The gradient vector \( g(.) \) is based on (3.19) with \( x_i^T = [\theta_i^{-1} \frac{\partial g}{\partial \beta}]_{\| \beta = \hat{\beta}} \).

As argued by a Referee, another important issue is to perform specification test for Poisson regression models and not only tests for overdispersion. The IM test can be used to perform specification tests, including a test for overdispersion, for Poisson models. Another possibility is to use Lee’s (1986) LM tests based on Pearson family or series expansion. These test statistics utilized up to third order moments. The polynomial approach can in principle be used to derive a test of higher order moment restrictions imposed by the Poisson model.

5.4.2. Variance tests in mixed Poisson models

It is of interest to apply tests for the variance assumption in mixed Poisson regression models. For instance, one might be interested to see whether the dispersion parameter in the variance function depends on exogenous variables. Consider testing for ‘heteroskedasticity’ in the double Poisson regression model with dispersion parameter \( \phi \); see Gurmu and Trivedi (1993). The approximate double Poisson model is based on (3.9) in which the dispersion parameter \( \phi \) may depend upon a set of \((q + 1 \times 1)\) vector of exogenous variables \( z_i \), the first element assumed to be unity, according to \( \phi_i = h(z_i^T \xi) \).

Here \( \xi \) is a \((q + 1 \times 1)\) parameter vector and \( h(.) \) is assumed to possess first and second order derivatives. Letting \( z_i^T \xi = \xi_0 \) so that \( h(\xi_0) = \sigma^2 \) is constant under the null hypothesis, the variance test for the double Poisson model is: \( H_0: \xi_1 = \xi_2 = \ldots = \xi_q = 0 \) or, alternatively, \( H_0 : \text{var}(y_i \mid x_i) = \sigma^2 \theta_i \). The specification \( \phi_i = h(z_i^T \xi) \) is similar to the one used by Breusch and Pagan (1979) for a test of heteroskedasticity in the linear regression model. An example of this representation is the logistic specification: \( \phi_i = 1/[1 + \exp(-z_i^T \xi)] \) suggested by Efron (1986).

The score test for heteroskedasticity in the double Poisson model is based on the statistic:

\[ \hat{\tau} = n^{-1} \sum \left[ \hat{\sigma}_i^{-2} \hat{d}_i - 1 \right] z_i, \]  

(5.8)
where \( \hat{d}_i = 2 \left[ y_i \cdot \log(y_i/\hat{\theta}_i) - (y_i - \hat{\theta}_i) \right] \) is the estimated deviance of the double Poisson model. This test can be obtained as one-half of the explained sum of squares from the regression of \( D \) on \( Z \), where \( D \) is an \((n \times 1)\) vector with typical element \((1 - \hat{\sigma}^{-2}\hat{d}_i)\) and \( Z \) is an \((n \times (q + 1))\) matrix with typical row \( z_i^t \).

Cameron (1991) proposed regression-motivated tests of 'heteroskedasticity' in models where the variance depends on the mean. The proposed tests require specification of the first two moments under the null and alternative hypotheses but, when the full distribution is specified, the optimal regression-based tests coincide with score tests for examples based on the LEF. Consider the application of these tests to overdispersed count models with conditional mean \( E(y_i \mid x_i) = \theta_i(x_i, \beta) \equiv \theta_i \) and conditional variance \( \text{var}(y_i \mid x_i) = v(\theta_i, \sigma^2) \). The tests involve testing the null hypothesis \( H_0: \text{var}(y_i \mid x_i) = v(\theta_i, \sigma^2) \) against the specific alternative \( H_1: \text{var}(y_i \mid x_i) = v(\theta_i, \sigma^2) + u(\theta_i, \sigma^2)'\xi \) where \( \xi \) is a \((q \times 1)\) parameter vector. In this case, the regression-based test can be obtained by testing the significance of \( \xi \) in the regression

\[
y_i^* = u(\theta_i, \sigma^2)'\xi + \nu_i, \tag{5.9}
\]

where \( y_i^* = (y_i - \theta_i)^2 - \partial v(\theta_i, \sigma^2) / \partial \theta_i \cdot (y_i - \theta_i) - v(\theta_i, \sigma^2) \) and \( \nu_i \) is a heteroskedastic error term. In implementation, \( \beta \) and \( \sigma^2 \) will be replaced by their consistent estimates \( \hat{\beta} \) and \( \hat{\sigma}^2 \). Finite sample properties of variance tests in overdispersed count models is yet to be investigated.

5.5. Tests of independence

In parametric multivariate count models it is sometimes of interest to apply tests of independence conditional on exogenous variables. One possible situation, illustrated by Cameron et al (1988) and King (1989a), involves a set of seemingly unrelated Poisson or negative binomial regressions with a nondiagonal contemporaneous covariance matrix. How can one test for dependence of errors in such models? Another situation, exemplified by Meghir and Robin (1992), involves a structural equation (e.g. expenditure) with an endogenous count variable as an explanatory variable (e.g. frequency of purchase) whose marginal distribution may be specified. Should the latter be treated as exogenous? Again there is reason for testing independence. Finally, in time series count models a specification test for serial independence is desirable (Li 1991).

Unfortunately, except in special cases such as the bivariate Poisson or the negative binomial it is often not possible to express such distributions in a flexible closed form. Wald and likelihood ratio type procedures are then not feasible for testing either the hypothesis of independence or the restricted hypothesis of zero correlation. The null of zero correlation is the usual starting point for testing independence, but in nongaussian models this is, in general, only a necessary, not sufficient, condition for independence. Using the key idea that under independence the joint pdf factorizes into a product
of marginals, Cameron and Trivedi (1993) developed score type tests of independence based on a series expansion of the unknown joint pdf of the observations. The leading term in the series is the product of the marginals; the remaining terms in the expansion are orthonormal polynomials of the univariate marginal densities. The idea behind the test is to measure the significance of the higher order terms in the expansion using estimates of the marginal models only.

In the univariate case, the orthogonal polynomial sequence \( \{P_n(y)\} \) for the random variable \( y \) with regular pdf \( f(y) \) satisfies the orthogonality condition: \( E[P_n(y)P_m(y)] = \delta_{mn}k_n, \ k_n \neq 0 \), where \( \delta_{mn} \) is the Kronecker delta, \( \delta_{mn} = 0 \) if \( m \neq n \), \( \delta_{mn} = 1 \) if \( m = n \). In the special case of an orthonormal polynomial sequence, \( k_n = 1 \). An orthonormal polynomial is derived from an orthogonal polynomial by dividing by its standard deviation.

For the specific cases of Poisson, Negbin1 (with overdispersion parameter \( \alpha \)) and Negbin2 (with overdispersion parameter \( \delta \)) the first order orthogonal polynomial is simply \( P_1(y \mid X) = y - \mu \); the second order polynomials, \( P_2^\alpha(y \mid X) \), are, respectively, \( (y - \mu)^2 - y \) (Poisson), \( (y - \mu)^2 - (2\alpha - 1)(y - \mu) - \alpha \mu \) (Negbin1), and \( (y - \mu)^2 - (1 + 2\delta \mu)(y - \mu) - (1 + \delta \mu)\mu \) (Negbin2), where \( \mu = E(y \mid X) = \exp(X^\beta) \).

Now, let \( g(y_1, y_2) \) be a bivariate p.d.f. of continuous random variables \( y_1 \) and \( y_2 \) with respective marginal distributions \( f_1(y_1) \) and \( f_2(y_2) \) whose corresponding orthonormal polynomial sequences are, respectively, \( Q_n(y_1) \) and \( R_n(y_2) \), \( n = 0, 1, \ldots \). Under mild regularity conditions the following series expansion is formally valid:

\[
g(y_1, y_2) = f_1(y_1) \cdot f_2(y_2) \left[ 1 + \sum_{n=1}^{\infty} \sum_{m=1}^{\infty} \rho_{nm}Q_n(y_1)R_m(y_2) \right], \tag{5.10}
\]

where \( \rho_{nm} = E[Q_n(y_1)R_m(y_2)] = \int \int Q_n(y_1) \cdot R_m(y_2) \cdot g(y_1, y_2) \cdot dy_1 \cdot dy_2 \); see Lancaster (1969, p.97, Theorem 4.1) for a proof, and Kocherlakota and Kocherlakota (1993).

In general, a test of independence in the bivariate case requires us to test \( H_0: \rho_{nm} = 0 \) (all \( n, m \)). This onerous task may be simplified in one of two ways. The null may be tested against an alternative in which dependence is restricted to be a function of a small number of parameters, usually just one. Or we may approximate the bivariate distribution by a series expansion with a smaller number of terms and then derive a score (LM) test of the null hypothesis \( H_0: \rho_{nm} = 0 \) (some \( n, m \)). For independence we require \( \rho_{nm} = 0 \) for all \( n \) and \( m \). By testing only a subset of the restrictions, the hypothesis of approximate independence is tested. If \( p = 2 \), this is equivalent to the null hypothesis \( H_0: \rho_{11} = \rho_{22} = \rho_{12} = \rho_{21} = 0 \). For general \( p \), the appropriate moment restriction is: \( E_0[Q_n(y_1)R_m(y_2)] = 0, n, m = 1, 2, \ldots, p \), where \( E_0 \) denotes expectation under the null hypothesis of independence of \( y_1 \) and \( y_2 \).

For orthogonal polynomials, a test of the null hypothesis of \( H_0: \rho_{nm} = 0 \) is based
on:

\[ r_{nm}^2 = \left( \frac{\sum_{t=1}^{T} Q_{n,t}^* R_{m,t}^*}{\sum_{t=1}^{T} \left( Q_{n,t}^* R_{m,t}^* \right)^2} \right)^{-1} \left( \sum_{t=1}^{T} Q_{n,t}^* R_{m,t}^* \right) \overset{\sim}{\sim} \chi^2(1). \]  (5.11)

Note that \( r_{nm}^2 \) can be computed as \( T \) times the uncentered \( R^2 \) (equals the proportion of the explained sum of squares) from the auxiliary regression of 1 on \( Q_{n,t}^* R_{m,t}^* \). For orthonormal polynomials, a test of the null hypothesis of \( H_0: \rho_{nm} = 0 \) is

\[ r_{nm}^2 = T \cdot \left( \sum_{t=1}^{T} Q_{n,t} R_{m,t} \right) \left( \left( \sum_{t=1}^{T} Q_{n,t}^2 \right) \left( \sum_{t=1}^{T} R_{m,t}^2 \right) \right)^{-1} \left( \sum_{t=1}^{T} Q_{n,t} R_{m,t} \right) \overset{\sim}{\sim} \chi^2(1). \]  (5.12)

The conditional moment test of independence consists of testing for zero correlation between all pairs of orthonormal polynomials. So the steps are as follows: First, specify the marginals and estimate their parameters. Then evaluate the orthonormal polynomials at the estimated parameter values. Finally calculate the tests.

Cameron and Trivedi's Monte Carlo investigations of these tests for bivariate count regression models show that the tests have the correct size and high power when the marginals are correctly specified but they over reject when the marginals are misspecified.

5.6. Size-adjusted tests

The tests outlined in the preceding sections are based on asymptotic theory. However, from the practical point of view we are most interested in finite sample properties of various test statistics. A number of studies reported above have investigated the size and power properties of various tests in count models and have shown that the actual size of the tests is less than the nominal size based on asymptotic theory. In particular, the OPREG variant of the LM or IM test shows a poor correspondence between the actual and the nominal size of a given test statistic. The intuition of a moment-based test comes from the fact that, under the null hypothesis, \( E(\hat{\tau}) = 0 \). The problem in small sample behavior of the test is that, under the null hypothesis, \( E(\hat{\tau}) \) is no longer zero in finite samples. This has motivated many authors to propose size-adjusted test statistics so that \( \hat{\tau} \) will be centered on zero in finite samples. A required adjustment in this case will be based on \( \hat{\tau}^* = [\hat{\tau} - E(\hat{\tau})] \) so that, instead of (5.2), the size-adjusted test statistic will now be based on the magnitude of \( \hat{\tau}^* = n^{-1} \sum [m(y_i, x_i, \gamma) - E(m(y_i, x_i, \gamma))] \). The basic rational is that \( \hat{\tau}^* \) has the same asymptotic properties as \( \hat{\tau} \), but provides better distributional approximations in small samples.

Dean and Lawless (1989) have proposed size-corrected score tests for overdispersion in the regular Poisson model. The approach have also been extended to tests of overdispersion in truncated and censored Poisson models (Gurmu and Trivedi 1992). For the
regular Poisson case, consider the unadjusted score test for overdispersion based on the statistic $\tilde{\tau} = n^{-1} \sum_i \left[ (y_i - \hat{\theta})^2 / y_i \right]$. Using the exponential specification $\theta_i = \exp(x_i' \beta)$, the size-adjusted test is based on the statistic:

$$\tilde{\tau}^* = n^{-1} \sum \left[ (y_i - \hat{\theta}_i)^2 / y_i + \hat{\theta}_i \hat{P}_{ii} \right], \quad (5.13)$$

where $P_{ii} = \theta_i x_i' \left( \sum_i x_i x_i' \right)^{-1} x_i$ is part of the diagonal term in the projection matrix for the Poisson model. In this case the approximate result that $E \left[ (y_i - \hat{\theta})^2 / y_i \right] \equiv -\theta_i P_{ii}$ constitutes the adjustment to be made to $\tilde{\tau}$.

Simulation studies of these authors suggest the size-adjustment procedure is effective in improving the small sample accuracy of the size and power of the relevant tests. The gain from size-adjustment is the greatest for the OPGM version of the tests. The above approach can in principle be applied to modifying any moment-based test statistic. The major difficulty in developing this type of size-correction is to determine $E(\tilde{\tau})$, particularly if the weights associated with the fundamental moment condition are stochastic and if samples are truncated or censored. Approximations to the finite sample distributions of tests based on Edgeworth-type expansions (Harris 1985; Honda 1988; Chesher and Spady 1991) have not been tried in count models.

6. EXAMPLE

6.1. General considerations

We shall begin by mentioning some general issues in count regression modelling, and then proceed with illustrations of various methods discussed earlier.

At the outset of an empirical investigation one might ask what purpose of the investigation is intended to serve. In some cases, the main interest is in modelling the conditional mean function and in making inferences about the statistical significance of some key parameters; e.g. the price sensitivity of the average number of doctor consultations. It is conceivable that different models and estimation methods may yield similar results in that respect, even though they might differ in the fit in the tails of the frequency distribution. In other cases, e.g. the number of airline or automobile accidents, the whole frequency distribution may be of interest because of its implications for (say) insurance; the probability of a large number of accidents may be of interest. If the object of the exercise is to make conditional predictions about the expected number of events, the focus would be on the conditional mean function. But if the focus is on the probability of a given number of events, the frequency distribution itself is relevant. In the former case, features such as overdispersion affect the prediction intervals, not the mean prediction itself, whereas in the latter case overdispersion will directly affect the
estimated probability. In the latter case, parametric methods are attractive, whereas in the former case robustness of the estimate may be more important. The modelling issues often cannot always be neatly separated, because robust modelling of the conditional mean function may call for parametric models; for example, the conditional mean may correspond to that for the ZIP or Poisson hurdles model, in which case one would also need to model the probability that a given observation belongs to one of two underlying ‘types’. Consequently, the attention given to features such as variance function modelling will vary on a case by case basis.

6.2. An Illustration

Some of the important ideas and techniques of earlier sections will be illustrated using a count data model for estimating a recreation demand function, due to Ozuna and Gomaz (1993). This example uses survey data on the number of recreational boating trips to Lake Somerville, East Texas, in 1980, denoted by V3. For the survey participants data are also available on the facility’s subjective quality ranking (SO), the respondent’s taste for water-skiing (SKI) and income (I), a set of travel cost variables to Lake Somerville (FC3, C3) and to competing or substitute boating attractions at Lakes Conroe and Houston (C1, C4). The data are a subset of that collected by Sellar et al through a survey administered to 2,000 registered leisure boat owners in 23 counties of East Texas. The frequency distribution of V3 based on 659 observations is given in Table 5.1.

Two noteworthy features of the Table 5.1 are the relatively long tail - 49 respondents reported taking 10 or more trips - and the presence of a strong mode at zero. There is also some clustering at 10 and 15 trips, creating a rough impression of multi-modality. Further, the presence of responses in ‘rounded’ categories like 20, 25, 30, 40 and 50 raises a suspicion that the respondents in these categories may not accurately recall the frequency of their visits.

In modelling this data set, a large number of issues could arise but because of space limitations we shall focus on only a few. These include the choice of the parametric family and estimation method, representation of unobserved heterogeneity in the sample, and evaluation of the fitted model. In single equation count models it has been a popular modelling strategy to settle with a variant of the negative binomial after the diagnostic checks reveal the presence of overdispersion. Our example is intended to show how this approach could be improved upon.

As a first approximation, one may begin, following Ozuna and Gomaz (1993), with the Poisson regression based on the conditional mean function

$$
\theta_i = E[V3_i] = \exp(\beta_0 + \sum_{i=1}^{7} \beta_i X_i)
$$

where the vector $X = (SO, SKI, I, FC3, C1, C3, C4)$. The results are given in Table
5.2. The ‘t-statistics’ of all coefficients except that of $C1$ are significant. However, the ‘robust’ versions of the t-ratios are much smaller, reflecting how the neglect of the of overdispersion inflates the Poisson t-ratios. Two measures of goodness of fit are also included. The first is the $G^2$ or the deviance measure defined earlier. If two models are nested, then the difference in the deviance associated with the models, is the log-likelihood ratio for the two models, and hence may be used to make the standard $\chi^2$ test of the implicit restrictions. A second measure is the Cameron and Windmeijer (1993) $R^2$ measure based on the Pearson residuals for the Poisson model, denoted $R^2_p$. Similar measures based on the deviance residuals for the Poisson model could also have been reported.

The first question concerns the deficiencies in the fitted model. A useful descriptive way of doing this is to compare the actual and fitted frequency distribution of $V3$. This can be done if we define cells centered on integer values. This is done in Table 5.3 for values of $V3 < 16$. Notice that the actual frequency of zeros (417) is considerably higher than the fitted value of 239. Second, the fitted model underpredicts the high counts, perhaps because of the curious clumping of the actual frequency distribution. This lack of fit could be reflected in significant values of specification test statistics. Consider for example, the regression based score tests of the null hypothesis of zero overdispersion. Regress the moment function ($\bar{\epsilon}_i^2 - V3_i$) on $\hat{\theta}_1$ and ($\bar{\epsilon}_i^2 - V3_i$) on $\hat{\theta}_2^2$, where $\bar{\epsilon}_i = V3_i - \hat{\theta}_1$.

The results, with heteroskedasticity robust ‘t-ratios’ shown in parenthesis, are as follows:

\[
\begin{align*}
\bar{\epsilon}_i^2 - V3_i & = 5.44 \hat{\theta}_1 \\
(2.09) \\
\bar{\epsilon}_i^2 - V3_i & = 1.45 \hat{\theta}_2^2 \\
(3.03)
\end{align*}
\]

There is clearly evidence of overdispersion in the data; the Poisson regression model was rejected against both the Negbin 1 and the Negbin 2 alternatives. The model was then reestimated using the Negbin 2 specification; Table 5.2 shows the result. Observe that allowing for overdispersion greatly increases the log-likelihood; the log-likelihood of the Poisson model was -1529.43; that of the Negbin model is -825.56, which reflects the importance of modeling overdispersion. There are also sizeable shifts in the size and the significance of several coefficients, a fact that is not easily reconciled with the idea that the conditional mean of the Poisson model is correctly specified. The income variable, $I$, and the cost variable $FC3$ became ‘insignificant’ once overdispersion was allowed for, while the coefficient of $C1$ changed sign from negative to (a priori correct) positive and becomes ‘significant’. The Negbin estimates are plausible in that they indicate substitution from other sites towards Lake Somerville as travel costs rise and away from Lake Somerville as its own travel costs rise. $SO$ and $SKI$ also have the a priori expected positive sign.
The presence of overdispersion, while consistent with the Negbin specification, does not in itself imply that the Negbin specification is adequate; rejection of the null against a specified alternative does not necessarily imply that the alternative is the correct one. Further examination of the fit of the model showed the predicted values of high counts were generally much lower than the actual values, regardless of whether the conditional mean function was estimated by nonlinear least squares, Poisson or Negbin 2 model. The deficiencies of the model including the poor fit in the right tail of the observed distribution could be interpreted in several different ways including the following: (a) the conditional mean function is misspecified; (b) the unobserved heterogeneity distribution is misspecified; and (c) the high counts reflect measurement errors. We consider alternative approaches for obtaining improvements based on these considerations.

The failure to account for high counts could reflect the need for additional nonlinearities in the conditional mean function. These can be introduced by including quadratic cost and income terms in the conditional mean function. Accordingly, three squared cost variables, C1SQ, C3SQ, C4SQ, and three cross-product variables, C1C3, C1C4, C3C4, were introduced into the conditional mean function. This could be justified by appealing to the possible presence of nonlinearities in the budget constraint, or simply in terms of a better approximation to the functional form. They produced a significant improvement in the fit of the model. The $R^2_p$ measure increases in value from 0.65 to 0.76. However, the very high counts are still underpredicted. Though some would be persuaded by this evidence that the Negbin model is a better specification, one might consider the following alternative.

As plausible alternatives to the models considered above, one could use either the zero-inflated Poisson model or the hurdle type model discussed in Section 3.4; they lead to changes in the conditional mean and the conditional variance specification. The sample under analysis may represent a mixture of at least two types, those who never choose boating as recreation and those that do, but some of the latter might simply happen not have had a positive number of boating trips in the sample period. The 'non-Poissonness' in the sample arises because the zeros come from two sources, not one; this is the zero-inflated Poisson model. In the hurdle type model the conditional mean for the zero observations is different from that for the nonzero observations. If this were so, the choice of the Poisson or the negative binomial hurdle model would lead to an improved specification. Table 5.2 provides parameter estimates for the Poisson hurdles specification, for the zero part and for the non-zero part. These show that the parameter estimates for the two parts are significantly different; the log-likelihood is now -1291, significantly higher than -1529 for the Poisson model, though not as high as for the negative binomial model. This suggests that a hurdles type specification which also models overdispersion could be an improvement. Accordingly, Table 5.4 provides estimates based on Negbin 2 hurdles. Again, the parameter estimates for the zeros and the positives are significantly different; the log-likelihood is now -725. In Table
5.4 we also include an estimate of the ZIP model, see (3.16), in which the probability of a nonzero count, denoted as \( \varsigma \) in (3.16), is further modelled as a logit function of three variables, \( ONE, SO, \) and \( I \). The results are once again plausible in that they suggest that higher the subjective ranking of Lake Somerville as a water-skiing facility, the greater the probability of a positive number of visits. The variable \( I \) does not seem to significantly affect that probability. The coefficients in the conditional mean part of the model are similar to those found earlier.

Table 5.4 also presents results based on estimation methods which make weaker distributional assumptions about the regression errors. The first of these is the non-linear least squares estimator; the standard and the robust heteroskedasticity corrected ‘t-ratios’ are given. The parameter estimates for the cost variables show substantial differences from those for the Negbin model in Table 5.2. In Ozuna and Gomaz (1993) the major focus was on the consumer surplus measure based on the coefficient of \( C3 \); this differs between the Negbin model and other models, as do others. How should one interpret such differences? If the conditional mean had been correctly specified, then given significant overdispersion, one would expect to see the ‘t-ratios’ to differ substantially between estimators, but the coefficient estimates less so. This is not what we observe. One interpretation of this result is that the conditional mean specification is suspect. We have already noted the failure of the models to account for high counts.

An illustrative modelling exercise such as this is necessarily open-ended. So we conclude with remarks about other plausible directions for respecifications. If one doubts the accuracy of the high counts, and/or one is not really interested in modelling high counts per se, one might treat counts larger than some specified value, say 15 or 20, as ‘missing’ values and then use the censored count model. This will reduce the importance of overdispersion. Suppose one’s objective is to model the data such that one can distinguish between the behavior of non-users (those with zero counts), moderate users and heavy users of recreational boating, one might consider recoding the data and then using a model like the ordinal probit.

7. CONCLUDING REMARKS

Recent econometric research on count data models has two major features. The first is that greater attention is paid to the way in which the data are generated and collected. Since this permits introduction of additional a priori information into the model, the approach has generated interesting and important variations of the basic mixed Poisson model. This line of research continues to have additional potential. The second development is the move away from strong distributional assumptions in estimation and testing. In cross sectional and time series work this has encouraged the application of robust ‘semi-parametric’ moment based estimation. There is still a relative dearth of applications that clearly demonstrate the relative merits of these alternative approaches.
References


61


### Table 5.1: Frequency distribution of recreational boating trips

<table>
<thead>
<tr>
<th>Trips</th>
<th>0</th>
<th>1</th>
<th>2</th>
<th>3</th>
<th>4</th>
<th>5</th>
<th>6</th>
<th>7</th>
<th>8</th>
<th>9</th>
<th>10</th>
<th>11</th>
<th>12</th>
<th>13</th>
<th>14</th>
</tr>
</thead>
<tbody>
<tr>
<td>Frequency</td>
<td>417</td>
<td>68</td>
<td>38</td>
<td>34</td>
<td>17</td>
<td>13</td>
<td>11</td>
<td>2</td>
<td>8</td>
<td>1</td>
<td>13</td>
<td>2</td>
<td>5</td>
<td>1</td>
<td>14</td>
</tr>
<tr>
<td>Trips</td>
<td>16</td>
<td>20</td>
<td>25</td>
<td>26</td>
<td>30</td>
<td>40</td>
<td>50</td>
<td>88</td>
<td></td>
<td></td>
<td></td>
<td></td>
<td></td>
<td></td>
<td></td>
</tr>
<tr>
<td>Frequency</td>
<td>1</td>
<td>3</td>
<td>3</td>
<td>1</td>
<td>3</td>
<td>3</td>
<td>1</td>
<td>1</td>
<td></td>
<td></td>
<td></td>
<td></td>
<td></td>
<td></td>
<td></td>
</tr>
</tbody>
</table>

### Table 5.2: Alternative parametric estimates of the boating trips model

<table>
<thead>
<tr>
<th>Variable</th>
<th>Poisson</th>
<th>Negbin</th>
<th>Poisson</th>
<th>Poisson</th>
</tr>
</thead>
<tbody>
<tr>
<td>ONE</td>
<td>0.264 (2.82; 0.61)</td>
<td>-1.12 (5.04)</td>
<td>-1.88 (9.30)</td>
<td>2.15 (19.2)</td>
</tr>
<tr>
<td>SO</td>
<td>0.471 (27.60; 9.66)</td>
<td>.722 (16.45)</td>
<td>0.815 (20.76)</td>
<td>.442 (1.86)</td>
</tr>
<tr>
<td>SKI</td>
<td>0.418 (7.31; 2.15)</td>
<td>.621 (4.38)</td>
<td>0.403 (2.97)</td>
<td>.467 (7.94)</td>
</tr>
<tr>
<td>I</td>
<td>-.111 (5.68; 2.21)</td>
<td>-.026 (0.64)</td>
<td>0.01 (.27)</td>
<td>-.097 (4.75)</td>
</tr>
<tr>
<td>FC3</td>
<td>.898 (11.37; 3.64)</td>
<td>.669 (1.48)</td>
<td>2.94 (.19)</td>
<td>.60 (7.55)</td>
</tr>
<tr>
<td>C1</td>
<td>-.003 (1.10; 0.23)</td>
<td>.048 (4.62)</td>
<td>.006 (.51)</td>
<td>.0014 (0.35)</td>
</tr>
<tr>
<td>C3</td>
<td>-.042 (25.4; 3.62)</td>
<td>-.092 (15.3)</td>
<td>-.052 (7.58)</td>
<td>-.036 (-17.9)</td>
</tr>
<tr>
<td>C4</td>
<td>0.0361 (13.3; 3.85)</td>
<td>.038 (4.43)</td>
<td>.046 (4.66)</td>
<td>.024 (6.87)</td>
</tr>
<tr>
<td>1/α</td>
<td>1.37 (9.24)</td>
<td></td>
<td></td>
<td></td>
</tr>
<tr>
<td>log-lik</td>
<td>-1529</td>
<td>-825</td>
<td>-277</td>
<td>-1014</td>
</tr>
<tr>
<td>$G^2$</td>
<td>2305</td>
<td>-</td>
<td>-</td>
<td>-</td>
</tr>
<tr>
<td>$R^2_p$</td>
<td>0.65</td>
<td>-</td>
<td>-</td>
<td>-</td>
</tr>
<tr>
<td>N</td>
<td>659</td>
<td>659</td>
<td>659</td>
<td>659</td>
</tr>
</tbody>
</table>

Notes: Figures in parenthesis are absolute 't-ratios'. The second 't-ratio' in the Poisson column is the 'robust' t-ratio.
Table 5.3: Actual and fitted cell frequencies from the Poisson regression

<table>
<thead>
<tr>
<th>Trips</th>
<th>Fitted</th>
<th>Actual</th>
<th>Cum. Freq.</th>
</tr>
</thead>
<tbody>
<tr>
<td>0</td>
<td>239</td>
<td>417</td>
<td>.36</td>
</tr>
<tr>
<td>1</td>
<td>198</td>
<td>68</td>
<td>.66</td>
</tr>
<tr>
<td>2</td>
<td>52</td>
<td>38</td>
<td>.74</td>
</tr>
<tr>
<td>3</td>
<td>37</td>
<td>34</td>
<td>.80</td>
</tr>
<tr>
<td>4</td>
<td>39</td>
<td>17</td>
<td>.86</td>
</tr>
<tr>
<td>5</td>
<td>22</td>
<td>13</td>
<td>.89</td>
</tr>
<tr>
<td>6</td>
<td>24</td>
<td>11</td>
<td>.93</td>
</tr>
<tr>
<td>7</td>
<td>11</td>
<td>2</td>
<td>.94</td>
</tr>
<tr>
<td>8</td>
<td>8</td>
<td>1</td>
<td>.96</td>
</tr>
<tr>
<td>9</td>
<td>4</td>
<td>13</td>
<td>.96</td>
</tr>
<tr>
<td>10</td>
<td>4</td>
<td>2</td>
<td>.97</td>
</tr>
<tr>
<td>11</td>
<td>4</td>
<td>5</td>
<td>.98</td>
</tr>
<tr>
<td>12</td>
<td>4</td>
<td></td>
<td></td>
</tr>
</tbody>
</table>

Table 5.4: Alternative estimates of the boating trips model

<table>
<thead>
<tr>
<th>Variable</th>
<th>Negbin Hurdles:</th>
<th>Nonlinear least squares</th>
<th>ZIP</th>
</tr>
</thead>
<tbody>
<tr>
<td></td>
<td>Zeros</td>
<td>Positives</td>
<td></td>
</tr>
<tr>
<td>ONE</td>
<td>-3.046 (2.52)</td>
<td>0.841 (1.97)</td>
<td>1.678 (8.53; 3.36)</td>
</tr>
<tr>
<td>SO</td>
<td>4.638 (2.43)</td>
<td>0.172 (2.25)</td>
<td>0.280 (6.99; 4.73)</td>
</tr>
<tr>
<td>SKI</td>
<td>-0.025 (0.02)</td>
<td>0.622 (3.14)</td>
<td>0.487 (3.90; 2.04)</td>
</tr>
<tr>
<td>I</td>
<td>0.026 (0.11)</td>
<td>-0.057 (0.778)</td>
<td>-0.140 (2.911; 2.01)</td>
</tr>
<tr>
<td>FC3</td>
<td>16.203 (0.97)</td>
<td>0.576 (0.87)</td>
<td>0.942 (8.56; 3.86)</td>
</tr>
<tr>
<td>C1</td>
<td>0.030 (0.28)</td>
<td>0.057 (2.89)</td>
<td>-0.034 (4.35; 1.92)</td>
</tr>
<tr>
<td>C3</td>
<td>-0.156 (1.62)</td>
<td>-0.078 (7.07)</td>
<td>-0.037 (8.74; 2.82)</td>
</tr>
<tr>
<td>C4</td>
<td>0.117 (1.40)</td>
<td>0.012 (0.84)</td>
<td>0.045 (6.13; 3.38)</td>
</tr>
<tr>
<td>1/α</td>
<td>5.609 (1.81)</td>
<td>1.700 (3.87)</td>
<td></td>
</tr>
<tr>
<td>I</td>
<td></td>
<td></td>
<td>-1.189 (1.10)</td>
</tr>
<tr>
<td>SO</td>
<td></td>
<td></td>
<td>-6.15 (5.67)</td>
</tr>
<tr>
<td>ONE</td>
<td></td>
<td></td>
<td>5.80 (5.81)</td>
</tr>
<tr>
<td>log-lik</td>
<td></td>
<td>-134 -591</td>
<td>-1337.7</td>
</tr>
<tr>
<td>N</td>
<td>659</td>
<td>659</td>
<td>659</td>
</tr>
</tbody>
</table>

Notes: Figures in parenthesis are absolute ‘t-ratios’. The second ‘t-ratio’ in the third column is the ‘robust’ t-ratio. The ‘t-ratios’ for the ZIP model are based on robust standard errors.
<table>
<thead>
<tr>
<th>Number</th>
<th>Author</th>
<th>Title</th>
<th>Date</th>
</tr>
</thead>
<tbody>
<tr>
<td>205</td>
<td>Charles A. Holt, Fernando Solis-Soberon</td>
<td>The Calculation of Equilibrium Mixed Strategies in Posted-Offer Auctions</td>
<td>1/90</td>
</tr>
<tr>
<td>206</td>
<td>Ronald Fischer, Leonard J. Mirman</td>
<td>Strategic Dynamic Interaction: Fish Wars</td>
<td>10/89</td>
</tr>
<tr>
<td>207</td>
<td>T. W. Epps, Eugene Seneta</td>
<td>Stock Prices as Branching Processes</td>
<td>2/90</td>
</tr>
<tr>
<td>208</td>
<td>Charles A. Holt, Douglas Davis</td>
<td>The Effects of Non-Binding Price Announcements on Posted-Offer Markets</td>
<td>4/90</td>
</tr>
<tr>
<td>210</td>
<td>Ronald Fischer</td>
<td>Income Distribution in the Dynamic Two Factor Trade Model</td>
<td>5/90</td>
</tr>
<tr>
<td>211</td>
<td>Ronald Fischer</td>
<td>Trade and Income Distribution in a Specific-Factor Model with Bequests</td>
<td>5/90</td>
</tr>
<tr>
<td>212</td>
<td>Jonathan Eaton</td>
<td>Sovereign Debt, Reputation, and Credit Terms</td>
<td>7/90</td>
</tr>
<tr>
<td>214</td>
<td>T. W. Epps</td>
<td>Characteristic Functions: Geometrical Interpretations and Applications to Statistical Inference</td>
<td>9/90</td>
</tr>
<tr>
<td>215</td>
<td>Jonathan Skinner</td>
<td>The Dynamic Efficiency Cost of Not Taxing Housing</td>
<td>9/90</td>
</tr>
<tr>
<td>216</td>
<td>Jonathan Skinner</td>
<td>If Agricultural Land Taxation is so Efficient, Why Is It So Rarely Used?</td>
<td>7/90</td>
</tr>
<tr>
<td>217</td>
<td>R. Anton Braun, Charles L. Evans</td>
<td>Seasonality and Equilibrium Business Cycle Theories</td>
<td>10/90</td>
</tr>
<tr>
<td>218</td>
<td>Charles Engel, Anthony Rodrigues</td>
<td>Tests of Mean-Variance-Efficiency of International Equity Markets</td>
<td>9/90</td>
</tr>
</tbody>
</table>
255  Shiferaw Gurmu     Testing for Overdispersion in Censored Poisson Regression Models 6/93
      Jacques-Francois Thisse
257  Satyajit Chatterjee Inventories, Production Smoothing, and B. Ravikumar 8/93
      Anticipated Demand Variations
258  Gerhard Gloom      Endogenous Public Policy and Multiple B. Ravikumar 9/93
      Equilibria
259  I. V. Evstigneev    Robust Insurance Mechanisms and the Shadow L. Mirman 7/94
      Prices of Information Constraints
      W. K. Klein Haneveld
260  Simon P. Anderson  Oligopolistic Competition and the Optimal Andre De Palma 7/94
      Provision of Products
      Yuri Nesterov
261  Shiferaw Gurmu     Recent Developments in Models of Event Pravin K. Trivedi 9/94
      Counts: A Survey