Count Data Models for Financial Data

A. Colin Cameron
Department of Economics
University of California
Davis, CA 95616

Pravin K. Trivedi
Department of Economics
Indiana University
Bloomington, IN 47405

January 9, 1996

Abstract

To appear in G.S. Maddala and C.R. Rao ed., *Handbook of Statistics: Statistical Methods in Finance*, North- Holland.

In some financial studies the dependent variable is a count, taking non-negative integer values. Examples include the number of takeover bids received by a target firm, the number of unpaid credit installments (useful in credit scoring), the number of accidents or accident claims (useful in determining insurance premia) and the number of mortgage loans pre-paid (useful in pricing mortgage-backed securities). Models for count data, such as Poisson and negative binomial are presented, with emphasis placed on the underlying count process and links to dual data on durations. A self-contained discussion of regression techniques for the standard models is given, in the context of financial applications.

Contents

1. Introduction

In count data regression, the main focus is the effect of covariates on the frequency of an event, measured by non-negative integer values or counts. Count models, such as Poisson and negative binomial, are similar to binary models, such as probit and logit, and other limited dependent variable models, notably tobit, in that the sample space of the dependent variable has restricted support. Count models are used in a wide range of disciplines. For an early application and survey in economics see Cameron and Trivedi (1986), for more recent developments see Winkelmann (1994), and for a comprehensive survey of the current literature see Gurmu and Trivedi (1994).

    The benchmark model for count data is the Poisson. If the discrete random variable Y is Poisson distributed with parameter λ, it has density $e^{-\lambda}\lambda^y/y!$, mean λ and variance λ. Frequencies and sample means and variances for a number of finance examples are given in Table 1. The data of Jaggia and Thosar (1993) on the number of takeover bids received by a target firm after an initial bid illustrate the preponderance of small counts in a typical application of the Poisson model. The data of Greene (1994) on the number of major derogatory reports in the credit history of individual credit card applicants illustrate overdispersion, i.e. the sample variance is considerably greater than the sample mean, compared to the Poisson which imposes equality of population mean and variance, and

excess zeros since the observed proportion of zero counts of .804 is considerably greater than the predicted probability of $e^{-.456} = .633$. The negative binomial distribution, defined below, can potentially accommodate this overdispersion. In fact, the negative binomial with mean .456 and variance 1.810 gives predicted probability of zero counts of .809. A related example is the data of Guillen (1994) who modeled the number of unpaid installments by creditors of a bank. The data of Davutyan (1989) on the annual number of bank failures has the added complication of being a time series. The data may be serially correlated, as the five largest counts are the last five observations in the latter sample period.

In econometric applications with count data, analysis focuses on the role of regressors X, introduced by specifying $\lambda = \exp(X'\beta)$, where the parameter vector β may be estimated by maximum likelihood. For example, the mean number of takeover bids for a firm may be related to the size of the firm.

There are important connections between count regressions and duration (or waiting time) models. These connections can be understood by studying the underlying stochastic process for the waiting time between events, which involves the three concepts of states, spells and events. A *state* is a classification of an individual or a financial entity at a point in time; a *spell* is defined by the state, the time of entry and time of exit; and an *event* is simply the instantaneous transition from one state to another state.

A regression model for *durations* involves the relationship between the (non-negative) length of the spell spent in a particular state and a set of covariates. Duration models are often recast as models of the *hazard rate*, which is the instantaneous rate of transition from one state to another. A *count* regression involves the relationship between the number of events of interest in a fixed time interval and a set of covariates.

Which approach is adopted in empirical work will depend not only on the research objectives but also on the form in which the data are available. Econometric models of durations or transitions provide an appropriate framework for modelling the duration in a given financial state; count data models provide a framework for modelling the frequency of the event per unit time period. This article differs from many treatments in emphasizing the connections between the count regression and the underlying process, and the associated links with duration analysis.

To fix concepts consider the event of mortgage prepayment, which involves exit from the state of holding a mortgage, and termination of the associated spell.

If the available data provide sample information on the complete or incomplete life of individual mortgages, for those that were either initiated or terminated at some date, together with data on the characteristics of the mortgage holders and mortgage contracts, a duration regression is a natural method of analyzing the role of covariates.[1] Now, it is often the case that data may not be available on individual duration intervals, but may be available on the frequency of a repeated event per some unit of time; e.g. the number of mortgages that were pre-paid within some calendar time period. Such aggregated data, together with information on covariates, may form the basis of a count data regression. Yet another data situation, which we do not pursue, is that in which one has sample information on a binary outcome, viz., whether or not a mortgage was terminated within some time interval. A binary regression such as logit or probit is the natural method for analyzing such data.

Further examples of duration models are: duration between the initiation of a hostile bid for the takeover of a firm and the resolution of the contest for corporate control; the time spent in bankruptcy protection; the time to bank failure; the time interval to the dissolution of a publicly traded fund; and the time interval to the first default on repayment of a loan. Several examples of count data models in empirical finance literature have already been given. We reiterate that for each example it is easy to conceive of the data arising in the form of durations or counts.

In section 2 we exposit the relation between econometric models of durations and of counts. A self-contained discussion of regression techniques for count data is given in section 3, in the context of financial applications. Concluding remarks are made in section 4.

2. Stochastic Process Models for Count and Duration Data

Fundamentally, models of durations and models of counts are duals of each other. This duality relationship is most transparent when the underlying data generating process obeys the strict assumptions of a stationary (memoryless) Poisson process. In this case it is readily shown that the frequency of events follows the Poisson distribution and the duration of spells follows the exponential distribution. For example, if takeover bids for firms follow a Poisson process, then the number of bids for a firm in a given interval of time is Poisson distributed, while the elapsed time between bids is exponentially distributed. In this special case econometric

models of durations and counts are equivalent as far as the measurement of the effect of covariates (exogenous variables) is concerned.

Stationarity is a strong assumption. Often the underlying renewal process exhibits dependence or memory. The length of time spent in a state, e.g. the time since the last takeover bid, may affect the chances of leaving that state; or the frequency of the future occurrences of an event may depend upon the past frequency of the same event. In such cases, the information content of duration and count models may differ considerably. However, it can be shown that either type of model can provide useful information about the role of covariates on the event of interest. The main focus in the remainder of the paper is on count data models.

2.1. Preliminaries

We observe data over an interval of length t. For nonstationary processes behavior may also depend on the starting point of the interval, denoted s. The random variables (r.v.'s) of particular interest are $N(s, s + t)$, which denotes the number of events occurring in $(s, s + t]$, and $T(s)$, which denotes the duration of time to occurrence of the next event given an event occurred at time s. The distribution of the *number of events* is usually represented by the probability density function

$$\Pr\{N(s, s + t) = r\}, \qquad r = 0, 1, 2, ...$$

The distribution of the *durations* is represented in several ways, including

$$
\begin{aligned}
F_{T(s)}(t) &= \Pr\{T(s) < t\} \\
S_{T(s)}(t) &= \Pr\{T(s) \geq t\} \\
f_{T(s)}(t) &= \lim_{dt \to 0} \Pr\{t \leq T(s) < t + dt\} \\
h_{T(s)}(t) &= \lim_{dt \to 0} \Pr\{t \leq T(s) < t + dt \mid T(s) \geq t\} \\
H_{T(s)}(t) &= \int_{s}^{s+t} h_{T(s)}(u) \, du
\end{aligned}
$$

where the functions F, S, f, h and H are called, respectively, the *cumulative distribution function, survivor function, density function, hazard function* and *integrated hazard function*.

For duration r.v.'s the distribution is often specified in terms of the survivor and hazard functions, rather than the more customary c.d.f. or density function, as they have a more natural physical interpretation. In particular, the hazard

function gives the instantaneous rate (or probability in the discrete case) of transition from one state to another given that it has not occurred to date, and is related to the density, distribution and survivor functions by

$$h_{T(s)}(t) = \frac{f_{T(s)}(t)}{F_{T(s)}(t)} = \frac{f_{T(s)}(t)}{1 - S_{T(s)}(t)}.$$

As an example, consider the length of time spent by firms under bankruptcy protection. Of interest is how the hazard varies with time and with firm characteristics. If the hazard function is decreasing in t, then the probability of leaving bankruptcy decreases the longer the firm is in bankruptcy protection, while if the hazard function increases with the interest burden of the firm, then firms with a higher interest burden are more likely to leave bankruptcy than are firms with a low interest burden.

Modeling of the hazard function should take into account the origin state and the destination state. Two-state models are the most common, but multi-state models may be empirically appropriate in some cases. For example, a firm currently under bankruptcy protection may subsequently either be liquidated or resume its original operations; these possibilities call for a three-state model.

2.2. Poisson Process

Define the constant λ to be the rate of occurrence of the event. A *(pure) Poisson process* of rate λ occurs if events occur independently with probability equal to λ times the length of the interval. Formally, as $t \to 0$

$$\Pr\{N(s, s + t) = 0\} = 1 - \lambda t + o(t)$$
$$\Pr\{N(s, s + t) = 1\} = \lambda t + o(t)$$

and $N(s, s + t)$ is statistically independent of the number and position of events in $(0, s]$. Note that in the limit the probability of 2 or more events occurring is zero, while 0 and 1 events occur with probabilities of, respectively, $(1 - \lambda t)$ and λt.

For this process it can be shown that the number of events occurring in the interval $(s, s + t]$, for nonlimit t, is *Poisson* distributed with mean λt and probability

$$\Pr\{N(s, s + t) = r\} = \frac{e^{-\lambda t}(\lambda t)^r}{r!} \qquad r = 0, 1, 2, ...$$

while the duration to the next occurrence of the event is exponentially distributed with mean λ^{-1} and density

$$f_{T(s)}(t) = \lambda e^{-\lambda t}.$$

The corresponding hazard rate $h_{T(s)}(t) = \lambda$ is constant and does not depend on the time since the last occurrence of the event, exhibiting the so-called memoryless property of the Poisson process. Note also that the distributions of both the counts and durations are independent of the starting time s.

Set $s = 0$, and consider a time interval of unit length. Then N, the mean number of events in this interval, has mean given by

$$E[N] = \lambda,$$

while the mean of T, the duration between events, is given by

$$E[T] = \frac{1}{\lambda}.$$

Intuitively, a high frequency of events per period implies a short average inter-event duration.

The conditional mean function for a regression model is obtained by parameterizing λ in terms of covariates X, e.g. $\lambda = \exp(X'\beta)$. Estimation can be by maximum likelihood, or by (nonlinear) regression which for more efficient estimation uses $Var(N) = \lambda$ or $Var(T) = (1/\lambda)^2$ for a Poisson process.

The Poisson process may not always be the appropriate model for data. For example, the probability of one occurrence may increase the likelihood of further occurrences. Then a Poisson distribution may overpredict the number of zeros, underpredict the number of nonzero counts, and have variance in excess of the mean.

2.3. Time-dependent Poisson Process

The *time-dependent Poisson process*, also called the non-homogeneous or non-stationary Poisson process, is a nonstationary point process which generalizes the (pure) Poisson process by specifying the rate of occurrence to depend upon the elapsed time since the start of the process, i.e. we replace λ by $\lambda(s+t)$.[2]

The counts $N(s, s+t)$ are then distributed as Poisson with mean $\Lambda(s, s+t)$, where

$$\Lambda(s, s+t) = \int_s^{s+t} \lambda(u) du$$

The durations $T(s)$ are distributed with survivor and density functions

$$S_{T(s)}(t) = \exp(-\Lambda(s, s+t))$$

$$f_{T(s)}(t) = \lambda(s+t) \exp(-\Lambda(s, s+t)).$$

Hence $h_{T(s)}(t) = \lambda(s+t)$, so that $\lambda(\cdot)$ is the hazard function. Also $H_{T(s)}(t) = \Lambda(s, s+t)$, so that $\Lambda(\cdot)$ is the integrated hazard function.

One convenient choice of functional form is the *Weibull*, $\lambda(s+t) = \lambda\gamma(s+t)^{\gamma-1}$, in which case $\Lambda(s, s+t) = \lambda[s+t]^\gamma - \lambda s^\gamma$. In this case, the time-dependent component of $\lambda(\cdot)$ enters multiplicatively with exponent $\gamma - 1$. The parameter γ indicates duration dependence; $\gamma > 1$ indicates positive duration dependence, which means the probability that the spell in the current state will terminate increases with the length of the spell. Negative duration dependence is indicated by $\gamma < 1$. The mean number of events in $(s, s+t]$ also depends on s, increasing or decreasing in s as $\gamma > 1$ or $\gamma < 1$. This process is therefore nonstationary. The case $\gamma = 1$ gives the pure Poisson process, in which case the Weibull reduces to the exponential. The standard parametric model for econometric analysis of durations is the Weibull. Regression models are formed by specifying λ to depend on regressors, e.g. $\lambda = \exp(X'\beta)$, while γ does not.

This is an example of the *proportional hazards* or *proportional intensity* factorization:

$$\lambda(t, X, \gamma, \beta) = \lambda_0(t, \gamma)g(X, \beta) \qquad (2.1)$$

where $\lambda_0(t, \gamma)$ is a baseline hazard function, and the only role of regressors is as a scale factor for this baseline hazard. This factorization simplifies interpretation, as the conditional probability of leaving the state for an observation with $X = X_1$ is $\frac{g(X_1, \beta)}{g(X_2, \beta)}$ times that when $X = X_2$. Estimation is also simpler, as the role of regressors can be separated from the way in which the hazard function changes with time. For single-spell duration data this is the basis of the partial likelihood estimator of Cox (1972a). When the durations of multiple spells are observed this leads to estimation methods where most information comes from the counts, see Lawless (1987). Similar methods can be applied to grouped count data. For example, Schwartz and Torous (1993) model the number of active mortgages that are terminated in a given interval of time.

2.4. Renewal Process

A *renewal process* is a stationary point process for which the durations between occurrences of events are independently and identically distributed (i.i.d.). The (pure) Poisson process is a renewal process, but the time-dependent process is not since it is not stationary.

For a renewal process $f_{T(s)}(t) = f_{T(s')}(t), \forall s, s'$, and it is convenient to drop the dependence on s. We define N_t as the number of events (renewals) occurring in $(0, t)$ which in earlier notation would be $N(0, t)$ and will have the same distribution as $N(s, s + t)$. Also define T_r as the time up to the r-th renewal.

Then

$$
\begin{aligned}
\Pr\{N_t = r\} &= \Pr\{N_t < r + 1\} - \Pr\{N_t < r\} \\
&= \Pr\{T_{r+1} > t\} - \Pr\{T_r > t\} \\
&= F_r(t) - F_{r+1}(t)
\end{aligned}
$$

where F_r is the cumulative distribution function of T_r.

The second line of the last equation array suggests an attractive approach to the derivation of parametric distributions for N_t based on (or dual to) specified distributions for durations. For example, one may want a count distribution that is dual to the Weibull distribution since the latter can potentially accommodate certain types of time dependence.[3] Unfortunately, the approach is often not practically feasible.

Specifically, T_r is the sum of r i.i.d. duration times whose distribution is most easily found using the (inverse) Laplace transform, a modification for non-negative r.v.'s of the moment generating function.[4] Analytical results are most easily found when the Laplace transform is simple and exists in a closed form. When the durations are i.i.d. exponentially distributed, N_t is Poisson distributed as expected. Analytical results can also be obtained when durations are i.i.d. Erlangian distributed, where the Erlangian distribution is a special case of the 2-parameter gamma distribution that arises when the first parameter is restricted to being a positive integer; see Feller (1966), Winkelmann (1993). For many standard duration time distributions, such as the Weibull, analytical expressions for the distribution of T_r and hence N_t do not exist. In principle a numerical approach could be used, but currently there are no studies along these lines.

Some useful asymptotic results are available. If the i.i.d. durations between

events have mean μ and variance σ^2, then the r.v.

$$Z = \frac{N_t - t/\mu}{\sigma\sqrt{t/\mu^3}} \overset{a}{\sim} N(0,1).$$

The *expected* number of renewals $E[N_t]$, called the renewal function, satisfies

$$E[N_t] = \frac{t}{\mu} + O(1)$$

as $t \to \infty$, so that a halving of the duration times will approximately double the mean number of renewals. Thus if a renewal process is observed for a long period of time, analysis of count data will be quite informative about the mean duration time. For a Poisson process the relationship is exact.

Parametric analysis of a renewal process begins with the specification of the distribution of the i.i.d. durations. Analysis is therefore straightforward if data on the duration lengths are available. Most econometric analysis of renewal processes focuses on the implications when spells are *incomplete* or *censored*. The observed data may be the backward recurrence time, i.e. the length of time from the last renewal to fixed time point t, or the forward recurrence time, i.e. the time from t to the next renewal, but not the duration of the completed spell which is the sum of the backward and forward recurrence times; see Lancaster (1990, p.94).

2.5. Other Stochastic Processes

There are many other stochastic processes that could potentially be applied to financial data. A standard reference for stochastic processes is Karlin and Taylor (1975). Like many such references it does not consider estimation of statistical models arising from this theory. A number of monographs by Cox do emphasize statistical applications, including Cox and Lewis (1966) and Cox (1962). The standard results for the Poisson process are derived in Lancaster (1990, pp. 86-87). Some basic stochastic process theory is presented in Lancaster (1990, chapter 5), where renewal theory and its implications for duration analysis is emphasized, and in Winkelmann (1994, chapter 2).

Markov chains are a subclass of stochastic processes that are especially useful for modelling count data. A *Markov chain* is a Markov process, i.e. one whose future behavior given complete knowledge of the current state is unaltered by additional knowledge of past behavior, that takes only a finite or denumerable

range of values, and can be characterized by the transition probabilities from one state (discrete value) to another. If these discrete values are non-negative integers, or can be rescaled to non-negative integer values, the Markov chain describes a probabilistic model for counts. This opens up a wide range of models for counts, as many stochastic processes are Markov chains. One example, a branching process, is considered in section 3.6.

3. Econometric Models of Counts

The Poisson regression is the common starting point for count data analysis, and is well motivated by assuming a Poisson process. Data frequently exhibit important "non-Poisson" features, however, including:

1. Overdispersion: the conditional variance exceeds the conditional mean, whereas the Poisson distribution imposes equality of the two.

2. Excess zeros: a higher frequency of zeros (or some other integer count) than that predicted by the Poisson distribution with a given mean.

3. Truncation from the left: small counts (particularly zeros) are excluded.

4. Censoring from the right: counts larger than some specified integer are grouped.

The use of Poisson regression in the presence of any of these features leads to a loss of efficiency (and sometimes consistency), incorrect reported standard errors, and a poor fit. These considerations motivate the use of distributions other than the Poisson. These models for count data are usually specified with little consideration of the underlying stochastic process.

For convenient reference, Table 2 gives some commonly used distributions and their moment properties. Each sub-section considers a class of models for count data, presented before consideration of applications and the stochastic data generating process. Table 3 provides a summary of applications from the finance literature and the models used, in the order discussed in the text.

3.1. Preliminaries

Typical data for applied work consist of N observations, the i-th of which is (y_i, X_i), $i = 1, \ldots, N$, where the scalar dependent variable y_i is the number of occurrences of the event of interest, and X_i is the $k \times 1$ vector of covariates that are thought to determine y_i. Except where noted we assume independence across observations. Econometric models for the counts y_i are nonlinear in parameters. Maximum likelihood (ML) estimation has been especially popular, even though closely related methods of estimation based on the first two moments of the data distribution can also be used.

Interest focuses on how the mean number of events changes due to changes in one or more of the regressors. The most common specification for the conditional mean is

$$E[y_i|X_i] = \exp(X_i'\beta) \tag{3.1}$$

where β is a $k \times 1$ vector of unknown parameters. This specification ensures the conditional mean is nonnegative and, using $\partial E[y_i|X_i]/\partial X_{ij} = \exp(X_i'\beta)\beta_j$, strictly monotonic increasing (or decreasing) in X_{ij} according to the sign of β_j. Furthermore, the parameters can be directly interpreted as semi-elasticities, with β_j giving the proportionate change in the conditional mean when X_{ij} changes by one unit. Finally, if one regression coefficient is twice as large as another, then the effect of a one-unit change of the associated regressor is double that of the other. Throughout we give results for this particular specification of the mean.

As an example, let y_i be the number of bids after the initial bid received by the i-th takeover target firm and S_i denote firm size, measured by book value of total assets of the firm in billions of dollars. Then Poisson regression of y_i on S_i using the same sample as Jaggia and Thosar (1993) yields a conditional mean $E[y_i|S_i] = \exp(0.499 + 0.037S_i)$, so that a one billion dollar increase in total assets leads to a 3.7 percent increase in the number of bids.

Sometimes regressors enter logarithmically in (3.1). For example, we may have

$$\begin{aligned} E[y_i|X_i] &= \exp(\beta_1 \log_e(X_{1i}) + X_{2i}'\beta_2) \\ &= X_{1i}^{\beta_1} \exp(X_{2i}'\beta_2) \end{aligned} \tag{3.2}$$

in which case β_1 is an elasticity. This formulation is particularly appropriate when X_{1i} is a measure of exposure, such as number of miles driven if modelling

the number of automobile accidents, in which case we expect β_1 to be close to unity.

3.2. Poisson, Negative Binomial and Inverse-Gaussian Models

3.2.1. Maximum Likelihood Estimation

The *Poisson regression model* assumes that y_i given X_i is Poisson distributed with density

$$f(y_i|X_i) = \frac{e^{-\lambda_i}\lambda_i^{y_i}}{y_i!}, \qquad y_i = 0, 1, 2, \ldots \tag{3.3}$$

and mean parameter $\lambda_i = \exp(X_i'\beta)$ as in (3.1). Given independent observations, the log-likelihood is

$$\log L = \sum_{i=1}^{n} \{y_i X_i'\beta - \exp(X_i'\beta) - \log y_i!\}. \tag{3.4}$$

Estimation is straightforward. The log-likelihood function is globally concave, many statistical packages have built-in Poisson ML procedures, or the Newton-Raphson algorithm can be implemented by iteratively reweighted OLS. The first-order conditions are

$$\sum_{i=1}^{n} (y_i - \exp(X_i'\beta))X_i = 0,$$

or that the unweighted residual $(y_i - \exp(X_i'\beta))$ is orthogonal to the regressors. Applying the usual ML theory yields $\hat{\beta}$ asymptotically normal with mean β and

$$Var(\hat{\beta}) = \left(\sum_{i=1}^{n} \exp(X_i'\beta)X_i X_i'\right)^{-1}, \tag{3.5}$$

using $E\left[\frac{\partial^2 \log L}{\partial\beta\partial\beta'}\right] = -\sum_{i=1}^{n} \exp(X_i'\beta)X_i X_i'$.

The Poisson distribution imposes equality of the variance and mean. In fact observed data are often overdispersed, i.e. the variance exceeds the mean. Then the Poisson MLE is still consistent if the mean is correctly specified, i.e. (3.1) holds, but it is inefficient and the reported standard errors are incorrect.[5]

More efficient parameter estimates can be obtained by ML estimation for a specified density less restrictive than the Poisson. The standard two-parameter

distribution for count data that can accommodate overdispersion is the *negative binomial*, with mean λ_i, variance $\lambda_i + \alpha\lambda_i^2$, and density

$$f(y_i|X_i) = \frac{\Gamma(y_i + \alpha^{-1})}{\Gamma(y_i + 1)\Gamma(\alpha^{-1})}\left(\frac{\alpha^{-1}}{\alpha^{-1} + \lambda_i}\right)^{\alpha^{-1}}\left(\frac{\lambda_i}{\alpha^{-1} + \lambda_i}\right)^{y_i}. \qquad y_i = 0, 1, 2, \ldots$$

(3.6)

The log-likelihood for mean parameter $\lambda_i = \exp(X_i'\beta)$ as in (3.1) equals

$$\log L = \sum_{i=1}^{n}\{\log\left(\frac{\Gamma(y_i + \alpha^{-1})}{\Gamma(y_i + 1)\Gamma(\alpha^{-1})}\right) - (y_i + \alpha^{-1})\log(1 + \alpha\exp(X_i'\beta)) + y_i\log\alpha + y_iX_i'\beta\}$$

(3.7)

There are alternative parameterizations of the negative binomial, with different variance functions. The one above is called the *Negbin 2* model by Cameron and Trivedi (1986), and is computed for example by LIMDEP. It nests as a special case the *Geometric*, which sets $\alpha = 1$. An alternative model, called *Negbin 1*, has variance $(1 + \alpha)\lambda_i$ which is linear rather than quadratic in the mean. This Negbin 1 model is seldom used and is not formally presented here. For both models estimation is by maximum likelihood, with $(\hat{\alpha}, \hat{\beta})$ asymptotic normal with variance matrix the inverse of the information matrix. Both models reduce to the Poisson in the special case where the overdispersion parameter α equals zero.

One motivation for the negative binomial model is to suppose that y_i is Poisson with parameter $\lambda_i v_i$ rather than λ_i, where v_i is unobserved individual heterogeneity. If the distribution of v_i is i.i.d. gamma with mean 1 and variance α, then while y_i conditional on λ_i and v_i is Poisson, conditional on λ_i alone it is negative binomial with mean λ_i and variance $\lambda_i + \alpha\lambda_i^2$ (i.e. Negbin 2). This unobserved heterogeneity derivation of the negative binomial assumes that the underlying stochastic process is a Poisson process. An alternative derivation of the negative binomial assumes a particular form of nonstationarity for the underlying stochastic process, with occurrence of an event increasing the probability of further occurrences. Cross section data on counts are insufficient on their own to discriminate between the two.

Clearly a wide range of models, called *mixture* models, can be generated by specifying different distributions of v_i. One such model is the *Poisson-Inverse Gaussian* model of Dean et. al. (1989), which assumes v_i has an inverse Gaussian distribution. This leads to a distribution with heavier tails than the negative bino-

mial. Little empirical evidence has been provided to suggest that such alternative mixture models are superior to the negative binomial.

Mixture models cannot model underdispersion (variance less than mean), but this is not too restrictive as most data is overdispersed. Parametric models for underdispersed data include the Katz system, see King (1989), and the generalized Poisson, see Consul and Famoye (1992).

When data are in the form of counts a sound practice is to estimate both Poisson and negative binomial models. The Poisson is the special case of the negative binomial with $\alpha = 0$. This can be tested by a likelihood ratio test, with -2 times the difference in the fitted log-likelihoods of the two models distributed as $\chi^2(1)$ under the null hypothesis of no overdispersion. Alternatively a Wald test can by performed, using the reported "t-statistic" for the estimated α in the negative binomial model, which is asymptotically normal under the null hypothesis of no overdispersion. A third method, particularly attractive if a package program for negative binomial regression is unavailable, is to estimate the Poisson model, construct $\hat{\lambda}_i = \exp(X_i' \hat{\beta})$, and perform the auxiliary OLS regression (without constant)

$$\{(y_i - \hat{\lambda}_i)^2 - y_i\}/ \hat{\lambda}_i = \alpha \, \hat{\lambda}_i + u_i. \tag{3.8}$$

The reported t-statistic for α is asymptotically normal under the null hypothesis of no overdispersion against the alternative of overdispersion of the Negbin 2 form. This last test coincides with the score or LM test for Poisson against negative binomial, but is more general as its motivation is one based on using only the specified mean and variance. It is valid against any alternative distribution with overdispersion of the Negbin 2 form, and it can also be used for testing underdispersion; see Cameron and Trivedi (1990). To test overdispersion of the Negbin 1 form, replace (3.8) with

$$\{(y_i - \hat{\lambda}_i)^2 - y_i\}/ \hat{\lambda}_i = \alpha + u_i. \tag{3.9}$$

3.2.2. Estimation Based on First Moment

To date we have considered fully parametric approaches. An alternative is to use regression methods that use information on the first moment, or the first and second moments, following Gourieroux, Montfort and Trognon (1984), Cameron and Trivedi (1986) and McCullagh and Nelder (1989). The simplest approach

15

is to assume that (3.1) holds, estimate β by the inefficient but nonetheless consistent Poisson MLE, denoted $\hat{\beta}$, and calculate correct standard errors. This is particularly easy if it is assumed that the variance is a multiple τ of the mean

$$Var(y_i|X_i) = \tau \exp(X_i'\beta) \tag{3.10}$$

which is overdispersion of the Negbin 1 form. Then for the Poisson MLE

$$Var(\hat{\beta}) = \tau \left(\sum_{i=1}^{n} \exp(X_i'\beta)X_iX_i' \right)^{-1}, \tag{3.11}$$

so that correct standard errors (or t-statistics) can be obtained from those reported by a standard Poisson package by multiplying (or dividing) by $\sqrt{\hat{\tau}}$, where

$$\hat{\tau} = \frac{1}{n-k} \sum_{i=1}^{n} \frac{(y_i - \exp(X_i' \hat{\beta}))^2}{\exp(X_i' \hat{\beta})}. \tag{3.12}$$

This can often be directly calculated from computer output, as it is simply the Pearson statistic (3.19) divided by the degrees of freedom. If $\hat{\tau} = 4$, for example, the reported t-statistics need to be deflated by a factor of two.

If instead the variance is quadratic in the mean, i.e.

$$Var(y_i|X_i) = \exp(X_i'\beta) + \alpha(\exp(X_i'\beta))^2 \tag{3.13}$$

use

$$\begin{aligned} Var(\hat{\beta}) &= \left(\sum_{i=1}^{n} \exp(X_i'\beta)X_iX_i' \right)^{-1} \left(\sum_{i=1}^{n} \{\exp(X_i'\beta) + \alpha(\exp(X_i'\beta))^2\} X_iX_i' \right) \tag{3.14} \\ & \quad \left(\sum_{i=1}^{n} \exp(X_i'\beta)X_iX_i' \right)^{-1}, \end{aligned}$$

evaluated at a consistent estimate of α such as

$$\hat{\alpha} = \sum_{i=1}^{n} \exp(X_i' \hat{\beta}))^2 \{(y_i - \exp(X_i' \hat{\beta}))^2 - \exp(X_i' \hat{\beta})\} / \sum_{i=1}^{n} (\exp(X_i' \hat{\beta}))^4. \tag{3.15}$$

16

Finally, a less restrictive approach is to use the Eicker-White robust estimator

$$Var(\hat{\beta}) = \left(\sum_{i=1}^{n} \exp(X_i'\beta)X_iX_i' \right)^{-1} \left(\sum_{i=1}^{n}(y_i - \exp(X_i'\beta))^2 X_iX_i' \right) \left(\sum_{i=1}^{n} \exp(X_i'\beta)X_iX_i' \right)^{-1}) \tag{3.16}$$

which does not assume a particular model for the conditional variance.

Failure to make such corrections when data are overdispersed leads to over-statement of the statistical significance of regressors.

3.2.3. Estimation Based on First Two Moments

The previous sub-section used information on the second moment only in cal-culating the standard errors. Directly using this information in the method of estimation of β can improve efficiency.

When the variance is a multiple of the mean, the most efficient estimator using only (3.1) and (3.10) can be shown to equal the Poisson MLE, with correct standard errors calculated using (3.11) and (3.12).

When the variance is quadratic in the mean, the most efficient estimator using only (3.1) and (3.13) solves the first-order conditions

$$\sum_{i=1}^{n} \frac{(y_i - \exp(X_i'\beta))}{\exp(X_i'\beta) + \hat{\alpha} (\exp(X_i'\beta))^2} \exp(X_i'\beta))X_i = 0, \tag{3.17}$$

where the estimator $\hat{\alpha}$ is given in (3.15), and has asymptotic variance

$$Var(\hat{\beta}) = \left(\sum_{i=1}^{n} \{\exp(X_i'\beta) + \alpha(\exp(X_i'\beta))^2\}^{-1}(\exp(X_i'\beta))^2 X_iX_i' \right)^{-1}. \tag{3.18}$$

Such estimators, based on the first two moments, are called *quasi-likelihood* estima-tors in the statistics literature and quasi-generalized pseudo-maximum likelihoods estimators by Gourieroux, Montfort and Trognon (1984).

Finally, we note that an adaptive semi-parametric estimator which requires specification of only the first moment, but is as efficient as any estimator based on knowledge of the first two moments, is given by Delgado and Kniesner (1990).

17

3.2.4. Model Evaluation

An indication of the likely magnitude of underdispersion and overdispersion can be obtained by comparing the sample mean and variance of the dependent count variable, as subsequent Poisson regression will decrease the conditional variance of the dependent variable somewhat but leave the average of the conditional mean unchanged (the average of the fitted means equals the sample mean as Poisson residuals sum to zero if a constant term is included). If the sample variance is less than the sample mean, the data will be even more underdispersed once regressors are included, while if the sample variance is more than twice the sample mean the data are almost certain to still be overdispersed upon inclusion of regressors.

Formal tests for overdispersion and underdispersion, and for discrimination between Poisson and negative binomial, have been given in section 3.2.1. The choice between negative binomial models with different specification of the variance function, e.g. Negbin 1 and Negbin 2, can be made on the basis of the highest likelihood. The choice between different non-nested mixture models can also be made on the basis of highest likelihood, or using Akaike's information criterion if models have different numbers of parameters.

A more substantive choice is whether to use a fully parametric approach, such as negative binomial, or whether to use estimators that use information on only the first and second moments. In theory, fully parametric estimators have the advantage of efficiency but the disadvantage of being less robust to model departures, as even if the mean is correctly specified the MLE for count data models (aside from the Poisson and Negbin 2) will be inconsistent if other aspects of the distribution are misspecified. In practice, studies such as Cameron and Trivedi (1986) and Dean et. al. (1989) find little difference between ML estimators and estimators based on weaker assumptions. Such potential differences can be used as the basis for a Hausman test; see, for example, Dionne and Vanasse (1992). And for some analysis, such as predicting count probabilities rather than just the mean, specification of the distribution is necessary. There are a number of ways to evaluate the performance of the model. A standard procedure is to compare the *Pearson Statistic*

$$P = \sum_{i=1}^{n} \frac{(y_i - \exp(X_i' \hat{\beta}))^2}{v(X_i, \hat{\beta}, \hat{\alpha})^{1/2}}, \tag{3.19}$$

where $v(X_i, \beta, \alpha) = Var(y_i|X_i)$, to $(n-k)$, the number of degrees of freedom. This is useful for testing the adequacy of the Poisson, where $v(X_i, \beta, \alpha) = \exp(X_i'\beta)$.

18

But its usefulness for other models is more limited. In particular, if one specifies $v(X_i, \beta, \alpha) = \alpha \exp(X'_i\beta)$, and estimates α by (3.12), then P always equals $(n-k)$.

Cameron and Windmeijer (1995) propose various R-squareds for count data models. For the Poisson model their preferred deviance-based R-squared measure is

$$R^2_{DEV,P} = \frac{\sum_{i=1}^n y_i \log(\exp(X'_i \hat{\beta})/\bar{y})}{\sum_{i=1}^n y_i \log(y_i/\bar{y})} \tag{3.20}$$

where $y \log y = 0$ when $y = 0$. If a package reports the log-likelihood for the fitted model, this can be computed as $(l_{fit} - l_0)/(l_y - l_0)$ where l_{fit} is the log-likelihood for the fitted model, l_0 is the log-likelihood in the intercept-only model, and l_y is the log-likelihood for the model with mean equal to the actual value, i.e. $l_y = \sum_{i=1}^n y_i \log(y_i) - y_i - \log(y_i!)$ which is easily calculated separately. This same measure is applicable to estimation of the model with overdispersion of the form (3.10). For ML estimation of the negative binomial with overdispersion of the form (3.13), i.e. Negbin 2, the corresponding R-squared measure is

$$R^2_{DEV,NB2} = 1 - \frac{\sum_{i=1}^n y_i \log(y_i/\hat{\lambda}_i) - (y_i + \hat{\alpha}^{-1}) \log((y_i + \hat{\alpha}^{-1})/(\hat{\lambda}_i + \hat{\alpha}^{-1}))}{\sum_{i=1}^n y_i \log(y_i/\bar{y})) - (y_i + \hat{\alpha}^{-1}) \log((y_i + \hat{\alpha}^{-1})/(\bar{y} + \hat{\alpha}^{-1}))} \tag{3.21}$$

where $\hat{\lambda}_i = \exp(X'_i \hat{\beta})$.

A crude diagnostic is to calculate a fitted frequency distribution as the average over observations of the predicted probabilities fitted for each count, and to compare this to the observed frequency distribution. Poor performance on this measure is reason for rejecting a model, though good performance is not necessarily a reason for acceptance. As an extreme example, if only counts 0 and 1 are observed and a logit model with constant term is estimated by ML, it can be shown that the average fitted frequencies exactly equal the observed frequencies.

3.2.5. Some Applications to Financial Data

Examples 1-4 illustrate, respectively, Poisson (twice), negative binomial and mixed Poisson-inverse Gaussian.

Example 1: Jaggia and Thosar (1993) model the number of bids received by 126 U.S. firms that were targets of tender offers during the period 1978-1985 and were actually taken over within 52 weeks of the initial offer. The dependent count variable y_i is the number of bids after the initial bid received by the target firm, and takes values given in Table 1. Jaggia and Thosar find that the number of bids increases with defensive actions taken by target firm management (legal defense via lawsuit and invitation of bid by friendly third party), decreases with the bid premium (bid price divided by price 14 working days before bid), initially increases and then decreases in firm size (quadratic in size), and is unaffected by intervention by federal regulators. No overdispersion is found using (3.8).

Example 2: Davutyan (1989) estimates a Poisson model for data summarized in Table 1 on the annual number of bank failures in the U.S. over the period 1947 to 1981. This reveals that bank failures decrease with increases in overall bank profitability, corporate profitability, and bank borrowings from the Federal Reserve Bank. No formal test for the Poisson is undertaken. The sample mean and variance of bank failures are, respectively, 6.343 and 11.820, so that moderate overdispersion may still be present after regression and t-statistics accordingly somewhat upwardly biased. More problematic is the time series nature of the data. Davutyan tests for serial correlation by applying the Durbin-Watson test for autocorrelation in the Poisson residuals, but this test is inappropriate when the dependent variable is heteroskedastic. A better test for first-order serial correlation is based on the first-order serial correlation coefficient, r_1, of the standardized residual $(y_t - \hat{\lambda}_t)/\sqrt{\hat{\lambda}_t} : Tr_1^2$ is asymptotically $\chi^2(1)$ under the null hypothesis of no serial correlation in y_t, where T is the sample size; see Cameron and Trivedi (1993). Time series regression models for count data are in their infancy; see Gurmu and Trivedi (1994) for a brief discussion.[5]

Example 3: Dionne and Vanasse (1992) use data on the number of accidents with damage in excess of $250 reported to police during August 1982 - July 1983 by 19013 drivers in Quebec. The frequencies are very low, with sample mean of 0.070. The sample variance of 0.078 is close to the mean, but the Negbin 2 model is preferred to Poisson as the dispersion parameter is statistically significant, and the chisquare goodness-of-fit statistic is much better. The main contribution of this paper is to then use these cross-section negative binomial parameter estimates to derive predicted claims frequencies, and hence insurance premia, from data on different individuals with different characteristics and records. It is assumed

that the number of claims $(y_{i1}, \ldots, y_{iT})$ by individual i over time periods $1, \ldots, T$ are independent Poisson with means $(\lambda_{i1} v_i, \ldots, \lambda_{iT} v_i)$ where $\lambda_{it} = \exp(X'_{it}\beta)$ and v_i is a time invariant unobserved component that is gamma distributed with mean 1 and variance α.[6] Then the optimal predictor at time $T+1$ of the number of claims of the i-th individual, given knowledge of past claims, current and past characteristics (but not the unobserved component v_i) is $\exp(X'_{i,T+1}\beta) \left[\frac{1/\alpha + \bar{Y}_i}{1/\alpha + \bar{\lambda}_i} \right]$, where $\bar{Y}_i = \frac{1}{T} \sum_{t=1}^{T} y_{it}$ and $\bar{\lambda}_i = \frac{1}{T} \sum_{t=1}^{T} \exp(X'_{it}\beta)$. This is evaluated at the cross-section negative binomial estimates $(\hat{\alpha}, \hat{\beta})$. This is especially easy to implement when the regressors are variables such as age, sex and marital status whose changes over time are easily measured.

Example 4: Dean et. al. (1989) analyze data published in Andrews and Herzberg (1985) on the number of accident claims on third party motor insurance policies in Sweden during 1977 in each of 315 risk groups. The counts take a wide range of values - the median is 10 while the maximum is 2127 - so there is clearly a need to control for the size of risk group. This is done by defining the mean to equal $T_i \exp(X'_i\beta)$, where T_i is the number of insured automobile-years for the group, which is equivalent to including $\log T_i$ as a regressor and constraining its coefficient to equal unity, see (3.2). Even after including this and other regressors, the data are overdispersed. For Poisson ML estimates the Pearson statistic is 485.1 with 296 degrees of freedom, which for overdispersion of form (3.10) implies using, (3.12), that $\hat{\tau} = 1.638$, considerably greater than 1. Dean et. al. control for overdispersion by estimating by ML a mixed Poisson-inverse Gaussian model, with overdispersion of form (3.13). These ML estimates are found to be within one percent of estimates from solving (3.17) that use only the first two moments. No attempt is made to compare the estimates with those from a more conventional negative binomial model.

3.3. Truncated, Censored and Modified Count Models

In some cases only individuals who experience the event of interest are sampled, in which case the data are *left-truncated at zero* and only positive counts are observed. Let $f(y_i|X_i)$ denote the untruncated parent density, usually the Poisson or Negbin 2 defined in (3.3) or (3.6). Then the truncated density, which normalizes by $1 - f(0|X_i)$, the probability of the conditioning event that y_i exceeds zero, is $\frac{f(y_i|X_i)}{1 - f(0|X_i)}$, $y_i = 1, 2, 3, \ldots$, and the log-likelihood function is

$$\log L = \sum_{i:y_i>0} \log f(y_i|X_i) - \log(1 - f(0|X_i)). \tag{3.22}$$

Estimation is by maximum likelihood. For the Poisson model, $f(0|X_i) = \exp(-\exp(X_i'\beta))$, while for the Negbin 2 model, $f(0|X_i) = -\alpha^{-1}\log(1 + \alpha\exp(X_i'\beta))$. One could in principle estimate the model by nonlinear regression on the truncated mean, but there is little computational advantage to doing this rather than maximum likelihood. Other straightforward variations, such as left-truncation at a point greater than zero and right-truncation, are discussed in Grogger and Carson (1991) and Gurmu and Trivedi (1992).

More common than right-truncation is *right-censoring*, when counts above a maximum value, say m, are recorded only as a category m or more. Then the log-likelihood function is

$$\log L = \sum_{i:y_i<m} \log f(y_i|X_i) + \sum_{i:y_i\geq m} \log(1 - \sum_{j=0}^{m} f(j|X_i)). \tag{3.23}$$

Even if the counts are completely recorded, it may be the case that not all values for counts come from the same process. In particular, the process for zero counts may differ from the process for positive counts, due to some threshold for zero counts. An example for continuous data is the sample selectivity model used in labor supply, where the process determining whether or not someone works, i.e. whether or not hours are positive, differs from the process determining positive hours. Similarly for count data, the process for determining whether or not a credit installment is unpaid may differ from the process determining the number of unpaid installments by defaulters. *Modified count models* allow for such different processes. We consider modification of zero counts only, though the methods can be extended to other counts.

One modified model is the *hurdle* model of Mullahy (1986). Assume zeros come from the density $f_1(y_i|X_i)$, e.g. Negbin 2 with regressors X_{1i} and parameters α_1 and β_1, while positives come from the density $f_2(y_i|X_i)$, e.g. Negbin 2 with regressors X_{2i} and parameters α_2 and β_2. Then the probability of a zero value is clearly $f_1(0|X_i)$, while to ensure that probabilities sum to 1, the probability of a positive count is $\frac{1-f_1(0|X_i)}{1-f_2(0|X_i)}f_2(y_i|X_i)$, $y_i = 1, 2, \ldots$ The log-likelihood function is

$$\log L = \sum_{i:y_i=0} \log f_1(0|X_i) + \sum_{i:y_i>0} \{\log(1-f_1(0|X_i)) - \log(1-f_2(0|X_i)) + \log(f_2(y_i|X_i))\}.$$
$$\tag{3.24}$$

An alternative modification is the *with zeros* model, which combines binary and count processes in the following way. If the binary process takes value 0, an event that occurs with probability $f_1(0|X_i)$, say, then $y_i = 0$. If the binary process takes value 1, an event that occurs with probability $1 - f_1(0|X_i)$, then y_i can take count values $0, 1, 2, \ldots$ with probabilities $f_2(y_i|X_i)$ determined by a density such as Poisson or negative binomial. Then the probability of a zero value is $f_1(0|X_i) + (1 - f_1(0|X_i))f_2(0|X_i)$, while the probability of a positive count is $(1 - f_1(0|X_i))f_2(y_i|X_i)$, $y_i = 1, 2, \ldots$ The log-likelihood is

$$\log L = \sum_{i:y_i=0} \log\{f_1(0|X_i) + (1 - f_1(0|X_i))f_2(0|X_i)\} \qquad (3.25)$$

$$+ \sum_{i:y_i>0} \{\log(1 - f_1(0|X_i)) + \log f_2(y_i|X_i)\}. \qquad (3.26)$$

This model is also called the zero inflated counts model, though it is possible that it can also explain too few zero counts. This model was proposed by Mullahy (1986), who set $f_1(0|X_i)$ equal to a constant, say β_1, while Lambert (1992) and Greene (1994) use a logit model, in which case $f_1(0|X_i) = (1 + \exp(-X'_{1i}\beta_1))^{-1}$.

Problems of too few or too many zeros (or other values) can be easily missed by reporting only the mean and variance of the dependent variable. It is good practice to also report frequencies, and to compare these with the fitted frequencies.

Example 5: Guillen (1994) analyzes the number of unpaid installments for a sample of 4691 individuals granted credit by a Spanish bank. The raw data exhibit considerable overdispersion, with a mean of 1.581 and variance of 10.018. This overdispersion is still present after inclusion of regressors on age, marital status, number of children, net monthly income, housing ownership, monthly installment, credit card availability, and the amount of credit requested. For the Negbin 2 model $\hat{\alpha} = 1.340$. Interest lies in determining bad credit risks, and a truncated Negbin 2 model (3.22) is separately estimated. If the process determining zero counts is the same as that determining positive counts, then estimating just the positive counts leads to a loss of efficiency. If instead the process determining zero counts differs from that determining positive counts, then estimating the truncated model is equivalent to maximizing a subcomponent of the hurdle log-likelihood (3.24) with no efficiency loss.[7]

Example 6: Greene (1994) analyzes the number of major derogatory reports (MDR), a delinquency of sixty days or more on a credit account, of 1319 individual applicants for a major credit card. MDR's are found to decrease with increases

in the expenditure-income ratio (average monthly expenditure divided by yearly income), while age, income, average monthly credit card expenditure and whether the individual holds another credit card are statistically insignificant. The data are overdispersed, and the Negbin 2 model is strongly preferred to the Poisson. Greene also estimates the Negbin 2 with zeros model, using logit and probit models for the zeros with regressors on age, income, home ownership, self-employment, number of dependents, and average income of dependents. A with zeros model may not be necessary, as the standard Negbin 2 model predicts 1070 zeros, close to the observed 1060 zeros. The log-likelihood of the Negbin 2 with zeros model of -1020.6, with 7 additional parameters, is not much larger than that of the Negbin 2 model of -1028.3, with the former model preferable on the basis of Akaike's information criterion. Greene additionally estimates a count data variant of the standard sample selection model for continuous data.

3.4. Exponential and Weibull for Duration Data

The simplest model for duration data is the *exponential*, the duration distribution implied by the pure Poisson process, with density $\lambda e^{-\lambda t}$ and constant hazard rate λ. If data are completely observed, and the exponential is estimated when a different model such as Weibull is correct, then the exponential MLE is consistent if the mean is still correctly specified, but inefficient, and usual ML output gives incorrect standard errors. This is similar to using Poisson when negative binomial is correct. A more important reason for favoring more general models than the exponential, however, is that data are often incompletely observed, in which case incorrect distributional choice can lead to inconsistent parameter estimates. For example, observation for a limited period of time may mean that the longer spells are not observed to their completion. The restriction of a constant hazard rate is generally not appropriate for econometric data, and we move immediately to analysis of the Weibull, which nests the exponential as a special case. Our treatment is brief, as the focus of this paper is on counts rather than durations. Standard references include Kalbfleisch and Prentice (1980), Kiefer (1988) and Lancaster (1990).

The *Weibull* is most readily defined by its hazard rate $\lambda(t)$, or $h(t)$ in earlier notation, which equals $\lambda \gamma t^{\gamma-1}$. A regression model is formed by specifying λ to depend on regressors, viz. $\lambda = \exp(X'\beta)$, while γ does not. The hazard for

observation i is therefore

$$\lambda_i(t_i|X_i) = \gamma t_i^{\gamma-1} \exp(X_i'\beta), \qquad (3.27)$$

with corresponding density

$$f_i(t_i|X_i) = \gamma t_i^{\gamma-1} \exp(X_i'\beta) \exp(-t_i^\gamma \exp(X_i'\beta)). \qquad (3.28)$$

The conditional mean for this process is somewhat complicated

$$E(t_i|X_i) = (\exp(X_i'\beta))^{-1/\gamma} \Gamma(1 + 1/\gamma). \qquad (3.29)$$

Studies usually consider the impact of regressors on the hazard rate rather than the conditional mean. If $\beta_j > 0$ then an increase in X_{ij} leads to an increase in the hazard and a decrease in the mean duration, while the hazard increases (or decreases) with duration if $\gamma > 1$ (or $\gamma < 1$).

In many applications durations are only observed to some upper bound. If the event does not occur before this time the spell is said to be *incomplete*, more specifically *right-censored*. The contribution to the likelihood is the probability of observing a spell of at least t_i, or the survivor function

$$S_i(t_i|X_i) = \exp(-t_i^\gamma \exp(X_i'\beta)). \qquad (3.30)$$

Combining, the log-likelihood when some data is incomplete is

$$\log L \;=\; \sum_{i:complete} \{\log\gamma + (\gamma-1)\log t_i + X_i'\beta - t_i^\gamma \exp(X_i'\beta)\} \qquad (3.31)$$

$$+ \sum_{i:incomplete} -t_i^\gamma \exp(X_i'\beta), \qquad (3.32)$$

and γ and β are estimated by ML.

With incomplete data, the Weibull MLE is inconsistent if the model is not correctly specified. One possible misspecification is that while t_i is Weibull, the parameters are γ and $\lambda_i \upsilon_i$ rather than γ and λ_i, where υ_i is unobserved individual heterogeneity. If the distribution of υ_i is i.i.d. gamma with mean 1 and variance α, this leads to the *Weibull-gamma* model with survivor function,

$$S_i(t_i|X_i) = [1 + t_i^\gamma \exp(X_i'\beta)]^{-1/\alpha}, \qquad (3.33)$$

from which the density and log-likelihood function can be obtained in the usual manner.

The standard general model for duration data is the *proportional hazards* or *proportional intensity* model, introduced in (2.1). This factorizes the hazard rate as

$$\lambda_i(t_i, X_i, \gamma, \beta) = \lambda_0(t_i, \gamma) \exp(X_i'\beta), \tag{3.34}$$

where $\lambda_0(t_i, \gamma)$ is a baseline hazard function. Different choices of $\lambda_0(t_i, \gamma)$ correspond to different models, e.g. the Weibull is $\lambda_0(t_i, \gamma) = \gamma t_i^{\gamma-1}$ and the exponential is $\lambda_0(t_i, \gamma) = 1$. The only role of regressors is as a scale factor for this baseline hazard. The factorization of the hazard rate also leads to a factorization of the log-likelihood, with a subcomponent not depending on the baseline hazard, which is especially useful for right-censored data. Define $R(t_i) = \{j|t_j \geq t_i\}$ to be the risk set of all spells which have not yet been completed at time t_i. Then Cox (1972a) proposed the estimator which maximizes the partial likelihood

$$\log L = \sum_{i=1}^{n} \left\{ X_i'\beta - \log \left[\sum_{j \in R(t_i)} \exp(X_j'\beta) \right] \right\}. \tag{3.35}$$

This estimator is not fully efficient, but has the advantage of being consistent with correct standard errors those reported by a ML package, regardless of the true functional form of the baseline hazard.

Example 7: Bandopadhyaya (1993) analyzes data on 74 U.S. firms that were under chapter 11 bankruptcy protection in the period 1979-90. 31 firms were still under bankruptcy protection, in which case data is incomplete, and ML estimates of the censored Weibull model (3.31) are obtained. The dependent variable is the number of days in bankruptcy protection, with mean duration (computed for complete and incomplete spells) of 714 days. The coefficient of interest amount outstanding is positive, implying an increase in the hazard and decrease in mean duration of bankruptcy protection. The other statistically significant variable is a capacity utilization measure, also with positive effect on the hazard. The estimated $\hat{\alpha} = 1.629$ exceeds unity, so that firms are more likely to leave bankruptcy protection the longer they are in protection. The associated standard error, 0.385, leads to a "t-statistic" for testing the null hypothesis of exponential, $\alpha = 1$, equal to 1.63 which is borderline insignificant for a one-sided test at 5 percent. The

26

Weibull model is preferred to the exponential and the log-logistic on grounds that it provided the "best fit".

Example 8: Jaggia and Thosar (1995) analyze data on 161 U.S. firms that were the targets of tender offers contested by management during 1978-85. In 26 instances the tender offer was still outstanding, and the data censored. The dependent variable is the length of time in weeks from public announcement of offer to the requisite number of shares being tended, with mean duration (computed for complete and incomplete spells) of 18.1 weeks. The paper estimates and performs specification tests on a range of models. Different models give similar results for the relative statistical significance of different regressors, but different results for how the hazard rate varies with time since the tender offer. Actions by management to contest the tender offer, mounting a legal defense and proposing a change in financial structure, are successful in decreasing the hazard and increasing the mean duration time to acceptance of the bid, while competing bids increase the hazard and decrease the mean. The preferred model is the Censored Weibull-gamma (3.33). The estimated hazard, evaluated at $X_i = \bar{X}$, initially increases rapidly and then decreases slowly with t, whereas the Weibull gives a monotone increasing hazard rate. A criticism of models such as Weibull-gamma is that they assume that all spells will eventually be complete, whereas here some firms may never be taken over. Jaggia and Thosar give a brief discussion of estimation and rejection of the split-population model of Schmidt and Witte (1989) which allows for positive probability of no takeover. This study is a good model for other similar studies, and uses techniques readily available in LIMDEP.

3.5. Poisson for Grouped Duration Data

A leading example of state transitions in financial data is the transition from the state of having a mortgage to mortgage termination either by pre-payment of the mortgage debt or by default. Practically this is important in pricing mortgage-backed securities. Econometrically this involves modeling the time interval between a mortgage loan origination and its pre-payment or default. Specific interest attaches to the shape of the hazard as a function of the age of the mortgage and the role of covariates. The Cox proportional hazards (PH) model for durations has been widely used in this context (Green and Shoven (1986), Lane et al (1986), Baek and Bandopadhyaya (1994)). One can alternatively analyze grouped duration data as counts (Schwartz and Torous (1993)).

Example 9: Green and Shoven (1986) analyze terminations between 1975 and 1982 of 3,938 Californian 30-year fixed rate mortgages issued between 1947 and 1976. 2,037 mortgages were paid-off. Interest lies in estimating the sensitivity of mortgage prepayments to the differential between the prevailing market interest rate and the fixed rate on a given mortgage, the so-called "lock-in magnitude". The available data are quite limited, and an imputed value of this lock-in magnitude is the only regressor, so that other individual specific factors such as changes in family size or income are ignored. (The only individual level data that the authors had was the length of tenure in the house and an imputed measure of the market value of the house.) The transition probability for a mortgage of age a_i, where $a_i = t_i - t_{0i}$ and t_{0i} denotes mortgage origination date, is given by $\lambda_i(a_i, X, \beta) = \lambda_0(a_i, \gamma_i) \exp(X'\beta)$. The authors used the Cox partial likelihood estimator to estimate $(\beta, \gamma_i, \ i = 1, .., 30)$; the (nonparametric) estimate of the sequence $\{\gamma_i, \ i = 1, 2, ..\}$, somewhat akin to estimates of coefficients of categorical variables corresponding to each mortgage age, yields the baseline hazard function. The periods 1975-78 and 1978-82 are treated separately to allow for a possible structural change in the β coefficient following a 1978 court ruling which prohibited the use of due-on-sale clauses for the sole purpose of raising mortgage rates. The authors were able to show the sensitivity of average mortgage prepayment period to interest rate changes.

Example 10: Schwartz and Torous (1993) offer an interesting alternative to the Green-Shoven approach, combining the Poisson regression approach with the proportional hazard structure. Their Freddie Mac data on 30-year fixed rate mortgages over the period 1975 to 1990, has over 39,000 pre-payments and over 8,500 defaults. They use monthly grouped data on mortgage pre-payments and defaults, the two being modelled separately. Let n_j denote the number of known outstanding mortgages at the beginning of the quarter j, y_j the number of prepayments in that quarter, and $X(j)$ the set of time-varying covariates. Let $\lambda(a, X(j), \beta) = \lambda_0(a, \gamma) \exp(X(j)'\beta)$ denote the average monthly prepayment rate expressed as a function of exogenous variables $X(j)$, and a baseline hazard function $\lambda_0(a, \gamma)$. Then the expected number of quarterly prepayments will be $n_j \cdot \lambda_0(a, \gamma) \exp(X(j)'\beta)$, and ML estimation is based on the Poisson density

$$f(y_j \mid n_j, X(j)) = \frac{[n_j \cdot \lambda_0(a, \gamma) \exp(X(j)'\beta)]^{y_j} \exp(-n_j \cdot \lambda_0(a, \gamma) \exp(X(j)'\beta)}{y_j!}$$

(3.36)

The authors use dummy variables for region, quarter, and the age of mortgage in years at the time of pre-payment. Other variables include loan to value ratio at origination, refinancing opportunities and regional housing returns. Their results indicate significant regional differences and a major role for refinancing opportunities.

3.6. Other Count Models

U.S. stock prices are measured in units of one-eighth dollar (or tick), and for short time periods should be explicitly modelled as integer. For the six stocks studied in detail by Hausman, Lo and MacKinlay (1994), 60 percent of same-stock consecutive trades had no price change and a further 35 percent changed by only one tick. Even daily closing prices can experience changes of only a few ticks. This discreteness in stock prices is generally ignored, though some studies using continuous pricing models have allowed for it (Gottlieb and Kalay (1985) and Ball (1988)).

One possible approach is to model the price level (measured in number of ticks) as a count. But this count will be highly serially correlated, and time series regression models for counts are not yet well developed. More fruiful is to model the price change (again measured in number of ticks) as a count, though the standard count models are not appropriate as some counts will be negative.

A model that permits negative counts is the *ordered probit* model, presented for example in Maddala (1983). Let y_i^* denote a latent (unobserved) r.v. measuring the propensity for price to change, where $y_i^* = X_i'\beta + \varepsilon_i$, ε_i is $N(0, \sigma_i^2)$ distributed, and usually $\sigma_i^2 = 1$. Higher values of y_i^* are associated with higher values j of the actual discrete price change y_i in the following way: $y_i = j$ if $\alpha_j < y_i^* \leq \alpha_{j+1}$. Then some algebra yields

$$\begin{aligned} \Pr\{y_i = j\} &= \Pr\{\alpha_j - X_i'\beta < \varepsilon_i \leq \alpha_{j+1} - X_i'\beta\} \\ &= \Phi(\tfrac{\alpha_{j+1} - X_i'\beta}{\sigma_i}) - \Phi(\tfrac{\alpha_j - X_i'\beta}{\sigma_i}) \end{aligned} \tag{3.37}$$

Let d_{ij} be a dummy variable equal to one if $y_i = j$ and zero if $y_i \neq j$. The log-likelihood function can be expressed as

$$\log L = \sum_{i=1}^{n} \sum_j d_{ij} \log[\Phi(\frac{\alpha_{j+1} - X_i'\beta}{\sigma_i}) - \Phi(\frac{\alpha_j - X_i'\beta}{\sigma_i})] \tag{3.38}$$

This model can be applied to nonnegative count data, in which case $j = 0, 1, 2, \ldots, \max(y_i)$. Cameron and Trivedi (1986) obtained qualitatively similar results regarding the importance and significance of regressors in their application when ordered probit was used rather than Poisson or negative binomial. For discrete price change data that may be negative, Hausman et. al. (1992) use the ordered probit model, with $j = -m, -m + 1, \ldots, 0, 1, 2, \ldots, m$, where the value m is actually m or more, and $-m$ is actually $-m$ or less. Parameters to be estimated are then parameters in the model for σ_i^2, the regression parameters β, and the threshold parameters $\alpha_{-m+1}, \ldots, \alpha_m$, while $\alpha_{-m} = -\infty$ and $\alpha_{m+1} = \infty$.

Example 11: Hausman et. al. (1992) use 1988 data on time-stamped (to nearest second) trades on the New York and American Stock Exchanges for one hundred stocks, with results reported in detail for six of the stocks. Each stock is modelled separately, with one stock (IBM) having as many as 206,794 trades. The dependent variable is the price change (measured in units of $1/8) between consecutive trades. The ordered probit model is estimated, with $m = 4$ for most stocks. Regressors include the time elapsed since the previous trade, the bid/ask spread at the time of the previous trade, three lags of the price change and three lags of the dollar volume of the trade, while the variance σ_i^2 is a linear function of the time elapsed since the previous trade and the bid/ask spread at the time of the previous trade. This specification is not based on stochastic process theory, though arithmetic Brownian motion is used as a guide. Hausman et. al. conclude that the sequence of trades affects price changes and that larger trades have a bigger impact on price.

Example 12: Epps (1993) directly models the discrete stock price level (rather than change) as a stochastic process. It is assumed that the stock price at discrete time t, P_t, is the realization of a Galton-Watson process, a standard branching process, with the complication that the number of generations is also random. The conditional density (or transition probabilities) of P_t given P_{t-1} is easy to represent analytically, but difficult to compute as it involves convolutions. This makes estimation difficult if not impossible. Epps instead uses an approximation to model the (continuous) normalized price change $y_t = (P_t - P_{t-1})/\sqrt{P_{t-1}}$ which can be shown to be a realization of the Poisson compound-events distribution. Epps (1993) analyses daily individual stock closing price data from 1962 to 1987, with separate analysis for each of 50 corporations and estimation by a method of moments procedure. Advantages of the model include its prediction of a thick tail distribution for the conditional distribution of returns.

4. Concluding Remarks

The basic Poisson and negative binomial count models (and other Poisson mixture models) are straightforward to estimate with readily available software, and in many situations are appropriate. Estimation of a Poisson regression model should be followed by a formal test of underdispersion or overdispersion, using the auxiliary regressions (3.8) or (3.9). If these tests reject equidispersion, then standard errors should be calculated using (3.11), (3.14) or (3.16). If the data are overdispersed it is better to instead obtain ML estimates of the Negbin 2 model (3.6). However, it should be noted that overdispersion tests have power against other forms of model misspecification, for example the failure to account for excess zeros.

A common situation in which these models are inadequate is when the process determining zero counts differs from that determining positive counts. This may be diagnosed by comparison of fitted and observed frequencies. Modified count models, such as the hurdles or with zeros model, or models with truncation and censoring are then appropriate.

This study has emphasized the common basis of count and duration models. When data on both durations and counts are available, modelling the latter can be more informative about the role of regressors, especially when data on multiple spells for a given individual are available or when data are grouped. Grouping by a uniform time interval is convenient but sometimes the data on counts will not pertain to the same interval. One may obtain time series data on the number of events for different time intervals. Such complications can be accommodated by the use of proportional intensity Poisson process data regression models (Lawless (1987)).

The assumptions of the simplest stochastic processes are sometimes inadequate for handling financial data. An example is the number of transactions or financial trades that may be executed per small unit of time. Independence of events will not be a convincing assumption in such a case, so renewal theory is not appropriate. One approach to incorporating interdependence is use of modulated renewal processes (Cox (1972b)). For time series data on durations, rather than counts, Engle and Russell (1994) introduce the autoregressive conditional duration model which is the duration data analog of the GARCH model. This model is successful in explaining the autocorrelation in data on the number of seconds between consecutive trades of IBM stock on the New York Stock Exchange.

Time series count regression models are relatively undeveloped, except the pure time series case which is very limited. In fact, techniques for handling most of the standard complications considered by econometricians, such as simultaneity and selection bias, are much less developed for count data than they are for continuous data. A useful starting point is the survey by Gurmu and Trivedi (1994).

Acknowledgement

References

Andrews, D.F. and Herzberg, A.M. (1985). *Data.* Springer-Verlag, New York.

Baek, I-M., and A. Bandopadhyaya (1994). The Determinants of the Duration of Commercial Bank Debt Renegotiation for Sovereigns. *Journal of Banking and Finance*, to appear.

Ball, C. A. (1988). Estimation Bias Induced by Discrete Security Prices. *Journal of Finance* **43**, 841-865.

Bandopadhyaya, A. (1994). An Estimation of the Hazard Rate of Firms Under Chapter 11 Protection. *Review of Economics and Statistics*, 346-350.

Cameron, A.C. and P.K. Trivedi (1986). Econometric Models Based on Count Data: Comparisons and Applications of Some Estimators and Tests. *Journal of Applied Econometrics* **1** (1), 29-54.

Cameron, A.C. and P.K. Trivedi (1990). Regression Based Tests for Overdispersion in the Poisson Model. *Journal of Econometrics* **46** (3), 347-364.

Cameron, A.C. and P.K. Trivedi (1993). Tests of Independence in Parametric Models with Applications and Illustrations. *Journal of Business and Economic Statistics* **11**, 29-43.

Cameron, A.C. and F. Windmeijer (1995). R-Squared Measures for Count Data Regression Models with Applications to Health Care Utilization. *Journal of Business and Economic Statistics,* to appear.

Consul, P.C. and F. Famoye (1992). Generalized Poisson Regression Model. *Communications in Statistics: Theory and Method* **21** (1) 89-109.

Cox, D.R. (1962). *Renewal Theory.* Methuen, London.

Cox, D.R. (1972a). Regression Models and Life Tables. *Journal of the Royal Statistical Society, Series B.* **34**, 187-220.

Cox, D.R. (1972b). The Statistical Analysis of Dependencies in Point Pro-

cesses. in P.A.W. Lewis ed., *Stochastic Point Processes.* John Wiley and Sons, New York.

Cox, D.R. and P.A.W. Lewis (1966). *The Statistical Analysis of Series of Events.* Methuen, London.

Davutyan, N. (1989). Bank Failures as Poisson Variates. *Economic Letters* **29** (4), 333-338.

Dean, C., J. F. Lawless, and G. E. Wilmot (1989). A Mixed Poisson- Inverse Gaussian Regression Model. *Canadian Journal of Statistics* **17** (2). 171-181.

Delgado, M.A. and T.J. Kniesner (1990). Monte Carlo Comparisons and Estimates of an Econometric Model of Worker Absenteeism. Working Papers in Economics No. 90-018, Indiana University, Bloomington.

Dionne, G., and C. Vanasse (1992). Automobile Insurance Ratemaking in the Presence of Asymmetric Information. *Journal of Applied Econometrics* **7** (2), 149-166.

Engle, R.F. and J.R. Russell (1994). Forecasting Transaction Rates: The Autoregresive Conditional Duration Model. Working Paper No. 4966, National Bureau of Economic Research, Cambridge, Massachusetts.

Epps, W. (1993). Stock Prices as a Branching Process. Department of Economics, University of Virginia, Charlottesville.

Feller, W. (1966). *An Introduction to Probability Theory,* vol II. New York: Wiley.

Gottlieb, G. and A. Kalay (1985). Implications of the Discreteness of Observed Stock Prices. *Journal of Finance* **40** (1), 135-153.

Gourieroux, C., A. Montfort and A. Trognon (1984). Pseudo Maximum Likelihood Methods: Applications to Poisson Models. *Econometrica* **52** (3), 681-700.

Green, J. and J. Shoven (1986). The Effects of Interest Rates on Mortgage Prepayments. *Journal of Money, Credit and Banking* **18** (1), 41-59.

Greene, W. H. (1994). Accounting for Excess Zeros and Sample Selection in Poisson and Negative Binomial Regression Models. Discussion Paper EC-94-10, Department of Economics, New York University, New York.

Grogger, J.T. and R.T. Carson (1991). Models for Truncated Counts. *Journal of Applied Econometrics* **6** (3), 225-238.

Guillen, M. (1994). Count Data Models for a Credit Scoring System. Bank of Spain, unpublished manuscript.

Gurmu, S. and P.K. Trivedi (1992). Overdispersion Tests for Censored Poisson Regression Models. *Journal of Econometrics* **54**, 347-370.

Gurmu, S. and P.K. Trivedi (1994). Recent Developments in Models of Event Counts: A Survey. Discussion Paper No.261, Thomas Jefferson Center, University of Virginia, Charlottesville.

Hausman, J.A., A.W. Lo and A.C. MacKinlay (1992). An Ordered Probit Analysis of Transaction Stock Prices. *Journal of Financial Economics* 31, 319-379.

Jaggia, S., and S. Thosar (1993). Multiple Bids as a Consequence of Target Management Resistance: A Count Data Approach. *Review of Quantitative Finance and Accounting.* December, 447-457.

Jaggia, S., and S. Thosar (1995). Contested Tender Offers: An Estimate of the Hazard Function. *Journal of Business and Economic Statistics* **13** (1), 113-119.

Kalbfleisch J. and R. Prentice (1980). *The Statistical Analysis of Failure Time Data.* John Wiley and Sons, New York.

Karlin, S. and H. Taylor (1975). *A First Course in Stochastic Processes*, 2nd. Ed.. Academic Press, New York.

Kiefer, N.M. (1988). Econometric Duration Data and Hazard Functions. *Journal of Economic Literature* **26** (2), 646-679.

King, G. (1989). Variance Specification in Event Count Models: From Restrictive Assumptions to a Generalized Estimator. *American Journal of Political Science* **33**, 762-784.

Lambert, D. (1992). Zero-Inflated Poisson Regression with an Application to Defects in Manufacturing. *Technometrics* **34**, 1-14.

Lancaster, T. (1990). *The Econometric Analysis of Transition Data.* Cambridge University Press, Cambridge.

Lane, W., S. Looney and J. Wansley (1986). An Application of the Cox Proportional Hazard Model to Bank Failures. *Journal of Banking and Finance* **18** (4), 511-532.

Lawless, J.F. (1987). Regression Methods for Poisson Process Data. *Journal of the American Statistical Association* **82** (399), 808-815.

Maddala, G.S. (1983). *Limited-Dependent and Qualitative Variables in Econometrics.* Cambridge University Press, Cambridge.

McCullagh, P. and J.A. Nelder (1989). *Generalized Linear Models*, 2nd Ed. Chapman and Hall, London.

Mullahy, J. (1986). Specification and Testing of Some Modified Count Data Models. *Journal of Econometrics* **33** (3), 341-365.

Schmidt, P. and A. Witte (1989). Predicting Criminal Recidivism using Split

Population Survival Time Models. *Journal of Econometrics* **40** (1), 141-159.

Schwartz, E. S. and W.N. Torous (1993). Mortgage Prepayment and Default Decisions: A Poisson Regression Approach. *AREUEA Journal: Journal of the American Real Estate Institute* **21** (4), 431-449.

Winkelmann, R. (1993). A Count Data Model for Waiting TImes. Discussion Paper, University of Munich.

Winkelmann, R. (1994). *Count Data Models: Econometric Theory and an Application to Labor Mobility.* Springer-Verlag, Berlin.

Footnotes

[1] A spell may be in progress (incomplete) at the time of sampling. Inclusion of such censored observations in regression analysis is a key feature of duration models.

[2] The process begins at time 0, while the observed time interval starts at time s.

[3] The rate of occurrence for a renewal Weibull process is determined by the time since the previous event, when it is "renewed". For a time-dependent Weibull process it is instead determined by the time since the start of the process.

[4] If $F(t)$ is the distribution function of a random variable T, $T > 0$, then the Laplace transform of F is $L(s) = \int_0^\infty e^{-st} dF(t) = E(e^{-sT})$. If $T = t_1 + t_2 + ... + t_n$, then the Laplace transform of T is $L(s) = \prod_{i=1}^n L_i(s)$. Laplace transforms have a property of uniqueness in the sense that to any transform there corresponds a unique probability distribution.

[5] This is entirely analogous to the consequences of estimating the linear regression model by MLE under the assumption of normality and homoskedastic error, when in fact the error is non-normal and heteroskedastic but still has mean zero so that the conditional mean is correctly specified.

[6] This implies that in each time period the claims are Negbin 2 distributed.

[7] The hurdle log-likelihood (3.24) is additive in f_1 and f_2, the f_2 subcomponent equals (3.22) and the information matrix is diagonal if there are no common parameters in f_1 and f_2.

Table 1: Frequencies for Some Count Variables

Author	Jaggia-Thosar	Greene	Guillen	Davutyan
Count Variable	Takeover Bids	Derogatory	Credit	Bank
	after first	Credit Reports	Defaults	Failures
Sample Size	126	1319	4691	40
Mean	1.738	0.456	1.581	6.343
Variance	2.051	1.810	10.018	11.820
Counts...				
0	9	1060	3002	0
1	63	137	502	0
2	31	50	187	2
3	12	24	138	7
4	6	17	233	4
5	1	11	160	4
6	2	5	107	4
7	1	6	80	1
8	0	0	59	3
9	0	2	53	5
10	1	4	41	3
11	0	4	28	0
12	0	1	34	0
13	0	0	10	0
14	0	1	13	1
15	0	0	11	0
16	0	0	4	0
17	0	0	5	1
>17	0	0	24^a	5^b

a/ The large counts are 17 (5 times), 18 (8), 19 (6), 20 (3), 22 (1), 24 (1), 28 (1), 29 (1), 30 (1), 34 (1).

b/ The large counts are 42 (1), 48 (1) 79 (1), 120 (1), 138 (1).

Table 2: Standard Parametric Count Distributions and Their Moments

Family	Density	Count	Mean; Variance
Poisson	$f(y) = \frac{\exp(-\lambda)\cdot\lambda^y}{y!}$	$y = 0, 1, ...$	$\lambda; \lambda$
Negative Binomial	$f(y) = \frac{\Gamma(y+v)}{\Gamma(v)\Gamma(y+1)} \left(\frac{v}{\lambda+v}\right)^v \left(\frac{\lambda}{\lambda+v}\right)^y$	$y = 0, 1, ...$	$\lambda; \lambda + \frac{1}{v}\lambda^2$
Positive Counts	$f(y \mid y \geq 0) = \frac{f(y)}{1-F(0)}$	$y = 1, 2, ...$	Vary with f
Hurdles	$f(y) = f_1(0), y = 0$	$y = 0$	Vary with f_1, f_2
	$= \frac{1-f_1(0)}{1-f_2(0)} \cdot f_2(y)$	$y = 1, 2, ...$	
With Zeroes	$f(y) = f_1(0) + (1 - f_1(0)) \cdot f_2(y)$	$y = 0$	Vary with f_1, f_2
	$= (1 - f_1(0)) \cdot f_2(y)$	$y = 1, 2, ...$	

Table 3: Finance Applications

Example	Dependent Variable	Model
1. Jaggia and Thosar	Bids received by target firm	Poisson
2. Davutyan	Bank Failures per year	Poisson
3. Dionne and Vanasse	Accidents per person	Negative Binomial
4. Dean et. al.	Accident claims	Inverse Gaussian - Poisson
5. Guillen	Unpaid instalments	Truncated Negative Binomial
6. Greene	Derogatory credit reports	With Zeros Negative Binomial
7. Bandopadyaya	Time in bankruptcy protection	Censored Weibull
8. Jaggia and Thosar	Time to tender offer accepted	Censored Weibull-gamma
9. Green and Shoven	Mortgage prepayments	Proportional hazards
10. Schwartz and Torous	Mortgage prepayment or default	Grouped proportional hazards
11. Hausman et. al.	Stock price change	Ordered probit
12. Epps	Normalized stock price change	Poisson Compound-events