

# Count Models Based on Weibull Interarrival Times

**Blake McSHANE**

Department of Statistics, Wharton School, University of Pennsylvania, Philadelphia, PA 19104

**Moshe ADRIAN**

Department of Mathematics, University of Maryland, College Park, MD 20742

**Eric T. BRADLOW and Peter S. FADER**

Department of Marketing, Wharton School, University of Pennsylvania, Philadelphia, PA 19104  
([ebradlow@wharton.upenn.edu](mailto:ebradlow@wharton.upenn.edu))

The widespread popularity and use of both the Poisson and the negative binomial models for count data arise, in part, from their derivation as the number of arrivals in a given time period assuming exponentially distributed interarrival times (without and with heterogeneity in the underlying base rates, respectively). However, with that clean theory come some limitations including limited flexibility in the assumed underlying arrival rate distribution and the inability to model underdispersed counts (variance less than the mean). Although extant research has addressed some of these issues, there still remain numerous valuable extensions. In this research, we present a model that, due to computational tractability, was previously thought to be infeasible. In particular, we introduce here a generalized model for count data based upon an assumed Weibull interarrival process that nests the Poisson and negative binomial models as special cases. The computational intractability is overcome by deriving the Weibull count model using a polynomial expansion which then allows for closed-form inference (integration term-by-term) when incorporating heterogeneity due to the conjugacy of the expansion and a commonly employed gamma distribution. In addition, we demonstrate that this new Weibull count model can (1) model both over- and underdispersed count data, (2) allow covariates to be introduced in a straightforward manner through the hazard function, and (3) be computed in standard software.

KEY WORDS: Closed-form inferences; Hazard models; Polynomial expansions.

## 1. INTRODUCTION

The widespread popularity of the Poisson model for count data arises, in part, from its derivation as the number of arrivals in a given time period assuming exponentially distributed interarrival times. But of the thousands of other count models that have been developed over the years (see Wimmer and Altmann 1999 for an excellent synthesis), very few share this straightforward connection between a count model and its timing model equivalent. The connection between a count model and a timing process is more than just a theoretical nicety: in many different contexts, it is useful—if not essential—for a researcher to be able to estimate a model using one form (timing or counting) but apply it using the other. As but one example, marketing managers frequently collect interarrival time data (often in the form of a recency question) but want to make predictions of the number of arrivals (purchases) that a particular customer is likely to make over the next year.

Furthermore, the Poisson count model is truly valid only in the case where the data of interest support the restrictive assumption of *equidispersion*, that is, where the variance of the data equals the mean. Statisticians have recognized this limitation for many years, and now routinely use models that allow for *overdispersion* (i.e., datasets marked by a fatter, longer right tail than the Poisson can accommodate). A heterogeneous gamma-Poisson model (i.e., the negative binomial or NBD) is generally the first count model invoked for this common situation. But what about datasets with the opposite problem, namely, *underdispersion*? Statisticians have acknowledged and addressed this issue in different ways (King 1989; Cameron and

Johansson 1997; Cameron and Trivedi 1998), but with the possible exception of a count model featuring gamma-distributed interarrival times proposed by Winkelmann (1995), none of these underdispersed count models (to the best of our knowledge) offers the conceptual elegance and usefulness of the Poisson-exponential connection.

Winkelmann (1995) readily admitted the limitations of his gamma-based model. Among other reasons, he commented on the inability to obtain a closed-form hazard function for the gamma, which makes the incorporation of explanatory variables an ad hoc process when compared to the standard Poisson or NBD “regression” models. He pointed out that “the Weibull distribution is preferred in duration analysis for its closed-form hazard function. . .” but did not pursue such a model. The development and exploration of such a model is the main objective of the present article.

Before we develop our Weibull count model, we first set the stage by laying out the main properties that the Weibull count model developed here embodies.

- (1) The model generalizes (nests) the most commonly used extant models such as the Poisson and the NBD as special cases; thus, when a simple structure is sufficient, the researcher will see it through the estimated model parameters. Furthermore, standard inferential procedures

(e.g., the likelihood ratio test) can be used to compare different specifications.

- (2) The model handles both overdispersed and underdispersed data, both of which are likely to be seen in practice.
- (3) Researchers who believe that the interarrival times of their dataset are Weibull distributed now have a corresponding counting model to use.
- (4) The model is computationally feasible to work with: it is estimable without requiring a formal programming language or time-consuming simulation-based methods.
- (5) The model allows for the incorporation of person-level heterogeneity reflecting the fact that individuals' interarrival rates may vary quite substantially across the population.
- (6) The mechanism required to incorporate covariate effects is clear and simple. This process is consistent with standard "proportional-hazards" methods, which represent the dominant paradigm for ordinary single-event timing models.

In this article we derive a new model for count data that satisfies these six criteria in the following ways. First, our count model is based upon an assumed Weibull interarrival process, which nests the exponential as a well-known special case. Second, we demonstrate that the Weibull count model, via the shape parameter being less than, equal to, or greater than 1, can capture overdispersed, equidispersed, and underdispersed data, respectively. Third, the Weibull interarrival time story is richer than the exponential story, because it allows for nonconstant hazard rates (duration dependence). Fourth, we implement the model entirely in standard software. This is accomplished by deriving our model using a polynomial expansion (which can be expressed in closed form). See Bradlow, Hardie, and Fader (2000), Everson and Bradlow (2002), and Miller, Bradlow, and Dayaratna (2006) for similar polynomial expansion solutions for the negative binomial, beta-binomial, and binary logit models, respectively. Fifth, and related to the previous point, once the model is expressed as a closed-form sum of polynomial terms, we can easily introduce a conjugate mixing distribution (the gamma distribution) to capture the underlying dispersion in incidence rates across individuals. This ensures that our model nests the NBD in addition to the Poisson. Finally, we will demonstrate that we can use the proportional-hazards approach to introduce covariates in a very natural manner.

The remainder of this article is laid out as follows. In the next section, we provide a more detailed description of the major ways in which other researchers have extended basic count models (but rarely with an eye toward maintaining a known interarrival timing process). Section 3 contains the derivation of our Weibull count model, focusing on the polynomial expansion that leads to the closed-form benefits. In Section 4 we re-analyze the same data used by Winkelmann (1995) and provide a set of results comparing a sequence of nested models, the most complicated of which has an underlying Weibull arrival process, heterogeneous baseline rates, and covariates. Through the sequence of models that we fit, we are able to ascertain which aspects of the model are most critical. Finally, we provide some concluding remarks and areas for future research in Section 5.

## 2. PRIOR RELATED RESEARCH

The primary way in which this research contributes to the literature on count data is by generalizing the underlying interarrival timing model to allow for greater flexibility in its hazard function, which is how flexible forms of dispersion are accounted for (as described later). For example, Winkelmann (1995) offered a careful analysis of a counting model based on gamma-distributed interarrival times and discussed the relationship between the nature (i.e., slope) of the timing model hazard function and the type of dispersion seen in the equivalent count data. In particular, if we denote the mean of the interarrival distribution by  $\mu$ , the variance by  $\sigma^2$ , and the hazard function by

$$h(t) = \frac{f(t)}{1 - F(t)},$$

where  $f(t)$  and  $F(t)$  are the density and cumulative probability functions, respectively, we say that the distribution has negative duration dependence if  $dh(t)/dt < 0$  and positive duration dependence if  $dh(t)/dt > 0$ . If the hazard function is monotonic, then

$$\frac{dh(t)}{dt} > 0 \Rightarrow \frac{\sigma}{\mu} < 1,$$

$$\frac{dh(t)}{dt} = 0 \Rightarrow \frac{\sigma}{\mu} = 1,$$

$$\frac{dh(t)}{dt} < 0 \Rightarrow \frac{\sigma}{\mu} > 1$$

(see Barlow and Proschan 1965, p. 33). These three cases correspond to count data characterized by underdispersion, equidispersion, and overdispersion, respectively.

Focusing on nonconstant hazard rates (as above) is but one way in which researchers have extended count models; we discuss some other methods briefly. Another way to capture the same kinds of patterns seen in duration-dependent models is to assume that the probability of an event occurring depends on the *number of events* that have occurred previously, as opposed to the *arrival time* of the most recent event (duration dependence). These models are said to display contagion. For instance, they have been studied in the literature on accident proneness (Arbous and Kerrich 1951; Feller 1943). For more information, one can reference Gurland and Sethuraman (1995) for a contagious discrete-time model that leads to the negative binomial in which an occurrence increases and a nonoccurrence decreases the probability of a future occurrence. Other models for occurrence dependence have been developed by Mullahy (1986) and Gouieroux and Visser (1997). One can also make the assumption that successive events are independent but the process intensity varies as a function of time. This class of models is known as nonhomogeneous Poisson processes and is described in Lawless (1987). We believe that a promising area for future research would be a comparison of both forms of dependence (duration and occurrence), although here we focus only on the former.

Beyond an explicit focus on any kind of time dependence, there are many other distributions that have been formulated to be able to accommodate underdispersed as well as overdispersed data. Researchers such as Bening and Korolev (2002), Cameron and Trivedi (1998), King (1989), and Shmueli,

Minka, Kadane, Borle, and Boatwright (2005) have proposed and discussed a wide variety of generalized count models that can handle overdispersion and underdispersion. But few (if any) offer the benefits or elegance of something like the Poisson-exponential connection. In the next section we lay out our model that fully respects this connection and also offers a great deal of flexibility in being able to capture a range of count data dispersion patterns.

### 2.1 A Modeling Framework

Much extant research on count data has been focused on extending the basic Poisson model (denoted here as model [0]) to allow for hyperdispersion via a nonconstant hazard rate. The basic ways in which hyperdispersion has been accounted for include (model [1]) adding covariates to the model, (model [2]) incorporating individual-level heterogeneity for the baseline rates, and (model [3]) both [1] and [2]. In particular, if we let

$$[X_{it}|\lambda_i] \sim \text{Poisson}(\lambda_i \exp(Z'_{it}\beta)), \quad (1)$$

a proportional-hazards framework (Cox 1972), where  $X_{it}$  is a nonnegative integer (count) for unit  $i = 1, \dots, I$  on its  $t = 1, \dots, T_i$ th observation,  $\lambda_i$  is the baseline rate for unit  $i$ ,  $Z_{it} = (Z_{it1}, \dots, Z_{itP})$  is a vector of covariates that describe each individual, and  $\beta' = (\beta_1, \dots, \beta_P)$  is a vector of covariate slopes: model [0] is obtained by setting  $\lambda_i = \lambda$  for all  $i$  and  $Z'_{it}\beta = 0$  (an intercept only); model [1] is obtained by setting  $\lambda_i = \lambda$  for all  $i$  (the Poisson regression model); model [2] is obtained by setting  $P = 1$ ,  $Z_{it}\beta = 0$  and letting  $\lambda_i \sim g(\lambda_i|\theta)$  (when  $g$  is the gamma distribution, then model [2] integrated over the distribution of  $\lambda_i$  is the negative binomial distribution); and model [3] is as given in (1) where again  $\lambda_i \sim g(\lambda_i|\theta)$ . Model [3] is also sometimes referred to as the NegBin II model or a random-intercepts Poisson regression model. Later in Section 4, we compare the results of models [0]–[3] to those derived in this research.

What is of interest to note is that all of these extensions use the Poisson model (with associated exponential interarrival times) as their kernel. That is, these extensions to the model have not been done at the core unit of analysis, that is, the underlying arrival time distribution, but instead work strictly with the count model. What we do in this research is to enhance the flexibility of the arrival time model to account for richer patterns. In particular, instead, we assume that the underlying arrival time distribution for  $Y_{ik}$ , the  $k$ th arrival for unit  $i$ , follows a Weibull distribution with density given by

$$f(Y_{ik} = y|\lambda_i, \beta, c) = \lambda_i c y^{c-1} \exp(-\lambda_i y^c). \quad (2)$$

Later, when we introduce covariates into the model, we do it through the hazard function:

$$h(t) = \lambda c t^{c-1}, \quad (3)$$

which is monotonically increasing for  $c > 1$ , monotonically decreasing for  $c < 1$ , and constant (and equal to  $\lambda$ ) when  $c = 1$ .

Using the standard proportional-hazards framework, we then boost this “baseline” hazard [given in (3)] by a weighted vector of the covariates  $h(t) = h_0(t) \exp(\beta'Z)$ , and then rely on the well-known relationship between the hazard function and the cdf:

$$F(t) = 1 - \exp\left(-\int (h(u) du)\right)$$

to arrive at the Weibull regression model

$$f(Y_{ik} = y|\lambda_i, \beta, c) = \lambda_i \exp(Z'_{it}\beta) c y^{c-1} \exp(-\lambda_i \exp(Z'_{it}\beta) y^c). \quad (4)$$

We note that when  $c = 1$ , (4) simplifies to a heterogeneous exponential arrival time model with covariates that leads to count models [0]–[3] above.

Thus, directly analogous to models [0]–[3] which are based on an exponential interarrival time, our interest lies in looking at various reduced-form specifications of the model given in (4). Specifically, we denote as model [4] the Weibull model without heterogeneity and without covariates (model [0] analog) such that  $\lambda_i = \lambda$  and  $Z'_{it}\beta = 0$ . We label model [5] as the Weibull regression model (without heterogeneity) such that  $\lambda_i = \lambda$ . Model [6] is the model (to be discussed in Sec. 3.2) in which we allow for heterogeneity in baseline rates  $\lambda_i$  but do not include covariates ( $Z'_{it}\beta = 0$ ). Finally, model [7] is the fully parameterized model that includes heterogeneity and covariates. All eight of these models will be fit and results compared in Section 4.

It is important to note that whereas the baseline Weibull model proposed here in (3) allows for increasing, constant, or decreasing hazard, the individual-level hazard function is always monotonic. The way in which this restriction is addressed here is via the inclusion of time-varying covariates,  $Z_{it}$ , as in (4), and with the gamma mixing distribution. These additions provide a great deal of flexibility to accommodate virtually any kind of observable (i.e., aggregate) hazard function. Alternatively, one can start with a nonmonotonic hazard function at the individual level, for example, using a generalized Weibull model (Mudholkar, Srivastava, and Kollia 1996). Whereas such a distribution would provide even greater flexibility, we are cautious about taking this step. When heterogeneity and/or covariates are also included in the model, it becomes difficult to sort out the various effects, and computational concerns arise as well. Our experience suggests that there is already enough flexibility with the proposed model, and further generalizations— involving difficult trade-offs between modeling flexibility, parameter identification, and data requirements—should be approached with great care.

### 3. BASIC THEORY AND DEFINITIONS

Before discussing the Weibull count model itself, we describe the general framework utilized to derive the model that is based upon the relationship between interarrival times and their count model equivalent. Let  $Y_n$  be the time from the measurement origin at which the  $n$ th event occurs. Let  $X(t)$  denote the number of events that have occurred up until time  $t$ . The relationship between interarrival times and the number of events is

$$Y_n \leq t \iff X(t) \geq n.$$

We can restate this relationship by saying that the amount of time at which the  $n$ th event occurred from the time origin is less than or equal to  $t$  if and only if the number of events that have occurred by time  $t$  is greater than or equal to  $n$ .

We therefore have the following relationships that allow us to derive our Weibull count model  $C_n(t)$ :

$$C_n(t) = P(X(t) = n) = P(X(t) \geq n) - P(X(t) \geq n + 1) \\ = P(Y_n \leq t) - P(Y_{n+1} \leq t). \tag{5}$$

If we let the cumulative density function (cdf) of  $Y_n$  be  $F_n(t)$ , then  $C_n(t) = P(X(t) = n) = F_n(t) - F_{n+1}(t)$ . In the case where the measurement time origin (and thus the counting process) coincides with the occurrence of an event, then  $F_n(t)$  is simply the  $n$ -fold convolution of the common interarrival time distribution which may or may not have a closed-form solution. Based upon (5), we derive our Weibull count model next based upon a polynomial expansion of  $F(t)$ .

### 3.1 Weibull Count Model

We derive the basic Weibull count model, model [4] from earlier, by assuming that the interarrival times are independent and identically distributed Weibull with probability density function (pdf)  $f(t) = \lambda c t^{c-1} e^{-\lambda t^c}$  ( $c, \lambda \in R^+$ ), and corresponding cdf  $F(t) = 1 - e^{-\lambda t^c}$ , which simplifies to the exponential model when  $c = 1$ .

The challenge in deriving the Weibull count model arises in the need to be able to evaluate convolutions of the form  $\int_0^t F(t-s)f(s) ds$ . Whereas this integral is easily solved for the exponential density as well as the gamma with an integer-value shape parameter (a.k.a. the Erlang distribution), it does not have a proper solution for the Weibull. Thus, our approach is to handle this integral (and derive the Weibull count model as a whole) using a Taylor series expansion of the Weibull density.

In particular, the Taylor series approximations obtained by expanding the exponential pieces ( $e^{-\lambda t^c}$ ), respectively, for both the cdf and pdf of the Weibull are

$$F(t) = \sum_{j=1}^{\infty} \frac{(-1)^{j+1} (\lambda t^c)^j}{\Gamma(j+1)} \tag{6}$$

and

$$f(t) = \sum_{j=1}^{\infty} \frac{(-1)^{j+1} c_j \lambda^j t^{cj-1}}{\Gamma(j+1)}. \tag{7}$$

Utilizing, as in (5), that  $C_n(t) = F_n(t) - F_{n+1}(t)$ , we obtain the following recursive relationship that we utilize in deriving the Weibull count model:

$$C_n(t) = \int_0^t F_{n-1}(t-s)f(s) ds - \int_0^t F_n(t-s)f(s) ds \\ = \int_0^t C_{n-1}(t-s)f(s) ds. \tag{8}$$

Before proceeding to develop the general solution to the problem, we note that  $F_0(t)$  is 1 for all  $t$  and  $F_1(t) = F(t)$ . Therefore, we have  $C_0(t) = F_0(t) - F_1(t) = e^{-\lambda t^c} = \sum_{j=0}^{\infty} \frac{(-1)^j (\lambda t^c)^j}{\Gamma(j+1)}$ . Using the recursive formula in (8), we can therefore compute  $C_1(t)$ :

$$C_1(t) = \int_0^t C_0(t-s)f(s) ds$$

$$= \int_0^t \left( \sum_{j=0}^{\infty} \frac{(-1)^j (\lambda(t-s)^c)^j}{\Gamma(j+1)} \right) \\ \times \left( \sum_{k=1}^{\infty} \frac{(-1)^{k+1} c k \lambda^k s^{ck-1}}{\Gamma(k+1)} \right) ds \\ = \sum_{j=0}^{\infty} \sum_{k=1}^{\infty} \frac{(-1)^j (-1)^{k+1} (\lambda)^j (\lambda)^k}{\Gamma(j+1) \Gamma(k+1)} \int_0^t c k (t-s)^{cj} s^{ck-1} ds \\ = \sum_{j=0}^{\infty} \sum_{k=1}^{\infty} \frac{(-1)^j (-1)^{k+1} (\lambda)^j (\lambda)^k}{\Gamma(j+1) \Gamma(k+1)} \\ \times \frac{(t)^{cj} (t)^{ck} \Gamma(cj+1) \Gamma(ck+1)}{\Gamma(cj+ck+1)}. \tag{9}$$

Then, by using a change of variables  $m = j$  and  $l = m + k$ , we obtain

$$= \sum_{l=1}^{\infty} \left( \sum_{m=0}^{l-1} \frac{(-1)^m (-1)^{l-m+1} (\lambda)^m (\lambda)^{l-m}}{\Gamma(m+1) \Gamma(l-m+1)} \right) \\ \times \frac{(t)^{cm} (t)^{cl-cm} \Gamma(cm+1) \Gamma(cl-cm+1)}{\Gamma(cm+cl-cm+1)} \\ = \sum_{l=1}^{\infty} \frac{(-1)^{l+1} (\lambda t^c)^l}{\Gamma(cl+1)} \left( \sum_{m=0}^{l-1} \frac{\Gamma(cm+1) \Gamma(cl-cm+1)}{\Gamma(m+1) \Gamma(l-m+1)} \right) \\ = \sum_{l=1}^{\infty} \frac{(-1)^{l+1} (\lambda t^c)^l \alpha_m^l}{\Gamma(cl+1)},$$

where

$$\alpha_m^l = \sum_{m=0}^{l-1} \frac{\Gamma(cm+1) \Gamma(cl-cm+1)}{\Gamma(m+1) \Gamma(l-m+1)}.$$

This suggests a general form for  $C_n(t)$ , namely,  $\sum_{l=n}^{\infty} \frac{(-1)^{l+n} (\lambda t^c)^l \alpha_n^l}{\Gamma(cl+1)}$ , which is confirmed by

$$C_{n+1}(t) = \int_0^t C_n(t-s)f(s) ds \\ = \int_0^t \left( \sum_{j=n}^{\infty} \frac{(-1)^{j+n} (\lambda(t-s)^c)^j \alpha_j^n}{\Gamma(cj+1)} \right) \\ \times \left( \sum_{k=1}^{\infty} \frac{(-1)^{k+1} c k \lambda^k s^{ck-1}}{\Gamma(k+1)} \right) ds \\ = \sum_{j=n}^{\infty} \sum_{k=1}^{\infty} \frac{(-1)^{j+n} (-1)^{k+1} (\lambda)^j (\lambda)^k \alpha_j^n}{\Gamma(cj+1) \Gamma(k+1)} \\ \times \int_0^t c k (t-s)^{cj} s^{ck-1} ds \\ = \sum_{j=n}^{\infty} \sum_{k=1}^{\infty} \frac{(-1)^{j+n} (-1)^{k+1} (\lambda)^j (\lambda)^k \alpha_j^n}{\Gamma(cj+1) \Gamma(k+1)} \\ \times \frac{(t)^{cj} (t)^{ck} \Gamma(cj+1) \Gamma(ck+1)}{\Gamma(cj+ck+1)}$$

$$\begin{aligned}
 &= \sum_{l=n+1}^{\infty} \frac{(-1)^{l+n+1} (\lambda t^c)^l}{\Gamma(cl+1)} \left( \sum_{m=n}^{l-1} \alpha_m^n \frac{\Gamma(cl-cm+1)}{\Gamma(l-m+1)} \right) \\
 &= \sum_{l=n+1}^{\infty} \frac{(-1)^{l+1} (\lambda t^c)^l \alpha_l^{n+1}}{\Gamma(cl+1)}, \tag{10}
 \end{aligned}$$

where  $\alpha_l^{n+1} = \sum_{m=n}^{l-1} \alpha_m^n \Gamma(cl-cm+1)/\Gamma(l-m+1)$ .

Therefore, we have the main result of this article, the Weibull count model:

$$P(N(t) = n) = C_n(t) = \sum_{j=n}^{\infty} \frac{(-1)^{j+n} (\lambda t^c)^j \alpha_j^n}{\Gamma(cj+1)}, \tag{11}$$

$n = 0, 1, 2, \dots,$

where  $\alpha_j^0 = \Gamma(cj+1)/\Gamma(j+1)$ ,  $j = 0, 1, 2, \dots$ , and  $\alpha_j^{n+1} = \sum_{m=n}^{j-1} \alpha_m^n \Gamma(cj-cm+1)/\Gamma(j-m+1)$ , for  $n = 0, 1, 2, \dots$ , for  $j = n+1, n+2, n+3, \dots$ .

We note in addition that the expectation of this count model is

$$E(N) = \sum_{n=1}^{\infty} \sum_{j=n}^{\infty} \frac{n(-1)^{j+n} (\lambda t^c)^j \alpha_j^n}{\Gamma(cj+1)}, \tag{12}$$

with variance given by

$$\begin{aligned}
 \text{Var}(N) &= E(N^2) - (E(N))^2 \\
 &= \sum_{n=2}^{\infty} \sum_{j=n}^{\infty} \frac{n^2 (-1)^{j+n} (\lambda t^c)^j \alpha_j^n}{\Gamma(cj+1)} \\
 &\quad - \left( \sum_{n=1}^{\infty} \sum_{j=n}^{\infty} \frac{n(-1)^{j+n} (\lambda t^c)^j \alpha_j^n}{\Gamma(cj+1)} \right)^2. \tag{14}
 \end{aligned}$$

In fact, the moment generating function (MGF) is also readily obtained. In particular, let  $M_I(u)$  denote the MGF where  $I$  is the Weibull counting random variable in the pdf. Then,

$$\begin{aligned}
 M_I(u) &= E(e^{iu}) \\
 &= \sum_{i=0}^{\infty} e^{iu} \sum_{j=i}^{\infty} \frac{(-1)^{j+i} (\lambda t^c)^j \alpha_j^i}{\Gamma(cj+1)} \\
 &= \sum_{i=0}^{\infty} e^{iu} \sum_{j=i}^{\infty} \frac{(-1)^{j+i} (\lambda t^c)^j \alpha_j^i}{\Gamma(cj+1)} \\
 &= \sum_{i=0}^{\infty} \sum_{j=i}^{\infty} \frac{e^{iu} (-1)^{j+i} (\lambda t^c)^j \alpha_j^i}{\Gamma(cj+1)}, \tag{15}
 \end{aligned}$$

and correspondingly, to obtain the moments we note that

$$\begin{aligned}
 \frac{d^n}{du^n} M_I(u) &= \frac{d^n}{du^n} \sum_{i=0}^{\infty} \sum_{j=i}^{\infty} \frac{e^{iu} (-1)^{j+i} (\lambda t^c)^j \alpha_j^i}{\Gamma(cj+1)} \\
 &= \sum_{i=n}^{\infty} \sum_{j=i}^{\infty} \frac{i^n e^{iu} (-1)^{j+i} (\lambda t^c)^j \alpha_j^i}{\Gamma(cj+1)}. \tag{16}
 \end{aligned}$$

Hence, based on our polynomial expansion, we obtain closed-form expressions for the density as well as its moments.

### 3.2 The Benefits of the Weibull Count Model

We now revisit the properties listed in Section 1, point by point (and provided in italics below), to describe both those aspects that the basic Weibull count model (without covariates and without heterogeneity) given in (11) provides, and those that require extensions.

- (1) *The model generalizes (nests) the most commonly used extant models such as the Poisson and the NBD as special cases; thus, when a simple structure is sufficient, the researcher will clearly see it through the estimated model parameters. Furthermore, standard inferential procedures (e.g., the likelihood ratio test) can be used to compare different specifications.*

We note that when we set  $c = 1$  and  $t = 1$  in (11), we do in fact get the Poisson count model as  $P(N(t) = n) = \sum_{j=n}^{\infty} (-1)^{j+n} (\lambda)^j \alpha_j^n / \Gamma(j+1)$ , a standard result. With regard to the negative binomial model, we discuss this with respect to item 5 below, when  $\lambda$  is allowed to vary across the population.

- (2) *The model handles both overdispersed and underdispersed data, both of which are likely to be seen in practice.*

Through extensive simulations (because the result is unavailable in closed form), we have verified that for  $0 < c < 1$ , the probability mass function associated with the Weibull count model displays overdispersion, whereas for  $c > 1$ , underdispersion is displayed. That is, the underlying interarrival times have a decreasing (increasing) hazard for  $0 < c < 1$  ( $c > 1$ ). Thus, negative duration dependence is associated with overdispersion, positive duration dependence with underdispersion (Winkelmann 1995). A lack of duration dependence leads to the Poisson distribution with equal mean and variance.

As one demonstration of these findings, Figures 1 and 2 display probability histograms for the Weibull and Poisson count models with different parameter values. Both the Weibull and the Poisson were intentionally chosen to have identical means

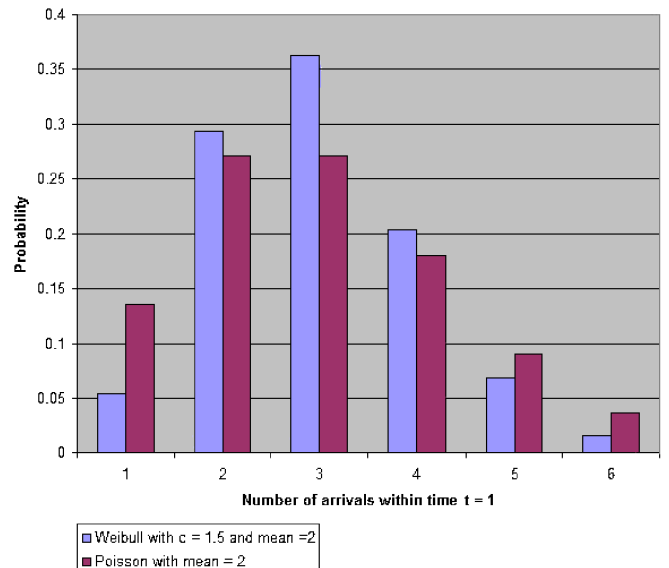


Figure 1. Poisson and Weibull models displaying underdispersion.

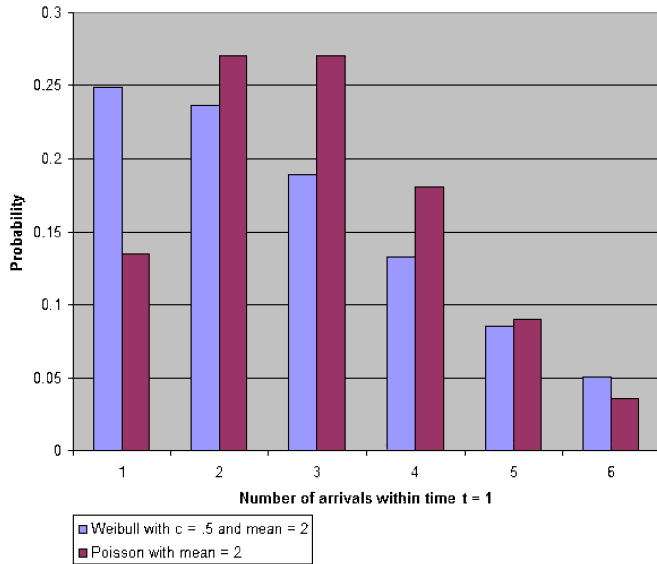


Figure 2. Poisson and Weibull models displaying overdispersion.

(set to 2); yet their dispersion is quite different. In Figure 1, we have the probability histograms for an underdispersed Weibull with parameters  $c = 1.5$  and  $\lambda = 2.93$ , and a Poisson with  $\lambda = 2$ . The variance of the Weibull count model in this case is .880. In Figure 2, we have the probability histograms for an overdispersed Weibull with parameters  $c = .5$  and  $\lambda = 1.39$ , and again the Poisson with  $\lambda = 2$ . The variance of the Weibull count model in this case is 3.40, which is greater than the mean, as expected.

(3) *Researchers who believe that the interarrival times of their dataset are Weibull distributed now have a corresponding counting model to use.*

As (11) is derived from the Weibull timing model, the link between the timing model and its counting model equivalent is maintained. Hence, in those cases where an analysis of the interarrival times (if the data are available) suggests that a more flexible timing model is needed, it can now be incorporated via its count model equivalent. Furthermore, in those cases where one only has count data, but would like to make forecasts of the next arrival time, this can now be done given the timing and count model link that is now achieved.

(4) *The model is computationally feasible to work with: it is estimable without requiring a formal programming language or time-consuming simulation-based methods.*

Although the summations shown in the expressions above may seem a bit daunting at first, they are easy to manage from an operational standpoint. We will demonstrate in Section 4 that the model is tractable enough that we perform parameter estimation in standard software, and in addition that the results are not particularly sensitive to the number of terms that are used in the summation, beyond a certain point, which can be identified through empirical testing.

(5) *The model allows for the incorporation of person-level heterogeneity reflecting the fact that individuals' interarrival rates may vary quite substantially across the population.*

One nice feature of the model presented in (11) is that introducing heterogeneity across units in their rate parameters,  $\lambda_i$ , is straightforward. If, as is standard in many timing models, we assume that the underlying rates are drawn from a gamma distribution  $\lambda_i \sim \text{gamma}(r, \alpha)$ , we can increase the model flexibility at the expense of only one additional model parameter and also, as per item 1, when  $c = 1$  nest the negative binomial model. Thus, when we combine our polynomial expansion Weibull count model in (11) with a gamma mixing distribution, we get a count model that nests the Poisson and negative binomial.

In particular, the derivation of the heterogeneous Weibull count model, model [6] from Section 2.1, is given as follows:

$$\begin{aligned}
 P(N(t) = n) &= \int_0^\infty \left[ \sum_{j=n}^\infty \frac{(-1)^{j+n} (\lambda_i t^c)^j \alpha_j^n}{\Gamma(cj + 1)} \right] g(\lambda_i | r, \alpha) d\lambda_i \\
 &= \int_0^\infty \left[ \sum_{j=n}^\infty \frac{(-1)^{j+n} (\lambda_i t^c)^j \alpha_j^n}{\Gamma(cj + 1)} \right] \\
 &\quad \times \frac{\alpha^r (\lambda_i)^{r-1} e^{-\alpha \lambda_i}}{\Gamma(r)} d\lambda_i \\
 &= \sum_{j=n}^\infty \frac{(-1)^{j+n} (t^c)^j \alpha_j^n}{\Gamma(cj + 1)} \int_0^\infty \lambda_i^j \frac{\alpha^r (\lambda_i)^{r-1} e^{-\alpha \lambda_i}}{\Gamma(r)} d\lambda_i \\
 &= \sum_{j=n}^\infty \frac{(-1)^{j+n} (t^c)^j \alpha_j^n}{\Gamma(cj + 1)} \frac{\Gamma(r + j)}{\Gamma(r) \alpha^j}. \tag{17}
 \end{aligned}$$

This expression is simply a weighted sum of the  $j$ th moments of the gamma distribution around zero,  $\Gamma(r + j)/\Gamma(r)\alpha^j$ , as  $\lambda_i^j$  enters the polynomial approximated likelihood in a linear way. Hence, the conjugacy of the gamma mixing distribution and the polynomial approximated likelihood is directly obtained.

(6) *The mechanism required to incorporate covariate effects is clear and simple. This process is consistent with standard "proportional-hazards" methods, which represent the dominant paradigm for ordinary single-event timing models.*

Now that we have the closed-form solution for the heterogeneous count model with an underlying Weibull interarrival process, we extend it to allow for the inclusion of covariates, that is, models [5] and [7] from Section 2.1. We define the Weibull regression model, without heterogeneity, as

$$\begin{aligned}
 P(N(t) = n) &= \sum_{j=n}^\infty \frac{(-1)^{j+n} (\lambda e^{x_i' \beta} t^c)^j \alpha_j^n}{\Gamma(cj + 1)} \\
 &= \sum_{j=n}^\infty \left( \frac{(-1)^{j+n} (\lambda t^c)^j \alpha_j^n}{\Gamma(cj + 1)} \right) (e^{x_i' \beta})^j, \tag{18}
 \end{aligned}$$

where  $x_i'$  denotes the covariate vector for unit  $i$  and  $\beta$  a set of covariate slopes. In an analogous manner, we derive model [7], our most complex model which allows for Weibull interarrival times, covariate heterogeneity, and parameter heterogeneity and is given by

$$P(N(t) = n) = \sum_{j=n}^\infty \frac{(-1)^{j+n} (t^c)^j \alpha_j^n}{\Gamma(cj + 1)} \frac{\Gamma(r + j)}{\Gamma(r) \alpha^j} (e^{x_i' \beta})^j, \tag{19}$$

after integrating over  $\lambda_i \sim \text{gamma}(r, \alpha)$ .

We next describe an application of these models using a dataset initially described and analyzed by Winkelmann (1995) that is an underdispersed count dataset with covariates.

#### 4. TESTING AND RESULTS

Besides the derivation of the Weibull count model, and the extensions to include covariates and heterogeneity, an additional goal of this research was to provide an empirical demonstration of our model. Is the polynomial expansion derived here empirically tractable, and will it provide improved results compared to existing methods? Remarkably enough, the computational approach for our class of models, including the computation of bootstrap standard errors (Efron 1982), was conducted entirely in Microsoft Excel, an aspect we believe makes our approach widely accessible. The spreadsheet calculates the first hundred  $\alpha$  coefficients, and then uses Solver to maximize the likelihood with respect to the data; it is available upon request.

Specifically, to compute the standard errors of coefficients under the series of models, we utilized a bootstrap procedure in which 100 replicate datasets of identical size to the original data for each model were generated by sampling individual respondent covariate-count outcome pairs,  $(Z_i, N_i)$ , with replacement. The results reported for the standard errors are the standard deviation of the coefficients across those samples. We note that the bootstrapping procedure can be implemented using a weighted likelihood approach where each observation pair's weight in the likelihood is the number of times that it appears in the replicate sample. This equivalence of using a weighted likelihood approach to compute bootstrap standard errors is not specific to this model, so it can be utilized in a large number of research domains, and can be applied in software packages (e.g., a spreadsheet) that contain little more than random number generation and function maximizing capabilities. In addition, bootstrap standard errors were compared to standard errors computed using numerical estimates of the gradient and Hessian. The standard errors were of comparable magnitude in all cases, contained no general pattern, and were roughly 20–30% bigger on average using the bootstrap, reflecting the potential asymmetric and heavier tailed models utilized here.

We apply our series of models to a dataset initially (and more fully) described by Winkelmann (1995) which contains as a dependent variable the number of children born to a random sample of females. A number of explanatory variables,  $Z_i$ , are available including the female's general education (measured

as the number of years of school), a series of dummy variables for post-secondary education (either vocational training or university), nationality (German or not), rural or urban dwelling, religious denomination (Catholic, Protestant, and Muslim, with other or none as reference group), and continuous variables for year of birth and age at marriage.

This dataset was chosen for a number of reasons. First, the article by Winkelmann (1995) acted as a motivation for this research; hence utilizing the identical dataset made sense. Second, for this dataset, the variance of the number of births is less than the mean (2.3 versus 2.4); thus we have an opportunity to demonstrate the ability of the Weibull family of count models to handle underdispersion. Finally, as Winkelmann (1995) already contained the results for the Poisson regression model (model [1] here) and the gamma-based count model that he derived in that article, we already had results that would both let us confirm the accuracy of our computational approach and provide a strong benchmark (the gamma-based model) to which we can compare the Weibull.

Before presenting the results, we note (as is standard in extant Weibull timing model research) that we reparameterized our Weibull count model from its regular form  $(r, \alpha, c)$  to a parameterization given by  $(1/r, r/\alpha, c)$ . This has been shown to have (and we confirm here that there are) multiple benefits in model implementation, including (1) greater stability in the parameter estimation process, (2) parameter estimates that are not at the boundary of the parameter space (thus enabling likelihood ratio tests for model comparison), and (3) standard errors of coefficients that are more stable and meaningful than those associated with the direct estimation of  $r, \alpha$ , and  $c$ .

Tables 1 and 2 list the results of the basic models (without covariates) and the regression models, respectively. We note that the log-likelihood values computed using our count model approach, for both the regular Poisson (LL = -2186.8) and Poisson regression (LL = -2101.8), are identical to those in table 1 of Winkelmann (1995, p. 471), thus verifying the accuracy of our polynomial expansion approach. In addition, the last column in Table 2, the results of the gamma count regression model, is taken directly from table 1 of Winkelmann (1995, p. 471). We describe our findings with respect to the models first without and then with covariates.

In Figure 3 we plot the actual and fitted values for the Poisson, the NBD, the Weibull, and the heterogeneous Weibull, noting that, for this dataset, the Poisson and the NBD are indistinguishable. Whereas all of the models tend to underfit at two children and overfit for values near two, a result also seen in

Table 1. Basic model results for total marital fertility

Variable	Model							
	Poisson		NBD		Weibull		Het. Weibull	
	Coef	SE	Coef	SE	Coef	SE	Coef	SE
$\lambda$	2.380	.042			2.635	.099		
$c$					1.116	.042	1.436	.053
$1/r$			.000	.009			.200	.012
$r/\alpha$			2.384	.047			3.625	.156
Log likelihood	-2,186.8		-2,186.8		-2,180.4		-2,165.6	

Table 2. Regression model results for total marital fertility

Variable	Model									
	Poisson		NBD		Weibull		Het. Weibull		Gamma	
	Coef	SE	Coef	SE	Coef	SE	Coef	SE	Coef	SE
German	-.200	.050	-.198	.040	-.229	.062	-.268	.054	-.190	.060
Years of schooling	.033	.004	.034	.002	.038	.010	.062	.015	.032	.027
Vocational training	-.153	.038	-.152	.033	-.173	.047	-.181	.040	-.144	.037
University	-.155	.098	-.155	.087	-.179	.125	-.264	.076	-.146	.130
Catholic	.218	.046	.218	.042	.244	.066	.242	.038	.206	.059
Protestant	.113	.057	.113	.043	.125	.069	.118	.046	.107	.063
Muslim	.548	.064	.551	.052	.640	.089	.673	.037	.523	.070
Rural	.059	.031	.062	.027	.068	.038	.071	.042	.055	.032
Year of birth	.002	.001	.003	.000	.002	.001	.003	.003	-.002	.002
Age at marriage	-.030	.003	-.030	.000	-.030	.003	-.034	.006	-.290	.006
$\lambda$	3.150	.264			4.050	.331				
$c$					1.236	.045	1.362	.061		
$1/r$			.000	.000			.061	.008		
$r/\alpha$			3.130	.131			3.604	.093	1.439	.233
Log likelihood	-2,101.8		-2,101.8		-2,077.0		-2,067.5		-2,078.2	

Winkelmann (1995) even when covariates were included, the heterogeneous Weibull minimizes error and maximizes likelihood. Given that the location of the error falls at the value of two children, a number of children seen as ideal by many, and this error seems consistent across models, we might conclude that contraceptive practices or cultural norms have “in-

terfered” with the underlying independence and identically distributed assumption. Moreover, we would expect the Weibull models to perform best relative to the Poisson and the NBD on datasets that are underdispersed, which it does. However, for this dataset, the mean is 2.4 and the variance is 2.3, so the data are only very slightly underdispersed, thus explaining the similarity of the fitted values.

The basic models show that both Weibull count models have significantly better log-likelihoods than the Poisson and the NBD. The latter two models are identical for this dataset, because the observed underdispersion will drive the NBD heterogeneity to zero (the corresponding values of  $r$  and  $\alpha$  obtained from  $1/r$  and  $r/\alpha$  are extremely large). The presence of gamma heterogeneity around the Poisson process would overdispersion, not underdispersion, the fertility counts, so it would not help in this case. Interestingly, once one utilizes the Weibull timing model instead of the exponential, the need for heterogeneity now arises (LL = -2165.5 for the heterogeneous Weibull as compared to -2180.4 for the nonheterogeneous). In fact, we hypothesize that because the Weibull model indicates duration dependence ( $c$  significantly greater than 1), this needs to be counterbalanced by heterogeneity to provide an adequate fit. It is somewhat unusual to encounter a situation in which the move to a more flexible individual-level model leads to a greater degree of heterogeneity than a more restricted specification. Although we observe this pattern for our data, we do not know if it holds in general; it is an interesting area for future research.

The results in Table 1 provide initial evidence that duration dependence plays a distinctly different role when compared to heterogeneity. It is valuable to have a model that can distinguish between these two factors. If the underlying dataset were instead overdispersed, one could use the heterogeneous Weibull count model to determine whether the “non-Poisson” dispersion effects were coming from the underlying timing process or from cross-sectional differences. This can be a valuable contribution of our model.

Notice finally that the value of  $c$  in the nonheterogeneous Weibull model is 1.116, slightly more than two standard errors

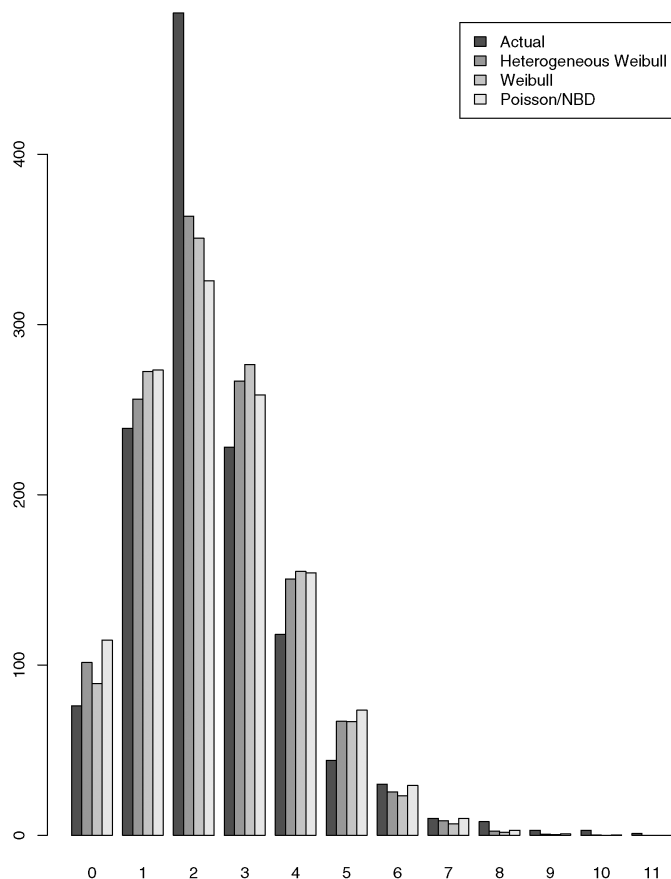


Figure 3. Fertility data compared to fitted values.



above 1.0, and for the heterogeneous model it is 1.436—almost eight standard errors above 1.0. This is consistent with our earlier discussion result that when  $c$  is greater than 1, the Weibull count model's variance is less than the mean—underdispersion. It also indicates that the “arrival process” for babies is not completely random. A mother is unlikely to have a baby immediately after the birth of a previous child (which fits the laws of nature quite well), but the odds (or hazard) of delivering another child steadily increase thereafter. An anonymous reviewer suggested that a Weibull model with a “blocked” period reflecting that women cannot have children within a certain time frame after birth would be a more realistic empirical model, and we agree that this is an interesting area for future research.

Turning our attention to the models with covariates, we first note that the two Weibull regression models provide the best fits, that is, a slight improvement in log-likelihood for the Weibull model without heterogeneity and a significant improvement for the heterogeneous Weibull model, compared to the Poisson and Winkelmann's gamma count model. The values of  $c$  for the Weibull regression and heterogeneous Weibull regression models are comparable to the models with no covariates, and still significantly greater than 1.0. The coefficients for the covariates show very small differences across the models. The coefficients of all variables are identical in sign to those in Winkelmann (1995), are extremely stable across the class of models, and have comparable standard errors such that the variables that are significant coincide in both sets of models (the only difference of note is that the year-of-birth and age-of-marriage variables were centered in Winkelmann, and not here, hence the difference in size of the coefficients; the Poisson regression models as indicated by the log-likelihoods are the same).

## 5. CONCLUSIONS

In this research, we have derived and provided an empirical demonstration for an entirely new class of count models derived from a Weibull interarrival time process. The new model has many nice features such as its closed-form nature, computational simplicity, the ability to nest both the Poisson and NBD models, and the ability to bring in both heterogeneity and covariates in a natural way. The key to the derivation is the use of a Taylor series expansion to get around the fact that, unlike the exponential or gamma distributions, there is no simple way to obtain a convolution of two (or more) Weibulls.

From an empirical standpoint, we showed that the Weibull count model without heterogeneity offers a slight improvement in log-likelihood when compared to the gamma count model of Winkelmann (1995) and a dramatic improvement over extant models commonly used. When including heterogeneity in the Weibull model (both with and without covariates), the improvement is even greater, suggesting that the improved effects of adding a flexible timing model and heterogeneity may be complementary. Admittedly it is impossible to generalize from one dataset, but these results provide encouraging signs about the model's validity and usefulness. More importantly, the model provides a sizeable improvement over the more traditional Poisson/NBD model (with and without covariates). This may have important implications in many cases, because most

researchers have always turned to heterogeneity as the first explanation/correction for datasets that do not conform well to the simple assumption of Poisson counts (and, implicitly, exponential interarrival times). Now researchers have a very plausible second explanation available (i.e., Weibull interarrival times) and they can further explore it using conventional techniques such as proportional hazards for covariates and a parametric mixing distribution for heterogeneity. This is a powerful combination of old and new methods that has substantial promise for a wide variety of application areas.

## ACKNOWLEDGMENT

The authors thank Rainer Winkelmann for useful suggestions and comments and for generously providing us the data used in this study.

[Received January 2006. Revised February 2007.]

## REFERENCES

- Arbous, A. G., and Kerrich, J. E. (1951), “Accident Statistics and the Concept of Accident-Prone,” *Biometrics*, 7, 341–433.
- Barlow, R. E., and Proschan, F. (1965), *Mathematical Theory of Reliability*, New York: Wiley.
- Bening, V. E., and Korolev, V. Y. (2002), *Generalized Poisson Models and Their Applications in Insurance and Finance*, Leiden, Netherlands: Brill Academic Publishers.
- Bradlow, E. T., Hardie, B. G. S., and Fader, P. S. (2002), “Bayesian Inference for the Negative Binomial Distribution via Polynomial Expansions,” *Journal of Computational and Graphical Statistics*, 11, 189–201.
- Cameron, A. C., and Johansson, P. (1997), “Count Data Regression Using Series Expansion: With Applications,” *Journal of Applied Econometrics*, 12, 203–223.
- Cameron, A. C., and Trivedi, P. K. (1998), *Regression Analysis of Count Data*, Cambridge, U.K.: Cambridge University Press, pp. 59–188.
- Cox, D. R. (1972), “Regression Models and Life-Tables,” *Journal of the Royal Statistical Society, Ser. B*, 34, 187–220.
- Efron, B. (1982), “Bootstrap Methods: Another Look at the Jackknife,” *The Annals of Statistics*, 7, 1–26.
- Everson, P. J., and Bradlow, E. T. (2002), “Bayesian Inference for the Beta-Binomial Distribution via Polynomial Expansions,” *Journal of Computational and Graphical Statistics*, 11, 202–207.
- Feller, W. (1943), “On a General Class of ‘Contagious’ Distributions,” *The Annals of Mathematical Statistics*, 14, 389–400.
- Gourieroux, C., and Visser, M. (1997), “A Count Data Model With Unobserved Heterogeneity,” *Journal of Econometrics*, 79, 247–268.
- Gurland, J., and Sethuraman, J. (1995), “How Pooling Failure Data May Reverse Increasing Failure Rates,” *Journal of the American Statistical Association*, 90, 1416–1423.
- King, G. (1989), “Variance Specification in Event Count Models: From Restrictive Assumptions to a Generalized Estimator,” *American Journal of Political Science*, 33, 762–784.
- Lawless, J. F. (1987), “Regression Methods for Poisson Process Data,” *Journal of the American Statistical Association*, 82, 808–815.
- Miller, S. J., Bradlow, E. T., and Dayaratna, K. (2006), “Closed-Form Bayesian Inferences for the Logit Model via Polynomial Expansions,” *Quantitative Marketing and Economics*, 4, 173–206.
- Mudholkar, G. S., Srivastava, D. K., and Kollia, G. D. (1996), “A Generalization of the Weibull Distribution With Application to the Analysis of Survival Data,” *Journal of the American Statistical Association*, 91, 1575–1583.
- Mullahy, J. (1986), “Specification and Testing of Some Modified Count Data Models,” *Journal of Econometrics*, 33, 341–351.
- Shmueli, G., Minka, T. P., Kadane, J. B., Borle, S., and Boatwright, P. (2005), “A Useful Distribution for Fitting Discrete Data: Revival of the Conway–Maxwell–Poisson Distribution,” *Journal of the Royal Statistical Society, Ser. C*, 54, 127–142.
- Trivedi, P. K., and Cameron, A. C. (1996), “Applications of Count Data Models to Financial Data,” in *Handbook of Statistics*, Vol. 14, Amsterdam: North-Holland, Chap. 12, pp. 363–391.

- Trivedi, P. K., and Deb, P. (1997), "The Demand for Health Care by the Elderly: A Finite Mixture Approach," *Journal of Applied Econometrics*, 12, 313–332.
- Wimmer, G., and Altmann, G. (1999), *Thesaurus of Univariate Discrete Probability Distributions*, Germany: Stamm Verlag.
- Winkelmann, R. (1995), "Duration Dependence and Dispersion in Count-Data Models," *Journal of Business & Economic Statistics*, 13, 467–474.
- (2003), *Econometric Analysis of Count Data* (4th ed.), Heidelberg, NY: Springer.