Applied Stochastic Models in Business and Industry

Special Issue Paper

Received 6 December 2010, Revised 13 April 2011, Accepted 7 June 2011 Published online 8 September 2011 in Wiley Online Library

(wileyonlinelibrary.com) DOI: 10.1002/asmb.918

The COM-Poisson model for count data: a survey of methods and applications

Kimberly F. Sellers[a], Sharad Borle[b] and Galit Shmueli[c*†]

The Poisson distribution is a popular distribution for modeling count data, yet it is constrained by its equidispersion assumption, making it less than ideal for modeling real data that often exhibit over-dispersion or under-dispersion. The COM-Poisson distribution is a two-parameter generalization of the Poisson distribution that allows for a wide range of over-dispersion and under-dispersion. It not only generalizes the Poisson distribution but also contains the Bernoulli and geometric distributions as special cases. This distribution's flexibility and special properties have prompted a fast growth of methodological and applied research in various fields. This paper surveys the different COM-Poisson models that have been published thus far and their applications in areas including marketing, transportation, and biology, among others. Copyright © 2011 John Wiley & Sons, Ltd.

Keywords: regression model; overdispersion; underdispersion; Conway-Maxwell-Poisson; marketing; transportation; biology

1. Introduction

With the huge growth in the collection and storage of data as a result of technological advances, count data have become widely available in many disciplines. Although classic examples of count data are exotic in nature, such as the number of soldiers killed by horse kicks in the Prussian cavalry [1], the number of typing errors on a page, or the number of lice on heads of Hindu male prisoners in Cannamore, South India 1937-39 [2], today's count data are as mainstream as noncount data. Examples include the number of visits to a website, the number of purchases at a brick-and-mortar or an online store, the number of calls to a call center, or the number of bids in an online auction.

The most popular distribution for modeling count data has been the Poisson distribution. Applications using the Poisson distribution for modeling count data are wide ranging. Examples include Poisson control charts for monitoring the number of nonconforming items, Poisson regression models for modeling epidemiological and transportation data, and Poisson models for the number of bidder arrivals at an online auction site [3].

Although Poisson models are very popular for modeling count data, many real data do not adhere to the assumption of equidispersion that underlies the Poisson distribution (namely, that the mean and variance are equal). An early result has therefore been the popularization of the negative binomial distribution, which can capture over-dispersion. Although initially using the negative binomial distribution posed computational challenges [4], today there is no such issue, and the negative binomial distribution and regression are included in most statistical software packages. Hilbe [5] provides an extensive description of the negative binomial regression and its variants.

The negative binomial distribution provides a solution for over-dispersed data, that is, when the variance is larger than the mean. Over-dispersion takes place in various contexts, such as contagion between observations. The opposite case is that of under-dispersion, where the variance is smaller than the mean. Although the literature contains more examples of over-dispersion, under-dispersion is also common. Rare events, for instance, generate under-dispersed counts. Examples include the number of strike outbreaks in the UK coal mining industry during successive periods between 1948 and 1959

[a]Department of Mathematics and Statistics, Georgetown University, Washington, DC 20057, USA
[b]Department of Marketing, Jones Graduate School of Business, Rice University, Houston, TX 77251, USA
[c]Department of Decision, Operations and Information Technologies, Robert H. Smith School of Business, University of Maryland, College Park, MD 20742, USA
*Correspondence to: Galit Shmueli, Department of Decision, Operations and Information Technologies, Robert H. Smith School of Business, Indian School of Business, Gachibowli, Hyderabad 500 032, India and University of Maryland, College Park, MD 20742, USA.
†E-mail: gshmueli@rhsmith.umd.edu

and the number of eggs per nest for a species of bird [6]. In such cases, neither the Poisson nor the negative binomial distributions provide adequate approximations. Several distributions have been proposed for modeling both over-dispersion and under-dispersion. These include the weighted Poisson distributions of del Castillo and Perez-Casany [7] and the generalized Poisson (GP) distribution of Consul [8]. Both are generalizations of a Poisson distribution, where an additional parameter is added. The GP distribution has also been developed into a regression model ([9, 10]), control charts [11], and as a model for misreporting ([12, 13]). The shortcoming of the GP model, however, is its inability to capture some levels of dispersion because the distribution is truncated under certain conditions regarding the dispersion parameter and thus is not a true probability model [9].

In this work, we describe a growing stream of research and applications using a flexible two-parameter generalization of the Poisson distribution called the Conway–Maxwell–Poisson (COM-Poisson or CMP) distribution. The main advantage of this distribution is its flexibility in modeling a wide range of over-dispersion and under-dispersion with only two parameters, while possessing properties that make it methodologically appealing and useful in practice. In our opinion, these properties have lead to the growing interest and development in both methodological research (theoretical and computational) and applications using the COM-Poisson, both by statisticians and by nonstatisticians. Although the majority of the COM-Poisson related work has been ongoing in the last 10 years, the amount of interest that it has generated among researchers in various fields and the fast rate of methodological developments warrant a survey of the current state to allow those familiar with the COM-Poisson to gauge the scope of current affairs and for those unfamiliar with the COM-Poisson to get an overview of existing and potential developments and uses.

Historically, the distribution was briefly proposed by Conway and Maxwell [14] as a model for queuing systems with state-dependent service times. To the best of our knowledge, this early form was used in the field of linguistics for fitting word lengths. The great majority of the COM-Poisson development has begun over the last decade, with the initial major publication by Shmueli *et al.* [15]. The motivation for developing the statistical methodology for the COM-Poisson distribution in [15] arose from an application in marketing, where the purpose was to model the number of purchases by customers at an online grocery store (one of the earlier online grocery stores), where the data exhibited different levels of dispersion when examined by different product categories. The original development started from the point of allowing the ratio of consecutive probabilities $\mathrm{P}(Y = y - 1)/\mathrm{P}(Y = y)$ to be more flexible than a linear function in y, as dictated by a Poisson distribution.

Although the COM-Poisson distribution is a two-parameter generalization of the Poisson distribution, it has special characteristics that make it especially useful and elegant. For instance, it also generalizes the Bernoulli and geometric distributions and is a member of the exponential family in both parameters. Following the publication of [15], research involving the COM-Poisson distribution has developed in several directions by different authors and research groups. It has also been applied in various fields, including marketing, transportation, and epidemiology. The purpose of this paper is to survey the various COM-Poisson developments and applications, which have appeared in the literature in different fields, in an attempt to consolidate the accumulated knowledge and experience related to the COM-Poisson, to highlight its usefulness for solving different problems, and to propose possibilities for further methodological development needed for analyzing count data.

The paper is organized as follows. Section 2 describes the COM-Poisson distribution, its properties and estimation. In Section 3, we describe the different models that have evolved from the COM-Poisson distribution. These include regression models of various forms (e.g., constant dispersion, group-level dispersion, and observation-level dispersion), via different approaches (generalized linear model, Bayesian), and different estimation techniques (maximum likelihood, quasi-likelihood, MCMC) as well as other COM-Poisson-based models (such as cure-rate models). Section 4 surveys a variety of applications using the COM-Poisson, including marketing and electronic commerce, transportation, biology, and disclosure limitation. Section 5 concludes the manuscript with a discussion and future directions.

2. The COM-Poisson Distribution

Conway and Maxwell [14] originally proposed what is now known as the COM-Poisson distribution as a solution to handling queuing systems with state-dependent service rates. In their article, Conway and Maxwell derived the COM-Poisson distribution from a set of differential difference equations that describe a single queue-single server system with random arrival times, with Poisson inter-arrival times with parameter λ, first-come-first-serve policy, and exponential service times that depend on the system state having mean $= n^c\mu$, where n is the number of units in the system, $1/\mu$ is the mean service time for a unit when that unit is the only one in the system, and c is the 'pressure coefficient' indicating the degree to which the service rate of the system is affected by the system state. This distribution has been applied in some fields (mainly in linguistics; see Section 4). However, its statistical properties were not studied in a cohesive fashion until Shmueli *et al.* [15] did so, naming it the COM-Poisson (or CMP) distribution.

2.1. The probability distribution

The COM-Poisson probability distribution function has the form

$$P\left(Y=y\right)=\frac{\lambda^{y}}{\left(y!\right)^{\nu}Z(\lambda,\nu)},\quad y=0,1,2,\ldots,\lambda>0,\nu\geqslant 0$$

for a random variable Y, where $Z\left(\lambda,\ \nu\right)=\sum_{s=0}^{\infty}\lambda^{s}/(s!)^{\nu}$ is a normalizing constant; ν is considered the *dispersion parameter* such that $\nu>1$ represents under-dispersion, and $\nu<1$ over-dispersion. The COM-Poisson distribution generalizes not only the Poisson distribution ($\nu=1$) but also the geometric distribution ($\nu=0,\lambda<1$), and the Bernoulli distribution ($\nu\to\infty$ with probability $\lambda/(1+\lambda)$). Although there are no simple closed forms linking the parameters λ and ν to moments, there are several relationships that highlight the roles of each parameter and their effect on the distribution. One such formulation (derived by Ralph Snider at Monash University) is $\lambda=E(Y^{\nu})$ that displays λ as the expected value of the power-transformed counts, with power ν. Other formulations for the moments include the recursive form

$$E\left(Y^{r+1}\right)=\begin{cases}\lambda[E(Y+1)]^{1-\nu} & r=0\\ \lambda\frac{\partial}{\partial\lambda}E\left(Y^{r}\right)+E\left(Y\right)E(Y^{r}) & r>0,\end{cases}\tag{1}$$

and a form that presents the expected value and variance as derivatives with respect to $\log(\lambda)$:

$$E\left(Y\right)=\frac{\partial\log Z(\lambda,\nu)}{\partial\log\lambda}$$
$$\mathrm{Var}\left(Y\right)=\frac{\partial E\left(Y\right)}{\partial\log\lambda}.$$

The expected value and variance are approximated by

$$E\left(Y\right)\approx\lambda^{1/\nu}-\frac{\nu-1}{2\nu}\tag{2}$$

[15], and $\mathrm{Var}\left(Y\right)\approx(1/\nu)\lambda^{1/\nu}$ [16]. These approximations are accurate when $\nu\leqslant 1$ or $\lambda>10^{\nu}$ [15]. The COM-Poisson distribution has the moment generating function, $M_{Y}\left(t\right)=E\left(e^{Yt}\right)=Z(\lambda e^{t},\nu)/Z(\lambda,\ \nu)$, and probability generating function, $E\left(t^{Y}\right)=Z(\lambda t,\ \nu)/Z(\lambda,\nu)$.

In terms of computation, the infinite sum $Z\left(\lambda,\ \nu\right)=\sum_{s=0}^{\infty}\lambda^{s}/(s!)^{\nu}$, which is involved in computing moments and other quantities, might not appear elegant computationally; however, from a practical perspective, it is easily approximated to any level of precision. Minka *et al.* [17] addressed computational issues and provided useful approximations and upper bounds for $Z(\lambda,\nu)$ and related quantities. In practice, the infinite sum can be approximated by truncation. The upper bound on the truncation error from using only the first $k+1$ counts ($s=0,\ldots,k$) is given by [17] as

$$\frac{\lambda^{k+1}}{(k+1)!^{\nu}\left(1-\varepsilon_{k}\right)}$$

where $\varepsilon_{k}>\lambda(j+1)^{\nu}$ for all $j>k$.

When $\nu>1$ (under-dispersion), the elements in the sum quickly decrease, requiring only a small number of summations and hence do not pose any computational challenge. For $\nu<1$, where the truncation of the infinite sum must use multiple values to achieve reasonable accuracy, the following asymptotic form is useful:

$$Z(\lambda,\ \nu)=\frac{\exp\left(\nu\lambda^{1/\nu}\right)}{\lambda^{(\nu-1)/(2\nu)}(2\pi)^{(\nu-1)/2}\sqrt{\nu}}\left(1+O\left(\lambda^{-1/\nu}\right)\right),$$

Minka *et al.* [17] comment that this formula is accurate when $\lambda>10^{\nu}$.

Nadarajah [18] extended the computational focus and derivation, obtaining exact expressions for the CDF and moments for integer values of >1 (i.e., for special cases involving under-dispersion; note that the author mistakenly refers to $\nu>1$ as over-dispersion). Although these approximations are helpful, one can easily compute the values of COM-Poisson moments (to a high degree of accuracy), the normalizing constant, and other quantities for non-integer values of by using computational packages such as **compoisson** in *R*.

Shmueli *et al.* [15] also showed that the COM-Poisson distribution belongs to the exponential family in both parameters and determined the sufficient statistics for a COM-Poisson distribution, namely $S_1=\sum_{i=1}^{n}Y_i$ and $S_2=\sum_{i=1}^{n}\log(Y_i!)$

where $Y_1, \ldots, Y_n$ denotes a random sample of n COM-Poisson random variables (see also Exercise 5 in Chapter 6 of [19]). An example of the usefulness of the sufficient statistics is given in Section 4, in the context of data disclosure.

Taking a Bayesian approach to the COM-Poisson distribution, Kadane *et al.* [20] used the exponential family structure of the COM-Poisson to establish a conjugate prior density of the form

$$h(\lambda, \nu) = \lambda^{a-1} e^{-\nu b} Z^{-c} \kappa(abc), \tag{3}$$

where $\lambda > 0$ and $\nu \geqslant 0$, and $\kappa(a, b, c)$ is the normalization constant; the posterior has the same form with $a' = a + S_1$, $b' = b + S_2$, and $c' = c + n$. They showed that for this density to be proper, a, b and c must satisfy the condition

$$\frac{b}{c} > \log\left(\left\lfloor \frac{a}{c} \right\rfloor !\right) + \left(\frac{a}{c} - \left\lfloor \frac{a}{c} \right\rfloor\right) \log\left(\left\lfloor \frac{a}{c} \right\rfloor + 1\right).$$

The Bayesian formulation is especially useful when prior information is available. To facilitate conveying prior information more easily in terms of the conjugate prior, [20] developed an online data elicitation program where users can choose a, b, c based on the predictive distribution that is presented graphically.

2.2. *COM-Poisson as a weighted Poisson distribution*

Several authors have noted that the COM-Poisson belongs to the family of weighted Poisson distributions. In general, a random variable Y is defined to have a weighted Poisson distribution if the probability function can be written in the form

$$P(Y = y) = \frac{\mathrm{e}^{-\lambda} \lambda^y w_y}{W y!}, \quad y = 0, 1, 2, \ldots; \quad \lambda > 0,$$

where $W = \sum_{s=0}^{\infty} \mathrm{e}^{-\lambda} \lambda^s w_s / s!$ is a normalizing constant [7]. Ridout and Besbeas [6] and Rodrigues *et al.* [21] note that the COM-Poisson distribution can be viewed as a weighted Poisson distribution with weight function, $w_y = (y!)^{1-\nu}$. Kokonendji *et al.* [22] mention that weighted Poisson distributions are widely used for modeling data with partial recording, in cases where a Poisson variable is observed or recorded with probability w_y, when the event $Y = y$ occurs. Further, presenting the COM-Poisson as a weighted Poisson distribution allows deriving knowledge regarding the types of dispersion that the model can capture. For instance, Kokonendji *et al.* [22] show that the COM-Poisson is closed by 'pointwise duality' for all $\nu \in [0, 2]$; that is, for a given COM-Poisson distribution with $\nu_1 \in [0, 2]$, there exists another COM-Poisson distribution with $\nu_2 = 2 - \nu_1 \in [0, 2]$, which is its pointwise dual distribution. The meaning of the closed pointwise duality for all $\nu \in [0, 2]$ in the COM-Poisson case is that any value within this range is guaranteed to account for either over-dispersion or under-dispersion of the same magnitude.

Ridout and Besbeas [6] compare the COM-Poisson with an alternative weighted Poisson where the weights have the form

$$w_y = \begin{cases} e^{-\beta_1(\lambda - y)}, & y \leqslant \lambda \\ e^{-\beta_2(\lambda - y)}, & y > \lambda \end{cases}.$$

Under-dispersion exists for $\beta_1, \beta_2 > 0$, over-dispersion holds for $\beta_1, \beta_2 < 0$, and equidispersion is achieved when $\beta_1 = \beta_2 = 0$. They called this the three-parameter exponentially weighted Poisson (or two-parameter exponentially weighted Poisson for $\beta_1 = \beta_2 = \beta$) and compared goodness of fit in several applications where the data display under-dispersion. For one dataset, which described the number of strikes over successive periods in the UK, the two-parameter exponentially weighted Poisson was found to outperform the COM-Poisson, but the COM-Poisson still provides a good fit. In data describing clutch size (i.e., number of eggs per nest), zero-truncated versions of the distributions are considered because the dataset is severely under-dispersed with a variance-to-mean ratio of 0.10. All distributions considered performed somewhat poorly, including the zero-truncated two-parameter or three-parameter exponentially weighted Poisson and the COM-Poisson distributions.

2.3. *Parameter estimation*

Shmueli *et al.* [15] presented three approaches for estimating the two parameters of the COM-Poisson distribution, given a set of data. The first is a weighted least squares (WLS) approach, which takes advantage of the form

$$\log[\mathrm{P}(Y = y - 1)/\mathrm{P}(Y = y)] = -\log\lambda + \nu \log(y),$$

and which relies on fitting a linear relationship to the ratios of consecutive count proportions; $\mathrm{P}(Y = y)$ is estimated by the proportion of y values in the data. WLS is needed to correct for the nonzero covariance and the nonconstant variance of the ratios. A plot of the ratios versus the counts, y, can give an initial indication of the slope and intercept, displaying the level of dispersion compared to a Poisson case (slope $= 1$). See [15] for further details and an example.

The second estimation approach is a maximum likelihood approach. Maximum likelihood estimates are easily derived for the COM-Poisson distribution because of its membership in the exponential family. The log-likelihood function can be written as

$$\log L\left(y_1, \ldots, y_n | \lambda, \nu\right) = \log \lambda \sum\nolimits_{i=1}^{n} y_i - \nu \sum\nolimits_{i=1}^{n} \log(y_i!) - n \log Z\left(\lambda, \nu\right), \tag{4}$$

and the maximum likelihood estimation can thus be achieved by iteratively solving the set of normal equations

$$\mathrm{E}\left(Y\right) = \bar{Y}, \text{ and } E(\log\left(Y!\right)) = \overline{\log Y!}$$

or by directly optimizing the likelihood function using optimization software. For example, using the R software, $\hat{\lambda}$ and $\hat{\nu}$ can be obtained by maximizing the likelihood function directly via the functions **nlminb** or **optim** that perform constrained optimization, under the constraint $\nu \geqslant 0$; alternatively, the likelihood function can be rewritten as a function of $\log(\hat{\nu})$ so that the likelihood function can be maximized via an unconstrained optimization function such as **nlm** (in R). In either case, the WLS or even Poisson regression estimates can be used as initial estimates.

Finally, a third approach is Bayesian estimation. Because the COM-Poisson has a conjugate prior, estimation is simple and straightforward once the hyperparameters a, b, c in Equation (3) are specified.

2.4. Further extensions

Shmueli *et al.* [15] discussed several extensions of the COM-Poisson distribution. Among them are zero-inflated and zero-deflated COM-Poisson distributions, which extend the COM-Poisson by including an extra parameter that captures a contaminating process that produces more or less zeros. In data with no zero counts, the authors discuss a shifted COM-Poisson distribution (also illustrated for modeling word lengths, where there are no zero-length words).

Another extension presented in [15] is the COM-Poisson-binomial distribution. The latter arises as the conditional distribution of a COM-Poisson variable, conditional on a sum of two COM-Poisson variables with possibly different λ parameters but the same ν. The COM-Poisson-binomial distribution generalizes the binomial distribution, allowing for more flexible variance magnitudes compared with the binomial variance. It can be interpreted as the sum of dependent Bernoulli variables with a specific joint distribution (see [15] for details). This idea can be further extended to a COM-Poisson-multinomial distribution.

3. Regression models

Modeling count data as a response variable in a regression-type context is common in many applications [23]. The most common model is Poisson regression, yet it is limiting in its equidispersion assumption. Much attention has focused on the case of data over-dispersion (e.g., [23, 24], and [5]), which arises in practice because of experimental design issues and/or variability within groups. Many of the proposed approaches (e.g., [25]), however, cannot be applied to address under-dispersion and/or have restrictions that make such approaches unfavorable [4]. The restricted GP regression, for example, can effectively model data over-dispersion or under-dispersion; however, the dispersion parameter is bounded in the case of under-dispersion such that it is not a true probability model [9].

Recall the log-likelihood function of the COM-Poisson distribution given in Equation (4). The COM-Poisson's exponential family structure allows for various regression models to be considered to describe the relationship between the explanatory variables and the response. Because the expected value of a COM-Poisson does not have a simple form, different researchers have chosen various link functions to describe the relationship between $\mathrm{E}(Y)$ and the set of covariates via $\boldsymbol{X}'\boldsymbol{\beta} (= \beta_0 + \beta_1 X_1 + \cdots + \beta_p X_p)$. The following sections discuss several proposed COM-Poisson regression models.

3.1. COM-Poisson model with constant dispersion

Sellers and Shmueli [26] took a generalized linear model approach and used the link function $\eta(\mathrm{E}(\boldsymbol{Y})) = \log \boldsymbol{\lambda}$ for modeling the relationship between $\mathrm{E}(\boldsymbol{Y})$ and $\boldsymbol{X}'\boldsymbol{\beta}$. This choice of link function, while an indirect function of $\mathrm{E}(Y)$, has two advantages: first, it leads to a regression model that generalizes the common Poisson regression (with link function $\log \boldsymbol{\lambda}$) as well as the logistic regression (with link function $\log \boldsymbol{\lambda} = \log \boldsymbol{p}/(1 - \boldsymbol{p})$). Including two well-known regression models as special cases is useful theoretically and practically. Logistic regression is a special case of the COM-Poisson regression when $\nu \to \infty$. Sellers and Shmueli [26] showed that in practice, fitting binary response data with the COM-Poisson regression produces results identical to a logistic regression. The second advantage of using $\log \boldsymbol{\lambda}$ as the link function is that it leads to elegant estimation, inference, and diagnostics. This result highlights the lesser role that the conditional mean plays when considering count distributions of a wide variety of dispersion levels. Unlike Poisson or linear regression, where the conditional mean is central to estimation and interpretation, in the COM-Poisson regression model, we must take into account the entire conditional distribution.

The formulation allows for estimating $\boldsymbol{\beta}$ and an unknown constant parameter ν via a set of normal equations. As in the case of maximum likelihood estimation of the COM-Poisson distribution (Section 2.3), the normal equations can be solved iteratively. Sellers and Shmueli [26] proposed an appropriate iterative re-WLS procedure to obtain the maximum likelihood estimates for $\boldsymbol{\beta}$ and ν. Alternatively, the maximum likelihood estimates can be obtained by using a constrained optimization function over the likelihood function (with constraint $\nu \geqslant 0$) or an unconstrained optimization function over the likelihood written in terms of $\log \nu$. Standard errors of the estimated parameters are derived using the Fisher Information matrix. Sellers and Shmueli [26] used this formulation to derive a dispersion test, which tests whether a COM-Poisson regression is warranted over an ordinary Poisson regression for a given set of data. In terms of inference, large sample theory allows for normal approximation when testing hypotheses regarding individual explanatory variables. For small samples, however, a parametric bootstrap procedure is proposed. The parametric bootstrap is achieved by resampling from a COM-Poisson distribution with parameters $\hat{\boldsymbol{\lambda}} = \boldsymbol{x}'\hat{\boldsymbol{\beta}}$ and $\hat{\nu}$, where $\hat{\boldsymbol{\beta}}$ and $\hat{\nu}$ are estimated from a COM-Poisson regression on the full dataset. The resampled datasets include new response values accordingly. Then, for each resampled dataset, a COM-Poisson regression is fitted, thus producing new associated estimates, which can then be used for inference.

With respect to computing fitted values, [26] describes two ways of obtaining fitted values: using estimated means via the approximation, $\hat{y}|x = \hat{\lambda}^{1/\hat{\nu}} - [(\hat{\nu}-1)/2\hat{\nu}]$, or using estimated medians via the inverse CDF for $\hat{y}|x$ and $\hat{\nu}$. Note that the latter is used in logistic regression to produce classifications with a cut-off of 0.5 (whereas choosing a different percentile from the median will correspond to a different cut-off value). Finally, a set of diagnostics is proposed for evaluating goodness of fit and detecting outliers. Measures of leverage, Pearson residuals, and deviance residuals are described and illustrated. The *R* package, **COMPoissonReg** (available on CRAN), contains procedures for estimating the COM-Poisson regression coefficients and standard errors under the constant dispersion assumption, as well as computing diagnostics, the dispersion test, and even simulating COM-Poisson data and performing the parametric bootstrap.

A different COM-Poisson regression formulation was suggested and used by Boatwright *et al.* [27], as a marginal model of purchase timing, in a larger household-level model of the joint distribution of purchase quantity and timing for online grocery sales. Their Bayesian specification allowed for a λ parameter with cross-sectional as well as temporal variation via a multiplicative model, $\lambda_{ij} = \lambda_i \delta_1^{x_{1ij}}, \delta_2^{x_{2ij}}, \ldots, \delta_k^{x_{kij}}$, where i denotes a household and j is the temporal index; x_{1ij}, $x_{2ij}, \ldots, x_{kij}$ are time-varying covariates measured in logarithmic form. The estimation involves specifying appropriate independent priors over various parameters and using an MCMC framework. Additional details regarding this work are provided in Section 4.2.

Lord *et al.* [28] independently proposed a Bayesian COM-Poisson regression model to address data that are not equidispersed. Their formulation used an alternate link function, $\log(\lambda^{1/\nu})$, which is an approximation to the mean under certain conditions. Their choice was aimed at allowing the interpretation of the coefficients in terms of their impact on the mean. They used noninformative priors in modeling the relationship between the explanatory and response variables and performed parameter estimation via MCMC. Comparing goodness-of-fit and out-of-sample prediction measures, [28] empirically showed the similarity in performance of the COM-Poisson and negative binomial regression for modeling over-dispersed data, thereby highlighting the advantage of the COM-Poisson over the negative binomial regression in its ability to not only adequately capture over-dispersion but also under-dispersion and low counts.

Jowaheer and Khan [29] proposed a quasi-likelihood approach to estimate the regression coefficients, arguing that the maximum likelihood approach is computationally intensive (their work precedes Sellers and Shmueli [26]). Their method requires the first two moments of the COM-Poisson distribution, and accordingly, the authors use the approximation provided in Equation (2) (noting that [29] present this equation with a typographical error) and $\mathrm{Var}(Y) = \lambda^{1/\nu}/\nu$, as well as the moments recursion provided in Equation (1) to build the iterative scheme via the Newton–Rhapson method. The resulting estimators $\left(\beta^{QL}, \nu^{QL}\right)$ are consistent and $\sqrt{n}\left(\left(\beta^{QL}, \nu^{QL}\right) - (\beta, \nu)\right)^T$ is asymptotically normal as $n \to \infty$ [29]. Although the authors note only a negligible loss in efficiency, one must still be cautious in using this derivation, as the expected value representation used here is accurate under a constrained space for ν and λ. When $\nu > 1$ and $\lambda \leqslant 10^{\nu}$, this comparison requires further study.

3.2. COM-Poisson model with group-level dispersion

Because the COM-Poisson can accommodate a wide range of over-dispersion and under-dispersion, a group-level dispersion model allows the incorporation of different dispersion levels within a single dataset. The alternative of modeling all groups with a single level of dispersion causes information loss and can lead to incorrect conclusions.

Sellers and Shmueli [30] introduced an extension of their COM-Poisson regression formulation [26], which allows for different levels of dispersion across different groups of observations. Their formulation uses the link functions,

$$\log(\lambda) = \beta + \sum_{j=1}^{p} \beta_j X_j'$$
$$\log(\nu) = \gamma + \sum_{k=1}^{K-1} \gamma_k G_k',$$

where G_k is a dummy variable corresponding to one of K groups in the data.

Estimating the $\boldsymbol{\beta}$ and $\boldsymbol{\gamma}$ coefficients is carried out by maximizing the log-likelihood, where the log-likelihood for observation i is given by

$$\log \mathrm{L_i}(\lambda_\mathrm{i}, \nu_\mathrm{i}|y_\mathrm{i}) = y_\mathrm{i}\log(\lambda_\mathrm{i}) - \nu_\mathrm{i}\log(y_\mathrm{i}!) - \log Z(\lambda_\mathrm{i}, \nu_\mathrm{i}),$$

where

$$\log(\lambda_\mathrm{i}) = \beta_0 + \beta_1 x_{\mathrm{i}1} + \cdots + \beta_\mathrm{p} x_\mathrm{ip} \doteq \boldsymbol{X}_\mathrm{i}\boldsymbol{\beta} \text{ and}$$
$$\log(\nu_\mathrm{i}) = \gamma_0 + \gamma_1 G_{\mathrm{i}1} + \cdots + \gamma_{\mathrm{K}-1} G_{\mathrm{i,K}-1} \doteq \boldsymbol{G}_\mathrm{i}\boldsymbol{\gamma}.$$

Because the COM-Poisson distribution belongs to the exponential family, the appropriate normal equations for $\boldsymbol{\beta}$ and $\boldsymbol{\gamma}$ can be derived. Using the Poisson estimates, $\boldsymbol{\beta}^{(0)}$ and $\boldsymbol{\gamma}^{(0)} = \mathbf{0}$, as starting values, coefficient estimation can again be achieved via an appropriate iterative re-WLS procedure or by using existing nonlinear optimization tools (e.g., **nlm** or **optim** in *R*) to directly maximize the likelihood function. The associated standard errors of the estimated coefficients are derived in an analogous manner to that described in [26].

3.3. COM-Poisson model with observation-level dispersion

An even more flexible model in terms of dispersion is to allow dispersion to differ for the different observations. [16] suggested that such a model would be useful in modeling power outages. [31] independently proposed the model that allows for the dispersion parameter, ν_i, to vary with observation i and considered a relationship between ν_i and the explanatory variables in the $(p+1)$-dimensional row vector, $\boldsymbol{X}_\mathrm{i}$. Accordingly, the log-likelihood for observation i is given as

$$\log \mathrm{L_i}(\lambda_\mathrm{i}, \nu_\mathrm{i}|y_\mathrm{i}) = y_\mathrm{i}\log\lambda_\mathrm{i} - \nu_\mathrm{i}\log(y_\mathrm{i}!) - \log \mathrm{Z}(\lambda_\mathrm{i}, \nu_\mathrm{i}),$$

where

$$\log\lambda_\mathrm{i} = \beta_0 + \beta_1 x_{\mathrm{i}1} + \cdots + \beta_\mathrm{p} x_\mathrm{ip} \doteq \boldsymbol{X}_\mathrm{i}'\boldsymbol{\beta} \text{ and}$$
$$\log\nu_\mathrm{i} = \gamma_0 + \gamma_1 x_{\mathrm{i}1} + \cdots + \gamma_\mathrm{p} x_\mathrm{ip} \doteq \boldsymbol{X}_\mathrm{i}'\boldsymbol{\gamma}.$$

As in previous models, appropriate normal equations can be derived for $\boldsymbol{\beta}$ and $\boldsymbol{\gamma}$, and coefficient estimation can be achieved via an appropriate iterative re-WLS procedure or by using existing nonlinear optimization tools.

3.4. Other models

COM-Poisson cure-rate model

Rodrigues *et al.* [21] used the COM-Poisson distribution to establish a cure-rate model. By letting Y denote the minimum time-to-occurrence among all competing causes, the survival function is defined as

$$S_p(Y) = P(Y \geqslant y) = \sum_{m=0}^{\infty} P(M=m)[S(y)]^m,$$

where $S(y) = 1 - F(y)$ is the independent and identically distributed survival function for the j competing causes, $W_j(j = 1, 2, \ldots, m)$, and M (i.e., the number of competing causes) is a weighted Poisson random variable as defined in Section 1.2. Thus, letting M follow a COM-Poisson distribution implies the survival function $S_p(y) = Z(\eta S(y), \nu)/Z(\eta, \nu)$, where $\eta = \exp(\theta)$. Analogous to the flexibility of the COM-Poisson distribution, the COM-Poisson cure-rate model allows for flexibility with regard to dispersion. In particular, it encompasses the promotion time cure model ($\nu = 1$) and the mixture cure model ($\nu \to \infty$).

COM-Poisson model with censoring

Censored count data arise in applications such as surveys, where the possible answers to a question are, e.g., 0, 1, 2, 3, 4+. Extensions of the Poisson and negative binomial regression models exist for the case of censoring, as is a GP model [10].

Sellers and Shmueli [32] introduced a COM-Poisson model that allows for right-censored count data. To incorporate censoring, an indicator variable δ_i is used to denote a censored observation; an observation is either completely observed ($\delta_i = 0$) or censored ($\delta_i = 1$). The likelihood function is therefore given by

$$\begin{aligned}\log L &= \sum_{i=1}^{n} (1-\delta_i) \log P\,(Y_i = y_i) + \delta_i \log P\,(Y_i \geqslant y_i) \\ &= \sum_{i=1}^{n} (1-\delta_i)[y_i \log \lambda_i - \nu \log(y_i!) - \log Z(\lambda_i, \nu)] + \delta_i \log P(Y_i \geqslant y_i),\end{aligned}$$

where $P(Y_i \geqslant y_i) = 1 - \sum_{s=0}^{y_i-1} \lambda_i^s/(s!)^\nu Z(\lambda_i, \nu)$. One can analogously modify these equations to allow for left censoring or interval censoring. The authors compare the predictive power (i.e., the performance of the model in terms of predicting new observations) of the censored COM-Poisson model with other censored regressions applied to an under-dispersed dataset, and find that the censored COM-Poisson and censored GP models perform comparably well, with the COM-Poisson model obtaining the best predictive scores in a majority of cases.

4. Applications

4.1. Linguistics

Early applications of the COM-Poisson distribution have been in linguistics for the purpose of fitting word lengths. Theory in linguistics suggests that the most 'elementary' form of a word length distribution follows the difference equation [33],

$$P(Y = y+1) = \frac{a}{(y+1)^b} P(Y = y);$$

that is, words of successive length (e.g., lengths y and $y+1$) are represented in texts in a proportional way [34]. Hence, the COM-Poisson distribution, which adheres to this form, has been used for modeling word length. Wimmer *et al.* [33] used it to model the number of syllables in the Hungarian dictionary and in Slovak poems. Best [34] notes that the COM-Poisson distribution best models the German alphabetical lexicon (Viëtor) as well as a German frequency lexicon (Kaeding). Among 135 texts, in all but four texts the COM-Poisson was the best fitting model among a set of models (including the hyper-Poisson and hyper-Pascal).

4.2. Marketing and eCommerce

The Poisson and the negative binomial distributions remain popular tools when it comes to applications involving count data in Marketing and eCommerce. Following the revival of COM-Poisson distribution, however, there have been several applications in Marketing and eCommerce using the COM-Poisson. These applications have benefited from the property of COM-Poisson that allows for added flexibility to model under-dispersed as well as over-dispersed data.

Boatwright *et al.* [27] used the COM-Poisson distribution to develop a joint (quantity and timing) model for grocery sales at an online retailer. The number of weeks between purchases was modeled as a COM-Poisson variable, namely

$$w_{ij} \sim \begin{cases} \left(1-\pi_{ij}\right) \text{COM-Poisson}\left(\lambda_{ij}, \nu\right), & \text{for } w_{ij} > 0 \\ \pi_{ij}, & \text{for } w_{ij} = 0, \end{cases}$$

where $w_{ij} = 0, 1, \ldots$ measures the inter-purchase time in weeks rounded off to the nearest week, i indexes the household, and j indexes the temporal purchases; π_{ij} is the probability that $w_{ij} = 0$. A multiplicative form for λ_{ij} was used, $\lambda_{ij} = \lambda_i \delta_1^{x_{1ij}} \delta_2^{x_{2ij}} \ldots, \delta_k^{x_{kij}}$, where $x_{1ij}, x_{2ij}, \ldots, x_{kij}$ were time-varying covariates measured in logarithmic form. Household heterogeneity in the expected inter-purchase time was incorporated by allowing a hierarchy over $\lambda_i \sim \text{gamma}(a, b)$. Independent gamma priors were specified for the δ, ν, and a parameters, and an inverse gamma prior was specified for the b parameter. The model was estimated using an MCMC sampler. Borle *et al.* [35] used a similar joint model to study the impact of a large-scale reduction in assortment at an online grocery retailer. They found that the decline in shopping frequency resulted in a greater loss than did the reduction in purchase quantities and that the impact of assortment cut varied widely by category, with less-frequently purchased categories more adversely affected.

In another application of the COM-Poisson, [36] used the distribution to model quarterly T-shirt sales across 196 stores of a large retailer. The data consisted of quarterly sales of T-shirts (in eight colors and four sizes) across these stores along with information on the extent of stockouts in these shirts. The authors built a demand model (in the presence of stockouts) and studied the impact of each stock keeping unit (SKU) on sales. An imputation procedure was used to impute demand when stockouts were observed. The results showed that many items (T-shirts) affected category sales over and above their own sales volume. After deconstructing the role of a stockout of individual items into three effects (namely, lost own sales, substitution to other items, and the category sales impact), they found that the category impact has the largest magnitude. Interestingly, the disproportionate impact of individual items on category sales was not restricted to top selling items, for almost every single item affected category sales.

Kalyanam *et al.* [36] also performed some robustness checks to validate their model; these tests were primarily a comparison of results under alternative model specifications. The alternative models considered were a Poisson model, a geometric model, a COM-Poisson model without imputation (using a censored likelihood), a normal regression model, and a 'Rule of thumb' approach. In terms of log-likelihood measures, the COM-Poisson models (with and without imputation) performed best. In terms of predictive ability (as measured by the median absolute deviation statistic), the COM-Poisson and Poisson models were somewhat similar in performance.

In yet another interesting application of the COM-Poisson model, [37] used it to study customer behaviors at a US automotive services firm. The study was carried out to evaluate the effects of participation in a satisfaction survey and examine the role of customer characteristics and store-specific variables in moderating the effects of participation. The data for this study came from a longitudinal field study of customer satisfaction conducted by the US automotive services firm. The data contained two groups of customers, namely, one group that was surveyed for customer satisfaction, whereas the other group was not surveyed by the firm. Four customer behaviors were studied: (i) the number of promotions redeemed by a customer on each visit; (ii) the number of automotive services purchased on each visit; (iii) the time since the last visit in days (i.e., inter-purchase time); and (iv) the amount spent during each visit. Results revealed a substantial positive relationship between satisfaction survey participation and the four customer behaviors studied. Two of the four behaviors (the number of services bought per visit and the number of promotions redeemed per visit) were modeled using a COM-Poisson distribution. In particular, the number of services bought was modeled as a COM-Poisson variable with a one-unit location shift, that is, a shifted COM-Poisson variable. Figure 1 below shows bar charts of these variables across the entire data.

As seen from the bar charts, there is over-dispersion in the number of services bought, whereas the number of promotions redeemed is under-dispersed. This is also borne out by the estimated ν parameters for these two variables ($\hat{\nu}_{\text{services}} = 0.61$, $\hat{\nu}_{\text{promotions}} = 4.69$, respectively). This is an example where flexibility of the COM-Poisson in accounting for over-dispersion as well as under-dispersion in the data is very useful. The same count distribution can be used to model these two variables. An alternative count distribution such as the negative binomial may have performed equally well in modeling 'number of services bought'; however, it would have been a poor choice in modeling 'number of promotions redeemed'.

An eCommerce application of COM-Poisson is the study of the extent of multiple and late bidding in eBay online auctions. Borle *et al.* [38] empirically estimated the distribution of bid timings and the extent of multiple bidding in a large set of eBay auctions, using bidder experience as a mediating variable. The extent of multiple bidding (the number of times a bidder changes his/her bids in a particular auction) was modeled as a COM-Poisson distribution. The data consisted of over 10,000 auctions from 15 consumer product categories. The two estimated metrics (extent of late bidding and the

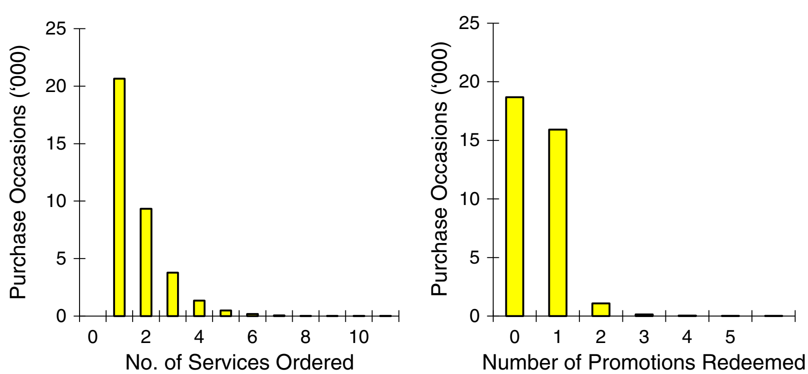

Figure 1. Distribution of the number of services bought per visit (left panel) and the number of promotions redeemed per-visit (right panel). Frequencies are in thousands.

extent of multiple bidding) allowed the authors to place these product categories along a continuum of these metrics. The analysis distinguished most of the product categories from one another with respect to these metrics, implying that product categories, after controlling for bidder experience, differ in the extent of multiple bidding and late bidding observed in them.

Apart from these applications, there have also been applications of the COM-Poisson in another important area of marketing, namely, that of customer lifetime value estimation (the monetary worth of a customer to a firm); see [39]. The customer lifetime value estimation becomes important because it is used as a metric in many marketing decisions that a firm makes, hence any improvements in its estimation has direct benefits to these decisions. In the marketing literature, the 'lifetime of a customer' in these models is typically measured either as a continuous time or in terms of 'number of lifetime purchases'. Accordingly, a continuous or a count distribution is used to model lifetime, respectively. Singh *et al.* [39] propose a modeling framework using data augmentation; the framework allows for a multitude of lifetime value models to be proposed and estimated. As a demonstration, the authors estimate two extant models in the literature and also propose and estimate three new models. One of the proposed models uses a COM-Poisson distribution to model the lifetime purchases. Compared with the other similar model that uses the beta-geometric distribution as the count distribution, the proposed COM-Poisson model improved the customer lifetime value predictions across a sample of 5000 customers by about 42% (a median absolute deviation statistic of \$62.46 as compared with \$107.72 using the beta-geometric).

4.3. *Transportation*

[28] and [40] used a COM-Poisson regression model to analyze motor vehicle crash data to link the number of crashes to the entering flows at intersections or on segments. [28] analyzed two datasets via several negative binomial and COM-Poisson generalized linear models: one dataset contained crash data from signalized four-legged intersections in Toronto, Ontario; the second dataset contained information from rural four-lane divided and undivided highways in Texas. The results of this study found that the COM-Poisson models performed as well as the negative binomial models with regard to goodness-of-fit and predictive performance. The authors have thus argued in favor of the COM-Poisson model given its flexibility in handling under-dispersed data as well. [40] analyzed an under-dispersed crash dataset from 162 railway-highway crossings in South Korea during 1998–2002. COM-Poisson models were found to produce better statistical performance than Poisson and gamma models.

4.4. *Biology*

Ridout and Besbeas [6] considered various forms of a weighted Poisson distribution (of which, they note, the COM-Poisson distribution is one) to model under-dispersed data; to illustrate this, they modeled the clutch size (i.e., number of eggs per nest) for a species of bird. The clutch size data were collected for the linnet in the UK between 1939 and 1999. The data exhibited strong under-dispersion where the variance-to-mean ratio was 0.10; see [6] for details. The COM-Poisson model reportedly 'performed poorly' with regard to goodness-of-fit (obtaining a chi-squared statistic equaling 4222.8 with 3 DOF) yet did considerably better than many other weighted Poisson distributions that were considered by the authors.

4.5. *Disclosure limitation*

Count data arise in various organizational settings. When the release of such data is sensitive, organizations need information-disclosure policies that protect data confidentiality while still providing data access. Kadane *et al.* [41] used the COM-Poisson as a tool for disclosure limitation. They showed that, by disclosing only the sufficient statistics (S_1 and S_2) and sample size n of a COM-Poisson distribution fitted to a confidential one-way count table, the exact cell counts $(f_j,\ j = 0, 1, 2, \ldots, J)$ as well as the table size (J) is sufficiently masked. The masking is obtained via two results: (i) usually many count tables correspond to the disclosed sufficient statistics and (ii) the various possible count tables are equally likely to be the undisclosed table. Moreover, finding the various solutions requires solving a system of linear equations, which are under-determined for tables with more than three cells, and can be computationally prohibitive for even small tables. They illustrated the proposed policy with two examples. The 'small' example consisted of a one-way table with counts of 0–5 which, when given only the sufficient statistics, produced 14 possible solutions. The larger and more realistic example consisted of a one-way table with 0–11 counts (the number of injuries in 10,000 car accidents in 2001) which, when given only the sufficient statistics, produced more than 80,000,000 possible solutions. The proposed policy is the first to deal with count data by releasing sufficient statistics.

5. Discussion

The COM-Poisson distribution is a flexible distribution that generalizes several classical distributions (namely, the Poisson, geometric, and Bernoulli) via its dispersion parameter. As a result, it bridges data distributions that demonstrate under-dispersion, equidispersion, and over-dispersion. Because it generalizes three well-known distributions, and some regression formulations generalize two popular models (logistic regression and Poisson regression), the COM-Poisson offers more than just a new model for count data. Its ability to handle different dispersion types and levels makes it useful in applications where the level of dispersion might vary, yet a single analytical framework is preferred.

In terms of fit, the COM-Poisson appears to fit data at least as well as competing two-parameter models. For instance, [15] reports that although the GP of [8] is flexible in terms of modeling over-dispersion and has simple expressions for the normalizing constant and moments, it cannot handle under-dispersion and is not in the exponential family, which makes analysis more difficult. Numerical studies show that for every COM-Poisson distribution with $0.75 < \nu < 1$ (or so), there is a GP distribution with very similar form. For $\nu < 0.75$, however, the two families differ markedly. In the linguistics applications, [34] compared the COM-Poisson with several alternatives and found it a better overall alternative for modeling word lengths. [28] shows that it performs at least as well as a negative binomial for over-dispersed data.

From a theoretical point of view, the COM-Poisson has multiple properties that make it favorable for methodological development. These properties have likely led to the fast growth of papers in disparate areas developing COM-Poisson-based models.

As to computational considerations, although the COM-Poisson does not offer simple closed-form formulas for moments and includes the infinite sum as a normalizing constant, in practice these only rarely are a cause for concern. Computationally, the infinite sum can be truncated while achieving precision to any pre-specified level. In cases of over-dispersion where the sum might require many more terms to reach some accuracy, approximations are available. Software packages, such as the R packages **compoisson** and **COMPoissonReg**, include Z-function calculations and offer users an easy way to estimate parameters and compute moments numerically. The unavailability of a simple formula for the mean does make interpreting coefficients in a regression model more complicated than in linear or multiplicative models where the mean and variance/dispersion are either independent or identical (such as linear or Poisson regression). As with other popular and useful regression models, the solution is to use marginal analysis and interpret coefficients by examining the conditional mean or median [26].

As illustrated by the breadth of work surveyed here, the COM-Poisson distribution has swiftly grown in interest, both with regard to methodological and applied work. Yet, there remain numerous opportunities for continued research and investigation with regard to this flexible distribution. One open question regards the predictive power of the COM-Poisson model. Although the COM-Poisson has been consistently shown to perform well in terms of fit, the few papers that report predictive power results using the COM-Poisson do not convey a consistent picture. In some cases, the predictive power of the COM-Poisson is shown to be similar to that of a Poisson distribution or the GP distribution, whereas in others, the COM-Poisson outperforms these distributions. Predicting individual values requires defining point predictions (e.g., means or medians) as well as predictive intervals. Evaluating the predictive performance in the count data context is also nontrivial [42] (e.g., how is a prediction error defined?).

Meanwhile, in almost every application where the Poisson distribution plays an important role, there is an opportunity to expand the toolkit to consider over-dispersion and under-dispersion, which are frequently the case with real data. Examples of such areas include the fields of reliability, data tracking and process monitoring, and time series analysis of count data. Even within the context of queuing (or more generally, stochastic processes) where the COM-Poisson originated, there has been very little in the way of further developing and applying the COM-Poisson distribution. In addition to the many fields where Poisson distributions are assumed, there is a growing number of fields where more and more count data are becoming available. Thus, there remain many opportunities for significant contribution to expanding the scope and usefulness of statistical modeling of count data, taking the parsimonious COM-Poisson approach.

Acknowledgements

The authors thank Jay Kadane from Carnegie Mellon University for his insightful comments and discussion regarding this manuscript.

References

1. von Bortkewicz L. *Das Gesetz der Kleinen Zahlen*. Teubner: Leipzig, 1898.
2. Bliss CI, Fisher RA. Fitting the negative binomial distribution to biological data. *Biometrics* 1953; **9**:176–199.

3. Etzion H, Pinker E, Seidmann A. Analyzing the simultaneous use of auctions and posted prices for online selling. *Manufacturing and Service Operations Management* 2006; **8**:68–91. DOI: 10.1287/msom.1060.0101.
4. McCullagh P, Nelder J. *Generalized Linear Models, 2nd edition*. Chapman & Hall/CRC: New York, NY, 1997.
5. Hilbe JM. *Negative Binomial Regression*. Cambridge University Press: New York, NY, 2008.
6. Ridout MS, Besbeas P. An empirical model for underdispersed count data. *Statistical Modelling* 2004; **4**:77–89. DOI: 10.1191/1471082X04st064oa.
7. del Castillo J, Pérez-Casany M. Overdispersed and underdispersed Poisson generalizations. *Journal of Statistical Planning and Inference* 2005; **134**:486–500. DOI: 10.1016/j.jspi.2004.04.019.
8. Consul PC. *Generalized Poisson Distributions: Properties and Applications*. Marcel Dekker: New York, NY, 1989.
9. Famoye F. Restricted generalized Poisson regression model. *Communications in Statistics – Theory and Methods* 1993; **22**:1335–1354. DOI: 10.1080/03610929308831089.
10. Famoye F, Wang WR. Censored generalized Poisson regression model. *Computational Statistics and Data Analysis* 2004; **46**:547–560. DOI: 10.1016/j.csda.2003.08.007.
11. Famoye F. Statistical control charts for shifted generalized Poisson distribution. *Statistical Methods and Applications* 2007; **3**:339–354. DOI: 10.1007/BF02589023.
12. Neubauer G, Djuras G. A generalized Poisson model for underreporting. *Proceedings of the 23rd International Workshop on Statistical Modelling*, Eilers PHC (ed.). Universiteit Utrecht: Netherlands, 2008; 368–373.
13. Pararai M, Famoye F, Lee C. Generalized Poisson–Poisson mixture model for misreported counts with an application to smoking data. *Journal of Data Science* 2010; **8**:607–617.
14. Conway RW, Maxwell WL. A queuing model with state dependent service rates. *Journal of Industrial Engineering* 1962; **12**:132–136.
15. Shmueli G, Minka TP, Kadane JB, Borle S, Boatwright P. A useful distribution for fitting discrete data: revival of the Conway–Maxwell–Poisson distribution. *Applied Statistics* 2005; **54**:127–142. DOI: 10.1111/j.1467-9876.2005.00474.x.
16. Guikema SD, Coffelt JP. A flexible count data regression model for risk analysis. *Risk Analysis* 2008; **28**:213–223. DOI: 10.1111/j.1539-6924.2008.01014.x.
17. Minka TP, Shmueli G, Kadane JB, Borle S, Boatwright P. Computing with the COM-Poisson distribution. *Technical Report Series*, Carnegie Mellon University Department of Statistics, Pennsylvania, 2003.
18. Nadarajah S. Useful moment and CDF formulations for the COM-Poisson distribution. *Statistical Papers* 2009; **50**:617–622. DOI: 10.1007/s00362-007-0089-9.
19. Faraway JJ. *Extending the Linear Model with R: Generalized Linear, Mixed Effects and Nonparametric Models*. Chapman & Hall / CRC: New York, NY, 2005.
20. Kadane JB, Shmueli G, Minka TP, Borle S, Boatwright P. Conjugate analysis of the Conway–Maxwell–Poisson distribution. *Bayesian Analysis* 2005; **1**:363–374. DOI: 10.1214/06-BA113.
21. Rodrigues J, de Castro M, Cancho VG, Balakrishnan N. COM-Poisson cure rate survival models and an application to a cutaneous melanoma data. *Journal of Statistical Planning and Inference* 2009; **139**:3605–3611. DOI: 10.1016/j.jspi.2009.04.014.
22. Kokonendji CC, Mizere D, Balakrishnan N. Connections of the Poisson weight function to overdispersion and underdispersion. *Journal of Statistical Planning and Inference* 2008; **138**:1287–1296. DOI: 10.1016/j.jspi.2007.05.028.
23. Barron D. The analysis of count data: overdispersion and autocorrelation. *Sociological Methodology* 1992; **22**:179–220.
24. Breslow N. Tests of hypotheses in overdispersed Poisson regression and other quasi-likelihood models. *Journal of the American Statistical Association* 1990; **85**:565–571.
25. Ismail N, Jemain A. Handling overdispersion with negative binomial and generalized Poisson regression models. *Casualty Actuarial Society Forum*, 2007; 103–158.
26. Sellers K, Shmueli G. A flexible regression model for count data. *Annals of Applied Statistics* 2010; **4**:943–961. DOI: 10.1214/09-AOAS306.
27. Boatwright P, Borle S, Kadane JB. A model of the joint distribution of purchase quantity and timing. *Journal of the American Statistical Association* 2003; **98**:564–572. DOI: 10.1198/016214503000000404.
28. Lord D, Guikema SD, Geedipally SR. Application of the Conway–Maxwell–Poisson for analyzing motor vehicle crashes. *Accident Analysis & Prevention* 2008; **40**:1123–1134.
29. Jowaheer V, Khan NAM. Estimating regression effects in COM-Poisson generalized linear model. *World Academy of Science, Engineering and Technology* 2009; **53**:1046–1050.
30. Sellers K, Shmueli G. Data dispersion: Now you see it... now you don't. Robert H. Smith School Research Paper No. RHS 06-122. http://ssrn.com/abstract=1612755 [21 May 2010].
31. Sellers K, Shmueli G. A regression model for count data with observation-level dispersion. *In Proceedings of the 24th International Workshop on Statistical Modelling*, Booth JG (ed.). Cornell University: New York, 2009; 337–344.
32. Sellers K, Shmueli G. Predicting censored count data with COM-Poisson regression. Robert H. Smith School Research Paper No. RHS-06-129. http://ssrn.com/abstract=1702845 [October 29, 2010].
33. Wimmer G, Kohler R, Grotjahn R, Altmann G. Toward a theory of word length distributions. *Journal of Quantitative Linguistics* 1994; **1**:98–106. DOI: 10.1080/09296179408590003.
34. Best K-H. Probability distributions of language entities. *Journal of Quantitative Linguistics* 2001; **8**:1–11. DOI: 10.1076/jqul.8.1.1.4091.
35. Borle S, Boatwright P, Kadane JB, Nunes J, Shmueli G. Effect of product assortment changes on consumer retention. *Marketing Science* 2005; **24**:616–622. DOI: 10.1287/mksc.1050.0121.
36. Kalyanam K, Borle S, Boatwright P. Deconstructing each item's category contribution. *Marketing Science* 2007; **26**:327–341. DOI: 10.1287/mksc.1070.0270.
37. Borle S, Dholakia U, Singh S, Westbrook R. The impact of survey participation on subsequent behavior: an empirical investigation. *Marketing Science* 2007; **26**:711–726. DOI: 10.1287/mksc.1070.0268.
38. Borle S, Boatwright P, Kadane JB. The timing of bid placement and extent of multiple bidding: an empirical investigation using ebay online auctions. *Statistical Science* 2006; **21**:194–205. DOI: 10.1214/088342306000000123.
39. Singh S, Borle S, Jain D. A generalized framework for estimating customer lifetime value when customer lifetimes are not observed. *Quantitative Marketing and Economics* 2009; **7**:181–205. DOI: 10.1007/s11129-009-9065-0.

40. Lord D, Geedipally SR, Guikema SD. Extension of the application of Conway–Maxwell–Poisson models: analyzing traffic crash data exhibiting underdispersion. *Risk Analysis* 2010; **30**:1268–1276. DOI: 10.1111/j.1539-6924.2010.01417.x.
41. Kadane JB, Krishnan R, Shmueli G. A data disclosure policy for count data based on the COM-Poisson distribution. *Management Science* 2006; **52**:1610–1617. DOI: 10.1287/mnsc.1060.0562.
42. Czado C, Gneiting T, Held L. Predictive model assessment for count data. *Biometrics* 2009; **65**:1254–1261. DOI: 10.1111/j.1541-0420.2009.01191.x.

Applied Stochastic Models in Business and Industry

(wileyonlinelibrary.com) DOI: 10.1002/asmb.919 Published online in Wiley Online Library

'The COM-Poisson model for count data: a survey of methods and applications' by K. Sellers, S. Borle and G. Shmueli

Sellers *et al.* are to be congratulated on a comprehensive paper describing a prime example where real life problems and theoretical developments merge to generate important methodological contributions. Count data, in today's data rich environment, have created new challenges for handling over-dispersion and under-dispersion that are addressed by the Conway–Maxwell–Poisson (COM-Poisson) two-parameter generalization of the Poisson distribution.

This brief discussion paper provides some comments organized in three sections. We begin with a discussion of COM-Poisson in a general scientific setting. The second section discusses possible extension of COM-Poisson to ordinal and nominal data, in one or more dimensions. The third section is focused on diagnostic evaluation of the COM-Poisson with a bootstrapped-based approach.

1. Statistical models versus phenomenological or mechanistic models

The Conway–Maxwell–Poisson (COM-Poisson) model is a statistical model describing, parametrically, behavior of count data with possible under-dispersion or over-dispersion relative to the Poisson model. In a sense, it is a descriptive model that can fit a wide range of phenomena and represent a specific population frame. Deming [3] labeled such studies "enumerative studies" as opposed to analytic studies where the objective is to predict future behavior. Analytic studies require a deep understanding of cause and effect relationships, something Deming called a theory of profound knowledge. Shmueli [12] expanded on this concept and provided characteristics of statistical analysis focused on explaining causal phenomena versus predicting future outcomes.

To exploit COM-Poisson in the context of analytic studies and make predictions in unobserved conditions, one needs to combine insights gained from the model with other modeling efforts. For example, Abeles *et al.* [1], among others, studied the extracellularly recorded spike trains in the brain that contain clusters of several spikes, separated by small inter-spike intervals. Such clusters represent sudden epochs of elevated firing rate because of a neuron's intrinsic dynamics, a response to bi-stable network behavior, or oscillations traveling through a region of the brain. These are called 'bursts' or 'up states', depending on the context. Analysis of bursting or up state data requires identification of when the bursts occur, their number, and their duration. To analyze such data, Abeles *et al.* [1] proposed a Markov-modulated two-state Poisson process two-state model for neural spike trains where spiking activity follows two homogeneous Poisson processes, one for each state. In their model, the state transitions from nonbursting to bursting and bursting to nonbursting occur according to a hidden Markov models. This approach handles over-dispersion and provides some insights on the underlying mechanism of brain activity.

Combining COM-Poisson with such mechanistic or phenomenological models can contribute to the understanding of such processes and lead to 'profound knowledge' that supports the formulation of focused interventions such as therapeutic protocols or damage prevention schemes. Our comment here is therefore that COM-Poisson can be used, in conjunction with other modeling efforts, for deriving predictions in yet untested conditions.

2. Outliers in contingency tables

Over-dispersion of count data is related to occurrence of outliers in one or more dimensions of ordinal or nominal data. Fuchs and Kenett [4] proposed the M-test for identifying outliers in contingency tables of one and two dimensions. The M-test is based on adjusted residuals and critical values derived from a Bonferroni-based tight bound. In the univariate case, cells with significant adjusted residuals stand out. In a sense, they represent over-dispersion of ordinal or nominal data.

Sellers *et al.* present in Section 3.2 of their paper an application of COM-Poisson to group level dispersion, a structure combining qualitative classifications (groups) with observations (counts). Groups characterized by two or more classification variables present an underlying two-way or higher-order contingency table. Further investigations of this relationship could provide interesting innovations to the literature on contingency tables.

Another extension of COM-Poisson can be obtained by considering the (scarce) literature on multivariate outliers for ordinal data. Riani *et al.* [10] define bivariate and multivariate outliers for ordinal variables as units presenting an unusual combination of the categories (or values) of the variables. To identify outliers in ordinal data, they evoke the ordinal regression model:

$$link(\gamma_{ij}) = \alpha_j - \left[b_1 x_{i1} + b_2 x_{i2} + \cdots + b_p x_{ip}\right],\ j = 1, \ldots k-1, \text{and } i = 1, \ldots, n,$$

where γ_{ij} is the cumulative distribution function for the j-th category of the i-th case and $link(\gamma_{ij})$ is the 'link function', which is a transformation of the cumulative probabilities of the ordered dependent variable that allows for estimation of the model [9].

The most commonly used links are Logit, Probit, Negative log–log, Complementary log–log, and Cauchit (Cauchy). Although the negative log–log link function is recommended when the probability of the lower category is high, the complementary log–log is particularly suitable when the probability of higher category is high. Finally, although the logit (probit) link is more suitable when a dependent variable is normally distributed, the Cauchit link is mainly used when extreme values are present in the data. The parameter α_j is the threshold value of the j-th category. The α_j terms are like the intercept in a linear regression, except that each j-th level of the response has its own value. The model is based on the assumption that there is a latent continuous outcome variable and that the observed ordinal outcome arises from discretizing the underlying continuum into k-ordered groups. The thresholds estimate these cut-off values. The number of regression coefficients is $p, b_1, \ldots, b_p$ are a set of regression coefficients, and $x_{i1}, x_{i2}, \ldots, x_{ip}$ are a set of p explanatory variables for the i-th subject.

In order to evaluate the goodness of fit of a model, the indexes that are typically used are the Cox and Snell R^2 , R^2_{CS}, and the pseudo Nagelkerke R^2, R^2_N, expressed as

$$R^2_{CS} = 1 - \left(\frac{l(0)}{l(\hat{\beta})}\right)^{\frac{2}{n}} \text{ and } R^2_N = \frac{R^2_{CS}}{\max(R^2_{CS})}$$

where $l(\hat{\beta})$ is the likelihood of the tested model and $l(0)$ is the likelihood of the initial model that does not contain explanatory variables (null model), and $\max(R^2_{CS}) = 1 - \{l(0)\}^{\frac{2}{n}}$. For more details and examples from a customer satisfaction survey, see Atkinson and Riani [2] and Riani *et al.* [10].

COM-Poisson can explain outliers from a modeling perspective. The link with the approach outlined might be worth investigating.

When the COM-Poisson group classification extends to two dimensions, the association between dimensions becomes of interest. An approach for detecting 2×2 contingency table with unusual patterns of association has been proposed by Kenett and Salini [7, 8]. This proposal exploits the simplex representation of 2×2 contingency tables and assesses association using a measure labeled the Relative Linkage Disequilibrium (RLD). RLD is measuring the distance of the point in a simplex corresponding to a 2×2 contingency table from the surface of independence, normalized by the maximum distance in the specific part of the simplex.

In technical terms, consider general count data X_1, X_2, X_3, X_4 in a 2×2 contingency table. where A and B represent

	B	ˆB
A	X_1	X_2
ˆA	X_3	X_4

grouping variables, X_1 represents counts in both group A and group B, X_2 represents counts in group A but not in group B, and so forth.

The four relative frequencies from a 2×2 contingency tables are $x_i = (X_i)/N, \sum_{i=1}^{4} X_i = N,\ 0 \leqslant x_i \leqslant 1, i = 1 \ldots 4$. Let $D = x_1 x_4 - x_2 x_3,\ f = x_1 + x_3$, and $g = x_1 + x_2$. It can be easily verified that

$$x_1 = fg + D$$

$$x_2 = (1-f)g - D$$

$$x_3 = f(1-g) - D$$

$$x_4 = (1-f)(1-g) + D$$

Or in vector notation, $\underline{X} = \underline{f} \otimes \underline{g} + D\underline{e} \otimes \underline{e}$, where $\underline{X} = (x_1, x_2, x_3, x_4)$, $\underline{f} = (f, 1-f)$, $\underline{g} = (g, 1-g)$, and $\underline{e} = (1, -1)$.
Under independence, $\alpha = x_1 x_4 / x_2 x_3 = 1$ and $D = 0$.

D is thus a measure assessing departure from independence. RLD is normalizing D by considering its maximum in the direction $\underline{e} \otimes \underline{e}$.

The computation of RLD is performed through the following algorithm:

If $D > 0$
then
if $x_3 < x_2$
  then $RLD = \frac{D}{D+x_3}$
  else $RLD = \frac{D}{D+x_2}$
else
if $x_1 < x_4$
  then $RLD = \frac{D}{D-x_1}$
  else $RLD = \frac{D}{D-x_4}$

RLD has been implemented as an interest measure in the ***arules*** R package available on CRAN [5]. By linking COM-Poisson to RLD, the M-test and other methods for identifying outliers in multivariate nominal or ordinal count data. As mentioned, COM-Poisson can extend to two-way or higher-order contingency tables.

3. Diagnostic bootstrapping of COM-Poisson

In Section 3 of their paper, Sellers *et al.* consider COM-Poisson regression models. In Sellers and Shmueli [11], overall goodness of fit criteria such as AIC and MSE are considered for assessing such models. They also provide a calculation of deviance residuals for identifying model fit and possible outliers. This section provides some inputs on a more in-depth assessment of the models goodness of fit by considering the airfreight breakage data used to demonstrate the application if COM-Poisson regression in Sellers and Shmueli [11]. The data consist of the number of ampules found broken upon arrival, in 10 air shipments, each carrying 1000 ampules on the flight. For each shipment i, we have the number of times the carton was transferred from one aircraft to another (Xi) and the number of ampules found broken upon arrival (Yi).

An intuitive and practical diagnostic tool based on bootstrapping for evaluating such models has been proposed by Kenett *et al.* [6]. The approach is very simple:

- Bootstrap the data and fit the model to the bootstrapped data.
- Calculate the standard error of the bootstrapped estimated model parameters.
- Compare the empirical bootstrap standard error to the theoretical standard errors derived from assumptions on the model.
- Discrepancies between the empirical and theoretical standard errors indicate lack of fit.

The bootstrap-based model diagnostic approach was applied to three models used in analyzing the airfreight breakage data. Like Sellers and Shmueli [11], the models considered were Poisson and COM-Poisson, a linear regression after log transforming the response.

The diagnostic check was performed by conducting 1000 bootstrap replicates of the data, thereby getting 1000 sets of model parameters.

The R code and its output are listed as follows:

```
Call:
boot(data = freight, statistic = bootGlmCmpCoeff, R = 1000, formula = as.formula
("broken ~ transfers"))
```

```
Bootstrap statistics:
original bias std. error prc. error
beta0 Poisson       2.3529495 -0.002584194  0.06808253 -0.0010982785
beta1 Poisson       0.2638422  0.008555700  0.04488155  0.0324273346
betao COM-Poisson 13.8247035  7.841636262 14.18763954  0.5672191279
beta1 COM-Poisson  1.4838223  0.823149124  1.40359449  0.5547491273
beta0 linear reg   2.3273482  0.001708580  0.07250285  0.0007341316
beta1 linear reg    .2800269  0.006146796  0.04881976  0.0219507308

Where:
linear reg -> log(y) = a + b*x
bias -> (bootstrap - model) estimate
prc. error -> (model std. error)/(bootstrap std. error)
```

Seems like the data under-dispersion affects negatively the bootstrap based diagnostic. This observation might be due to the properties of the asymptotic standard errors of the COM-Poisson and/or the small sample sizes.

Again, we would like to congratulate Kimberly F. Sellers, Sharad Borle, and Galit Shmueli for their comprehensive paper. It deals with a problem of growing relevance and presents novel methods for handling count data. Moreover, and as briefly discussed previously, it can be linked to many other areas of research and thereby open up even further opportunities for advancing the state of the art in this domain.

Appendix

```
R CODE:

  library(COMPoissonReg)
  # freight <- read.table("C:/CH01PR21.txt",header=FALSE)
  data(freight)

  # Compute lm estimates
  coef(lm(log(broken) ~ transfers, data=freight))

  # Compute Poisson estimates
  glm_model <- glm(broken ~ transfers, data=freight, family=poisson,
na.action=na.exclude) # beta estimates
  print("The standard Poisson estimates for the beta vector are")
  print(coef(glm_model))

  # Compute CMP estimates (under constant dispersion model)
  cmp_model = cmp(formula = broken ~ transfers, data=freight)
  print("The COM-Poisson estimates for the beta vector are")
  print(coef(cmp_model))
  print("The COM-Poisson estimate for the dispersion parameter nu is")
  print(nu(cmp_model))
  print("The associated standard errors for the betas in the constant
dispersion case are")
  print(sdev(cmp_model))

  cat("\n \n -- Boot -- \n \n")

  # a function for boot
  bootGlmCmpCoeff <- function(data, index, formula){
    lm_model <- lm(log(broken) ~ transfers, data=data[index,])
    library(COMPoissonReg)
    cmp_model = cmp(formula = formula, data= data[index,])
    cmp_model$glm_coefficients
```

```
  cmp_model$coefficients
  #sdev(cmp_model)[1:length(cmp_model$coefficients)]
  return(c(cmp_model$glm_coefficients,
    cmp_model$coefficients,
    coef(lm_model)))
}

library(boot)
BootGlmCmpCoeff <- boot(data = freight, bootGlmCmpCoeff,
  R = 1000,
  formula = as.formula("broken ~ transfers"))
```

RON S. KENETT
KPA Ltd. and University of Torino
E-mail: *ron@kpa-group.com*

References

1. Abeles M, Bergman H, Gat I, Meilijson I, Seidemann E, Thishby N, Vaadia E. Cortical activity flips among quasi stationary states. *Proceedings of the National Academy of Sciences, USA* 1995; **92**:8616–8620.
2. Atkinson AC, Riani M. *Robust Diagnostic Regression Analysis*. Springer Verlag: New York, 2000.
3. Deming WE. On the distinction between enumerative and analytic surveys. *Journal of the American Statistical Association* 1953; **48**(262): 244–255.
4. Fuchs C, Kenett RS. A test for detecting outlying cells in the multinomial distribution and two-way contingency tables. *Journal of the American Statistical Association* 1980; **75**:395–398.
5. Hahsler M, Grun B, Hornik K. The arules package: mining association rules and frequent itemsets, version 0.6-6, http://cran.r project.org/web/packages/arules/index.html, 2008.
6. Kenett RS, Rahav E, Steinberg D. Bootstrap analysis of designed experiments. *Quality and Reliability Engineering International* 2006; **22**:659–667.
7. Kenett RS, Salini S. Relative linkage disequilibrium applications to aircraft accidents and operational risks. *Transactions on Machine Learning and Data Mining* 2008; **1**(2):83–96.
8. Kenett RS, Salini S. Measures of association applied to operational risks. In *Operational Risk Management: A Practical Approach to Intelligent Data Analysis*, Kenett RS, Raanan Y (eds). John Wiley and Sons: Chichester, UK, 2010.
9. McCullagh P. Regression models for ordinal data. *Journal of the Royal Statistical Society* 1980; **42**:109–142.
10. Riani M, Torti F, Zani S. Outliers and robustness for ordinal data. In *Modern Analysis of Customer Satisfaction Surveys*, Kenett RS, Salini S (eds). John Wiley and Sons: Chichester, UK, 2011.
11. Sellers K, Shmueli G. A flexible regression model for count data. *Annals of Applied Statistics* 2010; **4**(2):943–961.
12. Shmueli G. To explain or to predict? *Statistical Science* 2010; **25**(3):289–310.

Applied Stochastic Models in Business and Industry

Discussion

(wileyonlinelibrary.com) DOI: 10.1002/asmb.937 Published online in Wiley Online Library

The Conway–Maxwell–Poisson model for analyzing crash data[‡]

1. Introduction

Sellers *et al*. [1] have provided an excellent summary about the flexible and unique properties of the COM-Poisson model both in terms of methodological advancements and applications. This discussion paper further expands on some of these topics, but focuses on the latest research on the use of the COM-Poisson for reducing the negative effects associated with motor vehicle collisions.

The appropriate selection of a model can be quite complex. The selection process is often guided by the study objectives (e.g., prediction versus relationships, etc.), inference goals (e.g., confidence intervals or parameter distribution), and the availability and quality of data. In highway safety (a subfield of transportation engineering), crash datasets (see Section 2) are often characterized by distinctive attributes not commonly found in other areas, such as biology, marketing or environmental engineering. Because of the large costs associated with the data collection procedure, crash datasets often contain a relatively small number of observations [2]. Regression models developed from datasets containing less than 30 observations are not uncommon. Furthermore, because crashes are rare events statistically speaking, many datasets are characterized by low sample mean values, which have been shown to provide biased or unreliable estimates when traditional regression models are used for risk analysis [3]. In short, crash data collected for analysis purposes are considered unique datasets in their own rights.

The application of the COM-Poisson model in highway safety research was initially fostered to examine the model properties for handling crash datasets. So far, the research on the COM-Poisson model has focused on under-dispersion, the observation-specific variance function, and its performance as a function of sample size and sample mean values. Each topic is discussed further below. But first, the characteristics related to crash data are briefly described in the next section.

2. Crash data

Motor vehicle crashes (e.g., collisions between two vehicles, vehicles running-off-the-road and hitting a tree, etc.) are ranked among the 10 leading causes of injuries across the world [4]. In the US, traffic crashes only fall behind cancer and heart disease as the total loss of human life (i.e., human-years) [5]. It is estimated that the economic societal cost caused by motor vehicle crashes was equal to $230.6 billion in 2000 [6]. Given the huge negative impact crashes have on society, the federal and state governments have placed a significant amount of resources on reducing the number and severity of crashes that plague the transportation network.

In highway safety, crash data are considered the best sources for estimating the safety performance of various components of the highway network, such as intersections, highway segments, bridges, and pedestrian crosswalks. These data can provide useful information about potential deficiencies located within the highway network and can consequently be used for allocating resources to implement countermeasures that would reduce the morbidity caused by these events. Crash data can also be used for evaluating the effects of these countermeasures or treatments. Because crash data are discrete, random and non-negative events, various statistical tools are need for conducting different kinds of safety analyses (e.g., identify hazardous sites; determine contributing factors that influence crashes; evaluate

[‡]*Discussion paper associated with 'The COM-Poisson Model for Count Data: A Survey of Methods and Applications' by Sellers, K., Borle, S., and Shmueli, G.*

countermeasures, etc.). Count data models are still considered the most popular tools for analyzing this type of data [7].

3. Under-dispersion

Although most count datasets are characterized by over-dispersion, it has been found that crash data can sometimes be plagued by under-dispersion. The under-dispersion can be generated by the data generating process (as discussed in [1]) or attributed to the modeling process [8, 9]. For the latter, the under-dispersion occurs when the observed count is modeled conditional upon its mean (i.e., marginal model). In some cases, the under-dispersion can be a sign of over-fitting, meaning that the model may contain too many variables. In highway safety, it has been noted that datasets characterized with low sample values sometimes lead to under-dispersion.

To handle under-dispersion, researchers in various fields have proposed alternative models. Sellers *et al*. [1] discussed some of them, including the weighted [10] and the generalized Poisson models [11]. There are also a few other models that have been proposed for handling under-dispersion. The first model is the gamma model, which can handle both over-dispersion and under-dispersion, similar to the COM-Poisson model. Two parameterizations have been proposed. The first parameterization makes use of the continuous gamma function [12], which means that the mean λ cannot be equal to zero (technically speaking, a continuous model or distribution should not be used to analyze count data). This obviously limits its applicability because, in many datasets, it may not be feasible for the mean to be equal to zero (see Ref. [13] about this assumption in highway safety). The second parameterization is based on the gamma waiting-time distribution, which allows for a monotonic increasing or decreasing function [14, 15]. For this parameterization to work, the observations are assumed to be dependent where the observation at time $t-1$ directly influences the observation at time t. For some datasets, this may be possible, but for most datasets, this is not realistic. For instance, a crash that occurred at time t cannot directly influence another one that will occur 6 months after the first event. The second model is known as the Double Poisson model [16]. Similar to the gamma and COM-Poisson models, it can also handle both over-dispersion and under-dispersion, by adding an extra term. This model has rarely been used by other researchers. It is currently being evaluated in the context of crash data analysis by the research team at Texas A&M University and some preliminary results seem to suggest that the distribution has some difficulties handling under-dispersion (compared with the COM-Poisson), but it appears to be more reliable for over-dispersed data.

This is where the COM-Poisson model offers more flexibility over the existing methods for handling under-dispersion. Lord *et al*. [9] have shown that the COM-Poisson performed better than the gamma-waiting time model for data characterized by under-dispersion. Using the data collected at 162 railway–highway crossings (i.e., train–vehicle collisions) located in South Korea between 1998 and 2002 [8], the COM-Poisson provided a better fit and predictive capabilities compared with the gamma-waiting time parameterization. Table I, taken from the original paper, summarizes the comparison analysis.

Table I. Parameter estimates with three different distributions (MLE) [25].

Variables	COM-Poisson	Poisson	Gamma
Constant	–6.657 (1.206)[a]	–5.326 (0.906)[a]	–3.438 (1.008)[a]
Ln(Traffic Flow)	0.648 (0.139)	0.388 (0.076)	0.230 (0.076)
Average daily railway traffic	—[b]	—	0.004 (0.0024)
Presence of commercial area	1.474 (0.513)	1.109 (0.367)	0.651 (0.287)
Train detector distance	0.0021 (0.0007)	0.0019 (0.0006)	0.001 (0.0004)
Time duration between the activation of warning signals and gates	—	—	0.004 (0.002)
Presence of track circuit controller	–1.305 (0.431)	–0.826 (0.335)	—
Presence of guide	–0.998 (0.512)	—	—
Presence of speed hump	–1.495 (0.531)	–1.033 (0.421)	–1.58 (0.859)
Shape Parameter (ν_0)	2.349 (0.634)	—	2.062 (0.758)
AIC	210.70	196.55	211.38
MPB	–0.007	0.004	0.179
MAD	0.348	0.359	0.459
MSPE	0.236	0.252	0.308

[a] Standard error.
[b] The variable was not statistically significant at the 15% level.
AIC, Akaike Information Criterion; MPB, mean pediction bias; MAD, mean absolute deviance; MSPE, mean squared predicted error.

4. Observation-specific variance function

As discussed by Sellers *et al*. [1], the COM-Poisson model is very flexible for modeling the dispersion. The model can be estimated using a constant dispersion or can allow the dispersion to vary for different observations (observation-specific). The latter parameterization in the context of a regression model was initially proposed by Guikema and Coffelt [17, 18]. The parameterization for the mean and the observation-specific shape parameter is described in Equations (1) and (2):

$$\ln(\mu) = \beta_0 + \sum_{i=1}^{p} \beta_i x_i \tag{1}$$

$$\ln(\nu) = \gamma_0 + \sum_{j=1}^{q} \gamma_j z_j \quad , \tag{2}$$

where x_i and z_j are covariates, with p covariates used in the centering link function and q covariates used in the shape link function (the sets of parameters used in the two link functions do not necessarily have to be identical); μ is a centering parameter that is approximately the mean of the observations in many cases (recall that $\mu = \lambda^{1/\nu}$, as discussed in [14]); and, ν is defined as the shape parameter of the COM-Poisson distribution and is used for estimating the variance.

Recent research in highway safety has shown that the dispersion parameter of the NB model can potentially be dependent upon the covariates of the model and could vary from one observation to another [19–21]. This characteristic has been shown to be observed when the mean function is mis-specified [22], although this may not be true in all cases [23]. When this occurs, the dispersion parameter of the model is considered to be structured (because the coefficients linking the dispersion/shape parameter to the covariates are statistically significant) meaning that it is dependent upon the characteristics of the data [21]. For instance, the dispersion parameter, hence the variance observed in the crash count, could vary geographically, but is not captured in the modeling process. The curious reader is referred to Ref. [21] for additional information about attributes associated with the observation-specific variance function of NB models.

Given this important characteristic, it became essential to examine whether the variance that varied across observations was the same between the NB and the COM-Poisson models. That is, are the observations with a large or small variance the same between both models? Geedipally and Lord [24] examined this characteristic using two crash datasets and found that the variance function estimated for both models were almost the same. For the NB, the variance was calculated using the traditional function $Var[Y_k] = \lambda_k + \alpha_k \lambda_k^2$, where μ_k is the estimated mean value and α_k is the estimated dispersion parameter specific for observation k. For the COM-Poisson, the variance can be estimated using the following relationship $Var[Y_k] \approx \mu_k/\nu_k$. Because this variance function is based on an approximation, the variance was estimated using the methodology proposed by Francis *et al*. [25]. For a given μ_k and ν_k, 100,000 observations were simulated and the sample mean and variance were calculated from the simulated data.

Figure 1, which is based on crash data collected at 868 signalized intersections in Toronto, Ont. in 1995, shows the comparison analysis for the variance estimated by the NB and COM-Poisson models. The models linked the number of crashes

Figure 1. Crash variance versus crash mean for the Texas data [38] (Note: x-axis and y-axis are formatted under a logarithmic scale).

to the entering traffic flows, defined as major and minor legs, at the intersections: : μ_k (or λ_k) $= \beta_0 F_{k\_major}^{\beta_1} F_{k\_minor}^{\beta_2}$; $\alpha_k = \delta_0 F_{k\_major}^{\delta_1} F_{k\_minor}^{\delta_2}$; and, $\nu_k = \gamma_0 F_{k\_major}^{\gamma_1} F_{k\_minor}^{\gamma_2}$. The study results showed that the COM-Poisson model was able to capture the same variance as for the NB model.

5. Model performance

For a model to be useful, it needs to be reliably and robustly estimated. In other words, the error and bias associated with the model's coefficients should be minimized. It is well-known that the coefficients of models can be strongly and

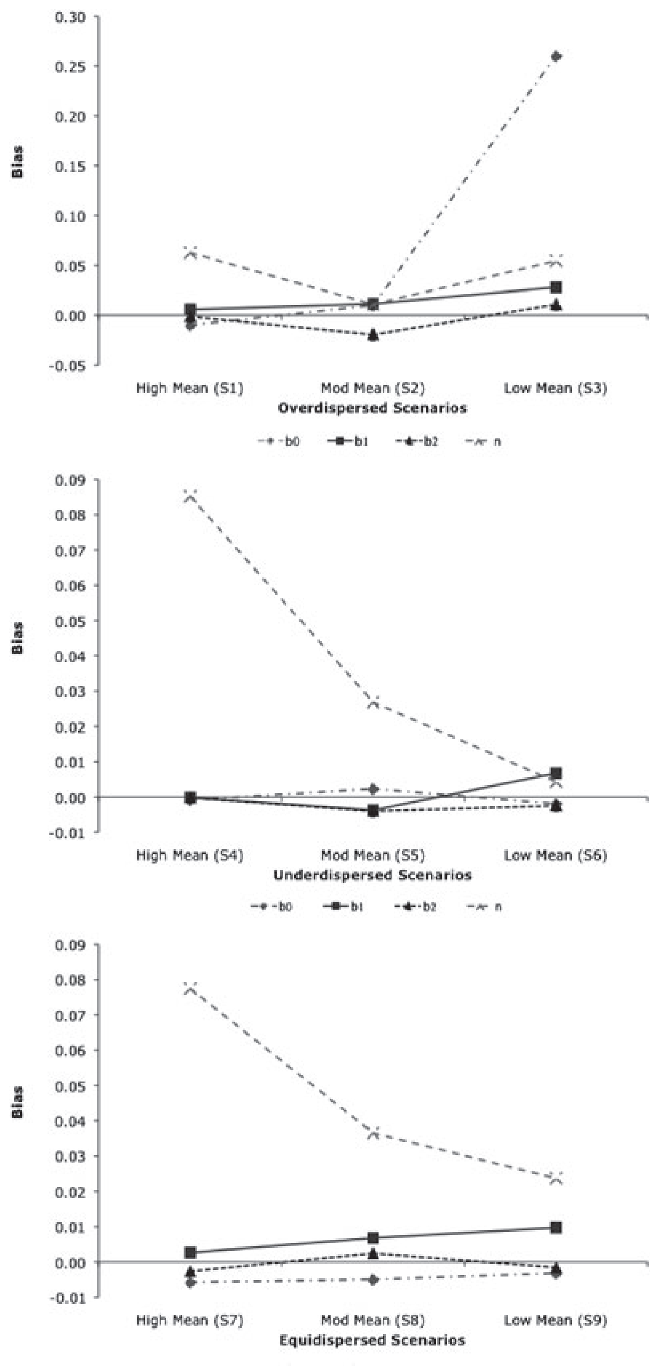

Figure 2. Prediction bias for parameter estimates under each scenario (absolute value). $\boldsymbol{\beta}_0$ diamonds, $\boldsymbol{\beta}_1$ squares, $\boldsymbol{\beta}_2$ triangles, $\boldsymbol{\nu}$ indicated by "x" on the line [25].

negatively influenced by small sample sizes and low sample means, characteristics often observed with crash data. Several researchers have noted that traditional models, such as the NB model, provide unreliable and biased estimates when they are estimated with such datasets [3, 26–29]. Although no model is immune to these problems, some models may perform better than others under extreme circumstances. For instance, it has been reported that the Poisson-lognormal model is less affected than the NB model for the problems described above [30].

The issue of model performance for the COM-Poisson model has recently been investigated by Francis *et al*. [25]. The authors analyzed the estimation accuracy of the model proposed by Guikema and Coffelt [17, 18] for datasets characterized by over-dispersion, equidispersion and under-dispersion with different mean values based on the maximum likelihood estimator. The analysis was based on 900 simulated datasets representing nine different mean-variance relationships. The functional form of the model was $\mu = \exp(\beta_0 + \beta_1 x_1 + \beta_2 x_2)$ where x_1, x_2 follows a uniform distribution on [0, 1]. The results showed that the COM-Poisson model was very accurate for all datasets with high and medium sample means irrespective of the type of and level dispersion observed in the data (see Figure 2). However, the model suffered from important limitations for moderate-mean and low-mean over-dispersed data, similar to the NB model, as the sample mean decreased. Interestingly, the model was much less affected when the data are under-dispersed, further suggesting that it is a viable alternative to the other models that have been proposed for modeling under-dispersion.

6. Concluding thoughts

Sellers *et al*. [1] have discussed how the COM-Poisson model has been successfully applied in a variety of fields and its flexibility for analyzing different types of data whether they are over-dispersed, under-dispersed or censored. Bearing in mind that the distribution was first used in a regression setting in the 2003–2005 time frame (see [1]), 'reintroduced' in 2005 [31, 32], and the GLM more fully developed a year later [17], what is very surprising is how quickly the distribution and the model have been used by analysts all around the world and further expanded by researchers over the last four or five years. A quick look at Scopus shows that more than 22 papers have been published in peer-reviewed journals and international conferences on the COM-Poisson (this number does not include forthcoming papers discussed above). All these papers have a combined citation record equal to 114 (at the time this discussion paper was written). This is no small feat, especially considering that a few existing models introduced several years ago (e.g., Double Poisson, etc.) have barely been used by scientists and researchers.

In highway safety, the COM-Poisson model has been applied for different conditions, such as for over-dispersed and under-dispersed data and when the variance is dependent upon the structure of the data. So far, the work on the model has been spearheaded by a small group of researchers, but others have recently shown an interest in using and further developing the COM-Poisson model for safety analyses. The work so far has shown that the COM-Poisson model offers a viable alternative to traditional models for exploring the unique characteristics associated with motor vehicle crashes. However, as discussed by Sellers *et al*. [1], there remain several opportunities for sustained theoretical and applied research related to the COM-Poisson distribution and the regression in highway safety and other fields.

Dominique Lord
Zachry Department of Civil Engineering, Texas A&M University,
3136 TAMU, College Station, TX 77843-3136, USA
E-mail: *d-lord@tamu.edu*

Seth D. Guikema
Department of Geography and Environmental Engineering, Johns Hopkins University
313 Ames Hall, 3400 N. Charles St., Baltimore, MD 21218, USA

References

1. Sellers K, Borle S, Shmueli G. The COM-Poisson Model for Count Data: A Survey of Methods and Application. *Applied Stochastic Models in Business and Industry* 2012.
2. Lord D, Bonneson JA. Calibration of Predictive Models for Estimating the Safety of Ramp Design Configurations. *Transportation Research Record* 1908; **2005**:88–95.
3. Lord D. Modeling motor vehicle crashes using Poisson-gamma models: examining the effects of low sample mean values and small sample size on the Estimation of the fixed dispersion parameter. *Accident Analysis and Prevention* 2006; **38**(4):751–766.
4. WHO. World Report on Road Traffic Injury Prevention. Eds. Peden et al. World Health Organization, Geneva, 2004. (accessed on Sept. 28,2011:http://www.who.int/violence_injury_prevention/publications/road_traffic/world_report/en/).
5. NHTSA. Traffic Safety Facts, 2004 Data. National Highway Traffic Safety Administration, Washington, D.C., 2005. (accessed on Sept. 8, 2011: http://www-nrd.nhtsa.dot.gov/Pubs/809911.PDF).

6. Blincoe L, Seay A, Zaloshnja E, Miller T, Romano E, Lutchter S, Spicer R. The Economic Impact of Motor Vehicle Crashes. *Report No. DOT HS 809 446*, National Highway Traffic safety Administration, Washington, D.C., 2000. (accessed on Sept. 8, 2011: http://www.cita-vehicleinspection.org/Portals/cita/autofore_study/LinkedDocuments/literature/NHTSA%20the%20economic%20impact%20of%20motor%20vehicle%20crashes%202000%20USA%202002.pdf).
7. Lord D, Mannering F. The Statistical Analysis of Crash-Frequency Data: A Review and Assessment of Methodological Alternatives. *Transportation Research - Part A* 2010; **44**(5):291–305.
8. Oh J, Washington SP, Nam D. Accident prediction model for railway-highway interfaces. *Accident Analysis and Prevention* 2006; **38**(2):346–356.
9. Lord D, Geedipally SR, Guikema S. Extension of the Application of Conway-Maxwell-Poisson Models: Analyzing Traffic Crash Data Exhibiting Under-Dispersion. *Risk Analysis* 2010; **30**(8):1268–1276.
10. del Castillo J, Pérez-Casany M. Overdispersed and underdispersed Poisson generalizations. *Journal of Statistical Planning and Inference* 2005; **134**:486–500.
11. Consul PC. *Generalized Poisson Distributions: Properties and Applications*. Marcel Dekker: New York, 1989.
12. Daniels S, Brijs T, Nuyts E, Wets G. Explaining variation in safety performance of roundabouts. *Accident Analysis and Prevention* 2010; **42**(6):1966–1973.
13. Lord D, Washington SP, Ivan JN. Poisson, Poisson-gamma and zero inflated regression models of motor vehicle crashes: balancing statistical fit and theory. *Accident Analysis and Prevention* 2005; **37**(1):35–46.
14. Cameron AC, Trivedi PK. *Regression Analysis of Count Data*. Cambridge University Press: Cambridge, U.K., 1998.
15. Winkelmann R. Duration Dependence and Dispersion in Count Data Models. *Journal of Business and Economic Statistics* 1995; **13**:467–474.
16. Efron B. Double exponential families and their use in generalized linear regression. *Journal of the American Statistical Association* 1986; **81**:709–721.
17. Guikema SD, Coffelt JP. An application of the Conway-Maxwell-Poisson generalized linear model. *Society for Risk Analysis Annual Meeting*, Baltimore, MD, December 2006.
18. Guikema SD, Coffelt JP. A flexible count data regression model for risk analysis. *Risk Analysis* 2008; **28**(1):213–223.
19. Heydecker BG, Wu J. Identification of Sites for Road Accident Remedial Work by Bayesian Statistical Methods: An Example of Uncertain Inference. *Advances in Engineering Software* 2001; **32**:859–869.
20. Hauer E. Overdispersion in modelling accidents on road sections and in Empirical Bayes estimation. *Accident Analysis and Prevention* 2001; **33**(6):799–808.
21. Miaou S-P, Lord D. Modeling Traffic-Flow Relationships at Signalized Intersections: Dispersion Parameter, Functional Form and Bayes vs Empirical Bayes. *Transportation Research Record* 1840; **2003**:31–40.
22. Mitra S, Washington SP. On the nature of over-dispersion in motor vehicle crash prediction models. *Accident Analysis and Prevention* 2007; **39**(3):459–468.
23. Geedipally SR, Lord D, Park B-J. Analyzing Different Parameterizations of the Varying Dispersion Parameter as a Function of Segment Length. *Transportation Research Record 2103* 2009:108–118.
24. Geedipally SR, Lord D. Examining the Crash Variances Estimated by the Poisson-Gamma and Conway-Maxwell-Poisson Models. *Transportation Research Record* **2011**, in press.
25. Francis RA, Geedipally SR, Guikema SD, Dhavala SS, Lord D, Larocca S. Characterizing the Performance of the Conway-Maxwell-Poisson Generalized Linear Model. *Risk Analysis* 2012. DOI: 10.1111/j.1539-6924.2011.01659.x.
26. Clark SJ, Perry JN. Estimation of the Negative Binomial Parameter κ by Maximum Quasi-Likelihood. *Biometrics* 1989; **45**:309–316.
27. Piegorsch WW. Maximum Likelihood Estimation for the Negative Binomial Dispersion Parameter. *Biometrics* 1990; **46**:863–867.
28. Dean CB. Modified Pseudo-Likelihood Estimator of the Overdispersion Parameter in Poisson Mixture Models. *Journal of Applied Statistics* 1994; **21**(6):523–532.
29. Wood GR. Generalised linear accident models and goodness of fit testing. *Accident Analysis and Prevention* 2002; **34**(4):417–427.
30. Lord D, Miranda-Moreno LF. Effects of Low Sample Mean Values and Small Sample Size on the Estimation of the Fixed Dispersion Parameter of Poisson-gamma Models for Modeling Motor Vehicle Crashes: A Bayesian Perspective. *Safety Science* 2008; **46**(5):751–770.
31. Shmueli G, Minka TP, Kadane JB, Borle S, Boatwright P. A useful distribution for fitting discrete data: revival of the Conway-Maxwell-Poisson distribution. *Journal of the Royal Statistical Society: Part C* 2005; **54**:127–142.
32. Kadane JB, Shmueli G, Minka TP, Borle S, Boatwright P. Conjugate analysis of the Conway-Maxwell-Poisson distribution. *Bayesian Analysis* 2006; **1**:363–374.

Applied Stochastic Models in Business and Industry

(wileyonlinelibrary.com) DOI: 10.1002/asmb.1923 Published online in Wiley Online Library

Rejoinder: The COM–Poisson Model for count data: A survey of methods and applications

We thank the two discussants for broadening the discussion of the Conway-Maxwell-Poisson (COM-Poisson) in different directions and for additional suggestions for extending the COM–Poisson.

Kenett's discussion offers further insights and background regarding applications of count data (e.g., spike trains in brain studies) and the contribution that descriptive models, such as the COM–Poisson distribution, can make to understand physical models or processes that exhibit data dispersion of various magnitudes and directions (over-dispersion and under-dispersion). Kenett mentions the potential of using the COM–Poisson 'in conjunction with other modeling efforts, for deriving predictions in yet untested conditions.' The usefulness of a model for description, however, does not guarantee its usefulness for out-of-sample prediction; and hence, we re-emphasize the open question regarding predictive power, given the mixed empirical evidence.

Section 2 of Kenett's discussion relates overdispersion to the presence of outliers in the data. As noted by Hilbe [1], however, there are numerous other causes of overdispersion, including omission of important explanatory predictors, failure to include interaction terms, or misspecification of the link function. Nonetheless, Kenett outlines an approach by which the COM–Poisson can be used to explain outliers. This discussion is interesting as it relates to the use of the COM–Poisson for the purposes of disclosure limitation, which is an application that we mentioned in the survey paper (Section 4.5). Disclosure limitation seeks to provide individual privacy while releasing sensitive data. Kadane *et al.* [2] use the sufficient statistics from a COM–Poisson distribution to mask the size and contents of a one-way table. In particular, this approach masks outlier information and Kenett's discussion supports the use of the COM–Poisson to model such sensitive data.

Kenett also discusses extending the COM–Poisson to two-way or higher-dimensional contingency tables. Sellers and Balakrishnan [3] address the matter of two-way tables by developing a bivariate COM–Poisson distribution that includes the bivariate Bernoulli, bivariate geometric, and bivariate Poisson distributions as special cases. Thus, the bivariate COM–Poisson distribution provides a flexible distribution for count data in the presence of data over-dispersion or under-dispersion.

The discussion paper by Lord and Guikema, meanwhile, provides additional insight on the COM–Poisson distribution, particularly when applied to motor vehicle collision (crash) data. The authors note certain characteristics of crash data, namely under-dispersion and small sample size, which make the COM–Poisson distribution an attractive choice. They also compare the COM–Poisson to other distributions and models that can handle over-dispersion or under-dispersion (e.g., the double Poisson distribution) that have been suggested in their field.

Section 4 of Lord and Guikema's discussion focuses on the COM–Poisson model with observation-specific variance. In general, observation-specific dispersion can be incorporated into a COM–Poisson model by allowing either of the parameters of the distribution to vary across observations. Further, this variation can be linked to covariates relevant to the application; for two such applications, see Boatwright *et al.* [4] and Borle *et al.* [5].

With respect to comparing the variance of count models such as the negative binomial and the COM–Poisson and comparing estimated variances of data using formulas of the variance, we note that the highly skewed nature of such count distributions (aside from special cases) makes measures such as percentiles more adequate and useful for evaluating performance or describing the data.

In Section 5, the authors reiterate the advantage of the COM–Poisson especially when data are under-dispersed. This is an important issue and many studies have pointed to this distinct advantage of the COM–Poisson.

Finally, Lord and Guikema highlight the usefulness of the COM–Poisson in practice by quantifying the large number of publications over a relatively short period. In our eyes, this is a clear indication of the importance of a close link between practical need and methodological development in the field of statistics. We look forward to further developments of the COM–Poisson that address real needs, as well as its application in new fields where count data are prevalent, which will hopefully stir even further needed methodological developments.

KIMBERLY F. SELLERS
Department of Mathematics and Statistics
Georgetown University
Washington, DC 20057, USA
E-mail: *kfs7@georgetown.edu*

SHARAD BORLE
Department of Marketing
Jones Graduate School of Business
Rice University
Houston, TX 77005, USA

GALIT SHMUELI
Indian School of Business
Hyderabad 500 032, India

References

1. Hilbe JM. *Negative Binomial Regression*. Cambridge University Press: New York, NY, 2008.
2. Kadane JB, Krishnan R, Shmueli G. A data disclosure policy for count data based on the COM–Poisson distribution. *Management Science* 2006; **52**:1610–1617. DOI: 10.1287/mnsc.1060.0562.
3. Sellers KF, Balakrishnan N. Bivariate Conway-Maxwell-Poisson distribution: formulation, properties, and inference, working paper, Georgetown University 2012.
4. Boatwright P, Borle S, Kadane JB. A model of the joint distribution of purchase quantity and timing. *Journal of the American Statistical Association* 2003; **98**:564–572. DOI: 10.1198/016214503000000404.
5. Borle S, Boatwright P, Kadane JB, Nunes J, Shmueli G. Effect of product assortment changes on consumer retention. *Marketing Science* 2005; **24**:616–622. DOI: 10.1287/mksc.1050.0121.