Rejoinder: The COM-Poisson Model for Count Data: A Survey of Methods and Applications

Kimberly F. Sellers
Department of Mathematics and Statistics
Georgetown University
Washington, DC 20057
USA

Sharad Borle
Department of Marketing
Jones Graduate School of Business
Rice University
Houston, TX 77251
USA

Galit Shmueli
Indian School of Business
Hyderabad 500 032
India

We thank the two discussants for broadening the discussion of the COM-Poisson in different directions, and for suggesting additional future directions for extending the COM-Poisson.

Kenett's discussion offers additional 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 further understand physical models or processes that exhibit data dispersion of various magnitudes and directions (over- 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 (2011) relates overdispersion to the presence of outliers in the data. As noted by Hilbe (2008), 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. (2006) 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 the discussion in Kenett (2011) 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 (2012) have addressed the matter of two-way tables by developing a bivariate Conway-Maxwell-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- 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- or under-dispersion (e.g. the double Poisson distribution) and 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. (2003) and Borle et al. (2005).

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 would like to 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 5.0, 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 of time. 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 methodological developments of the COM-Poisson that answer real needs, as well as its application in new fields where count data are prevalent, which will hopefully stir even further needed methodological developments.

References

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

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

Hilbe JM. *Negative Binomial Regression*. Cambridge University Press: New York, NY, 2008.

Sellers KF, Balakrishnan N. Bivariate Conway-Maxwell-Poisson Distribution: Formulation, Properties, and Inference, working paper, Georgetown University, 2012.

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.