Trent L. Lalonde
Spring 2014 Course Release
Joint Modeling of Mean and Dispersion Using Moment Methods

During the Spring 2014 semester I made significant progress according to my proposal, "Joint Modeling of Mean and Dispersion Using Moment Methods." The specific accomplishments were unexpected but followed the general goals of the original proposal, as outlined below. Through a combination of conference and invited presentations, manuscripts, book chapters, and original programs, this work has developed into a nice line of research.

The first proposed step was to develop a joint estimation method for the mean and dispersion of a (possibly non-normal) response using a generalized method of moments (GMM) approach. My intention was to use the joint generalized estimating equations (GEE) approaches as a model for extending GMM. Unfortunately, I have concluded that an extension of GMM similar to the GEE2 model proposed by Liang, Zeger, and Qaqish (1992) is infeasible. This is because the GEE2 estimation method relies on a user-defined correlation structure for the longitudinal responses. This correlation structure implicitly defines parameters for the dispersion model, through the parameters that describe the correlation. Because these correlation parameters are involved in estimation of mean parameters through the correlation structure in the GEE2 equations, the mean and dispersion are naturally connected. The GMM does not involve a correlation structure, and instead relied on moment conditions that are defined using mean parameters only. As such, the GMM estimation method does not allow for the same type of association between the mean estimation and the dispersion estimation.

The second through fourth proposed steps of this project depended on successful development of such a joint method, including implementing the method in R, comparing to existing likelihood methods using two longitudinal data sets, and preparing an R package. Due to the unforeseen hurdles in joint GMM development, these goals had to be adjusted. The following four tasks were completed during the release time.

(1) When a direct extension of existing joint estimation methods using GMM proved to be infeasible, I turned to another class of joint models, those for excess zeros of counts. While not involving joint estimation of both mean parameters and dispersion parameters, this class of models uses two sub-models to describe two characteristics of the behavior of mean counts. Similar to my original research interest, the two sub-models are comprised of two collections of model equations, distributions, and model parameters that need to be estimated jointly. This is very similar to the general structure of joint mean and variance modeling, with less direct associations between the two sub-models. I planned to use this specific type of joint modeling as a way to springboard to joint mean and dispersion estimation.

Specifically, "excess zeros" refers to count responses that show more "zero" observations than would typically be expected. Applications include rare events, for example unusual illnesses or diagnoses, or behaviors that are rarely exhibited. Most subjects would report
“zero” of these events. Two common models are applied for these cases, the ZIP model and the Hurdle model. Both models involve one sub-model that addresses the prevalence of observations of “zero” counts, along with a second sub-model that addresses the prediction of non-zero counts. The Hurdle Poisson (Hurdle) model contains an implicit assumption that individuals who report “zeros” are distinct from those who report positive counts. The Zero-Inflated Poisson (ZIP) model, on the other hand, allows all individuals in the studies the possibility of reporting “zero” counts.

Both the ZIP and Hurdle models have been extended to longitudinal data cases, using either likelihood-based methods (Markov Chain Monte Carlo estimation, MCMC) or the GEE method (currently ZIP GEE). As part of my Spring 2014 course release I performed a Monte Carlo simulation to compare these existing methods for longitudinal data, which has been accepted and will be presented at the Joint Statistical Meetings (JSM) in Boston, MA on August 3. The abstract for the JSM presentation is attached, and is also available at http://www.amstat.org/meetings/jsm/2014/onlineprogram/AbstractDetails.cfm?abstractid=312058. I have developed an extension of the ZIP GEE estimation method to a ZIP GMM method for longitudinal data, and have been invited to speak on this development at UC-Denver on November 5. The ZIP GMM is being prepared for submission to The Australian-New Zealand Journal of Statistics, a much stronger journal than the proposed Journal of Applied Statistics. The combination of the comparisons to be represented at JSM 2014 and the GMM extension to be presented at UC-Denver have been accepted as a roundtable discussion at the 2014 meeting of the American Public Health Association (APHA) in New Orleans, LA on November 17. The abstract for the APHA presentation is attached, and is also available at https://apaha.confex.com/apha/142am/webprogram/Paper299240.html. Through the winter I expect to extend the process of ZIP GMM to the joint estimation of mean and dispersion parameters, my original interest, and submit to JSM 2015 as an invited session.

(2) In addition to pursuing the joint models for excess zero counts, in casual conversation I learned that Michael Phillips was working on a Spring 2014 course release with data and research interests that align well with joint estimation of mean and dispersion. Specifically, he was interested in modeling both the center and the fluctuation in “craving” of individuals with a certain drug addiction. I have used his data set in place of the large national data sets proposed, as Michael’s data set is not widely available and thus should lead to a stronger reaction in review due to its novelty and the possibility of a unique contribution to the literature. Together we have applied many of the existing joint-glm estimation techniques and are preparing a manuscript of the results, possibly for Addiction.

(3) In addition to the paper summarizing the results of joint estimation of the mean and dispersion of “craving,” Michael and I have drafted a paper covering the options currently available for such joint-glm estimation. The paper reads as an instructional discussion of current modeling options, parameter explanations, and available software. It is intended for an audience of research practitioners and should provide a basis of knowledge and examples to replicate joint analysis of mean and dispersion, and is being prepared for Psychological Methods.
(4) Finally, based on conversations with colleagues external to UNC about my intentions and challenges in extending joint estimation of mean and dispersion parameters using GMM, I was invited to contribute a book chapter covering joint-logistic estimation of mean and variation. This chapter was completed in June of 2014, and the text as a whole is currently under review. The cover page of the chapter is attached; the entire chapter is available upon request.

Addressing my original proposal intentions, the joint GMM estimation will be pursued during the coming fall and spring semesters, with a strong outline now complete from the excess zero count models. Both will be implemented using R. Comparisons of existing methods appear in the JSM 2014 submission and in the second article with Michael Phillips. In place of the original tangible goals, this course release has led directly to a conference presentation, an invited talk, and a conference roundtable presentation; two manuscripts in progress with another UNC colleague and one single manuscript in progress; one book chapter contribution.
Activity Number: 28
Type: Contributed
Date/Time: Sunday, August 3, 2014: 2:00 PM to 3:50 PM
Sponsor: Biometrics Section
Abstract: #31558
Title: Model Comparisons for Correlated Count Responses with Excess Zeros and Time-Dependent Covariates
Authors: Trent L. Labods*
Company: University of Northern Colorado
Keywords: Longitudinal, Count Regression, Excess Zeros, Time-Dependent Covariates
Abstract:
Count responses often show an excess of zeros under the assumption of a Poisson distribution. Common modelling solutions include the zero-inflated Poisson model and the hurdle mixture model (Hu et al (2011); Murphy (1986); Lambert (1992)). Recently researchers have begun to consider the modeling options for clustered or correlated count responses with excess zeros (Kasahurtn et al, preprint). However, it has yet to be considered whether certain models are preferred for correlated count responses with excess zeros in the presence of time-dependent covariates. Time-dependent covariates have been shown to affect parameter estimates bias and efficiency in longitudinal analyses (Pope and Anderson (1994); Fitzmaurice (1995); Lai and Small (2007)). In this paper a comparison is made between the zero-inflated Poisson and the hurdle model for correlated count data with time-dependent covariates. Consideration is given to parameter estimates bias and hypothesis test results. An example data set is analyzed, using a longitudinal measure of the number of times of drug use as response.

Authors who are presenting talks have a * after their name.

Back to the full JSM 2014 program
Modeling Longitudinal Count Data with Excess Zeros and Time-Dependent Covariates: Application to Drug Use

Online Program

T. Lalonde, PhD, Department of Applied Statistics and Research Methods, University of Northern Colorado, Greeley,

When longitudinal count response data are collected, mixed Poisson regression models can give appropriate subject-specific risk results. When count responses show an excess of zeros, either a zero-inflated Poisson (ZIP) or hurdle mixture model can be employed to account for the inflation of zero counts. Recently researchers have begun to consider models with a combination of excess zeros and longitudinal observation. However, the performance of such models has not been studied for the case of longitudinal data with time-dependent covariates. Presence of time-dependent covariates has been shown to affect the efficiency of hypothesis tests for longitudinal data.

In this paper, a comparison is made between a mixed ZIP model and a mixed zero-inflated Conway-Maxwell-Poisson (COM-Poisson) model for longitudinal count data with excess zeros and time-dependent covariates. Consideration is given to results of hypothesis tests, parameter estimates, and parameter interpretations. An example longitudinal data set is modeled using instances of recreational drug use over the past year. A count response that commonly shows excess zeros. Time-dependent covariates include the social environment and the level of drug craving during data collection. The drug under consideration has recently been legalized region of data collection and recreational usage is of great interest for public policy. Based on researcher pretations, the mixed ZIP model is selected for the analysis, but benefits of both models are discussed.

Research Areas:
- Statistics, economics
- Health or related public policy
- And behavioral sciences

Research Objectives:
- Explore modeling options for longitudinal counts with excess zeros. Analyze longitudinal count data with excess zeros.

Keywords: Statistics, Drug Abuse

Author's disclosure statement:
I have conducted analyses on similar data, included in my presentations covering drug use over time. I have lectured and written on longitudinal, non-continuous responses and the application of methods of analysis.

Relevant financial relationships? No

I have financial interests in the American Public Health Association Conflict of Interest and Commercial Support Guidelines, and do not believe that any off-label or experimental uses of a commercial product or service discussed in my presentation.

Back to: 3016.0: Modeling, Estimation and Implementation Science
12.1 Questions

It is common to ask what factors influence the mean of the distribution of the responses or as we know it fitting regression type models to the mean is common. In fact, one may ask similar questions about the variability surrounding the mean. Thus, one may be interested to know whether the same or additional factors also affect the variance. Hence the natural questions may be:

1. How do we account for those factors (if appropriate) which may affect the variance when we answer questions concerning the mean?
2. How do we assess factors that affect the variability about the mean or the fluctuations of a response of interest? In other words, how do we model both the mean and the variance simultaneously?

After reading this chapter, one should be able to answer these questions.

12.2 Motivation

The modelling of correlated binomial data can be accomplished through a mean model if the interest is only on the mean and the dispersion is considered as a nuisance parameter. However, if the intraclass correlation is of interest then there is a need to apply a joint modeling of the mean and the dispersion. Efron (1986) was one of the