Contents lists available at ScienceDirect

Computational Statistics and Data Analysis

journal homepage: www.elsevier.com/locate/csda

Unified computational methods for regression analysis of zero-inflated and bound-inflated data

Yan Yang[a,*], Douglas Simpson[b]

[a] Department of Mathematics and Statistics, Arizona State University, Wexler Hall, Tempe, AZ 85287, USA

[b] Department of Statistics, University of Illinois, Illini Hall, 725 South Wright Street, Champaign, IL 61820, USA

ARTICLE INFO

Article history:
Received 23 August 2008
Received in revised form 10 November 2009
Accepted 22 December 2009
Available online 4 January 2010

Keywords:
EM
Generalized estimating equation
Louis method
Mixture model
Quasi-Newton
Tobit model
Two-part model

ABSTRACT

Bounded data with excess observations at the boundary are common in many areas of application. Various individual cases of inflated mixture models have been studied in the literature for bound-inflated data, yet the computational methods have been developed separately for each type of model. In this article we use a common framework for computing these models, and expand the range of models for both discrete and semi-continuous data with point inflation at the lower boundary. The quasi-Newton and EM algorithms are adapted and compared for estimation of model parameters. The numerical Hessian and generalized Louis method are investigated as means for computing standard errors after optimization. Correlated data are included in this framework via generalized estimating equations. The estimation of parameters and effectiveness of standard errors are demonstrated through simulation and in the analysis of data from an ultrasound bioeffect study. The unified approach enables reliable computation for a wide class of inflated mixture models and comparison of competing models.

1. Introduction

Bound-inflated data are prevalent in a wide variety of disciplines, such as health and safety studies, economics, finance and insurance risk analysis. Typically the responses are bounded below by zero with a significant mass of zero observations, resulting in data that are either discrete with too many zeros for a standard discrete distribution, or semi-continuous with positive continuous values combined with a substantial portion of zeros. In our collaborative research on ultrasound safety (O'Brien et al., 2006), groups of laboratory rabbits were exposed to focused ultrasound in both lungs to investigate the risk of ultrasound-induced hemorrhage. Due to the designed low to moderate acoustic pressure levels, about 80% of the observations were free of lesions while 20% exhibited lesions. The goal was to evaluate the effect of acoustic pressure and other factors based on the clustered, zero-inflated lesion size data to develop insight into the safe pressure levels for diagnostic clinical ultrasound.

A two-part model handles zeros and positive values, discrete or continuous, separately through two model components: a binary model for the occurrence of an event, and a zero-truncated Poisson or a log-normal model for the strictly positive size of the event conditional on its occurrence (Welsh et al., 1996; Zhou and Tu, 1999). For correlated counts with extra zeros, the zero-truncated Poisson and negative binomial models were extended by adding random effects to each model component (Yau and Lee, 2001; Min and Agresti, 2005). Dobbie and Welsh (2001) constructed generalized estimating equations (GEEs) with working correlation matrices for both components of the zero-truncated Poisson model. For semi-continuous longitudinal or clustered data, two-part random effects models were considered by Olsen and Schafer (2001)

[*] Corresponding author. Tel.: +1 480 965 6475; fax: +1 480 965 8119.
E-mail addresses: yy@math.asu.edu (Y. Yang), dgs@uiuc.edu (D. Simpson).

0167-9473/$ – see front matter. Published by Elsevier B.V.
doi:10.1016/j.csda.2009.12.012

and Tooze et al. (2002). Albert and Shen (2005) proposed a two-part latent process model, which was recently adapted to incorporate random effects as well with Bayesian inference (Ghosh and Albert, 2009). For Bayesian two-part models with random effects, see also Zhang et al. (2006).

A zero-inflated (ZI) latent mixture model adds the point mass at zero to a discrete or censored distribution also capable of producing zeros. Two sub-models are involved: a binary model for the partially observed mixture-component indicators, and a Poisson regression as in the ZI Poisson model (Lambert, 1992) or a left-censored normal as in the ZI Tobit model (Cragg, 1971). A left-censored log-normal mixed with the point mass at a positive lower limit of detection was introduced by Moulton and Halsey (1995). In the presence of correlation, random effects were incorporated into either one or both sub-models (Hall, 2000; Berk and Lachenbruch, 2002; Yau et al., 2003; Lee et al., 2006). Alternatively, Moulton et al. (2002) implemented GEEs with the working independence correlation; Hall and Zhang (2004) developed a GEE approach for the class of ZI exponential family models.

Although much work has been done on fitting ZI data, most derivations have relied on special features of the individual models. We extend the existing latent mixture models through development of a unified framework, the left-inflated mixture (LIM) models for both discrete and semi-continuous data with point inflation at an arbitrary lower bound. This class not only includes current models but is broader by allowing various survival distributions (e.g., censored extreme value, logistic and t distributions) that add flexibility for modeling semi-continuous data. For correlated data, we construct GEEs with the working independence likelihood and estimate the covariance matrix of parameter estimates by the sandwich formula.

The quasi-Newton and EM algorithms are used for common estimation of model parameters. To find asymptotic standard errors associated with the EM, we investigate the generalized Louis method that extends the method of Louis (1982) to dependent data. For the quasi-Newton algorithm, a simulation study is conducted to assess the adequacy of estimating the outer Hessian matrix in the sandwich formula with the approximate Hessian at convergence. The performance and computational speed of the two methods are also compared empirically.

The rest of this article is organized as follows. Section 2 defines the left-inflated mixture models through a latent variable representation. Section 3 concerns maximum likelihood estimation for independent data and generalized estimating equation analysis for correlated responses. Section 4 discusses computational optimization of the estimating criteria by the Broyden–Fletcher–Goldfarb–Shanno (BFGS) quasi-Newton and EM algorithms. Standard error estimation associated with each algorithm is discussed. Section 5 presents a simulation study that assesses and compares the two computational methods. The practical utility of our unified approach is illustrated with an ultrasound-induced lung hemorrhage study in laboratory animals in Section 6. Concluding comments are given in Section 7.

2. Left-inflated mixture models

Let $\boldsymbol{Y} = (\boldsymbol{Y}_1^T, \ldots, \boldsymbol{Y}_n^T)^T$ denote the multivariate response vector, where $\boldsymbol{Y}_i = (Y_{i1}, \ldots, Y_{im_i})^T$ is the response vector for subject i, and Y_{ij} is the jth measure on subject i, $i = 1, \ldots, n, j = 1, \ldots, m_i$. The Y_{ij} are assumed to be bounded below (on the left) by L with a nonzero probability of observations equal to L. The lower boundary L is assumed to be known from the application under study or given by objective methods when a lower detection limit exists (Moulton and Halsey, 1995).

Let $\boldsymbol{y} = (\boldsymbol{y}_1^T, \ldots, \boldsymbol{y}_n^T)^T$ be a realization of $\boldsymbol{Y}$. We consider left-inflated mixture models, in which the marginal distributions of the responses can be expressed as mixtures of distributions F_{ij} on $[L, \infty)$ and point masses concentrated at L. Here F_{ij} may be discrete or semi-continuous (as in the case of a left-censored distribution). The marginal densities have the form

$$\pi_{ij} f(y_{ij}) + \left(1 - \pi_{ij}\right) \delta_L(y_{ij}) = \begin{cases} 1 - \pi_{ij} + \pi_{ij} F_{ij}(L), & \text{if } y_{ij} = L \\ \pi_{ij} f(y_{ij}), & \text{if } y_{ij} > L \end{cases} \tag{1}$$

where $0 < \pi_{ij} \leq 1$ denotes the mixing weight, $F_{ij}(L) = F(Y_{ij} = L)$, $f(\cdot)$ is the frequency or density function for F, and $\delta_L(u)$ equals one if $u = L$ and zero otherwise.

Such models have convenient latent variable representations. Define the mixture-component indicator vector $\boldsymbol{W} = (\boldsymbol{W}_1^T, \ldots, \boldsymbol{W}_n^T)^T$, where $\Pr(W_{ij} = 1) = \pi_{ij}$. Introduce a random vector $\boldsymbol{Z} = (\boldsymbol{Z}_1^T, \ldots, \boldsymbol{Z}_n^T)^T$ whose marginal distributions match F_{ij} on $[L, \infty)$, but whose distributions on $(-\infty, L)$ are chosen for computational convenience. Finally, assume that W_{ij} and Z_{ij} are pairwise independent. Then

$$Y_{ij} \overset{\mathrm{d}}{=} \left(1 - W_{ij}\right) L + W_{ij} \left\{ L \cdot I(Z_{ij} \leq L) + Z_{ij} \cdot I(Z_{ij} > L) \right\} \tag{2}$$

for $i = 1, \ldots, n, j = 1, \ldots, m_i$, where "d" denotes in distribution and $I(A)$ is the indicator function for event A. If $L = 0$, then Eq. (2) yields a simplified representation for zero-inflated responses: $Y_{ij} \overset{\mathrm{d}}{=} W_{ij} Z_{ij} \cdot I(Z_{ij} > 0)$. For example, a ZI binomial random variable simplifies further to $W_{ij} Z_{ij}$ with $Z_{ij} \sim B(n_{ij}, \mu_{ij})$, while a ZI Tobit random variable has $Z_{ij} \sim N(\mu_{ij}, \sigma^2)$.

Consider regression models where both π_{ij} and μ_{ij} may depend on covariates through

$$\begin{aligned} h_1(\pi_{ij}) &= \boldsymbol{g}_{ij}^T \boldsymbol{\gamma} \\ h_2(\mu_{ij}) &= \boldsymbol{x}_{ij}^T \boldsymbol{\beta}, \end{aligned} \tag{3}$$

where h_1 and h_2 are known link functions, $\boldsymbol{g}_{ij}$ and $\boldsymbol{x}_{ij}$ are covariate vectors that may contain common variables, and $\boldsymbol{\gamma}$ and $\boldsymbol{\beta}$ are parameter vectors to be estimated. For discrete data μ_{ij} is generally the mean of Z_{ij}. In the case where Z_{ij} is binomial, we take $\mu_{ij} = E(Z_{ij})/n_{ij}$, following the standard practice in binomial response regression. For semi-continuous data μ_{ij} is the mean or other location parameters for the latent distribution of Z_{ij}.

A LIM model factors into a two-part model if the F_{ij} are continuous so that the non-degenerate mixture component has zero mass at the boundary. In this case the two-component distribution is marginally separable, and the two sets of responses can be modeled separately due to factorization of the working likelihood. For LIM models in general, however, W_{ij} and Z_{ij} are only partially observable. The likelihood is more complicated and more general computational methods are required.

3. Working independence and marginal analysis

This section considers computation of maximum likelihood estimates and standard errors for independent responses, and extends the computational methods to correlated data using a working independence marginal approach. The resulting generalized estimating equation analysis builds on ideas of Moulton et al. (2002), Lu et al. (2004) and Hall and Zhang (2004). Let $\boldsymbol{\theta}$ denote the vector containing all regression and scale parameters. If the model includes a scale parameter, as is common for semi-continuous data, then $\boldsymbol{\theta} = (\boldsymbol{\gamma}^T, \boldsymbol{\beta}^T, \tau)^T$, where $\tau = \log(\sigma)$; otherwise $\boldsymbol{\theta} = (\boldsymbol{\gamma}^T, \boldsymbol{\beta}^T)^T$. Using (1), the working independence log-likelihood is:

$$l(\boldsymbol{\theta}; \boldsymbol{y}) = \sum_{i=1}^{n} \sum_{j=1}^{m_i} \left[v_{ij} \log\{\pi_{ij} f(y_{ij})\} + (1 - v_{ij}) \log\{1 - \pi_{ij} + \pi_{ij} F_{ij}(L)\} \right], \tag{4}$$

where $v_{ij} = I(y_{ij} > L)$, and $F_{ij}(L) = \Pr(Z_{ij} \leq L)$.

The parameter vector $\boldsymbol{\theta}$ can be estimated by maximizing the working log-likelihood (4). The solution is a root of the estimating equation $\boldsymbol{u}(\hat{\boldsymbol{\theta}}; \boldsymbol{y}) = \boldsymbol{0}$, where $\boldsymbol{u}(\boldsymbol{\theta}; \boldsymbol{y}) = (\partial/\partial\boldsymbol{\theta}) l(\boldsymbol{\theta}; \boldsymbol{y}) = \sum_{i=1}^{n} \sum_{j=1}^{m_i} \boldsymbol{u}(\boldsymbol{\theta}; y_{ij})$. Assuming that the marginal models (1) and (3) are correctly specified, the marginal model parameter vector $\boldsymbol{\theta}_N$ is defined by the theoretical estimating equation

$$\frac{1}{N} \sum_{i=1}^{n} \sum_{j=1}^{m_i} E\{\boldsymbol{u}(\boldsymbol{\theta}_N; Y_{ij})\} = \boldsymbol{0},$$

where $N = \sum_{i=1}^{n} m_i$, and the expectation is taken over the true distribution for the underlying data. Suppose the sequence $\{\boldsymbol{\theta}_N\}$ converges to $\boldsymbol{\theta}$, then the theory of generalized method of moments (Hansen, 1982; Lu et al., 2004) implies that $\hat{\boldsymbol{\theta}}$ is consistent in estimating $\boldsymbol{\theta}$, and that $n^{1/2}(\hat{\boldsymbol{\theta}} - \boldsymbol{\theta}) \xrightarrow{\text{d}} N(\boldsymbol{0}, \boldsymbol{\Sigma})$. Further, the asymptotic covariance matrix $\boldsymbol{\Sigma}$ can be consistently estimated by the empirical sandwich estimator

$$\hat{\boldsymbol{\Sigma}} = \boldsymbol{A}_n^{-1} \boldsymbol{B}_n (\boldsymbol{A}_n^{-1})^T, \tag{5}$$

where $\boldsymbol{B}_n = \sum_{i=1}^{n} \{\sum_{j=1}^{m_i} \boldsymbol{u}(\hat{\boldsymbol{\theta}}; y_{ij})\}\{\sum_{j=1}^{m_i} \boldsymbol{u}(\hat{\boldsymbol{\theta}}; y_{ij})\}^T$ is sandwiched by inverses of the negative Hessian matrix $\boldsymbol{A}_n = -\sum_{i=1}^{n} \sum_{j=1}^{m_i} (\partial/\partial\boldsymbol{\theta}^T) \boldsymbol{u}(\boldsymbol{\theta}; y_{ij})|_{\boldsymbol{\theta}=\hat{\boldsymbol{\theta}}}$. The sandwich estimator captures correlations among repeated measures nonparametrically through $\boldsymbol{B}_n$. The consistency of $\hat{\boldsymbol{\Sigma}}$ does not require a correct specification of the covariance structure in the model (Diggle et al., 1994).

4. Computational algorithms

Two general approaches are compared to computing $\hat{\boldsymbol{\theta}}$: a quasi-Newton algorithm and an EM algorithm. We describe the two methods with a discussion on finding the estimated covariance matrix for $\hat{\boldsymbol{\theta}}$ in each case based on the sandwich formula (5). The following distributions of Z are implemented: normal, logistic, extreme value (left-skewed and right-skewed), t, Poisson, negative binomial, and binomial. Additional log-transformed models with the log-normal, log-logistic and Weibull distributions are implicitly included as well. The R code is available at http://math.la.asu.edu/~yy/limm.html/.

4.1. BFGS quasi-Newton optimization

A direct approach to estimating $\boldsymbol{\theta}$ is the quasi-Newton (QN) methods (e.g. Thisted, 1988). At each iteration, the objective function $l(\boldsymbol{\theta}; \boldsymbol{y})$ and its gradient vector $\boldsymbol{u}(\boldsymbol{\theta}; \boldsymbol{y})$ are evaluated, while the Hessian matrix is approximated by an updating formula. One widely used approximation is the BFGS method. A binary regression on $v_{ij} = I(y_{ij} > L)$ can provide initial values for $\boldsymbol{\gamma}$. Starting values for $\boldsymbol{\beta}$ or $(\boldsymbol{\beta}, \tau)$ may be obtained from a generalized linear model or a survival model on $\boldsymbol{y}$.

It appears to be most convenient to obtain $\boldsymbol{A}_n$ in Eq. (5) by the QN approximate Hessian at convergence. There does not yet seem to be a universal agreement on this. Chambers (1977, p. 148), for example, stated that “the great advantage is that many of the optimization procedures discussed [including QN] produce an estimate of $H(\hat{\boldsymbol{\theta}})$ [the covariance matrix] automatically, even when no explicit calculation of second derivatives takes place". Thisted (1988, p. 209), on the other hand, pointed out that "the matrix which plays the role of the Hessian may not be an adequate approximation to the Hessian itself, even after convergence has been attained". To investigate the proper use of the approximate Hessian for LIM models, we conduct a simulation study that compares the approximate and analytical Hessian matrices at convergence of the BFGS quasi-Newton algorithm. The details are presented in Section 5, where we observe evidence showing the adequacy of the approximate Hessian.

4.2. EM algorithm optimization

Quasi-Newton methods seek simultaneously an estimate for $\boldsymbol{\gamma}$ and one for $\boldsymbol{\beta}$ or $(\boldsymbol{\beta}, \tau)$ based on a locally quadratic approximation to the working likelihood (4). A potentially numerically more stable alternative is to work with a complete-data working likelihood for $(\boldsymbol{Y}, \boldsymbol{W})$ through the EM algorithm (Dempster et al., 1977), so that estimation is accomplished by solving standard sub-problems. The EM was used by Lambert (1992) for estimating the ZI Poisson model. Our formulation extends her approach to the class of LIM models.

Let $\boldsymbol{w} = (\boldsymbol{w}_1^T, \ldots, \boldsymbol{w}_n^T)^T$ be a realization of $\boldsymbol{W}$. The complete-data working log-likelihood of a LIM model consists of two separable parts: $l(\boldsymbol{\theta}; \boldsymbol{y}, \boldsymbol{w}) = l(\boldsymbol{\gamma}; \boldsymbol{y}, \boldsymbol{w}) + l(\boldsymbol{\beta}, \tau; \boldsymbol{y}, \boldsymbol{w})$, where

$$l(\boldsymbol{\gamma}; \boldsymbol{y}, \boldsymbol{w}) = \sum_{i=1}^{n} \sum_{j=1}^{m_i} \{w_{ij} \log \pi_{ij} + (1 - w_{ij}) \log(1 - \pi_{ij})\},$$

the log-likelihood for a binary regression with responses $\boldsymbol{w}$, and

$$l(\boldsymbol{\beta}, \tau; \boldsymbol{y}, \boldsymbol{w}) = \sum_{i=1}^{n} \sum_{j=1}^{m_i} w_{ij}\{v_{ij} \log f(y_{ij}) + (1 - v_{ij}) \log F_{ij}(L)\},$$

the log-likelihood for a weighted generalized linear model or left-censored regression with weights $\boldsymbol{w}$. With distinct parameters, $l(\boldsymbol{\gamma}; \boldsymbol{y}, \boldsymbol{w})$ and $l(\boldsymbol{\beta}, \tau; \boldsymbol{y}, \boldsymbol{w})$ can be evaluated separately for a maximization of $l(\boldsymbol{\theta}; \boldsymbol{y}, \boldsymbol{w})$. In the absence of a scale parameter, τ is omitted in the above expressions.

Consider the kth EM iteration for a LIM model. Because $l(\boldsymbol{\theta}; \boldsymbol{y}, \boldsymbol{w})$ is linear in the (partially) missing data w_{ij}, the *E step* that finds $Q(\boldsymbol{\theta} \mid \boldsymbol{\theta}^{(k)}) = E\{l(\boldsymbol{\theta}; \boldsymbol{Y}, \boldsymbol{W}) \mid \boldsymbol{y}, \boldsymbol{\theta}^{(k)}\}$, the conditional expectation of the complete-data working log-likelihood given the observed data $\boldsymbol{y}$ and current estimate $\boldsymbol{\theta}^{(k)}$, reduces to computing $\boldsymbol{w}^{(k)} = E(\boldsymbol{W} \mid \boldsymbol{y}, \boldsymbol{\theta}^{(k)})$. By Bayes' theorem,

$$w_{ij}^{(k)} = v_{ij} + (1 - v_{ij}) \cdot \left. \frac{\pi_{ij} F_{ij}(L)}{1 - \pi_{ij} + \pi_{ij} F_{ij}(L)} \right|_{\boldsymbol{\theta}=\boldsymbol{\theta}^{(k)}}$$

for $i = 1, \ldots, n, j = 1, \ldots, m_i$. After the w_{ij} are replaced with the $w_{ij}^{(k)}$, the *M step* maximizes $l(\boldsymbol{\theta}; \boldsymbol{y}, \boldsymbol{w}^{(k)})$ over $\boldsymbol{\theta}$, oftentimes by calling standard packages for fitting generalized linear models or survival models. The iterations may start with $\boldsymbol{\theta}^{(0)}$ or $\boldsymbol{w}^{(0)}$. Lambert (1992) proposed starting values for $\boldsymbol{\gamma}$ and $\boldsymbol{\beta}$ in the ZI Poisson model. We introduce simple initial values for $\boldsymbol{w} : w_{ij}^{(0)} = 1$ if $y_{ij} > L$, and $w_{ij}^{(0)} = 0.5$ if $y_{ij} = L$. This method has led to reasonably fast convergence in our simulation and data examples.

4.3. Generalized Louis method for EM standard errors

When the EM algorithm is used in fitting a LIM model, we adapt the method of Louis (1982) to compute $\boldsymbol{A}_n$ and modify $\boldsymbol{B}_n$ such that the sandwich formula (5) depends purely on the complete data. The resulting generalized Louis method simplifies the otherwise tedious and error-prone calculations based on the observed data alone.

Assuming working independence, the observed information matrix $\boldsymbol{A}_n$ for a LIM model can be obtained through the complete data $(\boldsymbol{y}, \boldsymbol{w})$ as follows:

$$\boldsymbol{A}_n = \boldsymbol{I}(\hat{\boldsymbol{\theta}}; \boldsymbol{y}, \hat{\boldsymbol{w}}) - \operatorname{cov}\{\boldsymbol{u}(\boldsymbol{\theta}; \boldsymbol{Y}, \boldsymbol{W}) \mid \boldsymbol{y}, \boldsymbol{\theta}\}|_{\boldsymbol{\theta}=\hat{\boldsymbol{\theta}}}, \tag{6}$$

where $\hat{\boldsymbol{w}} = E(\boldsymbol{W} \mid \boldsymbol{y}, \hat{\boldsymbol{\theta}}), \boldsymbol{u}(\boldsymbol{\theta}; \boldsymbol{y}, \boldsymbol{w}) = (\partial/\partial\boldsymbol{\theta}) l(\boldsymbol{\theta}; \boldsymbol{y}, \boldsymbol{w})$, and $\boldsymbol{I}(\boldsymbol{\theta}; \boldsymbol{y}, \boldsymbol{w}) = -(\partial/\partial\boldsymbol{\theta}^T)\boldsymbol{u}(\boldsymbol{\theta}; \boldsymbol{y}, \boldsymbol{w})$, a two-block diagonal matrix due to the decomposition of the working likelihood $l(\boldsymbol{\theta}; \boldsymbol{y}, \boldsymbol{w})$. In Eq. (6), $\boldsymbol{I}(\hat{\boldsymbol{\theta}}; \boldsymbol{y}, \hat{\boldsymbol{w}})$ is often referred to as the complete information while the conditional covariance matrix as the missing information that accounts for the uncertainty due to missing data. For LIM models, the missing information has an exact form:

$$\operatorname{cov}\{\boldsymbol{u}(\boldsymbol{\theta}; \boldsymbol{Y}, \boldsymbol{W}) \mid \boldsymbol{y}, \boldsymbol{\theta}\} = \sum_{i=1}^{n} \sum_{j=1}^{m_i} \operatorname{var}\left(W_{ij} \mid \boldsymbol{y}, \boldsymbol{\theta}\right) \begin{bmatrix} \zeta_{ij}^2 \boldsymbol{g}_{ij} \boldsymbol{g}_{ij}^T & \zeta_{ij} \psi_{ij} \boldsymbol{g}_{ij} \boldsymbol{x}_{ij}^T & \zeta_{ij} \varphi_{ij} \boldsymbol{g}_{ij} \\ \zeta_{ij} \psi_{ij} \boldsymbol{x}_{ij} \boldsymbol{g}_{ij}^T & \psi_{ij}^2 \boldsymbol{x}_{ij} \boldsymbol{x}_{ij}^T & \psi_{ij} \varphi_{ij} \boldsymbol{x}_{ij} \\ \zeta_{ij} \varphi_{ij} \boldsymbol{g}_{ij}^T & \psi_{ij} \varphi_{ij} \boldsymbol{x}_{ij}^T & \varphi_{ij}^2 \end{bmatrix}.$$

The terms ζ_{ij}, ψ_{ij} and φ_{ij} do not depend on w_{ij} and are given by

Table 1
Numbers of simulated trials (out of 2000 each) in which the observed information is not positive definite. QN1 is the quasi-Newton method starting from one EM step.

True model	$\Pr(Y=0)$	$\Pr(Z \le 0)$	QN	QN1	EM
Logit/Censored logistic	0.21–0.60	0.01–0.40	0	0	0
Logit/Censored logistic	0.51–0.90	0.10–0.60	9	7	1
Logit/Poisson	0.21–0.60	0.01–0.41	0	0	0
Logit/Poisson	0.50–0.90	0.10–0.59	0	0	0
Censored logistic	0.12–0.73	0.12–0.73	158	138	1
Poisson	0.11–0.72	0.11–0.72	246	255	1

- $\zeta_{ij} = 1/\{\pi_{ij}(1-\pi_{ij})\} \cdot (\partial/\partial\eta_{ij})\, h_1^{-1}(\eta_{ij})$,
- $\psi_{ij} = \{v_{ij} \cdot (\partial/\partial\mu_{ij}) \log f(y_{ij}) + (1-v_{ij}) \cdot (\partial/\partial\mu_{ij}) \log F_{ij}(L)\} \cdot (\partial/\partial\delta_{ij})\, h_2^{-1}(\delta_{ij})$,
- if the data are semi-continuous, then, additionally, $\varphi_{ij} = \{v_{ij} \cdot (\partial/\partial\sigma) \log f(y_{ij}) + (1-v_{ij}) \cdot (\partial/\partial\sigma) \log F_{ij}(L)\} \cdot \sigma$,

where $\eta_{ij} = \boldsymbol{g}_{ij}^T\boldsymbol{\gamma}$ and $\delta_{ij} = \boldsymbol{x}_{ij}^T\boldsymbol{\beta}$. Given $\boldsymbol{\theta} = \hat{\boldsymbol{\theta}}$, $\text{var}(W_{ij} \mid \boldsymbol{y}, \hat{\boldsymbol{\theta}}) = \hat{w}_{ij}(1-\hat{w}_{ij})$; the ζ_{ij}, ψ_{ij} and φ_{ij} are estimated by plugging in $\hat{\boldsymbol{\theta}}$. Derivations of the missing information for LIM models are provided in the Appendix.

For a mixture of two normals using the working independence likelihood, Lu et al. (2004) showed that the matrix $\boldsymbol{B}_n$ can also be found based on the complete data. This result nicely extends to the class of LIM models:

$$\boldsymbol{B}_n = \sum_{i=1}^{n} \left\{ \sum_{j=1}^{m_i} \boldsymbol{u}(\hat{\boldsymbol{\theta}}; y_{ij}, \hat{w}_{ij}) \right\} \left\{ \sum_{j=1}^{m_i} \boldsymbol{u}(\hat{\boldsymbol{\theta}}; y_{ij}, \hat{w}_{ij}) \right\}^T,$$

because $\boldsymbol{u}(\boldsymbol{\theta}; \boldsymbol{y}) = E\{\boldsymbol{u}(\boldsymbol{\theta}; \boldsymbol{Y}, \boldsymbol{W}) \mid \boldsymbol{y}, \boldsymbol{\theta}\}$ from Louis (1982) and $\boldsymbol{u}(\boldsymbol{\theta}; \boldsymbol{y}, \boldsymbol{w})$ is linear in the w_{ij}.

5. Simulation study

A simulation study is done to assess the adequacy of estimating $\boldsymbol{A}_n$ with the approximate Hessian at convergence of the BFGS quasi-Newton method, and to compare the QN and EM algorithms for estimating LIM models with independent responses. Key issues of comparison are convergence of the algorithm, post-optimization computation of the estimated covariance matrix, and computational speed. To examine the robustness against possible model misspecification, we also generate data from a purely censored or standard count model. To the best of our knowledge, no previous work on bound-inflated data has addressed these important computational issues. For a simulation study on the performance of the GEE approach assuming working independence for correlated data, see Lu et al. (2004).

Two ZI mixture models and their reduced one-component versions are evaluated: a logit/censored logistic model, a logit/Poisson model, a censored logistic model, and a Poisson model. Each model included a predictor x taking on $N = 100$ values uniformly spaced over the interval [0, 1]. Parameters of the ZI models were chosen so a simulated set contained about 20%–60% or 50%–90% zeros with 1%–40% or 10%–60% of the zeros being censored logistic or Poisson. Parameters of the censored or count model were selected to produce around 10%–70% zeros. The six simulations above were carried out in R with 2000 runs each.

Each set of data was fitted with a LIM model, estimated by both the QN and EM algorithms. Standard errors were obtained from the approximate and analytical Hessians in the QN method, and from Eq. (6) in the EM. Besides obtaining initial values as in Section 4, we also considered the QN starting from one EM step, referred to as QN1 in the sequel. All quasi-Newton and EM runs converged at the relative convergence tolerance of 1e−8. In a few runs the observed information was not positive definite. From Table 1 the problematic runs mostly occurred when the data were generated under the reduced model. The quasi-Newton approximate and exact observed information matrices, however, are always both positive definite or not.

Table 2 summarizes simulation behavior of parameter estimates in the logistic case (the intercepts behave similarly and not shown). When the true model is a LIM model, the median $\hat{\sigma}$ of the censored model is inflated due to excessive zeros. If the censored logistic is the true model, the median $\hat{\beta}_1$ and $\hat{\sigma}$ of the LIM model remain well-behaved while $\hat{\gamma}_1$ is large in the exponential scale, suggesting that a reduced censored model may suffice. In all simulations the quasi-Newton and EM parameter estimates are identical or similar. When the data are generated under a LIM model and contain a high amount of zeros, the QN1 and EM perform slightly better than the QN. The standard errors in the logistic case are given in Table 3, including only runs with a positive definite observed information. The standard errors obtained from the approximate and analytical Hessians in the quasi-Newton methods only differ at or after the third decimal place for each simulated set, likely due to floating-point errors. Regardless of the underlying model, the Monte Carlo and mean standard errors for $\hat{\beta}_1$ and $\hat{\sigma}$ closely agree with each other. When the true model is a LIM model, the two sets of standard errors for $\hat{\gamma}_1$ are mildly or greatly different, depending on the amount of zeros in the data. If the true model is the censored model, the two sets of standard errors for $\hat{\gamma}_1$ are large (as expected) and consistent in the quasi-Newton methods.

As parameter estimates of logistic regression on observable binary data can have infinite bias in finite samples (Rindskopf, 2002), $\boldsymbol{\gamma} = (\gamma_0, \gamma_1)^T$ for the partially observed component indicators is even more difficult to estimate with accurateness

Table 2
Median parameter estimates when the true model is the logit/censored logistic or purely censored logistic. QN1 is the quasi-Newton method with one EM step.

$\Pr(Y = 0)$	True value	Logit/Censored logistic			Censored logistic
		QN	QN1	EM	
0.21–0.60[a]	$\beta_1 = 2.0$	1.999	1.999	1.999	2.094
	$\sigma = 0.5$	0.483	0.483	0.483	0.732
	$\gamma_1 = 0.7$	0.694	0.694	0.692	–
0.51–0.90[a]	$\beta_1 = 1.3$	1.195	1.241	1.253	1.812
	$\sigma = 0.5$	0.458	0.458	0.459	0.758
	$\gamma_1 = 1.3$	1.463	1.376	1.343	–
0.12–0.73[b]	$\beta_1 = 1.5$	1.301	1.306	1.300	1.502
	$\sigma = 0.5$	0.456	0.455	0.455	0.488
	$\gamma_1 = \infty$	3.483	3.342	3.006	–

[a] The true model is the logit/censored logistic.
[b] The true model is the censored logistic.

Table 3
Monte Carlo (mean) standard errors based on runs with a positive definite observed information when the true model is the logit/censored logistic or purely censored logistic. QN1 is the quasi-Newton method with one EM step.

$\Pr(Y = 0)$	Parameter	Logit/Censored logistic			Censored logistic
		QN	QN1	EM	
0.21–0.60[a]	β_1	0.452 (0.432)	0.452 (0.432)	0.452 (0.433)	0.463 (0.463)
	$\log(\sigma)$	0.132 (0.131)	0.132 (0.131)	0.132 (0.131)	0.101 (0.105)
	γ_1	2.225 (1.503)	2.223 (1.500)	2.255 (1.502)	–
		1.334[c] (1.349)	1.334 (1.348)	1.333 (1.353)	–
0.51–0.90[a]	β_1	0.958 (0.801)	0.978 (0.794)	0.988 (0.788)	0.679 (0.681)
	$\log(\sigma)$	0.251 (0.230)	0.245 (0.230)	0.247 (0.230)	0.135 (0.171)
	γ_1	28.509 (8.164)	31.103 (8.127)	358.219 (3329.233)	–
		1.881[d] (1.802)	1.883 (1.840)	1.898 (1.830)	–
0.12–0.73[b]	β_1	0.467 (0.446)	0.470 (0.444)	0.463 (0.438)	0.330 (0.328)
	$\log(\sigma)$	0.137 (0.139)	0.137 (0.138)	0.137 (0.137)	0.110 (0.110)
	γ_1	52.058 (55.066)	50.881 (52.192)	815.917 (16353.057)	–

[a] The true model is the logit/censored logistic.
[b] The true model is the censored logistic.
[c] Excluding flagged runs with $c = 0.995$ (41, 41 and 40 runs flagged for QN, QN1 and EM, respectively).
[d] Excluding flagged runs with $c = 0.995$ (132, 122 and 128 runs flagged for QN, QN1 and EM, respectively).

Table 4
Median parameter estimates when the true model is the logit/Poisson or ordinary Poisson. QN1 is the quasi-Newton method with one EM step.

$\Pr(Y = 0)$	True value	Logit/Poisson			Poisson
		QN	QN1	EM	
0.21–0.60[a]	$\beta_1 = 1.6$	1.612	1.612	1.612	1.776
	$\gamma_1 = 0.7$	0.664	0.664	0.663	–
0.50–0.90[a]	$\beta_1 = 1.5$	1.531	1.531	1.533	2.298
	$\gamma_1 = 1.3$	1.296	1.294	1.292	–
0.11–0.72[b]	$\beta_1 = 1.9$	1.798	1.802	1.794	1.902
	$\gamma_1 = \infty$	4.026	4.277	3.509	–

[a] The true model is logit/Poisson.
[b] The true model is Poisson.

or precision. Similar simulation properties for $\boldsymbol{\gamma}$ were noted by Lambert (1992) for a logit/Poisson model. Large $|\hat{\gamma}_0|$ or $|\hat{\gamma}_1|$ cause some or all of the $\hat{\pi}_k$, $k = 1, \ldots, N$, to be nearly 0 or 1, making computation of the observed information unstable or, when the $\hat{\pi}_k$ are on the boundary, impossible. Suppose that a run is flagged if $\max(\hat{\pi}_k, \text{all } k) > c$ or if $\min(\hat{\pi}_k, \text{all } k) < 1 - c$, where $0 < c < 1$ is a constant close to 1. Table 3 provides the standard errors for $\hat{\gamma}_1$ after excluding the flagged runs with $c = 0.995$ when the true model is the logit/censored logistic. These simulation and mean standard errors have smaller sizes and are in a much better agreement. Parameter estimates and standard errors for $\hat{\beta}_1$ and $\hat{\sigma}$ are at most slightly affected without the flagged runs (not shown). The logit/Poisson and Poisson models show similar simulation properties as the logistic case; see Tables 4 and 5. In practice, close estimates from the quasi-Newton and EM methods provide evidence of well-behaved estimation. If most $\hat{\pi}_k$ are close to 1, a LIM model may be reduced to a censored or count model.

Table 5
Monte Carlo (mean) standard errors based on runs with a positive definite observed information when the true model is the logit/Poisson or ordinary Poisson. QN1 is the quasi-Newton method with one EM step.

$\Pr(Y=0)$	Parameter	Logit/Poisson			Poisson
		QN	QN1	EM	
0.21–0.60[a]	β_1	0.355 (0.356)	0.355 (0.356)	0.355 (0.356)	0.357 (0.287)
	γ_1	2.214 (1.464)	2.200 (1.463)	2.642 (1.469)	–
		1.308[c] (1.339)	1.308 (1.339)	1.306 (1.338)	–
0.50–0.90[a]	β_1	0.987 (0.859)	0.987 (0.859)	0.986 (0.859)	0.707 (0.552)
	γ_1	7.844 (1.972)	9.058 (2.004)	93.353 (209.370)	–
		1.750[d] (1.585)	1.750 (1.585)	1.743 (1.583)	–
0.11–0.72[b]	β_1	0.438 (0.460)	0.439 (0.459)	0.434 (0.451)	0.377 (0.377)
	γ_1	86.469 (90.554)	89.838 (103.926)	1026.066 (26464.889)	–

[a] The true model is the logit/Poisson.
[b] The true model is the Poisson model.
[c] Excluding flagged runs with $c = 0.995$ (37, 37 and 37 runs flagged for QN, QN1 and EM, respectively).
[d] Excluding flagged runs with $c = 0.995$ (33, 33 and 34 runs flagged for QN, QN1 and EM, respectively).

Table 6
Computing times in seconds. QN1 is the quasi-Newton method with one EM step.

True model	$\Pr(Y=0)$	Time/Run			Iter./Run			Time/Iter.		
		QN	QN1	EM	QN	QN1	EM	QN	QN1	EM
Logit/Censored logistic	0.21–0.60	0.15	0.14	0.69	13	11	35	0.0119	0.0129	0.0196
Logit/Censored logistic	0.51–0.90	0.32	0.28	3.14	41	35	163	0.0078	0.0081	0.0192
Logit/Poisson	0.21–0.60	0.12	0.12	0.59	10	10	25	0.0113	0.0122	0.0237
Logit/Poisson	0.50–0.90	0.13	0.12	1.54	14	12	66	0.0096	0.0100	0.0235
Censored logistic	0.12–0.73	0.60	0.59	4.15	84	71	196	0.0071	0.0083	0.0211
Poisson	0.11–0.72	3.92	4.29	3.98	775	827	154	0.0051	0.0052	0.0258

Table 6 compares the average CPU time and number of iterations per run required by fitting a LIM model. All computations were done on a desktop PC with Pentium dual-core 3.4 GHz processor and 3.25 GB RAM. The quasi-Newton methods generally demand much less computing time and fewer iterations to converge than the EM. When the data are generated from a LIM model, more computing time and iterations are needed as the proportion of zero responses increases. If the true model is a reduced one-component model, a LIM model approaches the boundary of the parameter space for the π_k and is more difficult to estimate, resulting in substantially more CPU time and iterations. The mean time per iteration is also shown in the table. It takes the EM longer to complete one iteration, so more iterations and more time per iteration both account for more time per run by the EM.

6. Application to ultrasound bioeffect study

We illustrate the computational methods with the clustered zero-bounded semi-continuous data from an ultrasound-induced lung hemorrhage study in laboratory rabbits (O'Brien et al., 2006). Ten or nine animals were randomly assigned to each of nine exposure groups and both lungs of each animal were exposed to a focused ultrasound beam ($n = 89$, $m_i = 2$). The response was depth of an ultrasound-induced lesion (in mm). The predictors were acoustic pressure (AP: 1.52–8.40 MPa) and exposure order (EO: 1 or 2). Marginally 80.3% of responses were zero (non-lesion), while most lesions had medium to large sizes. The data are available at http://math.la.asu.edu/~yy/limm.html.

The data were fitted by censored models, two-part models and LIM models, with potential correlations handled by generalized estimating equations using the working independence likelihood. For a LIM model, the logit and probit links for W and the normal, logistic, extreme value and t distributions for Z gave similar fits. Standard errors in each model were adjusted by the sandwich estimator to account for the within-rabbit correlation. We obtained the Hessian matrix in the sandwich formula for a LIM model by the approximate Hessian in the quasi-Newton method and by Louis method in the EM. These two approaches led to nearly identical parameter and standard error estimates, indicating well-behaved estimation.

Each type of models started with a full trend model, and the AIC was used for model selection. Table 7 reports the final models in the logistic case: the censored logistic, logit/log-logistic (a two-part model), and logit/censored logistic (a LIM model). The three models all show positive effects of acoustic pressure on the probability of lesion incidence and lesion size. The interaction between AP and EO in the two-part and LIM models implies that the positive ultrasound effect on lesion depth was stronger for the first exposed lungs than the second. Based on the AIC values, the LIM model gives the best fit. Consistent with the simulation, the LIM model has a much smaller estimate of the scale parameter than the censored model: $\log(\sigma) = -1.05$ compared with 0.30. The censored model thus requires a large scale parameter to achieve a big point mass at zero *and* to allow "extreme" positive values.

Table 7
Parameter estimates (Est.), standard errors (SE) and goodness of fit for the ultrasound data from the censored logistic, logit/log-logistic and logit/censored logistic models.

	Censored logistic			Logit/Log-logistic			Logit/Censored logistic		
	Est.	SE	p-value	Est.	SE	p-value	Est.	SE	p-value
Intercept (γ_0)		–		−5.266	0.800	<0.001	−5.119	0.891	<0.001
AP (γ_1)		–		0.679	0.119	<0.001	0.660	0.131	<0.001
Intercept (β_0)	−7.412	1.104	<0.001	−2.276	1.998	0.255	−2.804	2.029	0.167
AP (β_1)	1.003	0.139	<0.001	0.416	0.258	0.107	0.698	0.272	0.010
EO (β_2)		–		1.404	1.120	0.210	2.254	1.457	0.122
AP × EO(β_3)		–		−0.197	0.146	0.178	−0.326	0.197	0.098
log(σ)	0.300	0.106	0.005	−1.561	0.236	<0.001	−1.051	0.130	<0.001
Log-likelihood		−117.9			−102.4			−94.7	
AIC		241.9			218.8			203.4	
AIC w/o depth = 0.1		234.4			195.1			194.3	

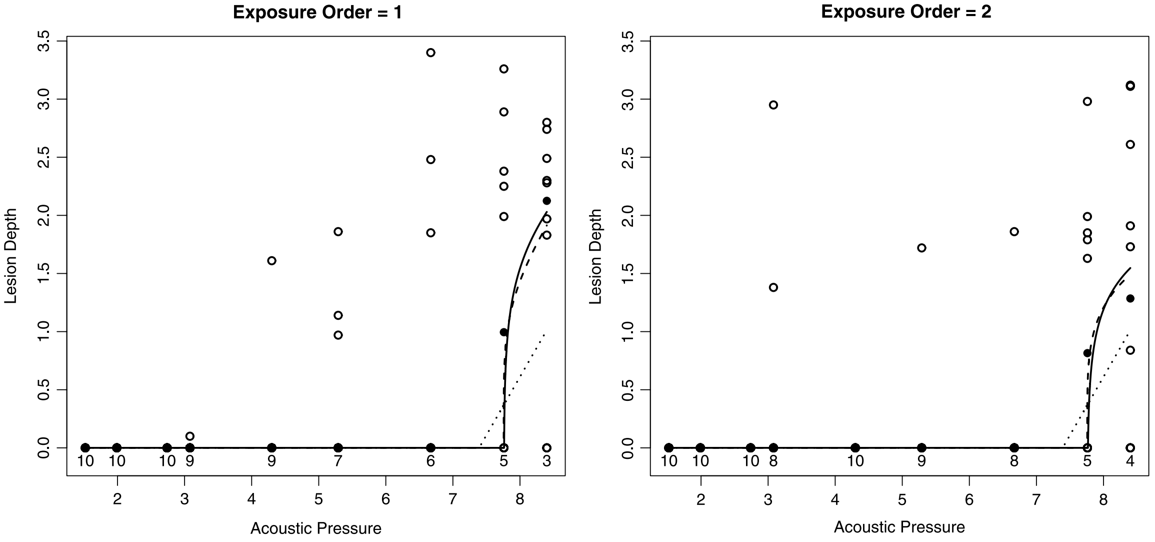

Fig. 1. Scatter plots of lesion depth versus acoustic pressure, superimposed by observed and fitted medians. Empty circles denote positive depth. Number symbols (located slightly below the $y = 0$ line) indicate how often a non-lesion occurs given exposure. Solid circles are observed median depth. Dotted median curves are estimated from the censored logistic model, dashed curves from the logit/log-logistic, and solid curves from the logit/censored logistic.

The two-part model does not fit as well as the LIM model. This is likely due to non-separability caused by the smallest positive lesion size 0.1 mm at AP = 3.08 MPa and EO = 1 (see Fig. 1). The AIC values of the models without that point are provided in the last row of Table 7. The decrease in AIC is much more dramatic for the two-part model than for the other two. It appears that adequacy of a two-part model depends heavily on separability between values at the boundary and those above, to which censored models and LIM models are more robust.

Visual summaries of the models are given in Fig. 1. The fitted median curves do not rise above zero until the exposure level is greater than 7 MPa. Beyond that the curves increase quickly as acoustic pressure goes up. The censored logistic model seems to have an attenuated ultrasound effect due to substantial amounts of zeros near the high end of the exposure range. The observed and estimated probabilities of zero are compared in Fig. 2. The probability of zero is 1 minus the risk of a lesion at a given exposure and is itself an important quantity to estimate. The fitted probability curves from the three models show similar decreasing trends as acoustic pressure increases, and they match the observed portions fairly well. The adequate fits in zero are not surprising as zeros dominate this data set.

7. Conclusions

We have developed the general class of left-inflated mixture models that unifies and expands on a number of existing latent mixture models for bound-inflated data. Parameter and standard error estimation methodology that uses the quasi-Newton and EM algorithms based on a GEE approach with the working independence likelihood has been implemented in the R language, so researchers can fit a wide range of bound-inflated mixture models with ease. We propose the generalized Louis method for computing EM standard errors, for which we have derived a simple explicit form of the missing information matrix. For the quasi-Newton method, our empirical results suggest that it is satisfactory to use the approximate Hessian at convergence to estimate the outer Hessian in the sandwich formula, which sheds some light on disagreements in the

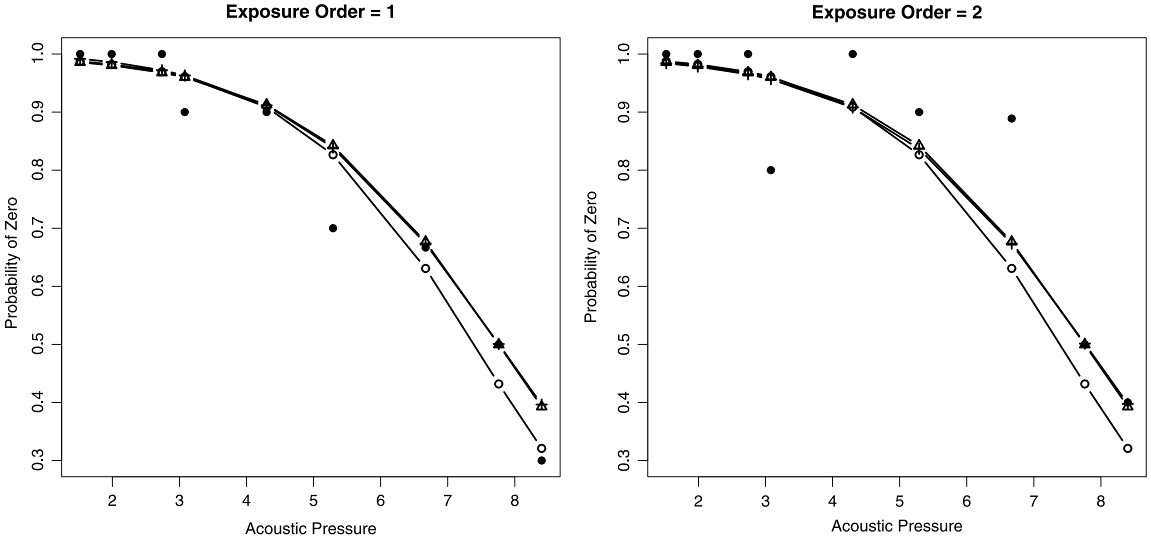

Fig. 2. Plots of observed and fitted probabilities of zero. Solid circles are observed proportions. Lines connecting empty circles are estimated from the censored logistic, lines connecting empty triangles from the logit/log-logistic, and lines connecting plus signs from the logit/censored logistic.

literature on this issue. The proposed unified framework can be readily extended to upper-bounded or double-bounded data with point inflation at the boundary value(s).

Simulation finds the two computational methods easily implemented, generally performing well, and collectively comparable. There is evidence showing that the EM algorithm is more numerically stable while the quasi-Newton converges faster with fewer iterations and less time per iteration. The EM may be preferred if the data are almost separable between values at the boundary and those exceeding it, or if each component of a LIM model has many parameters requiring estimation. When the data are barely separable and contain a high portion of values at the boundary, the quasi-Newton can save a great amount of time. A further issue, which does not appear to have been addressed previously, is the efficacy of bound-inflated modeling when a reduced one-component model suffices. Our simulation results indicate that, although the observed information may be non-positive definite due to the inflation probabilities approaching one, estimation of regression parameters remain well-behaved.

Acknowledgements

This research was supported in part by NIH grant EB02641. The authors thank the associate editor and two referees for helpful comments and Dr. William O'Brien, Jr. for permission to use the ultrasound exposure data.

Appendix. Missing information of left-inflated mixture models

In the following we obtain the form of the missing information for semi-continuous data. For discrete data, omit the expressions involving $\tau = \log \sigma$. For a shorthand notation write the complete-data score statistic $\boldsymbol{u}(\boldsymbol{\theta}; \boldsymbol{y}, \boldsymbol{w}) = (\boldsymbol{u}_{\boldsymbol{\gamma}}^{c\,T}, \boldsymbol{u}_{\boldsymbol{\beta}}^{c\,T}, u_{\tau}^{c})^{T}$, where $\boldsymbol{u}_{\boldsymbol{\gamma}}^{c} = (\partial/\partial\boldsymbol{\gamma})l(\boldsymbol{\gamma}; \boldsymbol{y}, \boldsymbol{w})$, $\boldsymbol{u}_{\boldsymbol{\beta}}^{c} = (\partial/\partial\boldsymbol{\beta})l(\boldsymbol{\beta}, \tau; \boldsymbol{y}, \boldsymbol{w})$, and $u_{\tau}^{c} = (\partial/\partial\tau)l(\boldsymbol{\beta}, \tau; \boldsymbol{y}, \boldsymbol{w})$; the superscript "$c$" stands for complete data. Assuming working independence, we have

$$\boldsymbol{u}_{\boldsymbol{\gamma}}^{c} = \sum_{i}\sum_{j} \frac{w_{ij} - \pi_{ij}}{\pi_{ij}(1 - \pi_{ij})} \cdot \frac{\partial h_1^{-1}(\eta_{ij})}{\partial \eta_{ij}} \cdot \boldsymbol{g}_{ij} = \sum_{i}\sum_{j} (w_{ij} - \pi_{ij})\zeta_{ij}\boldsymbol{g}_{ij}$$

$$\boldsymbol{u}_{\boldsymbol{\beta}}^{c} = \sum_{i}\sum_{j} w_{ij} \left\{ v_{ij} \cdot \frac{\partial \log f(y_{ij})}{\partial \mu_{ij}} + (1 - v_{ij}) \cdot \frac{\partial \log F_{ij}(L)}{\partial \mu_{ij}} \right\} \cdot \frac{\partial h_2^{-1}(\delta_{ij})}{\partial \delta_{ij}} \cdot \boldsymbol{x}_{ij} = \sum_{i}\sum_{j} w_{ij}\psi_{ij}\boldsymbol{x}_{ij}$$

$$u_{\tau}^{c} = \sum_{i}\sum_{j} w_{ij} \left\{ v_{ij} \cdot \frac{\partial \log f(y_{ij})}{\partial \sigma} + (1 - v_{ij}) \cdot \frac{\partial \log F_{ij}(L)}{\partial \sigma} \right\} \sigma = \sum_{i}\sum_{j} w_{ij}\varphi_{ij},$$

where $\eta_{ij} = \boldsymbol{g}_{ij}^{T}\boldsymbol{\gamma}$ and $\delta_{ij} = \boldsymbol{x}_{ij}^{T}\boldsymbol{\beta}$. Note that the ζ_{ij}, ψ_{ij} and φ_{ij} do not involve the missing data w_{ij}. The missing information is thus in the form of

$$\operatorname{cov}\{\boldsymbol{u}(\boldsymbol{\theta}; \boldsymbol{Y}, \boldsymbol{W}) \mid \boldsymbol{y}, \boldsymbol{\theta}\} = \begin{bmatrix} M_1 & M_4 & M_5 \\ M_4^{T} & M_2 & M_6 \\ M_5^{T} & M_6^{T} & M_3 \end{bmatrix},$$

where, after using the single subscript k, $k = 1, \ldots, N$, to replace the double subscript ij, $i = 1, \ldots, n, j = 1, \ldots, m_i$,

$$M_1 = \sum_k \text{var}\,(W_k \mid \boldsymbol{y}, \boldsymbol{\theta})\, \zeta_k^2 \boldsymbol{g}_k \boldsymbol{g}_k^T \qquad M_2 = \sum_k \text{var}\,(W_k \mid \boldsymbol{y}, \boldsymbol{\theta})\, \psi_k^2 \boldsymbol{x}_k \boldsymbol{x}_k^T$$

$$M_3 = \sum_k \text{var}\,(W_k \mid \boldsymbol{y}, \boldsymbol{\theta})\, \varphi_k^2 \qquad M_4 = \sum_k \text{var}\,(W_k \mid \boldsymbol{y}, \boldsymbol{\theta})\, \zeta_k \psi_k \boldsymbol{g}_k \boldsymbol{x}_k^T$$

$$M_5 = \sum_k \text{var}\,(W_k \mid \boldsymbol{y}, \boldsymbol{\theta})\, \zeta_k \varphi_k \boldsymbol{g}_k \qquad M_6 = \sum_k \text{var}\,(W_k \mid \boldsymbol{y}, \boldsymbol{\theta})\, \psi_k \varphi_k \boldsymbol{x}_k,$$

with $\text{var}(W_k \mid \boldsymbol{y}, \hat{\boldsymbol{\theta}}) = \hat{w}_k(1 - \hat{w}_k)$ and $\hat{w}_k = E(W_k \mid \boldsymbol{y}, \hat{\boldsymbol{\theta}})$.

References

Albert, P.S., Shen, J., 2005. Modelling longitudinal semicontinuous emesis volume data with serial correlation in an acupuncture clinical trial. Applied Statistics 54, 707–720.

Berk, K.N., Lachenbruch, P.A., 2002. Repeated measures with zeros. Statistical Methods in Medical Research 11, 303–316.

Chambers, J.M., 1977. Computational Methods for Data Analysis. Wiley, New York.

Cragg, J.H., 1971. Some statistical models for limited dependent variables with application to the demand for durable goods. Econometrica 39, 829–844.

Dempster, A.P., Laird, N.M., Rubin, D.B., 1977. Maximum likelihood from incomplete data via the EM algorithm (with discussion). Journal of the Royal Statistical Society. Series B 39, 1–38.

Diggle, P.J., Liang, K.Y., Zeger, S.L., 1994. Analysis of Longitudinal Data. Oxford University Press, Oxford, UK.

Dobbie, M.J., Welsh, A.H., 2001. Modelling correlated zero-inflated count data. Australian and New Zealand Journal of Statistics 43, 431–444.

Ghosh, P., Albert, P.S., 2009. A Bayesian analysis for longitudinal semicontinuous data with an application to an acupuncture clinical trial. Computational Statistics and Data Analysis 53, 699–706.

Hall, D.B., 2000. Zero-inflated Poisson and binomial regression with random effects: A case study. Biometrics 56, 1030–1039.

Hall, D.B., Zhang, Z., 2004. Marginal models for zero inflated clustered data. Statistical Modelling 4, 161–180.

Hansen, L., 1982. Large sample properties of generalized method of moments estimators. Econometrica 50, 1029–1054.

Lambert, D., 1992. Zero-inflated Poisson regression, with an application to defects in manufacturing. Technometrics 34, 1–14.

Lee, A.H., Wang, K., Scott, J.A., Yau, K.K.W., McLachlan, G.J., 2006. Multi-level zero-inflated Poisson regression modelling of correlated count data with excess zeros. Statistical Methods in Medical Research 15, 47–61.

Louis, T.A., 1982. Finding the observed information when using the EM algorithm. Journal of the Royal Statistical Society. Series B 44, 226–233.

Lu, S.E., Lin, Y., Shih, W.J., 2004. Analyzing excessive no changes in clinical trials with clustered data. Biometrics 60, 257–267.

Min, Y., Agresti, A., 2005. Random effect models for repeated measures of zero-inflated count data. Statistical Modelling 5, 1–19.

Moulton, L.H., Curriero, F.C., Barroso, P.F., 2002. Mixture models for quantitative HIV RNA data. Statistical Methods in Medical Research 11, 317–325.

Moulton, L.H., Halsey, N.A., 1995. A mixture model with detection limits for regression analyses of antibody response to vaccine. Biometrics 51, 1570–1578.

O'Brien, W.D., Yang, Y., Simpson, D.G., Frizzell, L.A., Miller, R.J., Blue, J.P., Zachary, J.F., 2006. Threshold estimation of ultrasound-induced lung hemorrhage in adult rabbits and comparison of thresholds in mice, rats, rabbits and pigs. Ultrasound in Medicine and Biology 32, 1793–1804.

Olsen, M.K., Schafer, J.L., 2001. A two-part random-effects model for semicontinuous longitudinal data. Journal of the American Statistical Association 96, 730–745.

Rindskopf, D., 2002. Infinite parameter estimates in logistic regression: Opportunities, not problems. Journal of Educational and Behavioral Statistics 27, 147–161.

Thisted, R.A., 1988. Elements of Statistical Computing: Numerical Computation. Chapman and Hall/CRC, Boca Raton, FL.

Tooze, J.A., Grunwald, G.K., Jones, R.H., 2002. Analysis of repeated measures data with clumping at zero. Statistical Methods in Medical Research 11, 341–355.

Welsh, A.H., Cunningham, R.B., Donnelly, C.F., Lindenmayer, D.B., 1996. Modelling the abundance of rare species: Statistical models for counts with extra zeros. Ecological Modelling 88, 297–308.

Yau, K.K.W., Lee, A.H., 2001. Zero-inflated Poisson regression with random effects to evaluate an occupational injury prevention programme. Statistics in Medicine 20, 2907–2920.

Yau, K.K.W., Wang, K., Lee, A.H., 2003. Zero-inflated negative binomial mixed regression modeling of over-dispersed count data with extra zeros. Biometrical Journal 45, 437–452.

Zhang, M., Strawderman, R.L., Cowne, M.E., Wells, M.T., 2006. Bayesian inference for a two-part hierarchical model: An application to profiling providers in managed health care. Journal of the American Statistical Association 101, 934–945.

Zhou, X.H., Tu, W.Z., 1999. Comparison of several independent population means when their samples contain log-normal and possibly zero observations. Biometrics 55, 645–651.