Marginal zero-inflated regression models for count data

Jacob Martin and Daniel B. Hall
Department of Statistics, University of Georgia, Athens, GA USA

ABSTRACT
Data sets with excess zeroes are frequently analyzed in many disciplines. A common framework used to analyze such data is the zero-inflated (ZI) regression model. It mixes a degenerate distribution with point mass at zero with a non-degenerate distribution. The estimates from ZI models quantify the effects of covariates on the means of latent random variables, which are often not the quantities of primary interest. Recently, marginal zero-inflated Poisson (MZIP; Long et al. [A marginalized zero-inflated Poisson regression model with overall exposure effects. Stat. Med. 33 (2014), pp. 5151–5165]) and negative binomial (MZINB; Preisser et al., 2016) models have been introduced that model the mean response directly. These models yield covariate effects that have simple interpretations that are, for many applications, more appealing than those available from ZI regression. This paper outlines a general framework for marginal zero-inflated models where the latent distribution is a member of the exponential dispersion family, focusing on common distributions for count data. In particular, our discussion includes the marginal zero-inflated binomial (MZIB) model, which has not been discussed previously. The details of maximum likelihood estimation via the EM algorithm are presented and the properties of the estimators as well as Wald and likelihood ratio-based inference are examined via simulation. Two examples presented illustrate the advantages of MZIP, MZINB, and MZIB models for practical data analysis.

ARTICLE HISTORY
Received 12 March 2016
Accepted 12 August 2016

KEYWORDS
Exponential dispersion family; generalized linear models; zero inflation; marginal models; EM algorithm

1. Introduction
Count data – both bounded and unbounded counts – regularly occur in a variety of disciplines. However, the standard modeling frameworks for such data, logistic regression and Poisson log-linear models, often fail to adequately account for the observed variability, a phenomenon known as overdispersion. In the common scenario that the data exhibit greater spread throughout their range (fat tails) than standard distributions would predict, a dispersion parameter can be helpful, either as a component of a model based on a more general two-parameter count distribution such as the negative binomial [13] or beta-binomial [4] distribution, or as an extra variance parameter in a model based only on moment assumptions and analyzed via quasilikelihood [24]. However, in other
cases, overdispersion may result from a more specific distributional violation: an excess frequency of zero values. This paper will focus on data that fall into the second scenario, which we refer to as excess zero (EZ) data. Many research disciplines encounter EZ data, such as manufacturing [12], horticulture [6], entomology [25], fish and wildlife conservation [15], and psychology [2]. The most common type of model for EZ data is the class of zero-inflated (ZI) regression models. Such models include the ZI Poisson (ZIP) [12] and ZI negative binomial (ZINB) [5] models for unbounded counts, and the ZI binomial (ZIB) [6] model for bounded counts. Continuous data with many zeros occur as well. While such data, known as semicontinuous data, can often be analyzed more easily than discrete EZ data, there are situations for which ZI models such as the zero-inflated tobit (ZIT) model prove useful (e.g. longitudinal semicontinuous data [18], data with both true zeros, zeros due to a detection limit [17]). ZI models are based on a mixture structure in which EZs are hypothesized to come from a degenerate distribution that produces zeros with probability one, mixed with a standard distribution that is responsible for the non-zero outcomes and possibly some of the zeros. Often these models include covariate effects, in which they become ZI regression models. In these cases, the dependence on covariates is built in through generalized linear model-like specifications for the non-degenerate distribution's mean and the mixing probability.

The goal of this paper is to formulate and explore an alternative to the ZI regression model in which the marginal mean of the response variable is modeled directly. In a ZI model, linear predictors are linked to the non-degenerate distribution's mean and the mixing probability. However, both these quantities are means for latent random variables rather than the observed response. This makes the interpretation of a ZI regression model awkward, especially if interest focuses on a covariate's direct effect on the mean response. To address this issue, we develop marginal ZI (MZI) models that retain the mixture structure of classical ZI regression, but parameterize it with a linear predictor that links directly to the marginal mean response. This approach makes model interpretation more straightforward and improves the efficiency of inferences on the marginal mean.

Among the situations in which the regression parameters of a ZI regression model are of direct interest are cases in which the model is motivated by an assumed latent class structure in the population generating the data. For instance, Heilbron [8] presented EZ data from self-reported counts of risky sexual behavior during the six months preceding data collection. In this case, it is plausible to assume that the surveyed population contains individuals who never engage in the behavior of interest (abstainers) as well as those who do. Here, we may be interested in covariate effects on the risky behavior rate among the non-abstinent subpopulation, as well as the effects of predictors of abstinence. For such effects, a ZIP or ZINB regression model is perfectly suited. However, there are many applications involving EZs where one is primarily interested in the overall mean response. For example, in Section 5 we analyze two examples involving EZ data, both arising from designed experiments from horticulture. In each case, there are no obvious latent sub-populations. Rather, as in many analyses of experimental data, the questions of primary interest focus on comparisons of the mean response across treatments. In these and many other examples, it is desirable that the regression function depend upon a single linear predictor that quantifies treatment and covariate effects. Such is the case in MZI models, which is, in contrast to ZI models, where two linear predictors must be resolved to summarize the overall effects of explanatory variables. Synthesizing results from a ZI model can be confusing and
cumbersome, and is often met with frustration by applied researchers who simply want to know whether the mean response differed significantly across treatments (and other explanatory factors) and by how much.

The importance of quantifying treatment and covariate effects on the marginal mean was recognized by Albert et al. [1] and Todem et al. [22], both of whom considered approaches for deriving marginal effects from fitted ZI models. However, these approaches involve additional calculations after fitting the model to obtain marginal effect estimates and delta method or bootstrap computations to obtain relevant standard errors. We find these approaches cumbersome and less appealing than modeling the quantities of interest directly in the original model specification.

The MZI models that we discuss are not entirely new. Recently, special cases of MZI regression— all involving unbounded count data and a log link for the marginal mean— have been considered by other authors. In particular, the marginal zero-inflated Poisson (MZIP) model was introduced by Long et al. [14] and extended in the form of the marginal zero-inflated negative binomial (MZINB) model by Preisser et al. [19] (see also Todem et al. [22], whose formulation subsumes these models and allows, in principle, non-degenerate distributions other than the Poisson and negative binomial for the unbounded counts). We provide a general formulation that encompasses additional MZI models— most notably the marginal zero-inflated binomial (MZIB) model for bounded counts—and give details regarding the EM fitting algorithm that have not appeared previously. Note that our terminology differs slightly from that of some previous authors in that we favor the term MZI model rather than marginalized ZI model. The latter term echoes the vocabulary of marginalized mixed-effect models, which were introduced by Heagerty [7] in the generalized linear mixed model context and have been considered in the ZI regression setting by Kassahun et al. [11]. While the goal of marginalized mixed-effect models is also to obtain covariate effects on the marginal mean, this typically involves specification of both conditional and marginal regression functions which are connected through an integral equation. Because we specify only a model for the marginal mean and there is no need for a marginalization operation to resolve marginal and conditional specifications, we prefer the term ‘marginal model’.

The remainder of the paper is organized as follows. Section 2 reviews ZI regression models. Section 3 introduces MZI regression models and discusses maximum likelihood estimation via the EM algorithm. In addition, this section catalogs the most important special cases of the MZI model and discusses some model interpretation issues. Section 4 presents the results of simulation studies that demonstrate the properties of ML estimators and likelihood-based inference in MZI regression, highlighting advantages over ZI regression. Examples involving both unbounded and bounded count data are used to illustrate the proposed class of models in Section 5. Section 6 provides some concluding remarks.

2. ZI regression

ZIP distributions have been in the statistical literature for more than a half century [3], but the first ZI regression model, the ZIP model, was developed by Lambert in 1992 [12]. The ZIP model is based on a finite mixture of a Poisson distribution and a distribution that is degenerate at zero. Denote the response vector as $y = (y_1, \ldots, y_n)'$, where $y_i$ is the
observed value of a random variable $Y_i$, where we assume independence among $\{Y_i\}$ across $i$ and

$$Y_i \sim \begin{cases} 0, & \text{with probability } \pi_i; \\ \text{Poisson}(\lambda_i), & \text{with probability } 1 - \pi_i. \end{cases} \tag{1}$$

The model allows $\lambda_i$ and $\pi_i$ to depend on covariates through the relationships

$$\log \left( \frac{\pi_i}{1 - \pi_i} \right) = g_i' \gamma,$$

$$\log (\lambda_i) = x_i' \beta,$$

where $x_i$ and $g_i$ are predictor vectors for the Poisson mean and mixing probability, respectively, and $\beta$ and $\gamma$ are the corresponding regression parameters. The coefficients are typically estimated via maximum likelihood. This approach can be implemented by direct optimization of the loglikelihood using, for example, gradient methods, or via the EM algorithm, which leads to iterative fitting of standard generalized linear models (GLMs) [6,12].

Extensions of the ZIP model are conceptually straightforward and can be accomplished by replacing the Poisson component of the model by another distribution suitable for the characteristics of the data. In particular, in a common situation in which the non-zero data exhibit overdispersion relative to a (truncated) Poisson distribution, a ZINB model [5] is useful. In this case, the non-degenerate component in Equation (1) becomes NB$(\lambda_i, \phi)$ and the same regression structure as in the ZIP model is retained. In the case of EZ bounded count data, the ZIB model [6] is useful, in which case the non-degenerate component becomes Bin$(m_i, \lambda_i)$, where $\lambda_i$ now represents the success probability and $m_i$ the number of trials constant for the binomial distribution. In this case, the regression structures of the model take the form

$$\log \left( \frac{\pi_i}{1 - \pi_i} \right) = g_i' \gamma,$$

$$\log \left( \frac{\lambda_i}{1 - \lambda_i} \right) = x_i' \beta.$$

In these and other ZI regression models, the methods of estimation and inference are largely the same as those originally developed for ZIP regression.

### 3. Marginal regression models

#### 3.1. Model formulation

One problem with ZI regression is that, in general, the resulting parameter estimates describe covariate effects on the latent count distribution, not the effects on the marginal mean, $E(Y|X)$. In many applications, interest focuses on the marginal mean, and while it is possible to compute the mean based on a standard ZI model, a covariate’s effect comes in through both $\pi$ and $\lambda$, complicating the interpretation of the model. Instead, it is desirable to model a covariate's effect on the mean directly while also accounting for EZs. To do so, we reparameterize the ZI model. Let $\pi_i$ and $\lambda_i$ be the probability of the response being
generated from the degenerate distribution and mean of the distribution, respectively, for the $i$th individual. Instead of a GLM-like specification for $\lambda_i$, as in standard ZI regression models, we assume $g(\mu_i) = x_i^T \beta$ for $g$ a link suitable for the range of $\mu_i$. In addition, a logistic model, $\log(\pi_i/1 - \pi_i) = g'(y)$ is assumed for the mixing probability as usual. The resulting class of models, which we refer to as MZI regression models, are still ZI mixtures of a degenerate distribution at zero with a standard exponential family distribution, but the mean of the non-degenerate distribution $\lambda_i$, and thus also its canonical parameter $\theta_i$, now depends upon both $\beta$ and $y$ through the relationship $\lambda_i = \mu_i/(1 - \pi_i)$. Writing the non-degenerate distribution's density function in standard exponential dispersion family form [10], the probability density function for subject $i$ becomes

$$ f(y_i; \gamma, \beta) = \left(1 - \pi_i(y)\right) \left[\frac{\pi_i(y)}{1 - \pi_i(y)} + \exp\left\{\frac{b[\theta_i(y, \beta)]}{a(\phi)} + c(\phi)\right\}\right]^{u_i} $$

$$ \times \left[\exp\left\{\frac{y_i[\theta_i(y, \beta)]}{a(\phi)} + c(y_i, \phi)\right\}\right]^{1-u_i}, \quad (2) $$

where $u_i = 1$ if $y_i = 0$ and 0 otherwise.

The joint loglikelihood for the model is the sum of log densities of the form Equation (2). A function optimizer can be used to find the MLE. However, this can be computationally taxing if the parameter dimension is large. Therefore, we also outline an EM algorithm that is similar to the one used for standard ZI models, although it is not as convenient in the MZI case because the complete data likelihood does not factor cleanly into separate components for $y$ and $\beta$.

The complete data loglikelihood is

$$ \ell^c(y, \beta; y, z) = \sum_{i=1}^n \left\{ z_i \log \left(\frac{\pi_i}{1 - \pi_i}\right) + \log(1 - \pi_i) + (1 - z_i) \log h(y_i; \Psi) \right\}, \quad (3) $$

where $h(y_i; \Psi)$ is the non-degenerate distribution, $\Psi$ is the vector of parameters, and

$$ z_i = \begin{cases} 1, & \text{if } y_i \text{ is from the degenerate distribution;} \\ 0, & \text{otherwise.} \end{cases} $$

At the $r+1$th EM iteration, the E step is taken by estimating the posterior mixing probabilities. Using Bayes’ theorem, we obtain

$$ z_i^{(r+1)} = \frac{\hat{\pi}_i^{(r)} u_i / (\hat{\pi}_i^{(r)} + (1 - \hat{\pi}_i^{(r)}) h(y_i; \hat{\pi}_i^{(r)}, \hat{\mu}_i^{(r)})}$$

(cf. [12, p. 4]). Then the parameters, $\Psi$, can be estimated iteratively via a two-part M step. The first part uses a generic function optimizer (e.g. the nlm() function in R) to maximize $\ell^c(y, \hat{\beta}^{(r)}; y, z^{(r+1)})$ to find $\hat{\gamma}^{(r+1)}$. After finding $\hat{\gamma}^{(r+1)}$, $\hat{\beta}^{(r+1)}$ is calculated by fitting a weighted GLM of the form $g^*(\lambda_i; (1 - \hat{\pi}_i^{(r+1)})) = g((1 - \hat{\pi}_i^{(r+1)}) \lambda_i) = x_i^T \beta$ with weights $1 - z_i^{(r+1)}$. Here, note that the link $g^*$ for this step is a function of $\lambda_i$ but also depends on $1 - \hat{\pi}_i^{(r+1)}$, which is treated as fixed. Its form matches that of $g$, the link function for $\mu_i$ in the MZI model specification, evaluated at $1 - \hat{\pi}_i^{(r+1)} \lambda_i$. The E and two-part M steps are iterated until the algorithm converges. The model is considered to have
converged when \( (\ell(\Psi^{(r)}; y, z) - \ell(\Psi^{(r+1)}; y, z))/\ell(\Psi^{(r)}; y, z) \) is less than a tolerance level, which is taken as \( 10^{-8} \) for the models fit in this paper.

The benefit of using this approach is that there is a reduction in the parameter dimension that is maximized using the function optimizer. Instead of maximizing \( \ell(\gamma, \beta; y) \) with respect to \((\gamma', \beta')'\) using unstructured optimization, only \( \ell_c(\gamma, \beta^{(r)}; y, z^{(r+1)}) \) needs to be optimized with respect to \( \gamma \) in this manner. The dimension of \( \beta \) is often larger than that of \( \gamma \), leading to a large decrease in computational complexity at each iteration of the fitting algorithm, although with the EM algorithm we can expect many iterations to be necessary. The information matrix can be found via direct differentiation of the observed data loglikelihood and is given in the appendix. At convergence, an estimated asymptotic variance–covariance matrix can be obtained from the inverse information matrix evaluated at the MLE \( \hat{\Psi} \).

### 3.2. Special cases

In this section, we provide some further details for the MZIB, MZIP and MZINB models, which are the most important special cases within the MZI class. In each case, we give the observed and complete data loglikelihoods as well as some details of the EM algorithm. The former function can be obtained as the sum over \( i \) of the log of Equation (2) after substitution of the specific form of the exponential density \( h \). The complete data loglikelihoods follow from Equation (3), again with substitution of the appropriate form of \( h \). In particular,

\[
h(y_i; \Psi) = \begin{cases} 
\binom{n_i}{y_i} \mu_i^{y_i} (1 - \pi_i - \mu_i)^{n_i - y_i} (1 - \pi_i)^{-n_i}, & \text{in the MZIB case;} \\
\exp\{-\mu_i/(1 - \pi_i)\} \mu_i^{y_i} (1 - \pi_i)^{y_i}, & \text{in the MZIP case;} \\
\frac{\Gamma(\phi + y_i)}{\Gamma(\phi)} \mu_i^{y_i} [(1 - \pi_i)\phi]^{\phi} [(1 - \pi_i)\phi + \mu_i]^{-(\phi + y_i)}, & \text{in the MZINB case.}
\end{cases}
\]

#### 3.2.1. MZIB model

For the MZIB model, the non-degenerate component is Bin\((m_i, \lambda_i)\) and the observed data loglikelihood is

\[
\ell(\gamma, \beta; y, m) = \sum_{i=1}^{n} \left\{ \log(1 - \pi_i) + u_i \log \left[ \frac{\pi_i}{1 - \pi_i} + (1 - \pi_i)^{-m_i} (1 - \pi_i - \mu_i)^{m_i} \right] \right.
\]

\[
+ (1 - u_i) \left[ y_i \log \mu_i + (m_i - y_i) \log(1 - \pi_i - \mu_i) - m_i \log(1 - \pi_i) \right],
\]

where \( m_i \) is the number of trials and \( \mu_i \) is the marginal probability of success for the \( i \)th observation. In this model, the link function \( g \) relating \( \mu \) to \( x' \beta \) typically will be the logit, but other links such as the probit and complementary log–log functions are also permissible. The complete data loglikelihood, on which the EM algorithm is based, takes
the form
\[
\ell^c(\mathbf{y}, \mathbf{\beta}; \mathbf{y}, \mathbf{m}) = \sum_{i=1}^{n} \left\{ \log(1 - \pi_i) + z_i \log \left( \frac{\pi_i}{1 - \pi_i} \right) + y_i \log \mu_i + (1 - z_i) \left[ (m_i - y_i) \log(1 - \pi_i - \mu_i) - m_i \log(1 - \pi_i) \right] \right\},
\]

For the \( r+1 \)th iteration of the EM algorithm, the steps are

1. (E Step) Calculate the posterior probabilities, \( z_i^{(r+1)} \), using (2).
2. (M Step) Update the estimates for \( \mathbf{\gamma} \) and \( \mathbf{\beta} \):
   (a) \( \hat{\mathbf{\gamma}}^{(r+1)} \) is found by maximizing \( \ell^c(\mathbf{y}, \mathbf{\beta}; \mathbf{y}, \mathbf{m}) \) with \( \mathbf{\beta} \) fixed at \( \mathbf{\beta}^{(r)} \).
   (b) \( \hat{\mathbf{\beta}}^{(r+1)} \) is calculated using a GLM with:
      • link \( g^{*} \{ \lambda_i; (1 - \hat{\pi}_i^{(r+1)}) \} = g \{(1 - \hat{\pi}_i^{(r+1)})\lambda_i \}
      • weights \( 1 - z_i^{(r+1)} \).
3. Repeat steps 1 and 2 until convergence.

### 3.2.2. MZIP Model

For the MZIP model, the non-degenerate component takes the form Pois(\( \lambda_i \)), leading to an observed data loglikelihood of the form
\[
\ell(\mathbf{y}, \mathbf{\beta}; \mathbf{y}) = \sum_{i=1}^{n} \left\{ \log(1 - \pi_i) + u_i \log \left[ \frac{\pi_i}{1 - \pi_i} + \exp \left( \frac{-\mu_i}{1 - \pi_i} \right) \right] + (1 - u_i) \left( y_i \log \mu_i - \mu_i \right) \right\},
\]
where again \( \mu_i \) is the marginal mean for the \( i \)th observation, and a log link for \( \mu \) is assumed.

In this case, the complete data loglikelihood is
\[
\ell^c(\mathbf{y}, \mathbf{\beta}; \mathbf{y}) = \sum_{i=1}^{n} \left\{ z_i \log \left( \frac{\pi_i}{1 - \pi_i} \right) - (y_i - 1) \log(1 - \pi_i) + \frac{1 - z_i}{1 - \pi_i} \left[ y_i(1 - \pi_i) \log(\mu_i) - \mu_i \right] \right\}.
\]

For the \( r+1 \)th iteration of the EM algorithm, the steps are

1. (E Step) Calculate the posterior probabilities, \( z_i^{(r+1)} \), using (2).
2. (M Step) Update the estimates for \( \mathbf{\gamma} \) and \( \mathbf{\beta} \):
   (a) \( \hat{\mathbf{\gamma}}^{(r+1)} \) is found by maximizing \( \ell^c(\mathbf{y}, \mathbf{\beta}; \mathbf{y}) \) with \( \mathbf{\beta} \) fixed at \( \mathbf{\beta}^{(r)} \).
   (b) \( \hat{\mathbf{\beta}}^{(r+1)} \) is calculated using a GLM with:
      • link \( g(\mu_i) = \mathbf{x}_i^\prime \mathbf{\beta} \)
      • response \( y_i^c \equiv y_i (1 - \hat{\pi}_i^{(r+1)}) \)
      • weights \( (1 - z_i^{(r+1)}) / (1 - \hat{\pi}_i^{(r+1)}) \).
3. Repeat steps 1 and 2 until convergence.
3.2.3. MZINB model

For the MZINB model, the non-degenerate component is NB($\lambda_i, \phi$), and the observed data loglikelihood is

$$
\ell(\mathbf{y}, \mathbf{\beta}, \phi; \mathbf{y}) = \sum_{i=1}^{n} \left\{ \log(1 - \pi_i) + u_i \log \left[ \frac{\pi_i}{1 - \pi_i} + \left( \frac{(1 - \pi_i)\phi}{\mu_i + (1 - \pi_i)\phi} \right)^\phi \right] + (1 - u_i) \right\} \left\{ y_i \log \mu_i + \phi \log[(1 - \pi_i)\phi] - (\phi + y_i)\log[\mu_i + (1 - \pi_i)\phi] + \log \frac{\Gamma(\phi + y_i)}{\Gamma(\phi)} \right\},
$$

As in the MZIP case, a log-linear model for $\mu$ is assumed. In this case, the complete data loglikelihood is

$$
\ell^c(\mathbf{y}, \mathbf{\beta}, \phi; \mathbf{y}) = \sum_{i=1}^{n} \left\{ \log(1 - \pi_i) + z_i \log \left( \frac{\pi_i}{1 - \pi_i} \right) + \right\} \left\{ 1 - \frac{z_i}{1 - \pi_i} \right\} \left\{ y_i(1 - \pi_i) \log \mu_i + (1 - \pi_i)\phi \log[(1 - \pi_i)\phi] - [(1 - \pi_i)\phi + y_i(1 - \pi_i)]\log[\mu_i + (1 - \pi_i)\phi] + \log \left( \frac{\Gamma(\phi + y_i)}{\Gamma(\phi)} \right) \right\}.
$$

For the $r+1$th iteration of the EM algorithm, the steps are

1. (E Step) Calculate the posterior probabilities, $z_i^{(r+1)}$, using (2).
2. (M Step) Update the estimates for $\mathbf{y}$ and $\mathbf{\beta}$:
   a. $((\hat{\mathbf{y}}^{(r+1)}), \hat{\phi}^{(r+1)})'$ is found by maximizing $\ell^c(\mathbf{y}, \mathbf{\beta}, \phi; \mathbf{y})$ with $\mathbf{\beta}$ fixed at $\mathbf{\beta}^{(r)}$.
   b. $\hat{\mathbf{\beta}}^{(r+1)}$ is calculated using a glm with:
      - link $g(\mu_i) = \mathbf{x}_i^{\prime} \mathbf{\beta}$
      - response $y_i^{c} \equiv y_i(1 - \hat{\pi}_i^{(r+1)})$
      - weights $(1 - z_i^{(r+1)})/(1 - \hat{\pi}_i^{(r+1)})$
      - fixed dispersion parameter $\phi_i^{c} \equiv \hat{\phi}^{(r+1)}(1 - \hat{\pi}_i^{(r+1)})$
3. Repeat steps 1 and 2 until convergence.

Here, the quantity $\phi_i^{c}$ is treated as a fixed dispersion parameter in the iteratively reweighted least-squares (IRLS) fitting algorithm used for $\mathbf{\beta}$. This can be done in some GLM software and IRLS algorithms, or can be programmed easily from scratch.

Alternatively, $\phi$ can be grouped with $\mathbf{\beta}$ in the M step and profiled during the M step for $(\beta', \phi')$. This would fix $\phi$ over each value in grid of values in its parameter space, allowing $\mathbf{\beta}$ to be updated via a standard GLM algorithm. This is akin to the approach is some negative binomial regression software (e.g. the glm.nb function in the MASS package for R), but offers no computational advantage over numerically optimizing $\ell^c(\mathbf{y}, \mathbf{\beta}, \phi; \mathbf{y})$ with respect to both $\mathbf{y}$ and $\phi$ as described above, which we find to be simpler to implement and faster computationally.
3.3. Model interpretability

Because of the functional relationship between $\mu_i$ and $\pi_i$, one should take care to use sensible specifications for these two quantities in an MZI model. In particular, while it is not incompatible to include a covariate in the model for $\pi_i$ but exclude its effect from the linear predictor for $\mu_i$, such models are best avoided. The reason for this recommendation is that if a covariate $x_i$ is specified to have a direct effect on $\pi_i$, then it will necessarily also have an effect on $\mu_i$ through the relationship $\mu_i = (1 - \pi_i)\lambda_i$. Thus, omitting $x_i$ from the model for $\mu_i$ implies that $\lambda_i$ changes with $x_i$ in just such a way as to nullify this covariate’s effect on $\pi_i$. Such a model seems highly implausible; therefore, for model interpretability’s sake we recommend that, in general, covariates in $G$ be a subset of those in $X$.

Also worth noting is the relationship between covariate effects on $\lambda_i$ and $\mu_i$ in ZI and MZI models involving log link. In particular, consider a ZI model with specifications

\[
\begin{align*}
\text{logit}(\pi_i) &= g'_y \gamma, \\
\log(\lambda_i) &= x'_i \beta.
\end{align*}
\]

The relationship between $\mu_i$, $\lambda_i$, and $\pi_i$ implies $\log(\mu_i) = \log(\lambda_i) + \log(1 - \pi_i) = x'_i \beta - v_i$, where $v_i = \log(1 + \exp(g'_y \gamma))$. Therefore, if the predictor $x_{ik}$ is not included in $g_i$, then $e^{\beta_k}$ quantifies the multiplicative effect of $x_{ik}$ on both $\lambda$ and $\mu$:

\[
\log \left( \frac{\mu_i}{\lambda_i} \right) = (c + 1)\beta_k + x'_{i2} \beta_2 - v_i - (c\beta_k + x'_{i2} \beta_2 - v_i)
\]

\[
= \beta_k = \log \left( \frac{\lambda'_i}{\lambda_i} \right),
\]

where, without loss of generality, $x_i = (x_i, x'_i)'$, and where $\mu'_i = E(y_i | x_{i2}, x_{ik} = c + 1)$, $\mu_i = E(y_i | x_{i2}, x_{ik} = c)$, and $\lambda'_i, \lambda_i$ are defined similarly, conditional on $z_i = 1$.

In particular, when the mixing probability is constant, then the ZI and MZI models are reparameterizations of each other. In this case, the regression coefficients $\beta$ in an MZIP or MZNBI model differ from those in the corresponding ZI model only through the intercept, $\beta_0$, which is shifted by the constant $v = \log(1 - \pi)$. The equivalence extends to models in which the mixing probability differs only across groups of subjects (that is, for an ANOVA-like specification for $g'_y \gamma$), provided that the group structure is also included in the specification of $x'_i \beta$. For example, in a ZI model with $\logit(\pi_{ij}) = g'_y$, and $\log(\lambda_{ij}) = \beta_j + x'_i \beta$, then a marginal group mean for group $j$ at fixed $x_i$ is given by $E(y_{ij} | x_i) = \exp((\beta_j - v_j) + x'_i \beta)$. For cases when the mixing probability depends on continuous covariates or for models that lack a log link (e.g. for bounded count data), however, the ZI and MZI model classes are not equivalent.

When MZI and ZI models are distinct, the advantage of the former class is that its parameters can be interpreted as covariate effects or differences in group means on the marginal mean response (on a link-transformed scale, as in all GLMs). In contrast, ZI model parameters pertain to the mean of the latent non-degenerate variable in the mixture underlying both model classes. While effects on the marginal mean can be estimated from ZI model parameters, there are problems with such quantities. First, in a ZI model,
there is typically no scale on which a covariate effect is constant; it depends on the values of other explanatory variables in the model. This vastly diminishes the interpretability of a ZI regression model when interest focuses on marginal mean effects. Secondly, when marginal mean effects are computed in a ZI model, these quantities are, in general, nonlinear functions of both \( \hat{\beta} \) and \( \hat{\gamma} \). While asymptotic normality for these quantities can often be justified via the delta method, the resulting asymptotic standard errors and normal-theory reference distributions for Wald inference can be expected to be poor relative to the more direct inference on marginal mean effects that is available for an MZI model. The latter consideration is not a concern when likelihood ratio inference is used, but Wald inference is often much easier in practice for these models because, for example, Wald confidence intervals are much less demanding computationally than profile likelihood intervals, and formulating and/or fitting the model under a null hypothesis of interest is sometimes difficult. We investigate the inference advantages of MZI models over traditional ZI regression in Section 4, but it is worth emphasizing the fact that the ZI model often does not yield constant covariate effects at all, over problems with inference on those effects in a ZI model.

4. Simulation

Simulation studies were performed to demonstrate the advantages of the MZI model over a ZI model, and to examine the properties of ML estimators and Wald and likelihood ratio-based inferences in these models. Following a similar formulation of Long et al. [14], data were generated from MZI models, all of which take the form

\[
\text{logit}(\pi_i) = \gamma_0 + \gamma_1 x_{1i} + \gamma_2 x_{2i},
\]

\[
g(\mu_i) = \beta_0 + \beta_1 x_{1i} + \beta_2 x_{2i},
\]

where \( x_1 \) is a vector of an equal number of zeros and ones and \( x_2 \sim \text{Uniform}(-1,1) \). Table 1 reports results for data generated from MZIB models where true parameter values are taken to be \( \gamma = (-1.386, 0.539, 0.25) \) and \( \beta = (-0.405, 0.811, -0.25) \). Properties of the estimates from correctly specified models fit to 10,000 data sets of size \( n = 100, 200, \) and 1000 are reported. These properties include percent relative median bias (PRMB), simulation standard deviation (SSD), the model-based standard error (MBSE), and model-based coverage probability (MBCP) of 95% Wald confidence intervals. Similar simulations were conducted for MZIP and MZINB models, but due to the similarity of the simulation scenarios and results to those reported by Long et al. [14] and Preisser et al. [19], these cases are relegated to an online supplement to this article available from the journal website. Calculations for the simulations and examples were done in R [20] using the zeroinfl function of the pscl package [9] for fitting ZI models and functions written by the authors for MZI model estimation and inference. R code for fitting the MZI models of Section 5 is included in the online supplement.

Results in Table 1 show reasonably small values of PRMB that decrease with the sample size, as expected. While there is some inaccuracy in the MBSEs at the smaller sample sizes, this diminishes with increasing sample size, and the Wald coverage probabilities are accurate for all choices of \( n \). For the MZIP and MZINB cases (not shown), the PRMB values at \( n = 100 \) are larger than those in Table 1, but the general pattern of results is similar, and these simulations suggest that ML estimation and Wald inference in MZI models perform reasonably well at moderate sample sizes.

Table 2 shows the coverage probabilities of Wald 95% confidence intervals of \( \mu_i = g^{-1}(\beta_0 + \beta_1 x_{1i} + \beta_2 x_{2i}) \). Ten thousand data sets were generated from the MZIP and MZINB models where \( \text{logit}(\pi_i) = 0.6 - 2x_{1i} + 0.25x_{2i}, \) \( \log(\mu_i) = 0.25 + \log(1.5)x_{1i} + 0.5x_{2i}, \) and
Table 1. Simulation results for ML estimation in the MZIB model.

| Sample size | Parameter | $\gamma_0$ | $\gamma_1$ | $\gamma_2$ | $\beta_0$ | $\beta_1$ | $\beta_2$ |
|-------------|-----------|-------------|-------------|-------------|----------|----------|----------|
| PRMB 100    |           | 0.83        | 1.47        | -0.20       | -0.24    | 0.86     | -0.82    |
| PRMB 200    |           | 0.55        | 0.75        | 1.14        | 0.24     | 0.68     | 0.99     |
| PRMB 1000   |           | 0.15        | 0.28        | 0.09        | 0.12     | 0.04     | 0.04     |
| SSD 100     |           | 0.386       | 0.509       | 0.334       | 0.148    | 0.284    | 0.220    |
| SSD 200     |           | 0.255       | 0.340       | 0.212       | 0.103    | 0.200    | 0.141    |
| SSD 1000    |           | 0.114       | 0.150       | 0.093       | 0.046    | 0.088    | 0.063    |
| MBSE 100    |           | 0.356       | 0.478       | 0.307       | 0.144    | 0.280    | 0.211    |
| MBSE 200    |           | 0.251       | 0.335       | 0.204       | 0.102    | 0.198    | 0.137    |
| MBSE 1000   |           | 0.112       | 0.148       | 0.093       | 0.046    | 0.088    | 0.063    |
| MBCP 100    |           | 0.950       | 0.949       | 0.941       | 0.946    | 0.948    | 0.949    |
| MBCP 200    |           | 0.953       | 0.954       | 0.946       | 0.948    | 0.948    | 0.946    |
| MBCP 1000   |           | 0.948       | 0.949       | 0.946       | 0.946    | 0.949    | 0.950    |

Note: PRMB, percent relative median bias; SSD, simulation standard deviation; MBSE, model-based standard error; and MBCP, model-based coverage probability.

Table 2. Comparing the coverage probabilities (width of interval) of the ZIP, ZINB, and ZIB models to their marginal equivalents.

| $x_2i$ | Size | ZIP | MZIP | ZINB | MZINB | ZIB | MZIB |
|--------|------|-----|------|------|-------|-----|------|
| -1     | 100  | 0.941(0.941) | 0.948(0.971) | 0.938(1.134) | 0.942(1.254) | 0.945(0.212) | 0.949(0.213) |
|        | 200  | 0.934(0.689) | 0.953(0.836) | 0.934(0.902) | 0.945(1.014) | 0.931(0.152) | 0.947(0.154) |
|        | 1000 | 0.853(0.290) | 0.948(0.351) | 0.915(0.376) | 0.950(0.423) | 0.873(0.068) | 0.947(0.069) |
| 0      | 100  | 0.946(0.942) | 0.948(0.950) | 0.941(1.161) | 0.940(1.155) | 0.949(0.191) | 0.951(0.191) |
|        | 200  | 0.949(0.668) | 0.950(0.670) | 0.945(0.832) | 0.945(0.826) | 0.945(0.136) | 0.950(0.136) |
|        | 1000 | 0.942(0.296) | 0.943(0.299) | 0.946(0.370) | 0.947(0.370) | 0.944(0.061) | 0.950(0.061) |
| 1      | 100  | 0.940(3.157) | 0.944(3.245) | 0.933(3.560) | 0.943(3.581) | 0.947(0.203) | 0.950(0.209) |
|        | 200  | 0.936(1.888) | 0.946(1.973) | 0.946(2.500) | 0.944(2.522) | 0.943(0.138) | 0.951(0.142) |
|        | 1000 | 0.903(0.815) | 0.951(0.853) | 0.929(1.070) | 0.951(1.085) | 0.914(0.062) | 0.951(0.063) |

$\phi = 3$ and from an MZIB model with logit$(\pi_i) = -1.386 + 0.539x_{1i} + .25x_{2i}$ and logit$(\mu_i) = -0.405 + 0.405x_{1i} + .25x_{2i}$, where $x$ is defined previously. Both MZI and ZI models were fitted to these data sets and the confidence interval for $\mu_i$ was calculated by $g^{-1} \{\log(\hat{\mu}_i) \pm 1.96 \times se(\hat{\mu}_i)\}$. For the ZI models, the delta method was used to find the standard error of $\hat{\mu}_i$ based on the estimated asymptotic covariance matrix of the ZI regression parameters, which is computed as the inverse observed information matrix evaluated at the ML estimates. Three different values of $x_{2i}$ were chosen when calculating the confidence interval for $\mu_i$, $-1$, $0$, and $1$. As expected, due to the linear relationship between the estimators for the ZI and MZI models with a log link when $x_{2i} = 0$, the ZIP and MZIP models gave almost identical coverage probabilities, and the same is true for the ZINB and MZINB models. Results are also similar when $x_{2i} = 0$ for the ZIB and MZIB models, but to a lesser degree because the MZI and ZI model equivalence does not hold true for the logit link. When $x_{2i} \neq 0$ (most noticeably for $x_{2i} = -1$), while the MZI model-based Wald intervals have close-to-nominal coverage, intervals for the ZI model have lower coverage, especially for samples sizes of 1000. As noted by Long et al. [14], the delta method for a ZI model strongly depends on the values of extraneous covariates and the larger the value of these covariates, the greater the effect on estimation of the standard error, with the effect increasing with the sample size.
Table 3. The power of a .05 significance level test for $H_{01} : \log(R) = 0$ and $H_{02} : \gamma_1 = \beta_1 = 0$.

| R   | Size | Wald | LRT | Wald | LRT | Wald | LRT | Wald | LRT | Wald | LRT | Wald | LRT | Wald | LRT | Wald | LRT |
|-----|------|------|-----|------|-----|------|-----|------|-----|------|-----|------|-----|------|-----|------|-----|
|     |      |      |     |      |     |      |     |      |     |      |     |      |     |      |     |      |     |
| $x^2_i = 1$ | | | | | | | | | | | | | | | | | | |
| 1   | 100  | 0.059 | 0.056 | 0.109 | 0.998 | 1.000 | 0.060 | 0.058 | 0.083 | 0.845 | 0.933 | 0.053 | 0.052 | 0.052 | 0.515 | 0.520 |
|     | 200  | 0.048 | 0.048 | 0.117 | 1.000 | 1.000 | 0.057 | 0.056 | 0.083 | 0.992 | 0.999 | 0.063 | 0.061 | 0.060 | 0.831 | 0.837 |
|     | 1000 | 0.049 | 0.047 | 0.248 | 1.000 | 1.000 | 0.056 | 0.056 | 0.191 | 1.000 | 1.000 | 0.050 | 0.050 | 0.061 | 1.000 | 1.000 |
| 1.5 | 100  | 0.468 | 0.463 | 0.489 | 0.585 | 0.608 | 0.379 | 0.367 | 0.346 | 0.374 | 0.393 | 0.220 | 0.215 | 0.178 | 0.700 | 0.705 |
|     | 200  | 0.720 | 0.714 | 0.790 | 0.898 | 0.898 | 0.577 | 0.577 | 0.595 | 0.611 | 0.629 | 0.384 | 0.372 | 0.294 | 0.961 | 0.961 |
|     | 1000 | 1.000 | 1.000 | 1.000 | 1.000 | 1.000 | 0.996 | 0.996 | 0.997 | 1.000 | 1.000 | 0.954 | 0.953 | 0.898 | 1.000 | 1.000 |
| 2   | 100  | 0.924 | 0.921 | 0.927 | 0.991 | 0.991 | 0.827 | 0.821 | 0.799 | 0.835 | 0.849 | 0.607 | 0.593 | 0.495 | 0.992 | 0.992 |
|     | 200  | 1.000 | 1.000 | 1.000 | 1.000 | 1.000 | 0.983 | 0.982 | 0.973 | 0.995 | 0.995 | 0.855 | 0.852 | 0.761 | 1.000 | 1.000 |
|     | 1000 | 1.000 | 1.000 | 1.000 | 1.000 | 1.000 | 1.000 | 1.000 | 1.000 | 1.000 | 1.000 | 1.000 | 1.000 | 1.000 | 1.000 | 1.000 |

Table 3 contains the results of simulations that examine tests of hypotheses that compare marginal means across two treatment groups. Here, we generated data from MZI models with $g(\mu_i) = \beta_0 + \beta_1 \text{trt}_i + \beta_2 x_i$, $i = 1, \ldots, n$, where trt$_i$ is a 0–1 indicator of which subjects received the treatment. In addition, we set logit($\gamma_i$) = $\gamma_0 + \gamma_1 \text{trt}_i + \gamma_2 x_i$ and the values of $\gamma$ and $\beta$ are the same as the simulations of Table 2 except $\beta_1$ takes three different values: 0, log(1.5), and log(2). Define $R$ as the mean response ratio in the two treatment groups; we are interested in testing the hypothesis $H_0: \log(R) = 0$. In MZI models this quantity does not depend upon $x_i$, and the hypothesis can be formulated and computed easily using either a Wald or likelihood ratio (LR) test of $H_{01} : \beta_1 = 0$. In this context, ZI regression models do not provide a convenient framework for inference on $R$. Nevertheless, when available explanatory variables include a treatment indicator and continuous covariate it is natural to consider ZI regression models of the form $g(\lambda_i) = \beta_0^* + \beta_1^* \text{trt}_i + \beta_2^* x_i$, logit($\pi_i$) = $\gamma_0^* + \gamma_1^* \text{trt}_i + \gamma_2^* x_i$. Unlike the corresponding MZI model, here the ratio of means depends upon $x_i$. In particular, the hypothesis $\beta_1^* = 0$ no longer addresses whether
the ratio of marginal means is 0 and is not a suitable substitute for $H_0$. Instead, we consider two hypothesis tests that one might conduct in the ZI model if one were interested in whether the marginal mean was equal across treatment groups. First, using the relationship $\mu_i = (1 - \pi_i)\lambda_i$ one can always estimate $R_i \equiv E(y_i | \text{trt}_i = 1, x_{2i})/E(y_i | \text{trt}_i = 0, x_{2i})$ and conduct a Wald test of the hypothesis $H_{02} : \log(R_i) = 0$ using the delta method to estimate the required nonlinear functions of the model’s regression parameters and their standard errors. Note that in the ZI model the ratio of marginal means depends on $x_{2i}$, hence the notation $R_i$ instead of $R$. Thus $H_{02}$ is not equivalent to $H_0$ and its target of inference varies with $x_{2i}$. Despite this non-equivalence, we examine the ZI regression-based Wald test of $H_{02}$ at $x_{2i} = -1, 0, 1$. Second, a natural way to assess the overall treatment effect in a ZI model is to consider models with and without the treatment factor in the set of explanatory variables in the model. That is, for ZI models we consider Wald and LR tests of $H_{03} : \gamma_1^* = \beta_1^* = 0$. Note that, again, $H_{03}$ is not equivalent to $H_0$ [1] because while $H_{03}$ implies $H_0$, the converse is not true. In particular, it is possible for the treatment to have positive effects on both $\lambda$ and $\pi$ that offset to give a net null effect on the marginal mean response.

In Table 3, data are generated from MZI models with $R = 1, 1.5, \text{ and } 2$ by setting $\beta_1$ to 0, $\log(1.5)$ and $\log(2)$, respectively. In the first case, $H_{01}$ is true, $H_{02}$ is true provided that $x_{2i} = 0$, and $H_{03}$ is false. In this setting, all of the tests for $H_{01}$, as well as tests for $H_{02}$ when $x_{2i} = 0$, have empirical sizes that are reasonably close to nominal (0.05) at the largest sample size; the Wald and LR tests of $H_{01}$, which are those that are based on fitted MZI models, show the best performance overall. When $x_{2i} \neq 0$, the Wald test of $H_{02}$ based on a ZI model often rejects at a much higher rate than .05, and its rejection rate increases with the sample size. The non-equivalence of $H_0$ and $H_{03}$ can be seen by the high rejection rates for tests of the latter hypothesis based on ZI models, when the data are generated with $\log(R) = 0$. Thus none of the tests based on ZI models that attempt to address this hypothesis are suitable. When $H_0$ is false, the power of the Wald and LRTs of $H_{01}$ are similar, with neither test dominating the other.

Collectively, our simulation results demonstrate that, except in special cases when MZI and ZI models are equivalent (see Section 3.3), the latter class of models does not provide a suitable framework for estimation and inference on the marginal mean.

5. Examples

5.1. Apple roots

In Ridout et al. [21], data from an experiment in which the number of roots produced by shoots of an apple tree cultivar were analyzed. Growing conditions were determined by the $2 \times 4$ factorial combinations of two factors: length of photoperiod (8 and 16 h) and concentration of the cytokinin BAP (2.2, 4.4, 8.8, and 17.6 µM). The design was unbalanced, with a total of 267 shoots propagated, where two treatments had 40 shoots, one had 39 shoots, four had 30 shoots, and one had 28 shoots propagated. The response variable is the number of roots grown per shoot. The data set contains a large number of zeros, but histograms of the response by photoperiod suggest that zero inflation is primarily of concern under the 16-h photoperiod (see Figure 1).
Figure 1. Histograms for the apple root data set by photoperiod.

Let $y_{ijk}$ be the number of roots for the $k$th shoot at the $i$th photoperiod and the $j$th BAP concentration. Also, $\pi_{ijk}$, $\lambda_{ijk}$, and $\mu_{ijk}$ are the mixing probability, latent mean, and overall mean for $y_{ijk}$, respectively. Three different ZIP and ZINB models for the data were fitted and compared:

Model 1: \[ \logit(\pi_{ijk}) = \gamma_{1i}, \quad \log(\lambda_{ijk}) = \beta_{1i} + \beta_{2} \log(BAP_{ij}); \]
Model 2: \[ \logit(\pi_{ijk}) = \gamma_{1i}, \quad \log(\lambda_{ijk}) = \beta_{1i} + \beta_{2i} \log(BAP_{ij}); \]
Model 3: \[ \logit(\pi_{ijk}) = \gamma_{1i} + \gamma_{2} \log(BAP_{ij}), \quad \log(\lambda_{ijk}) = \beta_{1i} + \beta_{2i} \log(BAP_{ij}). \]

In addition, corresponding MZIP and MZINB models were also fit that have linear predictors of the same form as Models 1–3. The parameter estimates for each model are found in Table 4.

Because LR tests and AIC statistics favor the NB-based models over the corresponding Poisson-based models, we restrict further attention to ZINB and MZINB models. From Model 1, it would appear that $\log(BAP)$ has little effect on the mean since the standard error is greater than $\hat{\beta}_{2}$. However, when separated into two different slopes for each photoperiod, $\log(BAP)$ does have a significant effect on the mean, one that is positive for photoperiod 8 and negative for photoperiod 16. Since by Wald and likelihood ratio tests, $\hat{\gamma}_{2}$ is not statistically different from zero in Model 3, and Model 2 has the lowest AIC, we conclude that $\log(BAP)$ has a negligible effect on the mixing probabilities and settle on the MZINB and ZINB versions of Model 2 for further analyses.
Table 4. Estimates and standard errors for the apple root data.

| Parameter | Est    | SE    | Est    | SE    | Est    | SE    | Est    | SE    | Est    | SE    |
|-----------|--------|-------|--------|-------|--------|-------|--------|-------|--------|-------|
| $\gamma_{11}$ | -4.275 | 0.747 | -4.275 | 0.747 | -4.256 | 0.733 | -4.234 | 0.840 | -4.235 | 0.835 |
| $\gamma_{12}$ | -0.071 | 0.178 | -0.071 | 0.178 | -0.073 | 0.178 | -0.051 | 0.448 | -0.052 | 0.440 |
| $\gamma_{2}$  | 1.933  | 0.075 | 1.919  | 0.076 | 1.974  | 0.088 | 1.794  | 0.088 | 1.780  | 0.089 |
| $\beta_{11}$  | 1.669  | 0.087 | 1.011  | 0.121 | 2.007  | 0.132 | 1.350  | 0.249 | 1.340  | 0.249 |
| $\beta_{12}$  | 0.019  | 0.035 | 0.019  | 0.035 | 0.091  | 0.042 | 0.091  | 0.042 | 0.091  | 0.042 |
| $\beta_{2}$   | 0.011  | 0.217 | 0.011  | 0.212 | 0.011  | 0.212 | 0.011  | 0.212 | 0.011  | 0.212 |
| $\beta_{21}$  | 1.944  | 0.094 | 1.932  | 0.095 | 1.794  | 0.108 | 1.795  | 0.108 | 1.781  | 0.109 |
| $\beta_{22}$  | 1.676  | 0.106 | 1.024  | 0.135 | 2.006  | 0.158 | 1.355  | 0.180 | 1.332  | 0.265 |
| $\log(\phi)$ | 2.551  | 0.330 | 2.550  | 0.026 | 2.688  | 0.358 | 2.688  | 0.024 | 2.686  | 0.024 |
| AIC        | 1251.116 | 1251.116 | 1242.832 | 1242.832 | 1244.829 | 1244.829 | 1244.829 | 1244.829 |

| Parameter | Est    | SE    | Est    | SE    | Est    | SE    | Est    | SE    | Est    | SE    |
|-----------|--------|-------|--------|-------|--------|-------|--------|-------|--------|-------|
| $\gamma_{11}$ | -4.481 | 0.925 | -4.481 | 0.925 | -4.371 | 0.825 | -4.326 | 0.914 | -4.327 | 0.910 |
| $\gamma_{12}$ | -0.084 | 0.179 | -0.084 | 0.179 | -0.084 | 0.179 | -0.035 | 0.453 | -0.037 | 0.447 |
| $\gamma_{2}$  | 1.944  | 0.094 | 1.932  | 0.095 | 1.794  | 0.108 | 1.795  | 0.108 | 1.781  | 0.109 |
| $\beta_{11}$  | 1.676  | 0.106 | 1.024  | 0.135 | 2.006  | 0.158 | 1.355  | 0.180 | 1.332  | 0.265 |
| $\beta_{12}$  | 0.012  | 0.044 | 0.012  | 0.044 | 0.090  | 0.052 | 0.090  | 0.052 | 0.090  | 0.052 |
| $\beta_{2}$   | 0.011  | 0.217 | 0.011  | 0.212 | 0.011  | 0.212 | 0.011  | 0.212 | 0.011  | 0.212 |
| $\beta_{21}$  | 1.944  | 0.094 | 1.932  | 0.095 | 1.794  | 0.108 | 1.795  | 0.108 | 1.781  | 0.109 |
| $\beta_{22}$  | 1.676  | 0.106 | 1.024  | 0.135 | 2.006  | 0.158 | 1.355  | 0.180 | 1.332  | 0.265 |
| $\log(\phi)$ | 2.551  | 0.330 | 2.550  | 0.026 | 2.688  | 0.358 | 2.688  | 0.024 | 2.686  | 0.024 |
| AIC        | 1236.835 | 1236.835 | 1231.757 | 1231.757 | 1233.743 | 1233.744 | 1233.743 | 1233.744 |

Note: In Model 3, the ZI and MZI models are not equivalent and parameter estimates are not directly comparable across these two model classes. In Models 1 and 2, ZI and MZI models are equivalent with the same parameter interpretations except for $\beta_{11}$ and $\beta_{12}$.

Because Model 2 has an ANOVA structure for $\pi$, the MZINB and ZINB models are equivalent as described in Section 3.3. In particular, for this model, we have

$$
\begin{pmatrix}
\beta_{11} \\
\beta_{12} \\
\beta_{21} \\
\beta_{22}
\end{pmatrix}_{\text{MZINB}} = 
\begin{pmatrix}
\beta_{11} - \log([1 + \exp(\gamma_{11})] \\
\beta_{12} - \log([1 + \exp(\gamma_{12})]) \\
\beta_{21} \\
\beta_{22}
\end{pmatrix}_{\text{ZINB}}.
$$

This relationship is satisfied by the parameter estimates in Table 4. Note that the estimated probability of zero inflation for photoperiod $= 8$ is so small ($\hat{\pi}_{1jk} = \log^{-1}(\hat{\gamma}_{11}) = 0.012$) that the difference in the estimates of $\beta_{11}$ is negligible for the MZINB and ZINB models. In contrast (and as seen in Figure 1), much greater evidence of zero inflation was found for photoperiod $= 16$ ($\hat{\pi}_{2jk} = 0.479$) leading to a much more substantial shift in $\hat{\beta}_{12}$ for the MZINB model relative to the ZINB model. Note also that while MZI and ZI results are quite similar for Model 3, these models are not equivalent because of the presence of a covariate effect in the model for $\pi$. The close similarity of the parameter estimates and AIC statistics for the two model types is simply because the estimate of $\gamma_{2}$ is very near 0.
5.2. Whitefly data

van Iersel et al. [23] report the results of a horticultural experiment to examine the effect of six methods of applying pesticide to control silverleaf whiteflies on greenhouse-raised poinsettia. Fifty-four plants were used in a randomized complete block design where experimental units containing three plants were randomly assigned to each treatment in three complete blocks. Each week a collection of whiteflies was placed on each plant (confined inside a cage clipped to a leaf of the plant) and then the number of surviving whiteflies was measured two days later. It was of interest to compare the survival probabilities across plants treated with different methods of pesticide application. Despite the large frequency of zero-valued responses in these data, the ZIB model is not the best framework for analyzing them because its estimates of $\beta$ quantify treatment effects on the conditional probability of survival given that the response comes from the binomial component of the mixture. In the context of the problem, one might interpret this as the conditional probability of survival given that the treatment is not certain to kill all whiteflies. In contrast, the $\beta$ coefficients in an MZIB model provide a more straightforward and desirable interpretation as direct treatment effects on the log odds of survival.

Although the original experimental design included repeated measures over time, for simplicity we analyze only the data from week 6, the middle week of the study. Extensions of MZI models to the longitudinal/clustered data setting will be considered elsewhere. The six treatments used in the experiment included four methods of sub-irrigated pesticide following 4, 2, 1 and 0 days without water (treatments 1–4, respectively), a treatment involving pesticide application via hand watering (treatment 5), and a control treatment in which no pesticide was used (treatment 6).

Following van Iersel et al. [23], we summed the responses for each trio of plants and analyzed a single bounded count per experimental unit. Let $\mu_{ij}$ represent the probability of survival for the $n_{ij}$ insects from the experimental unit assigned to the $j$th treatment in the $i$th block, which was modeled by $\text{logit}(\mu_{ij}) = \beta_0 + \beta_{1i} + \beta_{2j}$ where $\beta_0$ is the intercept, $\beta_{1i}, i = 1, 2, 3,$ is the effect of the $i$th block, and $\beta_{2j}, j = 1, 2, \ldots, 6,$ is the effect of the $j$th treatment. Two models with this specification for $\mu_{ij}$ were fit to the data, differing in their specification of the mixing probability: (a) $\text{logit}(\pi_{ij}) = \gamma_0$, and (b) $\text{logit}(\pi_{ij}) = \gamma_0 + \gamma_{1j}$. Models with both treatment and block effects on the mixing probability were not considered due to limited degrees of freedom and the sparse occurrence of zeros across blocks. A Wald test of $H_0 : \gamma_{11} = \cdots = \gamma_{16} = 0$ strongly favors model (b) ($P(\chi_5^2 > 60.27) \approx 0$).

Table 5 shows the estimated marginal probabilities for each block and treatment combination. Based on model (b), there are several sets of contrasts among the six treatment effects that one might choose to test. We tested $H_{0A} : \frac{1}{5} \sum_{j=1}^{5} \beta_{2j} = \beta_{26}$, $H_{0B} : \frac{1}{4} \sum_{j=1}^{4} \beta_{2j} - \beta_{25}$, and $H_{0C} : \beta_{21} = \cdots = \beta_{24}$ to compare active treatments to the control ($P(\chi_1^2 > 30.61) \approx 0$), subirrigation to hand watering ($P(\chi_1^2 > 8.29) = 0.004$), and the four levels of restricting water prior to subirrigation-based pesticide application ($P(\chi_5^2 > 3.71) = 0.294$), respectively. These hypotheses have straightforward interpretations, and we conclude that the odds of survival are significantly different (lower) when pesticide is applied, significantly different (lower) when subirrigation is used for pesticide delivery, and that the survival odds do not differ significantly depending on the length of water restriction prior to subirrigation with pesticide.
Table 5. $\hat{\mu}_{ij}$ for each block and treatment combination.

| Block | Treatment |
|-------|-----------|
| 1     | 0.011     | 0.009 | 0.000 | 0.039 | 0.160 | 0.742 |
| 2     | 0.045     | 0.040 | 0.000 | 0.148 | 0.453 | 0.926 |
| 3     | 0.021     | 0.018 | 0.000 | 0.073 | 0.273 | 0.850 |

An advantage of the MZIB model compared to the ZIB model is that it estimates survival odds ratios for distinct treatments that do not depend on block. Given the additive structure of the model, this feature is desirable. For example, using the MZIB model, we may estimate the ratio of survival odds for treatments 2 and 5 as $\exp(\hat{\beta}_{25} - \hat{\beta}_{22}) = 20.08$. In contrast, a ZIB model with linear predictors of the same form as our chosen MZIB model yields an estimated survival odds ratio for treatments 2 and 5 of $(\hat{\mu}_{i5}/1 - \hat{\mu}_{i5})(1 - \hat{\mu}_{i2}/\hat{\mu}_{i2}) = ((1 - \hat{\pi}_{i5})\hat{\lambda}_{i5}/(1 - (1 - \hat{\pi}_{i5})\hat{\lambda}_{i5}))((1 - (1 - \hat{\pi}_{i2})\hat{\lambda}_{i2})/(1 - \hat{\pi}_{i2}\hat{\lambda}_{i2}))$. Note that this quantity depends upon $i$; so the model, despite the additive structure of its linear predictors, essentially induces an interaction between treatments and blocks. The ratio of survival odds for treatments 2 and 5 for blocks 1, 2, and 3 for a ZIB model are 27.29, 19.88, and 25.53, respectively. Note that the average of the odd ratios from the ZIB model is not especially close to the odds ratio found from the MZIB model, which highlights the fact that these models are not merely reparameterizations of each other and should not be expected to agree closely.

6. Conclusion

ZI regression is a useful and appealing framework for analyzing count data with EZs, but it is awkward to obtain inferences on the marginal mean from such models. Often there is no scale on which covariate effects are constant, precluding succinct statements about the relationship between the mean response and a covariate of interest. In many contexts an MZI model will prove more convenient and interpretable. Such models retain the basic mixture structure of ZI regression, but provide direct estimates of marginal mean effects. In addition to providing direct estimates of interpretable quantities, MZI models allow more straightforward inferences, avoiding the need for delta method confidence intervals and tests. This is not only an advantage of convenience but, as shown in our simulation results, produces more accurate coverage and error rates. As the simulations show, the coverage probabilities for confidence intervals derived from the ZI models decrease as the sample size increases while the coverage of the intervals formed from the MZI models remain close to the desired level, which is consistent with the results from Long et al. [14]. When testing for a treatment effect on the marginal mean, only those using the results of the MZI model have the correct size and are properly targeted at the hypothesis of interest; therefore the ZI model should not be used for inference on the marginal mean of the response except in special cases when its parameters have direct marginal interpretations.

While in this paper we have focused on models for independent data (e.g. a cross-sectional study design), extensions to accommodate longitudinal/clustered data are desirable. For this purpose, a generalization of the MZI regression model class that is natural and tractable incorporates random cluster-specific random effects. Discussion of such an
extended mixed-effect, MZI class will appear elsewhere. While this paper gives details on several special cases of MZI models (MZIP, MZIB, MZINB), these are other examples we have not discussed. For instance, an MZI tobit model may be useful for semicontinuous data and can be seen as an alternative to the ZIT model [16].

Acknowledgments

We are grateful to three anonymous referees for constructive comments and for drawing our attention to related research of which we were unaware.

Disclosure statement

No potential conflict of interest was reported by the authors.

References

[1] J. M. Albert, W. Wang, and S. Nelson, *Estimating overall exposure effects for zero-inflated regression models with application to dental caries*, Statist. Methods Med. Res. 23 (2014), pp. 257–278.
[2] J. Catty, P. Lissouba, S. White, T. Becker, R. E. Drake, A. Fioritti, M. Knapp, C. Lauber, W. Rössler, T. Tomov, J. van Busschbach, D. Wiersma, and T. Burns, *Predictors of employment for people with severe mental illness: Results of an international six-centre randomised controlled trial*, Br. J. Psychiatry 192 (2008), pp. 224–231.
[3] B. H. Cohen, *Recall of categorized words lists*, J. Exp. Psychol. 66 (1963), pp. 227–234.
[4] M. Crowder, *Beta-binomial ANOVA for proportions*, J. R. Statist. Soc.: Ser. A (Appl. Stat.) 27 (1977), pp. 34–37.
[5] W. Greene, *Accounting for excess zeros and sample selection in Poisson and negative binomial regression models*, Report EC-94-10, Department of Economics, New York University, New York, 1994.
[6] D. B. Hall, *Zero-inflated Poisson and binomial regression with random effects: a case study*, Biometrics 56 (2000), pp. 1030–1039.
[7] P. J. Heagerty, *Marginally specified logistic-normal models for longitudinal binary data*, Biometrics 55 (1999), pp. 688–698.
[8] D. C. Heilbron, *Zero-altered and other regression models for count data with added zeros*, Biom. J. 36 (1994), pp. 531–547.
[9] S. Jackman, *pscl: Classes and Methods for R Developed in the Political Science Computational Laboratory, Stanford University*. Department of Political Science, Stanford University, Stanford, CA. R package version 1.4.92015.
[10] B. Jorgensen, *Exponential dispersion models*, J. R. Statist. Soc. Ser. B (Statist. Methodol.) (1987), pp. 127–162.
[11] W. Kassahun, T. Neyens, G. Molenberghs, C. Faes, and G. Verbeke, *Marginalized multilevel hurdle and zero-inflated models for overdispersed and correlated count data with excess zeros*, Stat. Med. 33 (2014), pp. 4402–4419.
[12] D. Lambert, *Zero-inflated Poisson regression, with an application to defects in manufacturing*, Technometrics 34 (1992), pp. 1–14.
[13] J. Lawless, *Negative binomial and mixed Poisson regression*, Can. J. Stat. 15 (1987), pp. 209–225.
[14] D. L. Long, J. S. Preisser, A. H. Herring, and C. E. Golin, *A marginalized zero-inflated Poisson regression model with overall exposure effects*, Stat. Med. 33 (2014), pp. 5151–5165.
[15] M. N. Maunder and A. E. Punt, *Standardizing catch and effort data: A review of recent approaches*, Fish. Res. 70 (2004), pp. 141–159.
[16] L. H. Moulton, F. C. Curriero, and P. F. Barroso, *Mixture models for quantitative HIV RNA data*, Statist. Methods Med. Res. 11 (2002), pp. 317–325.
Appendix

Denote the link functions and linear predictors of the MZI regression model as

\[ g(\pi_i) = \delta_i = g'(y), \]
\[ h(\mu_i) = \eta_i = x'\beta, \]

and let \( d(\theta_i, \phi) = -b(\theta_i)/a(\phi) + c(0, \phi) \) (cf. Equation (1)). It follows that

\[ P(Y_i = 0) = \frac{\pi_i}{1 - \pi_i} + \exp\{d(\theta_i, \phi)\} \]
\[ P(Y_i \neq 0) = \frac{\pi_i}{a(\phi)} + c(y_i, \phi), \]

and the loglikelihood of the regression model can be written as

\[ \ell(y, \beta, \phi; y) = \sum_{i=1}^{n} \left[ \log(1 - \pi_i) + u_i \log P(Y_i = 0) + (1 - u_i)P(Y_i \neq 0) \right] \]

The information matrix is given by

\[ I(y, \beta) = -E \begin{pmatrix} \frac{\partial^2 \ell(y, \beta, \phi; y)}{\partial \phi \partial \beta'} & \frac{\partial^2 \ell(y, \beta, \phi; y)}{\partial \beta \partial \phi'} \\ \frac{\partial^2 \ell(y, \beta, \phi; y)}{\partial \beta \partial \phi'} & \frac{\partial^2 \ell(y, \beta, \phi; y)}{\partial \phi \partial \phi'} \end{pmatrix}, \]

where

\[ \frac{\partial^2 \ell(y, \beta, \phi)}{\partial \Psi \partial v'} = \sum_{i=1}^{n} \left\{ u_i \left[ \frac{\partial^2 P(Y_i = 0)}{\partial \Psi \partial v'} P(Y_i = 0) - \frac{\partial P(Y_i = 0)}{\partial \Psi} \frac{\partial P(Y_i = 0)}{\partial v'} \right] P(Y_i = 0)^{-2} \right. \]
\[ + \left. (1 - u_i) \frac{\partial^2 P(Y_i \neq 0)}{\partial \Psi \partial v'} + \frac{\partial^2 \log(1 - \pi_i)}{\partial \Psi \partial v'} \right\} , \]

for \( \Psi, v \in \{y, \beta, \phi\} \).
Let \( r_{\mu i} = (\partial \theta_i / \partial \mu_j) (\partial \pi_i / \partial \delta_j), r_{\mu i} = (\partial \theta_i / \partial \mu_j) (\partial \mu_i / \partial \eta_i), t_i = (\partial \theta_i / \partial \pi_i) (\partial \theta_i / \partial \mu_i), \) and \( s_i = (\partial \mu_i / \partial \eta_i) (\partial \pi_i / \partial \delta_i); \) note that if the canonical links are used, \( r_{\pi i} = r_{\mu i} = s_i t_i = 1 \). The partial derivatives needed to compute the information matrix are as follows:

\[
\frac{\partial P(Y_i = 0)}{\partial \beta} = -a(\phi)^{-1} \exp[d(\theta_i, \phi)] \frac{\partial b(\theta_i)}{\partial \theta_i} r_{\mu i} x_i'
\]

\[
\frac{\partial P(Y_i = 0)}{\partial \gamma} = \left[ (1 - \pi_i)^{-1} - \exp[d(\theta_i, \phi)] a(\phi)^{-1} \frac{\partial b(\theta_i)}{\partial \theta_i} \frac{\partial \pi_i}{\partial \delta_i} \right] \frac{\partial \pi_i}{\partial \delta_i} g_i'
\]

\[
\frac{\partial P(Y_i = 0)}{\partial \phi} = - \exp[d(\theta_i, \phi)] \left[ a'(\phi) b(\theta_i) + c'(0, \phi) \right]
\]

\[
\frac{\partial^2 P(Y_i = 0)}{\partial \beta \partial \beta'} = a(\phi)^{-1} \exp[d(\theta_i, \phi)] \left[ \left( \frac{\partial b(\theta_i)}{\partial \theta_i} \right)^2 - \frac{\partial^2 b(\theta_i)}{\partial \theta_i^2} \right] \frac{\partial \pi_i}{\partial \delta_i} x_i'
\]

\[
\frac{\partial^2 P(Y_i = 0)}{\partial \gamma \partial \gamma'} = \left[ 2(1 - \pi_i)^{-2} - a(\phi)^{-1} \exp[d(\theta_i, \phi)] \right] \left[ -a(\phi)^{-1} \left( \frac{\partial b(\theta_i)}{\partial \theta_i} \frac{\partial \pi_i}{\partial \delta_i} \right)^2 + \frac{\partial^2 b(\theta_i)}{\partial \theta_i^2} \right] \frac{\partial \pi_i}{\partial \delta_i} g_i'
\]

\[
\frac{\partial^2 P(Y_i = 0)}{\partial \phi \partial \phi'} = a(\phi)^{-1} \exp[d(\theta_i, \phi)] \left[ a'(\phi) b(\theta_i) + c'(0, \phi) \right]
\]

\[
\frac{\partial^2 P(Y_i = 0)}{\partial \pi_i} = \exp[d(\theta_i, \phi)] \left[ \left( \frac{\partial b(\theta_i)}{\partial \theta_i} \right)^2 - \frac{\partial^2 b(\theta_i)}{\partial \theta_i^2} \right] t_i - \frac{\partial b(\theta_i)}{\partial \theta_i} \frac{\partial \pi_i}{\partial \delta_i} \frac{\partial^2 \pi_i}{\partial \delta_i^2} s_i g_i x_i'
\]

\[
\frac{\partial^2 P(Y_i = 0)}{\partial \pi_i^2} = a(\phi)^{-1} \exp[d(\theta_i, \phi)] \left[ a'(\phi) b(\theta_i) - a''(\phi) b(\theta_i) + c'(0, \phi) \right]
\]

\[
\frac{\partial^2 P(Y_i = 0)}{\partial \beta \partial \gamma} = a(\phi)^{-1} \exp[d(\theta_i, \phi)] \left[ a'(\phi) b(\theta_i) + c'(0, \phi) \right] \frac{\partial b(\theta_i)}{\partial \theta_i} r_{\pi i} g_i'
\]

\[
\frac{\partial^2 P(Y_i = 0)}{\partial \beta \partial \phi} = a(\phi)^{-1} \exp[d(\theta_i, \phi)] \left[ a'(\phi) b(\theta_i) - a''(\phi) b(\theta_i) + c'(0, \phi) \right] \frac{\partial b(\theta_i)}{\partial \theta_i} r_{\pi i} x_i'
\]

\[
\frac{\partial^2 P(Y_i = 0)}{\partial \gamma \partial \phi} = a(\phi)^{-1} \exp[d(\theta_i, \phi)] \left[ a'(\phi) b(\theta_i) + c'(0, \phi) \right] \frac{\partial b(\theta_i)}{\partial \theta_i} t_i g_i x_i'
\]

\[
\frac{\partial^2 P(Y_i = 0)}{\partial \phi \partial \beta} = a(\phi)^{-1} \exp[d(\theta_i, \phi)] \left[ a'(\phi) b(\theta_i) - a''(\phi) b(\theta_i) + c'(0, \phi) \right] \frac{\partial b(\theta_i)}{\partial \theta_i} r_{\pi i} x_i'
\]

\[
\frac{\partial^2 P(Y_i = 0)}{\partial \phi \partial \gamma} = a(\phi)^{-1} \exp[d(\theta_i, \phi)] \left[ a'(\phi) b(\theta_i) + c'(0, \phi) \right] \frac{\partial b(\theta_i)}{\partial \theta_i} t_i g_i x_i'
\]

\[
\frac{\partial^2 P(Y_i = 0)}{\partial \phi \partial \phi} = \left[ E(Y_i) - b(\theta_i) \right] a'(\phi) - a''(\phi) + c''(0, \phi)
\]

\[
\frac{\partial^2 P(Y_i = 0)}{\partial \phi \beta} = 0
\]

\[
\frac{\partial^2 P(Y_i = 0)}{\partial \phi \gamma} = 0
\]