



Published in final edited form as:

Biometrics. 2012 September ; 68(3): 975–982. doi:10.1111/j.1541-0420.2011.01737.x.

On the Efficiency of Score Tests for Homogeneity in Two-Component Parametric Models for Discrete Data

David Todem^{1,*}, Wei-Wen Hsu², and Kyung Mann Kim³

¹Department of Epidemiology and Biostatistics, Michigan State University, B601 West Fee Hall, East Lansing, Michigan 48824, U.S.A

²Department of Statistics and Probability, Michigan State University, A413 Wells Hall East Lansing, Michigan 48824, U.S.A

³Departments of Statistics and Biostatistics & Medical Informatics, University of Wisconsin-Madison, 600 Highland Ave., Madison, Wisconsin 53792, U.S.A

Summary

In many applications of two-component mixture models for discrete data such as zero-inflated models, it is often of interest to conduct inferences for the mixing weights. Score tests derived from the marginal model that allows for negative mixing weights have been particularly useful for this purpose. But the existing testing procedures often rely on restrictive assumptions such as the constancy of the mixing weights and typically ignore the structural constraints of the marginal model. In this article, we develop a score test of homogeneity that overcomes the limitations of existing procedures. The technique is based on a decomposition of the mixing weights into terms that have an obvious statistical interpretation. We exploit this decomposition to lay the foundation of the test. Simulation results show that the proposed covariate-adjusted test statistic can greatly improve the efficiency over test statistics based on constant mixing weights. A real-life example in dental caries research is used to illustrate the methodology.

Keywords

Binary regression; Finite dimensional parameter; Goodness-of-fit test; Identifiable models; Likelihood function; Negative binomial models; Overdispersion; Score function

1. Introduction

In discrete data, when an observation exhibits an excessive frequency, it is a common practice to use two-component mixture models that combine an atom distribution with a proper distribution. Well known members of this class of models include the so-called zero-inflated models for independent count data first studied by Mullahy (1986), Farewell and Sprott (1988), and Lambert (1992) and their extensions to correlated count data by Hall (2000). Zero-inflated models have been extensively studied in statistical research and applied to data from various disciplines including agriculture, econometrics, medicine, engineering, sociology, and behavioral sciences. Ridout, Demetrio, and Hinde (1998) provide an extensive review of this literature.

Two-component mixture models provide an interesting parametric framework to accommodate heterogeneity in a population. A prevailing concern, however, is whether the inherent heterogeneity is consistent with observed data. Many authors have examined this important issue using two-sided score tests. In a *Biometrics* paper, van den Broek (1995, pp. 738–748) has developed a two-sided score test for homogeneity in zero-inflated models for count data. Others have extended this test to situations where the nondegenerate distribution is a member of the exponential family (Deng and Paul, 2000) and to clustered data (Xiang et al., 2006). The original test by van den Broek (1995) and its extensions were constructed under alternatives to homogeneity that allow for negative mixing weights. Specifically, they were derived under the marginal mixture model that ignores its hierarchical representation. Although this approach uses score functions that are well defined at the null value, it is often unclear how constraints generated from this marginal model are accommodated under the alternative. Most importantly, existing testing procedures typically do not fully exploit the general structure of the mixing weights. A common restriction is that constant mixing weights are often assumed under heterogeneity. This is an important limitation as covariate-dependent mixing weights are typical in real applications of these models. From a technical perspective, tests that rely on constant mixing weights may lack power to detect heterogeneity if both inflation and deflation are present in the population.

The problem of inflation and deflation, typically at zero, arises quite naturally in various applications. A good example is provided by a very unique survey designed to collect oral health information on low-income African American children (0–5 years) and their main caregivers (14+ years), living in the city of Detroit (see Tellez et al., 2006). This article aimed at promoting oral health and reducing its disparities within this community through the understanding of determinants of dental caries. Dental caries were measured using scores which represent the cumulative severity of the disease for each surveyed participants. These scores have well-documented shortcomings regarding their ability to describe dental caries experiences, but they continue to be instrumental in evaluating and comparing the risks of dental caries across population groups (Lewsey and Thomson, 2004). Most importantly, they remain popular in dental caries research for their ability to conduct historical comparisons in population-based studies. Homogeneous models when applied to these data typically reveal an inflation of zeros (few children with no dental caries predicted than observed) for younger children and deflation of zeros (more children with no dental caries predicted than observed) for older children. For these data, an analysis based on constant mixing weights under the alternative may fail to detect heterogeneity in this population.

Health services research is another area of applications where inflation and deflation are often encountered. For example, Hur et al. (2002) presented a study on patients undergoing a partial colectomy operation from 123 Veterans Affairs (VA) medical centers in the National Surgical Quality Improvement Program. The primary goal of this program was to improve surgical care for veteran patients by developing performance measures for surgery in the VA system. One of the primary study outcomes was the number of postoperative complications. It was reported that patients that are relatively healthier before the surgical operation, as measured by the classification from the American Society of Anesthesiologists, will likely have no complications, although their sicker counterparts will be associated with higher numbers of complications. For such data, a homogenous model will exhibit an inflation at zero for relatively healthier patients and a deflation at zero for the sickest patients.

The examples above then draw into question the efficiency of testing procedures which assume constancy of the mixing weights. In this article, we suggest an extension of the score test of homogeneity to covariates. A recent illustration of this approach for zero-inflated models is given by Jansakul and Hinde (2009). These authors related the mixing weights to covariates using an identity link function. A limitation, however, is that the identity link

function is seldom used in practice for this class of models. Moreover, the structure of the marginal model is not properly integrated into the testing procedure. As a solution, we formulate and develop a score test of homogeneity based on an intuitive approach that accommodates the features of the marginal model. Specifically, we embed the structural constraints of the marginal model into the procedure. The technique is based on decomposing the mixing weights into terms that have an obvious statistical interpretation. One appealing feature of this decomposition is that it naturally incorporates covariates into the mixing weights. We exploit this decomposition to lay the foundation of the test. We show that this is a natural strategy to adopt when testing is carried out under the marginal model that ignores its hierarchical representation. The proposed test in its current formulation can be naturally used to study heterogeneity at any point of the population and for any nondegenerate parametric distributions that are not necessarily members of the exponential family.

The remainder of this article is organized as follows. In Section 2, we specify the homogeneity hypothesis using composite link functions and develop a score test that accommodates the features of the marginal model. In Section 3, the empirical performance of the proposed test is studied and its real-life applications are illustrated using dental caries counts in young children. Some remaining issues are discussed in Section 4. Additional technical details are contained in a Web supplementary materials file.

2. The Method

2.1 Hypothesis Formulation and Reparameterization

We consider a sample of n independent realizations $\{y_i\}_{i=1}^n$, of random variables $\{Y_i\}_{i=1}^n$ each drawn from a mixture of an atom distribution at a known point y^* and an unknown nondegenerate discrete distribution G governed by a finite dimensional parameter vector θ ,

$$\Pr(Y_i=y_i)= \begin{cases} \omega_i+(1-\omega_i)g_i(y^*) & \text{if } y_i=y^* \\ (1-\omega_i)g_i(y_i) & \text{if } y_i \neq y^*. \end{cases} \quad (1)$$

Here ω_i is an unknown mixing weight and $g_i(\cdot)$ is a probability mass associated with G . The mixture model in equation(1) allows for two types of representations, depending on the support of ω_i . Under its hierarchical representation, the mixing weight is a probability mass, requiring the distributional constraints $0 \leq \omega_i \leq 1, i = 1, \dots, n$. Under the marginal representation, however, the only distributional constraints are that $0 \leq \Pr(Y_i = y_i) \leq 1$, for all $i = 1, \dots, n$, under the mixture model equation (1), which result in constraints on mixing weights,

$$-g_i(y^*)/(1-g_i(y^*)) \leq \omega_i \leq 1, \quad i=1, \dots, n. \quad (2)$$

The marginal model can then be used to accommodate both inflation (positive ω_i) and deflation (negative ω_i) at y^* , a property not shared by its hierarchical counterpart. The marginal model, however, does not allow any hierarchical interpretation of the mixture model when the mixing weight is negative. For further discussions on the interpretation of these models when $y^* = 0$, see Heilbron (1994).

Suppose we are interested in evaluating the hypothesis of zero mixing weights. Under the marginal model that allows for negative mixing weights, one is typically interested in the two-sided hypotheses,

$$H_0: \omega_i = 0, \text{ for all } i \text{ vs. } H_1: \omega_i \neq 0, \text{ for some } i,$$

, where ω_i satisfies constraints in equation (2). To test these hypotheses, we consider a suitable transformation of ω_i that incorporates the constraints in equation (2) into the testing scheme. A natural transformation in light of these constraints is given by,

$$\omega_i = \frac{\pi_i - g_i(y^*)}{1 - g_i(y^*)}, \quad 0 \leq \pi_i \leq 1. \quad (3)$$

Evidently, the lower and upper bounds of ω_i in equation (2) are attained at points 0 and 1 of π_i , respectively. The quantity π_i has an obvious statistical interpretation. To see this, consider a binary version of Y_i denoted $\delta(Y_i)$ taking value 1 if $Y_i = y^*$ and 0 otherwise. The term π_i is simply the first moment $E\{\delta(Y_i)\} = \Pr(Y_i = y^*)$ under the marginal mixture model equation (1), which can be estimated very well if many independent copies $\delta(y_i)$ of $\delta(Y_i)$ are available. The transformation as specified in equation (3) then arises from equating the probability mass at y^* as predicted by the mixture model in equation (1) to the first moment of $\delta(Y_i)$. That is, $E\{\delta(Y_i)\} = \omega_i + (1 - \omega_i)g_i(y^*)$, where $g_i(y^*)$ is the probability $\Pr(Y_i = y^*)$ under the distribution G .

Based on the parameterization equation (3), we formally state the hypotheses as,

$$H_0: \pi_i = g_i(y^*), \text{ for all } i \text{ vs. } H_1: \pi_i \neq g_i(y^*), \text{ for some } i.$$

We exploit this reformulation to lay the foundation of the test. We consider a suitable parameterization of π_i and $g_i(y^*)$ that reduces the homogeneity hypothesis above to a problem involving a small number of parameters. For this, we assume that μ_i the mean of Y_i under the nondegenerate parametric distribution G is finite. We express the probability mass at y^* with respect to the distribution G as a function of μ_i , that is $g_i(y^*) = g(\mu_i)$, where $g(\cdot)$ is a bounded function between 0 and 1, which depends on θ and y^* . If the data can be arranged to form strata with a finite, possibly small, number of terms π_i and μ_i , the homogeneity hypothesis naturally reduces to a problem involving a small number of parameters. This stratified approach has its limitations when the number of strata or more generally when the number of covariates is increasing, however. Alternatively, regression techniques relating μ_i and π_i to some observed covariates can be used to address the curse of dimensionality. For this, consider two-column vector covariates x_i and z_i , of dimensions p and q , respectively, that are also observed alongside the responses y_i , $i = 1, 2, \dots, n$. We relate μ_i to covariates x_i as follows, $h(\mu_i) = x_i' \beta$, where β is a parameter vector of dimension $p \times 1$ and a subset of θ and $h(\cdot)$ is a monotone, differentiable, and invertible function. Likewise, we relate π_i to covariates z_i as follows, $\pi_i = f(z_i' \gamma)$, where γ is a parameter vector of dimension $q \times 1$ and $f(\cdot)$ is a differentiable function. We find a suitable function $f(\cdot)$ that translates the homogeneity hypothesis into an equality involving only parameters β and γ . This is summarized in the proposition below.

Proposition—Assuming the model $h(\mu_i) = x_i' \beta$ under the nondegenerate distribution G , a suitable choice for $f(\cdot)$ is given by $f(z_i' \gamma) = \tilde{g} \circ h^{-1}(z_i' \gamma)$, where \circ represents the composite function operator, and the homogeneity (null) hypothesis can be reformulated as,

$$H_0: \tilde{g} \circ h^{-1}(z'_i \gamma) = \tilde{g} \circ h^{-1}(x'_i \beta), \text{ for all } i, \quad (4)$$

which further reduces to a linear contrast involving only parameters β and γ .

The proof of this proposition is provided in the Web supplementary file. The function $f(\cdot)$ falls in the class of composite link functions which are not always invertible (Thompson and Baker, 1981). Below, we give an explicit expression of this function for commonly used nondegenerate parametric distributions G and link functions $h(\cdot)$. Specific results when $y^* = 0$ are given in Table 1.

- i. The two-component model when G is a Poisson distribution with mean μ_i . Here $\theta = \beta$ and $g(\tilde{\mu}_i) = \exp\{-\mu_i\} \{\mu_i\}^{y^*} \Gamma^{-1}(y^* + 1)$. Assuming the log link function $\log\{\mu_i\} = x'_i \beta$, a natural parameterization relating π_i to $z'_i \gamma$ is,

$$\pi_i = \exp\{-\exp\{z'_i \gamma\}\} \{\exp\{z'_i \gamma\}\}^{y^*} \Gamma^{-1}(y^* + 1).$$

- ii. The two-component model when G is a binomial distribution with success probability μ_i and planned number of trials m_i . Here $\theta = \beta$ and

$$\tilde{g}(\mu_i) = m_i C_{y^*} \mu_i^{y^*} \{1 - \mu_i\}^{m_i - y^*} \text{ with } m_i C_{y^*} = \frac{\Gamma(m_i + 1)}{\Gamma(y^* + 1) \Gamma(m_i - y^* + 1)}. \text{ Assuming the logit link function } \log\{\mu_i / (1 - \mu_i)\} = x'_i \beta, \text{ a natural parameterization for } \pi_i \text{ is then,}$$

$$\pi_i = m_i C_{y^*} \{1 + \exp\{-z'_i \gamma\}\}^{-y^*} \{1 + \exp\{z'_i \gamma\}\}^{y^* - m_i}.$$

- iii. The two-component model when G is a negative binomial distribution with mean μ_i and overdispersion parameter κ . Here $\theta = \{\beta, \kappa\}$ and

$$\tilde{g}(\mu_i) = \frac{\Gamma(y^* + \kappa^{-1})}{\Gamma(\kappa^{-1}) \Gamma(y^* + 1)} \{1 + \kappa \mu_i\}^{-1/\kappa} \{1 + \kappa^{-1} \mu_i^{-1}\}^{-y^*}. \text{ For the log link function } \log\{\mu_i\} = x'_i \beta, \text{ a natural parameterization for } \pi_i \text{ is then,}$$

$$\pi_i = \frac{\Gamma(y^* + \kappa^{-1})}{\Gamma(\kappa^{-1}) \Gamma(y^* + 1)} \{1 + \kappa \exp\{z'_i \gamma\}\}^{-1/\kappa} \{1 + \kappa^{-1} \exp\{-z'_i \gamma\}\}^{-y^*}.$$

2.2 A Covariate-Adjusted Score Test for Homogeneity

Without any loss of generality, we will develop the test for the case where $x_i = z_i$, although our method generalizes quite naturally for any x_i and z_i . Under this assumption, our homogeneity hypothesis then becomes, $H_0 : \beta = \gamma$. Let C be a selector matrix such that $C\theta = \beta$, the homogeneity hypothesis is then equivalent to, $H_0 : C\theta = \gamma$. Under the alternative model, the log-likelihood function, given observations $\{y_i\}_{i=1}^n$, as a function of θ and γ is given by,

$$\ell(\theta, \gamma) = \sum_{i=1}^n \left(\delta[y_i] \log[\pi_i] + [1 - \delta\{y_i\}] \log \left[\frac{\{1 - \pi_i\} g_i\{y_i\}}{1 - g_i\{y^*\}} \right] \right).$$

We assume the following change of variables, $\alpha = C\theta - \gamma$, with $\alpha = 0$ under the null hypothesis. We denote by $u(\theta, \alpha)$ the score function $\ell(\theta, C\theta - \alpha) / (\theta, \alpha)$. To construct the score test statistic and derive its limiting null distribution, we set the following conditions for the arguments in the sequel to be valid.

- C1** The support set of (θ, α) is compact.

- C2** Assume a nonsingular matrix $\mathcal{I}(\theta, \alpha)$ such that $n^{-1} u(\theta, \alpha)' (\theta, \alpha) = \mathcal{I}(\theta, \alpha) + o_p(1)$.
- C3** For all $a > 0$, $\sup_{\|\zeta\| \leq a} \{n^{-1/2}[u((\theta, \alpha) + n^{-1/2}\zeta) - u(\theta, \alpha)] + \mathcal{I}(\theta, \alpha)\zeta\} = o_p(1)$, where $o_p(1)$ represents the convergence in probability as $n \rightarrow \infty$.

Under Conditions C1–C3, standard asymptotic results give, $n^{-1/2} u(\theta, \alpha) \rightarrow_d \mathcal{N}(0, \mathcal{I}(\theta, \alpha))$, where \rightarrow_d represents the convergence in distribution as $n \rightarrow \infty$ (see Cox and Hinkley, 1974, pp. 311–343, for this large-sample size result). Assume the following decomposition,

$u(\theta, \alpha) = (u'_\theta(\theta, \alpha), u'_\alpha(\theta, \alpha))'$ where the respective entries are the first-order derivatives of the log-likelihood function with respect to θ and α . Let $\mathcal{I}_{\theta\theta}(\theta, \alpha)$, $\mathcal{I}_{\theta\alpha}(\theta, \alpha)$, and $\mathcal{I}_{\alpha\alpha}(\theta, \alpha)$ be the corresponding blocks in $\mathcal{I}(\theta, \alpha)$. The main building block of the test statistic is given by,

$$u_\alpha(\theta, \alpha) = \sum_{i=1}^n \left\{ \frac{\delta(y_i) - \pi_i}{\pi_i(1 - \pi_i)} \frac{\partial \pi_i}{\partial \alpha} \right\}.$$

Note that π_i in this expression depends on both θ and α because of the change of variables. The other building block $\mathcal{I}(\theta, \alpha)$ is in general tedious to compute. In the Web supplementary file, we give details of these calculations when the component G of the mixture model is a Poisson or a binomial distribution (two well known exponential family members) and assuming $y^* = 0$.

Let $b(\theta)$ define the score function with respect to θ under the null distribution G . We assume the existence of a root- n consistent estimator $\hat{\theta}$ of θ^* , the true value of θ under the null distribution G , such that $n^{1/2}(\hat{\theta} - \theta^*) = n^{-1/2} \mathcal{J}_{\theta\theta}^{-1}(\theta^*, 0) b(\theta^*) + o_p(1)$. The core of the score statistic to evaluate the hypothesis $\alpha = 0$ is given by $\hat{u}_\alpha = u_\alpha(\hat{\theta}, 0)$. In the Web supplementary file, we give an explicit form for this quantity when the component G is a Poisson or a binomial distribution.

The asymptotic distribution of \hat{u}_α can be derived by applying a standard Taylor series expansion coupled with the law of large numbers,

$$n^{-1/2} \hat{u}_\alpha = n^{-1/2} \sum_{i=1}^n \{u_{i,\alpha}(\theta^*, 0) - \mathcal{J}'_{\theta\alpha}(\theta^*, 0) \mathcal{J}_{\theta\theta}^{-1}(\theta^*, 0) b_i(\theta^*)\} + o_p(1),$$

where $u_{i,\alpha}(\theta^*, 0)$ and $b_i(\theta^*)$ are, respectively, the random contributions of subject i to the score functions $u_\alpha(\theta^*, 0)$ and $b(\theta^*)$. The right-hand side of this equation consists of sums of independent random vectors $v_i(\theta^*) = u_{i,\alpha}(\theta^*, 0) - \mathcal{J}'_{\theta\alpha}(\theta^*, 0) \mathcal{J}_{\theta\theta}^{-1}(\theta^*, 0) b_i(\theta^*)$, $i = 1, \dots, n$, for which the central limit theorem applies. That is, $n^{-1/2} \hat{u}_\alpha \rightarrow_d \mathcal{N}(0, \Lambda(\theta^*))$, where $\Lambda(\theta^*) = \lim_{n \rightarrow \infty} n^{-1} \sum_{i=1}^n E(v_i^{\otimes 2}(\theta^*))$, with $v_i^{\otimes 2}(\theta^*) = v_i(\theta^*) v_i'(\theta^*)$. A straightforward calculation gives $\Lambda(\theta^*) = \mathcal{J}_{\alpha\alpha}(\theta^*, 0) - \mathcal{J}'_{\theta\alpha}(\theta^*, 0) \mathcal{J}_{\theta\theta}^{-1}(\theta^*, 0) \mathcal{J}_{\theta\alpha}(\theta^*, 0)$ for which a consistent estimator is obtained by replacing θ^* by its estimator $\hat{\theta}$. We then construct a score test statistic as,

$$s_n = n^{-1} \hat{u}'_\alpha \Lambda^{-1}(\hat{\theta}) \hat{u}_\alpha \rightarrow_d \chi_r^2 \text{ under } H_0 \text{ as } n \rightarrow \infty,$$

where $r = p = q$. The null hypothesis is rejected for large values of s_n .

3. Numerical Studies

3.1 Simulation

The two representations of the mixture model in equation (1) have some technical implications on the data generating mechanism. When ω_i is positive, the mixture model maintains its hierarchical representation and data can then be generated using the usual two-stage process. For negative ω_i , however, the mixture model loses its hierarchical representation and the data can not be generated from the two-stage process. Instead, data are generated directly from the marginal distribution in equation (1) by inverting the cumulative distribution function (CDF) of a uniform distribution on the interval (0, 1).

We conducted a simulation study to evaluate the empirical performances of the proposed covariate-adjusted score test in small to moderate sample sizes. We compared these performances to those of the tests proposed by van den Broek (1995), and Jansakul and Hinde (2009). Throughout our simulations, we generated data from a two-component model with true mixing weights ω_i^* , $i = 1, \dots, n$, at $y^* = 0$, and a nondegenerate Poisson distribution with a simple mean $\mu_i^* = \exp\{\beta_0^* - 1.45x_i\}$, where the intercept β_0^* takes values in the set $\{-0.75, 0, 0.75\}$ and x_i is a covariate generated from a uniform distribution on the interval (0, 1). Throughout our simulations, we performed the proposed covariate-adjusted score test assuming the working mixing weight model $\omega_i = \{\pi_i - \exp\{-\mu_i\}\} \{1 - \exp\{-\mu_i\}\}^{-1}$ under the alternative, where $\pi_i = \exp\{-\exp\{\gamma_0 + \gamma_1 x_i\}\}$ and $\mu_i = \exp\{\beta_0 + \beta_1 x_i\}$. The test of van den Broek (1995) was performed assuming $\omega_i = \gamma_0$, and that of Jansakul and Hinde (2009), assuming $\omega_i = \gamma_0 + \gamma_1 x_i$. With these parameterizations, the null hypotheses to be evaluated were then given by: $H_0 : \gamma_0 = \beta_0, \gamma_1 = \beta_1$, for our formulation; $H_0 : \gamma_0 = 0$, for van den Broek's test; and $H_0 : \gamma_0 = 0, \gamma_1 = 0$, for Jansakul and Hinde's test. The maximum likelihood estimate $\hat{\beta}$ of the true value of $\beta = (\beta_0, \beta_1)'$ under the null was obtained from a homogeneous Poisson model with mean $\mu_i = \exp\{\beta_0 + \beta_1 x_i\}$. Finally, all simulations were replicated 1,000 times and for sample sizes 50, 100, and 200.

To investigate the empirical type I error rates of the tests, we generated data from the homogeneous Poisson model. The empirical type I error rates at 5% nominal level are reported in Table 2. All considered score tests have well controlled type I error rates even for a sample size as small as 50 when β_0^* is large. However, for smaller values of β_0^* (large values of $-\beta_0^*$) these tests tend to be slightly conservative. In other words, when the true mean μ_i^* nears zero they tend to reject the null hypothesis less often than anticipated and this conservativeness does not diminish with increasing sample sizes.

We investigated the empirical power of the tests to detect various forms of heterogeneity in the population. First, we generated data from a two-component mixture model with a constant mixing weight fixed at 0.25. For this constant mixing weight model, increasing sample sizes and separation of mixture components improve the power of detecting the alternatives under consideration for all tests considered (see Table 3). The test proposed by van den Broek (1995) appears to be more powerful than the two covariate-adjusted tests, especially when the mixture components are well separated. This is expected as data were generated from a positive and constant mixing weight model, for which the associated marginal model maintains its hierarchical representation. The loss of power was fairly minor for covariate-adjusted tests, however.

Second, we allowed the true mixing weight to depend on covariates using the linear, the logistic, and the proposed transformation (mixing weight models in rows 2, 3, and 4 of

Table 3). Overall, the power of the three score tests improves with the sample size, regardless of the true mixing weight. When the true mixing weight is a linear or a logistic function of the covariates, all considered tests appear to have comparable powers. But the power deteriorates as the mean of the non-degenerate Poisson model nears zero. This is not surprising as the true mixing weights in these schemes are bounded between 0 and 1, a condition for the marginal mixture model to maintain its hierarchical representation. When the true mixing weights allow for negative values (mixing weight model in row 4 of Table 3), the proposed score test statistic appears to be more powerful than the other tests. The deterioration of power appears to be greater for the test proposed by van den Broek (1995). This is expected as the constant mixing weight test averages the mixing weights over the space of covariates, which may greatly affect power if both deflation and inflation are present in the population.

From the previous simulation results, one referee was concerned that the proposed covariate adjusted test outperforms the other competitors only when the simulated data are generated from a model that allows for negative mixing weights. Following this referee's suggestion, we conducted further simulation studies in which data were generated from a two-component mixture model with positive covariate dependent weights

$\omega_i^* = \{1 + \exp\{15 - 30x_i\}\}^{-1}$ and $\omega_i^* = \Phi(-15 + 30x_i)$, $i = 1, \dots, n$, where $\Phi(\cdot)$ is the CDF of a standard normal variable. For these simulation schemes, the proposed test outperforms the other two competitors, especially when the two mixture components are well separated (see the last two simulation schemes in Table 3). Here the test proposed by Jansakul and Hinde (2009) adjusts for the covariate x_i in modeling the mixing weight ω_i , but fails to provide an adequate functional relationship between ω_i and x_i . In general, we expect this test to perform well when linearity holds and poorly otherwise. Following another referee's suggestion, another simulation study was conducted to compare the three tests when covariates x_i and z_i are not equal. Findings from this article (results where x_i is a subset of z_i are shown in Table 5, Web supplementary file) were similar to those that assume $x_i = z_i$.

In sum, incorporating covariates in the score tests for homogeneity can greatly improve efficiency. But this efficiency gain highly depends on the prespecified working model of the mixing weight and the behavior of the tests under model mis-specification. When the true mixing weight does not depend on covariates, our test may not be as efficient as the constant mixing weight test. However, given that the true model is usually unknown to the analyst, a general approach that assumes covariate dependent mixing weights appears to be the most conservative data analytic strategy.

3.2 Dental Caries Data

To illustrate our methodology, we considered children dental caries data from the Detroit study. We focused on scores representing the number of tooth surfaces that show signs of clinically detectable enamel lesions comprising both noncavitated and cavitated lesions. Using this stringent dental caries definition, three different outcomes were derived. These outcomes represent the number of decayed surfaces (DS), the number of decayed and filled surfaces (DFS), and the number of decayed, missing, and filled surfaces (DMFS). Although this survey is longitudinal in nature, our numerical computations are based on cross-sectional data of 897 children surveyed in the first wave of examinations conducted between 2002 and 2003. Covariates considered include Age (the child's age in years), SI (the child's sugar intake), and their multiplicative interaction Age * SI. A more detailed description of the study can be found elsewhere (see, for example, Tellez et al., 2006).

It is a common practice to use zero-inflated models to characterize dental caries scores (see, for example, Böhning et al., 1999). In addition, it is well known that dental caries data often

exhibit overdispersion in addition to zero inflation. For low-income children, however, it is possible that both zero inflation and deflation may be present, even when overdispersion is accounted for. To investigate this, as a simple analysis we discretized the age variable ($\text{Age} < 3$, $3 \leq \text{Age} < 4$, and $\text{Age} \geq 4$) and fit a homogeneous negative binomial model with age group specific means and a common overdispersion parameter to DS. Figure 1 presents an informal assessment of this model by comparing the observed proportions to the predicted probabilities for each age group. In this population, there is an inflation of zeros for children under the age of 3 and a deflation for children above the age of 4 (Bottom panel of Figure 1). A constant mixing weight test by averaging out the mixing weights over the age groups, may fail to detect heterogeneity in this population.

We evaluated the homogeneity hypothesis using the proposed covariate-adjusted score test. Specifically, we considered a two-component mixture model for which the nondegenerate distribution is a negative binomial model with mean $\mu_i = \exp\{\beta_0 + \beta_1 \text{Age}_i + \beta_2 \text{SI}_i + \beta_3 \text{Age}_i * \text{SI}_i\}$ and overdispersion parameter κ and the mixing weight ω_i is given by equation (3) with $\pi_i = (1 + \kappa \exp\{\gamma_0 + \gamma_1 \text{Age}_i + \gamma_2 \text{SI}_i + \gamma_3 \text{Age}_i * \text{SI}_i\})^{-1/\kappa}$. For comparison, the score test with constant mixing weight $\omega_i = \gamma_0$ of van den Broek (1995) and that proposed by Jansakul and Hinde (2009) with mixing weight $\omega_i = \gamma_0 + \gamma_1 \text{Age}_i + \gamma_2 \text{SI}_i + \gamma_3 \text{Age}_i * \text{SI}_i$ were also performed. With these parameterizations, the null hypotheses to be evaluated were then given by: $H_0 : \gamma_j = \beta_j, j = 0, 1, 2, 3$, for our formulation; $H_0 : \gamma_0 = 0$, for van den Broek's test; and $H_0 : \gamma_j = 0, j = 0, 1, 2, 3$, for Jansakul and Hinde's test. All these tests were conducted by replacing the nuisance parameter β by its maximum likelihood estimate under the null distribution. Results of this analysis are presented in Table 4.

Our test statistic and that of Jansakul and Hinde (2009) reject the homogeneity hypothesis for all outcomes at 5% significance level, supporting the hypothesis of heterogeneity. But our test provides a stronger evidence for heterogeneity in this population in view of the p-values. The constant mixing weight test, however, fails to reject the null for all outcomes. The observed value of the constant mixing weight test statistic is almost zero, suggesting that inflation and deflation at zero appear to be of the same magnitude, taking into account the random variation. This is a good example where the score test based on constant mixing weights is not powerful enough to capture heterogeneity in the data. But when adjusted by covariates, the score test greatly improves efficiency in detecting heterogeneity.

It is worth noting that failure to reject homogeneity does not give evidence that the zero-inflated negative binomial model provides a best fit for the data. Instead, such rejection only gives grounds for the zero-inflated negative binomial model to be further evaluated. As a final investigation, we then compared the zero-inflated negative binomial model to the zero-inflated Poisson model. A one-sided test (Silvapulle and Silvapulle, 1995; Ridout, Hinde, and Demetrio, 2001) of the overdispersion parameter of the negative binomial model was highly significant at 1% significance level, revealing a strong evidence for overdispersion in addition to zero inflation. This confounding of mixtures with overdispersion is not uncommon in practice as recognized by Lindsay and Roeder (1992) and Böhning et al. (1999).

4. Discussion

The goal of this article was not to argue for or against the marginal model, but rather to show how to conduct inferences for the mixing weights in this class of models. Although score tests for evaluating homogeneity in two-component models for discrete data have been well discussed in the literature, existing methodologies have relied primarily on restrictive assumptions. One common restriction is that constant mixing weights are often assumed. From a practical perspective, this is an important limitation as covariates dependent mixing

weights are typical in real applications of two-component regression models. More generally, limitations of existing methodologies result from the structure of the marginal model being simply ignored or at best not properly integrated into the testing procedures. In this article, we formulated and developed a score test of homogeneity based on an intuitive approach that accommodates the features of the marginal model. Specifically, our proposed test adopted a novel parameterization that allows the structural constraints of the mixing weight to be embedded in the testing scheme. Most importantly, this parameterization naturally incorporates covariates in the mixing weights. We showed that this is a natural strategy to adopt when testing is carried out under the marginal model that ignores its hierarchical representation.

The proposed test can be extended to refine the model specification. We generally conduct the test of homogeneity because it is easy to control its type I error. However, if the null hypothesis is rejected, one is typically interested in evaluating composite hypotheses. For example, one may be interested in testing for equality of some coefficients of the binary regression model and corresponding coefficients of the nondegenerate regression model. Although the details to evaluate such composite hypotheses warrant a separate investigation, the results presented here can be used to evaluate these specific hypotheses with a proper adjustment of the type I error. It is noteworthy that rejecting the homogeneity hypothesis against a specific two-component model does not necessarily imply that the latter model is appropriate. Instead, we advocate that such rejection only gives grounds for the two-component model to be further investigated. It is therefore crucial to study how plausible are inferences with respect to a candidate model under consideration that are of major substantive interest from the observed data standpoint.

Supplementary Material

Refer to Web version on PubMed Central for supplementary material.

Acknowledgments

The authors are grateful to Amid Ismail and Woosung Sohn for their permission to use the Detroit dental caries data. This work was supported by the first author's NCI/NIH K-award, 5K01CA131259-04 and its supplement 3K01CA131259-02S1 from the 2009 American Recovery and Reinvestment Act funding mechanism.

References

- Böhning D, Dietz E, Schlattmann P, Mendonca L, Kirchner U. The zero-inflated Poisson model and the decayed, missing and filled teeth index in dental epidemiology. *Journal of the Royal Statistical Society: Series A (Statistics in Society)*. 1999; 162:195–209.
- Cox, DR.; Hinkley, DV. *Theoretical Statistics*. London: Chapman & Hall Ltd; 1974.
- Deng D, Paul SR. Score tests for zero inflation in generalized linear models. *The Canadian Journal of Statistics/La Revue Canadienne de Statistique*. 2000; 28:563–570.
- Farewell V, Sprott D. The use of a mixture model in the analysis of count data. *Biometrics*. 1988; 44:1191–1194. [PubMed: 2466491]
- Hall DB. Zero-inflated Poisson and binomial regression with random effects: A case study. *Biometrics*. 2000; 56:1030–1039. [PubMed: 11129458]
- Heilbron D. Zero-altered and other regression models for count data with added zeros. *Biometrical Journal*. 1994; 36:531–547.
- Hur K, Hedeker D, Henderson W, Khuri S, Daley J. Modeling clustered count data with excess zeros in health care outcomes research. *Health Services and Outcomes Research Methodology*. 2002; 3:5–20.
- Jansakul N, Hinde J. Score tests for extra-zero models in zero-inflated negative binomial models. *Communications in Statistics-Simulation and Computation*. 2009; 38:92–108.

- Lambert D. Zero-inflated Poisson regression, with an application to defects in manufacturing. *Technometrics*. 1992; 34:1–14.
- Lewsey JD, Thomson WM. The utility of the zero-inflated poisson and zero-inflated negative binomial models: A case study of cross-sectional and longitudinal dmf data examining the effect of socio-economic status. *Community Dentistry and Oral Epidemiology*. 2004; 32:183–189. [PubMed: 15151688]
- Lindsay B, Roeder K. Residual diagnostics for mixture models. *Journal of the American Statistical Association*. 1992; 87:785–794.
- Mullahy J. Specification and testing of some modified count data models. *Journal of Econometrics*. 1986; 33:341–365.
- Ridout, M.; Demetrio, CGB.; Hinde, J. Proceedings of International Biometric Conference. Cape Town, South Africa: International Biometric Society; 1998. Models for count data with many zeros; p. 179-192.
- Ridout M, Hinde J, Demetrio CGB. A score test for testing a zero-inflated Poisson regression model against zero-inflated negative binomial alternatives. *Biometrics*. 2001; 57:219–223. [PubMed: 11252601]
- Silvapulle MJ, Silvapulle P. A score test against one-sided alternatives. *Journal of the American Statistical Association*. 1995; 90:342–349.
- Tellez M, Sohn W, Burt B, Ismail A. Assessment of the relationship between neighborhood characteristics and dental caries severity among low-income African-Americans: A multilevel approach. *Journal of Public Health Dentistry*. 2006; 66:30–36. [PubMed: 16570748]
- Thompson R, Baker RJ. Composite link functions in generalized linear models. *Applied Statistics*. 1981; 30:125–131.
- van den Broek J. A score test for zero inflation in a Poisson distribution. *Biometrics*. 1995; 51:738–743. [PubMed: 7662855]
- Xiang L, Lee A, Yau K, McLachlan G. A score test for zero-inflation in correlated count data. *Statistics in Medicine*. 2006; 25:1660–1671. [PubMed: 16143993]

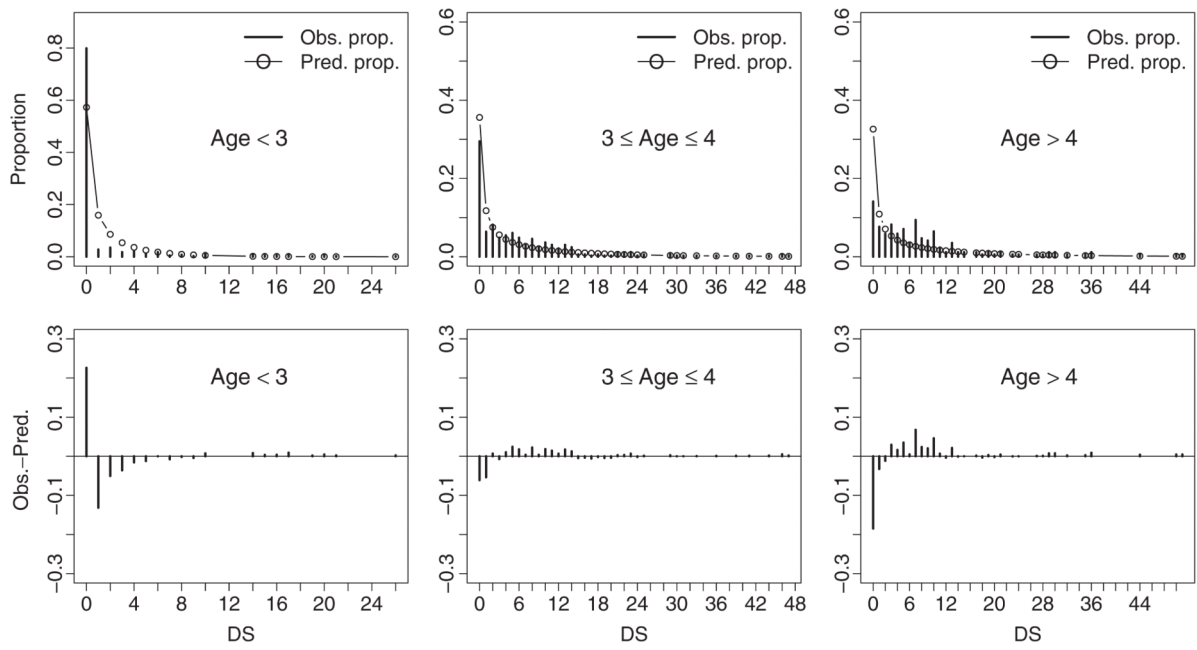


Figure 1. Observed and predicted proportions (Top panel) and observed–predicted proportions (Bottom panel) of the number of decayed surfaces (DS) for each age group. Fitted model is a negative binomial distribution with age group specific means.

Table 1

Natural parameterizations relating π_i to linear predictors $z'_i\gamma$ for commonly used two-component models and link functions relating μ_i to linear predictors $x'_i\beta$ when $y^* = 0$

Nondegenerate distribution G	Link function $h(\cdot)$ relating μ_i to $x'_i\beta$	Probability $g_i(0) = g(\tilde{\mu}_i)$	Parameterization relating π_i to $z'_i\gamma$
Poisson with mean μ_i	$\log\{\mu_i\}$	$\exp\{-\exp\{x'_i\beta\}\}$	$\exp\{-\exp\{z'_i\gamma\}\}$
Binomial with success probability μ_i and m_i trials	$\text{logit}\{\mu_i\}$	$\{1+\exp\{x'_i\beta\}\}^{-m_i}$	$\{1+\exp\{z'_i\gamma\}\}^{-m_i}$
	$\Phi^{-1}(\mu_i)$	$\{1+\Phi(x'_i\beta)\}^{m_i}$	$\{1+\Phi(z'_i\gamma)\}^{m_i}$
	$\log\{-\log\{\mu_i\}\}$	$\{1-\exp\{-\exp\{x'_i\beta\}\}\}^{m_i}$	$\{1-\exp\{-\exp\{z'_i\gamma\}\}\}^{m_i}$
Negative binomial with mean μ_i and over dispersion parameter κ	$\log\{\mu_i\}$	$\{1+\kappa\exp\{x'_i\beta\}\}^{-1/\kappa}$	$\{1+\kappa\exp\{z'_i\gamma\}\}^{-1/\kappa}$

Note: $\theta = \beta$ for the Poisson and the binomial processes;

$\theta = \{\beta, \kappa\}$ for the negative binomial process.

Table 2

Empirical size of score test statistics when the null model is a nondegenerate Poisson model with mean $\mu_i^* = \exp\{\beta_0^* - 1.45x_i\}$, $x_i \sim \text{Uniform}(0, 1)$, at 5% significance level

	<i>n</i> = 50		<i>n</i> = 100		<i>n</i> = 200				
	-0.75	0	-0.75	0	-0.75	0			
β_0^*	0.031	0.053	0.051	0.048	0.063	0.059	0.042	0.050	0.066
vdB test	0.031	0.053	0.051	0.048	0.063	0.059	0.042	0.050	0.066
J&H test	0.033	0.051	0.054	0.038	0.044	0.061	0.036	0.049	0.052
Prop. c-a test	0.036	0.048	0.048	0.040	0.046	0.057	0.033	0.045	0.060

Note: vdB, van den Broek, test with *df* = 1; J&H, Jansakul and Hinde, test with *df* = 2; Prop. c-a, Proposed covariate-adjusted, test with *df* = 2.

Table 3

Empirical power of score test statistics to detect various forms ω_i^* of heterogeneity coupled with a nondegenerate Poisson model with mean $\mu_i^* = \exp\{\beta_0^* - 1.45x_i\}$, $x_i \sim \text{Uniform}(0, 1)$, at 5% significance level

	n = 50			n = 100			n = 200		
	-0.75	0	0.75	-0.75	0	0.75	-0.75	0	0.75
β_0^*									
$\omega_i^* = 0.25$									
vdB test	0.060	0.096	0.313	0.086	0.147	0.578	0.122	0.294	0.895
J&H test	0.069	0.087	0.249	0.094	0.118	0.481	0.123	0.238	0.851
Prop. c-a test	0.071	0.086	0.247	0.094	0.122	0.478	0.124	0.241	0.849
$\omega_i^* = 0.45 + 0.15x_i$									
vdB test	0.100	0.221	0.698	0.156	0.388	0.960	0.219	0.684	1.000
J&H test	0.102	0.190	0.623	0.178	0.362	0.927	0.237	0.608	0.999
Prop. c-a test	0.105	0.191	0.617	0.179	0.363	0.933	0.235	0.603	0.999
$\omega_i^* = \frac{\exp\{-1.2+1.1x_i\}}{1+\exp\{-1.2+1.1x_i\}}$									
vdB test	0.059	0.109	0.373	0.090	0.213	0.704	0.124	0.343	0.953
J&H test	0.080	0.112	0.318	0.103	0.192	0.662	0.136	0.297	0.929
Prop. c-a test	0.087	0.112	0.317	0.105	0.194	0.655	0.135	0.292	0.924
$\omega_i^* = \frac{\pi_i^* - \exp\{-\mu_i^*\}}{1 - \exp\{-\mu_i^*\}}, \pi_i^* = \exp\{1.3 - 2.4x_i\}$									
vdB test	0.166	0.293	0.257	0.295	0.395	0.413	0.392	0.612	0.590
J&H test	0.236	0.454	0.386	0.383	0.628	0.544	0.506	0.847	0.731
Prop. c-a test	0.710	0.845	0.957	0.944	0.981	0.997	0.997	1.000	1.000
$\omega_i^* = \frac{\exp\{-15+30x_i\}}{1+\exp\{-15+30x_i\}}$									
vdB test	0.026	0.067	0.095	0.039	0.079	0.131	0.051	0.090	0.189
J&H test	0.054	0.107	0.166	0.082	0.141	0.271	0.090	0.180	0.448
Prop. c-a test	0.068	0.151	0.430	0.122	0.238	0.758	0.131	0.431	0.967
$\omega_i^* = \Phi(-15+30x_i)$									

β_0^*	$n = 50$			$n = 100$			$n = 200$		
	-0.75	0	0.75	-0.75	0	0.75	-0.75	0	0.75
vdB test	0.042	0.070	0.099	0.045	0.069	0.153	0.057	0.083	0.221
J&H test	0.051	0.097	0.160	0.075	0.114	0.278	0.104	0.167	0.465
Prop. c-a test	0.062	0.142	0.485	0.095	0.222	0.739	0.142	0.409	0.962

Note: vdB, van den Broek, test with $df = 1$; J&H, Jansakul & Hinde, test with $df = 2$; Prop. c-a, Proposed covariate-adjusted, test with $df = 2$.

Table 4

Score test statistics, degrees of freedom, and associated p-values for heterogeneity in dental caries data when the null model is a nondegenerate negative binomial distribution

Response	van den Broek test		Jansakul & Hinde test		Proposed cov.-adj. test	
	df	statistic	df	statistic	df	statistic
DS	1	0.0164	4	128.1512	4	151.8566
DFS	1	0.0083	4	123.2241	4	148.8391
DMFS	1	0.0009	4	126.7810	4	151.8784