Published in final edited form as:

Scand Stat Theory Appl. 2018 September ; 45(3): 465–481. doi:10.1111/sjos.12308.

A quasi-score statistic for homogeneity testing against covariate-varying heterogeneity

David Todem,

Michigan State University, East Lansing, USA

Wei-Wen Hsu, and

Kansas State University, Manhattan, USA

Jason P. Fine

University of North Carolina, Chapel Hill, USA

Summary

In statistical modeling, it is often of interest to evaluate non-negative quantities that capture heterogeneity in the population such as variances, mixing proportions and dispersion parameters. In instances of covariate-dependent heterogeneity, the implied homogeneity hypotheses are nonstandard and existing inferential techniques are not applicable. In this paper, we develop a quasi-score test statistic to evaluate homogeneity against heterogeneity that varies with a covariate profile through a regression model. We establish the limiting null distribution of the proposed test as a functional of mixtures of chi-square processes. The methodology does not require the full distribution of the data to be entirely specified. Instead, a general estimating function for a finite dimensional component of the model that is of interest is assumed but other characteristics of the population are left completely unspecified. We apply the methodology to evaluate the excess zero proportion in zero-inflated models for count data. Our numerical simulations show that the proposed test can greatly improve efficiency over tests of homogeneity that neglect covariate information under the alternative hypothesis. An empirical application to dental caries indices demonstrates the importance and practical utility of the methodology in detecting excess zeros in the data.

Keywords

Efficiency; Goodness-of-fit; Influence functions; Mixture models; Sensitivity analysis; Split samples; Testing at infinity

1. Introduction

Homogeneity testing is an essential exercise in statistical modeling involving formulations that capture heterogeneity in the population. Often, this exercise consists of evaluating non-negative heterogeneity quantities taking the form of variances, mixture proportions and dispersion parameters. When these non-negative quantities depend on covariates, the implied homogeneity hypotheses are nonstandard and existing inferential techniques are not applicable. This paper extends the literature by developing a procedure to evaluate homogeneity against alternatives that vary with covariates through a regression model, with

the test that ignores covariate information serving as a special case. The regression model relates the heterogeneity parameter to both categorical and continuous covariates, a formulation that is often encountered in real applications of this class of models.

To fix ideas, consider a population governed by an unknown law $dP_{\theta, \phi}$ indexed by a parameter vector $\theta \in \Theta \subset \mathbb{R}^p$ and a non-negative quantity $\phi \in \Phi \subseteq [0, \infty)$ representing the heterogeneity quantity, with the basic requirement that $0 \in \Phi$. The heterogeneity quantity ϕ varies with the subject's covariate profile through a regression model. Consistent with real applications, we assume the general regression model of the form $\phi(v, x) = g(\alpha^\top v + \gamma^\top x)$, where $g(\cdot)$ is a known, monotone increasing, differentiable and invertible function on domain $\mathbb{R}$ and codomain Φ , with $\phi(v, x)$ highlighting the dependence on v and x . Here, $v = (v_1, \dots, v_J)^\top$ represents the vector of indicator variables for a full factorial combination of some of the categorical covariates, with v_j taking value 1 if the sampling unit belongs to stratum $j = 1, \dots, J$, and 0 otherwise, with $\sum_{j=1}^J v_j = 1$, and $x = (x_1, \dots, x_q)^\top$ is a vector of continuous variables or indicators of categorical variables that do not contribute to the full factorial design, $\alpha = (\alpha_1, \dots, \alpha_J)^\top \in \bar{\mathbb{R}}^J$ and $\gamma = (\gamma_1, \dots, \gamma_q)^\top \in \Gamma \subseteq \mathbb{R}^q$ are unknown regression parameters associated with v and x . Here $\bar{\mathbb{R}}$ is the extended real line $[-\infty, \infty]$, which allows the elements of α to be on the infinite boundary. In contrast, the elements of γ belong to the set of real numbers. Because each stratum j has its own intercept α_j , $j = 1, \dots, J$, no constant is added in x under this parameterization. Of course in situations involving no stratification, that is $J = 1$, $v = v_1 = 1$, the overall model intercept is simply $\alpha = \alpha_1$.

Under the constraint $g(\alpha^\top v + \gamma^\top x) \geq 0$, for all v, x , we are interested in evaluating the hypotheses,

$$H_0 : g(\alpha^\top v + \gamma^\top x) = 0, \text{ for all } v, x, \quad \text{versus} \quad H_1 : g(\alpha^\top v + \gamma^\top x) > 0, \text{ for some } v, x. \quad (1)$$

Under the null hypothesis, the population is governed by the restricted law $dP_{\theta, 0}$. Using the logistic function $g(a) = \exp\{a\}/(1 + \exp\{a\})$ and the exponential function $g(a) = \exp\{a\}$, $a \in \mathbb{R}$, we provide below two motivating examples involving the evaluation of the mixture proportion and the dispersion parameter in popular finite dimensional models for count data.

Example 1: Testing for zero-inflation in two-component mixture models

Consider the class of zero-inflated models with probability law $dP_{\theta, \phi} = \phi dG_0 + (1 - \phi)dP_{\theta, 0}$, combining an atom dG_0 at 0 and an unknown non-degenerate discrete law $dP_{\theta, 0}$, with an unknown mixing probability $\phi \in \Phi = [0, 1)$. Typical examples of these finite mixture models include the models for uncorrelated data first studied by Mullahy (1986), Farewell and Sprott (1988) and Lambert (1992) and their extension to dependent data by Hall (2000). Assume that the mixture probability is related to covariates v and x through the regression model $\phi(v, x) = \exp\{\alpha^\top v + \gamma^\top x\} / \{1 + \exp\{\alpha^\top v + \gamma^\top x\}\}^{-1}$, $\alpha \in \bar{\mathbb{R}}^J$, $\gamma \in \Gamma$. In real applications of these mixture models, it is often of interest to evaluate the hypothesis $\exp\{\alpha^\top v + \gamma^\top x\} / \{1 + \exp\{\alpha^\top v + \gamma^\top x\}\}^{-1} = 0$, for all v, x , in hope of describing the data using the simpler homogeneous model $dP_{\theta, 0}$.

Example 2: Testing for overdispersion in Poisson models

In statistical analysis of count data, it is customary to use a negative binomial model $dP_{\theta, \phi}$ with mean $\mu(\theta)$ and dispersion parameter $\phi \in \Phi = [0, \infty)$, to accommodate the variation above and beyond that explained by a Poisson model $dP_{\theta, 0}$ (Dean, 1992). Assume that the dispersion parameter is related to covariates v and x through the log-linear model $\phi(v, x) = \exp\{a^\top v + \gamma^\top x\}$, $a \in \mathbb{R}^J$, $\gamma \in \Gamma$. Evaluating the homogeneity hypothesis $\exp\{a^\top v + \gamma^\top x\} = 0$, for all v, x , can be carried out in hope of describing the data using the simpler Poisson model.

Evaluating the homogeneity hypothesis in (1) against covariate-varying heterogeneity is technically nontrivial. Existing inferential procedures typically neglect covariate information under heterogeneity by restricting $\phi(v, x)$ for all v, x , to a constant under the alternative hypothesis (e.g., Dean, 1992; van den Broek, 1995). Specifically, this approach imposes the restrictions $J=1$ and $\gamma=0$, and evaluates the homogeneity hypothesis $g(a_1)=0$ against the alternative $g(a_1)>0$. For the two motivating examples, the homogeneity hypotheses become $\exp\{a_1\}=0$ or $\exp\{a_1\}\{1+\exp\{a_1\}\}^{-1}=0$, resulting in parameter a_1 being evaluated on the infinite boundary. Although valuable, such a restrictive approach may have reduced power in detecting heterogeneity that varies with covariates, and worse may lead to invalid inferences for the parameters of interest. When the working model for ϕ only involves a full factorial combination of the categorical covariates, in which case $\gamma=0$ and $J>1$, the homogeneity hypothesis then reduces to $g(a_j)=0$ for all $j=1, \dots, J$. The generalized score test for one-sided alternatives proposed by Silvapulle and Silvapulle (1995) can be used to assess homogeneity against stratum-specific heterogeneity. In the presence of x , however, the testing problem is nonstandard and it is unclear how to formulate the homogeneity hypothesis using parameters a and γ . In the context of zero-inflated models for count data, Jansakul and Hinde (2002, 2008) used the identity link function relating the probability of excess zeros to covariates to develop a score test of homogeneity. Under the identity link $g(a^\top v + \gamma^\top x) = a^\top v + \gamma^\top x$, the homogeneity hypothesis is equivalent to the constraints $a=0$ and $\gamma=0$. This test has the advantage of being computationally very feasible and appears to work well even if the regression structure under the alternative hypothesis is not correctly specified but contains the relevant covariates. A limitation, however, is that the identity link is seldom used for this class of model in real applications. And more importantly, it is unclear how this testing procedure incorporates the constraint that the mixing probability should lie in the $(0, 1)$ interval.

In this paper, we propose an approach that evaluates the hypothesis $g(a^\top v + \gamma^\top x) = 0$, for all v, x , using transformations that are consistent with real applications. We adopt a more flexible approach that confines γ to an unknown compact set while allowing a to lie in a non-compact (open) set. The homogeneity hypothesis then translates into testing infinitely large intercepts $a_1, \dots, a_J$. A complication, however, is that whenever these intercepts are infinitely large, γ becomes nonidentifiable, resulting in the so-called Davies problem (Davies, 1977). To address this complication, we regard the pivotal quantities for evaluating large intercepts as processes of a nonidentifiable parameter γ . The test statistic is then constructed by using functionals, such as the supremum function, of these processes over the whole support of γ . We characterize the limiting null distribution of this statistic as a

functional of mixtures of chi-square processes, which is rigorously approximated by a simple resampling procedure.

The paper proceeds as follows. In section 2, we make a connection between the evaluation of homogeneity and the Davies problem, and provide the basic assumptions of the model. We also derive a supremum quasi-score statistic for evaluating the homogeneity hypothesis, and discuss its limiting null distribution. Section 3 presents simulation studies to evaluate the finite sample performances of the proposed statistic for testing inflation in the class of zero-inflated models for discrete data, followed by an analysis of count data in dental caries research to illustrate the test's practical utility. Final remarks are provided in section 4.

2. Methodology

2.1. Specification of the homogeneity hypothesis

Because $\alpha^T v + \gamma^T x$ for all values of α and γ and covariates v and x covers the entire real space and that $g(\cdot)$ is a monotone increasing function, we have $g(-\infty) = 0$. Evaluating the homogeneity hypothesis in (1) then reduces to evaluating $\alpha^T v + \gamma^T x = -\infty$ for all v, x . We adopt the following working assumption.

Assumption—The space Γ of γ is compact, i.e. $\sup_{\gamma \in \Gamma} \|\gamma\| \leq \zeta$, for some $\zeta < \infty$.

Here and in the sequel, $\|\cdot\|$ represents the Euclidean norm of a vector. This assumption is merely technical and ensures that the null hypothesis only holds if $g(\alpha^T v) = 0$ for all v , which is equivalent to $\alpha_j = -\infty$, for all $j = 1, 2, \dots, J$. This holds because compactness is not assumed for the space of α . Sy and Taylor (2000, page 234) have adopted a similar assumption to evaluate a covariate-dependent cure rate in a two-component survival model. Inference was ad hoc, however, ignoring that setting the intercept to $-\infty$ renders slope parameters γ nonidentifiable, and that regularity conditions for the validity of the Wald test may not be satisfied.

For convenience, we consider the transformation $\psi_j = g(\alpha_j)$, $j = 1, \dots, J$ and define the cone $\Psi = \{\psi = (\psi_1, \dots, \psi_J) : \psi_j \geq 0, j = 1, 2, \dots, J\}$ and Ψ_+ a subset of Ψ for which $\psi_j > 0$, for some j . Whenever $\psi_j = 0$ ($\alpha_j = -\infty$) for all j , γ becomes nonidentifiable, leading to the homogeneity hypothesis,

$$H_0 : \psi = 0, \text{ for all } \gamma \in \Gamma.$$

When heterogeneity depends on covariates via a regression model, the implied hypotheses may not be typical and standard regularity conditions to conduct the test may not hold. In a stratified population, there are multiple non-negative constraints on ψ ; the parameter vector ψ under the null lies on the boundary of the parameter space Ψ ; the nuisance parameter γ is nonidentifiable under the null hypothesis $\psi = 0$; and most importantly there is an uncertainty regarding Γ the support space of the unidentified nuisance parameters. From a methodological standpoint, there have been several papers in the literature that examine related issues and complications. But these works often accommodate only one or two of

these complications. We are not aware of any testing procedure that addresses all these complications simultaneously.

The issue of parameters being on the boundary of the parameter space and that of one-sided hypotheses under alternatives have been extensively studied in the literature. A seminal work on the topic is that of Chernoff (1954) and more recent papers include those of Self and Liang (1987); Silvapulle and Silvapulle (1995); Andrews (2001) and Verbeke and Molenberghs (2003) and references therein. But these works often deal with the situation where nuisance parameters are fully identified under the null hypothesis.

The problem of unidentifiable nuisance parameters under the null has also been discussed in the literature. A popular approach for handling unidentifiable nuisance parameters consists of treating the pivotal quantity as an empirical process of these parameters (see, for example, Davies, 1977; Hansen, 1996; Andrews, 2001; Beg et al., 2001; Ritz and Skovgaard, 2005; Zhu and Zhang, 2006; and Song et al., 2009). This approach has been by far the most popular method for dealing with nonidentifiable parameters under the null. But in the current literature, it is often assumed that the support space of nonidentifiable parameters is known to the analyst. For example, Ritz and Skovgaard (2005) in their motivating example assumed that the nuisance parameter space is the full open parameter space $\mathbb{R}_+$. Likewise, Zhu and Zhang (2006) assumed that the set of nuisance parameter space is compact and well specified. In our testing problem, the support of nonidentifiable nuisance parameters is not known *a priori* beyond the compactness assumption. This is problematic because the empirical processes arguments often used for this type of problems may not hold in the absence of prior information about this support. A similar difficulty was faced by Beg et al. (2001) while evaluating the constancy of the conditional variance over time in the autoregressive conditional heteroscedasticity model for time series data. But the choice of the unidentified parameter set was ad hoc, ignoring any potential risk for misspecification.

2.2. The model specification

The working model considered in this paper concerns the class of finite dimensional parameter models indexed by $\xi = (\psi^\top, \gamma^\top, \theta^\top)^\top$. Vector $\psi \in \Psi$ is the parameter to evaluate, $\gamma \in \Gamma$ and $\theta \in \Theta$ represent the nonidentifiable and the identifiable nuisance parameters under the null hypothesis. Let y denote the response variable and z the vector containing all available covariates, including $\{v, x\}$. We assume that there are n random i.i.d copies $\{y_i, z_i\}_{i=1}^n$ of $\{y, z\}$ that could be used to conduct inferences on ξ . To construct the test, we assume that there is an estimating function for ξ , denoted $Q_n(\xi)$ such that $E\{Q_n(\xi)\} = 0$ for an interior point ξ of the support space, with the following smoothness condition.

Condition A1—We assume the existence of a nonsingular matrix G such that for any $\varrho > 0$,

$$\sup_{\|\dot{\xi}\| \leq \varrho} \|n^{-1/2}\{Q_n(\xi + n^{-1/2}\dot{\xi}) - Q_n(\xi)\} + G\dot{\xi}\| = o_p(1),$$

where $o_p(1)$ represents the convergence to 0 in probability as $n \rightarrow \infty$ and $G = \lim_{n \rightarrow \infty} -n^{-1} Q_n(\xi)/\xi$ w.p. 1. Additionally, we assume that $n^{-1/2} Q_n(\xi)$ converges in distribution to a central normal distribution.

A trivial example of function $Q_n(\xi)$ is the score function derived from parametric likelihood, in which case $G = \lim_{n \rightarrow \infty} n^{-1} \text{var}\{Q_n(\xi)\}$. Estimating functions derived from pseudo-likelihood (Besag, 1975) and composite likelihood (Lindsay, 1988) are other great examples of these smooth functions. This general smoothness assumption guarantees that the methodology will have wide applications including settings where likelihood functions are unavailable.

2.3. A quasi-score statistic for evaluating homogeneity

Let $v_n(\theta)$ be the component of $Q_n(\xi)$ corresponding to θ at $\psi = 0$. Let $\hat{\theta}$ be an estimator of θ under the null hypothesis. Based on the smoothness condition above, there exists a nonsingular matrix A such that $n^{1/2}(\hat{\theta} - \theta) = n^{-1/2} A^{-1} v_n(\theta) + o_p(1)$. Assume γ fixed and let $u_n(\gamma, \theta)$ denote the component of $Q_n(\xi)$ corresponding to ψ at $\psi = 0$ and by $u_n(\gamma, \hat{\theta})$ its estimated version when θ is replaced by $\hat{\theta}$. We also denote by $b_i(\gamma, \theta)$ and $d_i(\gamma, \theta)$ the contributions of subject i to $u_n(\gamma, \theta)$ and to the first-order derivative of the ψ -component of $Q_n(\xi)$ with respect to θ , evaluated at $\psi = 0$. Define $c_i(\gamma, \theta) = b_i(\gamma, \theta) - h(\gamma, \theta) A^{-1} a_i(\theta)$, where $a_i(\theta)$ is the contribution of subject i to the estimating function $v_n(\theta)$ and $h(\gamma, \theta) = E\{d_i(\gamma, \theta)\}$. Under condition A1, we have the following pointwise result, $n^{1/2} u_n(\gamma, \hat{\theta}) = n^{1/2} \sum_{i=1}^n c_i(\gamma, \theta) + \eta_n(\gamma)$, where $\eta_n(\gamma) = o_p(1)$ as $n \rightarrow \infty$, for each γ . This establishes a pointwise convergence of $n^{-1/2} u_n(\gamma, \hat{\theta})$ to a centered normal distribution with asymptotic covariance matrix $E\{c_1(\gamma, \theta)^{\otimes 2}\}$, for each γ . Here and throughout $c_1^{\otimes 2} = c_1 c_1^T$.

Because of the one-sided constraints $\psi \in \Psi$, we follow Silvapulle and Silvapulle (1995) in considering a modified quasi-score statistic of the form

$$s_n(\gamma) = \hat{u}_n(\gamma)^T \hat{\tau}(\gamma) \hat{u}_n(\gamma) - \inf_{\psi \in \Psi} \left\{ (\hat{u}_n(\gamma) - \psi)^T \hat{\tau}(\gamma) (\hat{u}_n(\gamma) - \psi) \right\}, \quad (2)$$

with $\hat{u}_n(\gamma) = n^{-1/2} u_n(\gamma, \hat{\theta})$ and $\hat{\tau}(\gamma)$ being an estimate of the inverse of $E\{c_1(\gamma, \theta)^{\otimes 2}\}$. Note that for fixed γ , $s_n(\gamma)$ is the likelihood ratio test of H_0 against H_1 based on the asymptotic normality of $n^{-1/2} u_n(\gamma, \hat{\theta})$. A sufficient condition that guarantees the existence of the inverse of $E\{c_1(\gamma, \theta)^{\otimes 2}\}$ is given below.

Condition A2—Assume that $E\{c_1(\gamma, \theta)^{\otimes 2}\}$ is uniformly positive definite over $\gamma \in \Gamma$ and $\theta \in \Theta$, in other words $\inf_{\gamma \in \Gamma, \theta \in \Theta} \text{eigmin}(E\{c_1(\gamma, \theta)^{\otimes 2}\}) > 0$, where $\text{eigmin}(\cdot)$ denotes the smallest eigenvalue.

In an unstratified population, $J = 1$, the quantity in (2) reduces to $s_n(\gamma) = \hat{u}_n(\gamma)^2 \hat{\tau}(\gamma) \mathbf{1}(\hat{u}_n(\gamma) > 0)$, where $\mathbf{1}(\cdot)$ represents the indicator function for an event. However when $J > 1$, the computation of $s_n(\gamma)$ for each fixed γ involves the minimization of the quadratic and convex function $(\hat{u}_n(\gamma) - \psi)^T \hat{\tau}(\gamma) (\hat{u}_n(\gamma) - \psi)$ in $\psi \in \Psi$. When all elements of $\hat{u}_n(\gamma)$ are positive,

this convex function is minimized at $\hat{u}_n(\gamma)$ and takes value 0. However when some components of $\hat{u}_n(\gamma)$ are negative, this function can easily be minimized numerically in practice using available optimization subroutines such as *optimize* in the *R* statistical computing environment.

If γ is known, the quantity in (2) is the score test statistic for one-sided alternatives proposed by Silvapulle and Silvapulle (1995). In general, however, $s_n(\gamma)$ is not a statistic due to unknown γ . To remove the dependence on γ , we introduce the supremum statistic,

$$T_n = \sup_{\gamma \in \Gamma} s_n(\gamma). \quad (3)$$

The null hypothesis is then rejected for large values of T_n .

2.4. Large sample properties

Let $\mathcal{B} = \{b(\gamma, \theta) : \gamma \in \Gamma, \theta \in \Theta, i = 1, \dots, n\}$ and $\mathcal{D} = \{d(\gamma, \theta) : \gamma \in \Gamma, \theta \in \Theta, i = 1, \dots, n\}$ be two function classes. We further adopt the following conditions:

Condition B1. Assume that Θ the support of θ is a compact set and θ the true value of θ is an interior point of Θ .

Condition B2. Function classes $\mathcal{B}$ and $\mathcal{D}$ are pointwise measurable and satisfy the uniform entropy condition; see van der Vaart and Wellner (2000a) for the definitions and related concepts.

Condition B3. Process $n^{-1} \sum_{i=1}^n c_i(\gamma_1, \theta) c_i(\gamma_2, \theta)^\top$ in $\gamma_1, \gamma_2 \in \Gamma$ converges almost surely to $\sigma_c^*(\gamma_1, \gamma_2) = E\{c_1(\gamma_1, \theta) c_1(\gamma_2, \theta)^\top\}$, uniformly over $\gamma_1, \gamma_2 \in \Gamma$.

Condition B4. Under the null $\psi = 0$, function $h(\gamma, \theta)$ is assumed uniformly bounded, i.e. $\sup_{\gamma \in \Gamma, \theta \in \Theta} \|h(\gamma, \theta)\|_2 < \infty$, where $\|\cdot\|_2$ denotes a matrix norm.

Condition *B1* defines the parameter space for the implied parameter θ . The entropy condition *B2* ensures that processes $u_n(\gamma, \theta)$ and $u_n(\gamma, \hat{\theta})$ are well behaved across all $\gamma \in \Gamma$. This condition is satisfied by functions which are uniformly bounded and uniformly Lipschitz of order $> \{\dim(\theta) + \dim(\gamma)\}/2$, where $\dim(\cdot)$ denotes the dimension of a vector. Conditions *B3* and *B4* give conditions under which uniform asymptotic results for $u_n(\gamma, \theta)$ and $u_n(\gamma, \hat{\theta})$ may be obtained. The two models discussed in the introduction meet these requirements and even more primitive sufficient conditions that imply these high level conditions.

Theory results on the pointwise convergence of $u_n(\gamma, \hat{\theta})$ for each $\gamma \in \Gamma$ are well established. Theorem 1 below gives a more general result on its uniform convergence across all $\gamma \in \Gamma$.

Theorem 1. Under Conditions B1–B4 and the null hypothesis, $\sup_{\gamma \in \Gamma} \|n^{-1} u_n(\gamma, \hat{\theta})\| = o_p(1)$ with $n \rightarrow \infty$, and the random process $n^{-1/2} u_n(\gamma, \hat{\theta})$ in γ converges in distribution to a centered Gaussian process with covariance kernel $\sigma_c^*(\gamma_1, \gamma_2)$, $\gamma_1, \gamma_2 \in \Gamma$.

Given that the set of discontinuity points of the supremum function has zero probability mass, the continuous mapping theorem can be used to retain the limit-preserving properties. This leads to the following corollary.

Corollary. Under conditions of Theorem 1, $T_n \rightarrow_d \sup_{\gamma \in \Gamma} s(\gamma)$, as $n \rightarrow \infty$, where the random process $s_n(\gamma)$ converges in distribution to the process $s(\gamma)$ as $n \rightarrow \infty$.

A well known result due to Shapiro (1988) states that $s(\gamma)$ is distributed as a mixture of chi-square distributions, for each $\gamma \in \Gamma$. That is, $\Pr(s(\gamma) \geq \nu) = \sum_{j=0}^J \pi_j(\gamma, \Psi) \Pr(\chi_j^2 \geq \nu)$, $\nu \geq 0$, where χ_j^2 is a chi-square random variable with j degrees of freedom (χ_0^2 being a zero point mass distribution) and $\pi_j(\gamma, \Psi)$, $j = 0, \dots, J$, are non negative mixing proportions such that $\sum_{j=0}^J \pi_j(\gamma, \Psi) = 1$. These weights depend on γ , Ψ and $E\{c_1(\gamma, \theta)^{\otimes 2}\}$ and can be expressed in a closed form for some special cases (Shapiro, 1988).

2.5. The resampling procedure

While for each fixed γ process $s(\gamma)$ follows a well known distribution under the null, the asymptotic distribution of $\sup_{\gamma \in \Gamma} s(\gamma)$ is quite complicated and analytically intractable. Davies (1977) has proposed upper bounds to significance levels for supremum test statistics, but these bounds may be of limited utility in practice (see Hansen, 1996 for a discussion on this issue). A simple nonparametric bootstrap may then be used to approximate the null distribution of the statistic (Efron and Tibshirani, 1993). Its validity follows from empirical processes arguments given in van der Vaart and Wellner (2000b, pages 345–359). A difficulty, however, is that the bootstrap resampling requires fitting the null model for each bootstrap sample, which may be computationally demanding. Alternatively, a simple resampling technique that perturbs the influence function using normal variates to approximate the true asymptotic distribution of the test statistic T_n can be used. This technique has been extensively used in the literature when the desired limiting distribution is hard if not impossible to derive analytically (see, for example, Parzen et al., 1994; Hansen, 1996; Zhu and Zhang, 2006; and Lee et al., 2011).

We operate conditionally on the sample $\{y_i, z_i\}_{i=1}^n$. Define $\tilde{u}_n(\gamma, \theta) = \sum_{i=1}^n c_i(\gamma, \theta) \zeta_i$ where $\{\zeta_1, \dots, \zeta_n\}$ are identical and independent standard normal variates. Process $\tilde{u}_n(\gamma, \hat{\theta})$ is similarly defined by replacing θ by its consistent estimator $\hat{\theta}$ under the null. Given observed data $\{y_i, z_i\}_{i=1}^n$, the random variation in $\tilde{u}_n(\gamma, \theta)$ and $\tilde{u}_n(\gamma, \hat{\theta})$ results from the randomness in ζ_i , $i = 1, \dots, n$. Theorems 2 and 3 below give a theoretical justification for the proposed resampling procedure.

Theorem 2. Under the null hypothesis and Conditions B1–B4, as $n \rightarrow \infty$, the unconditional distribution of the random process $n^{-1/2} u_n(\gamma, \hat{\theta})$ in γ is asymptotically equivalent to the conditional distribution of the random process $n^{-1/2} \tilde{u}_n(\gamma, \theta)$ in γ , given the sample data.

We further use asymptotic results on the conditional distribution of process $n^{-1/2} \tilde{u}_n(\gamma, \theta)$, $\gamma \in \Gamma$, given sample data to establish the desired limiting distribution of process $n^{-1/2} u_n(\gamma, \hat{\theta})$, $\gamma \in \Gamma$. This connection is established through the process $n^{-1/2} \tilde{u}_n(\gamma, \hat{\theta})$, $\gamma \in \Gamma$.

Theorem 3. Under the null hypothesis and Conditions B1–B4, as $n \rightarrow \infty$, the conditional distribution of the random process $n^{-1/2} \tilde{u}_n(\gamma, \hat{\theta})$ in γ is asymptotically equivalent to the conditional distribution of the random process $n^{-1/2} \tilde{u}_n(\gamma, \theta)$ in γ , given the sample data.

We give below details on the resampling procedure to obtain the large sample distribution of the proposed test T_n . For $k = 1, \dots, K$, we execute the following steps.

Step 1. We generate i.i.d standard normal variates $\{\zeta_1^{(k)}, \dots, \zeta_n^{(k)}\}$, with superscript (k) indexing replications.

Step 2. Given observed data $\{y_i, z_i\}_{i=1}^n$ and values of $\gamma \in \Gamma$, we calculate the statistic

$$\tilde{u}_n^{(k)}(\gamma, \hat{\theta}) = \sum_{i=1}^n c_i(\gamma, \hat{\theta}) \zeta_i^{(k)}, \text{ where } \hat{\theta} \text{ is a consistent estimator of } \theta \text{ under the null.}$$

Then we calculate $\tilde{s}_n^{(k)}(\gamma)$ by replacing $u_n(\gamma, \hat{\theta})$ by $\tilde{u}_n^{(k)}(\gamma, \hat{\theta})$ in expression (2).

Step 3. We calculate the artificial statistic $\tilde{T}_n^{(k)} = \sup_{\gamma \in \Gamma} \tilde{s}_n^{(k)}(\gamma)$.

By repeatedly generating the normal variates $\{\zeta_1, \dots, \zeta_n\}$, say K times, and repeating steps 2 and 3 for each generated sample, we obtain the empirical distribution of $\tilde{T}_n^{(k)}$, $k = 1, \dots, K$, which converges to the asymptotic distribution of T_n . By the Glivenko-Cantelli Theorem, when K is large, an approximate p-value of the test is $K^{-1} \sum_{k=1}^K \mathbf{1}(\tilde{T}_n^{(k)} \geq t_n)$ the proportion of these artificial statistics which exceed t_n the observed value of the statistic given data $\{y_i, z_i\}_{i=1}^n$.

2.6. Specification of the support set Γ

The calculation of the statistic T_n and artificial statistics $\tilde{T}_n^{(k)}$, $k = 1, \dots, K$, requires Γ the support set of γ to be specified. Ideally, the analyst may use historical information or unrelated data to determine this set. In practice, however, this information is often unavailable to the analyst. Alternatively, two strategies can be adopted. In the first strategy, the analyst may elect to choose plausible values of γ to conduct the test using the full sample. One would expect this approach to work well if the chosen set contains the true parameter and poorly otherwise. Due to this uncertainty, we recommend that a sensitivity analysis be conducted with respect to the choice of Γ , with a special focus on wider sets. Of course, wider sets may render the test unnecessarily conservative and thereby may sacrifice power. Most importantly, extreme values of γ may compromise the validity of the test by violating Assumption 2 that imposes Γ to be bounded. Another strategy consists of randomly splitting the sample into a smaller planning sample of size ρn and a larger analysis sample of size $(1 - \rho)n$, where $0 < \rho < 1$. The planning sample of size ρn is used to empirically approximate Γ , using say a 99% joint confidence region for γ under the alternative model, then discarded, and the remaining sample to conduct the test. This approach was recently illustrated by Heller et al. (2009) in the context of observational studies, where the planning sample is used to guide decisions about the design in a such a way that the study would be less sensitive to unobserved biases. In the sequel, Γ_ρ shall represent the set Γ obtained from

the planning sample of size pn , and by our convention, Γ_0 shall represent the set not related to the sample.

3. Numerical Studies

We illustrate the use of the proposed quasi-score statistic to evaluate homogeneity in zero-inflated models for discrete data (Lambert, 1992), a particular but representative special case. This class of models have a distribution of the form, $dP_{\theta,\phi} = \phi dG_0 + (1 - \phi)dP_{\theta,0}$, $\phi \in [0, 1)$, is an unknown mixing proportion, dG_0 is an atom at 0 and $dP_{\theta,0}$ is an unknown non-degenerate discrete distribution known up to a parameter θ . For the analysis of the dental caries data in Section 3.1, $dP_{\theta,0}$ is a negative binomial model. In the numerical experiment, however, $dP_{\theta,0}$ is simply assumed to be a Poisson model. In these numerical studies, the estimating function is the score function derived from parametric likelihood. We conduct the proposed hypothesis tests by replacing the parameter θ by its maximum likelihood estimate (MLE) $\hat{\theta}$ under the null. We also approximate Γ the space of γ by a fine grid to ease the computational burden. The choice of the size of the grid depends on the computational feasibility.

3.1. Dental caries data

We consider data from the Detroit Center for Research on Oral Health Disparities of the University of Michigan, Ann Arbor, which were reported by Tellez et al. (2006). There were 897 low-income African American children (0–5 years) surveyed between 2002–2003 with their DMFS (number of Decayed, Missing and Filled Surfaces) dental caries scores collected. These scores represent the cumulative severity of dental caries experience for each study participant. They have well documented shortcomings regarding their ability to provide a more refined measure of dental caries, but continue to be instrumental in evaluating and comparing the risks of dental caries across population groups and conducting historical comparisons (Lewsey and Thomson, 2004). Covariates considered for our numerical computations include the child's age in years (Age), the child's sugar intake (SI) and their multiplicative interaction (Age*SI).

Following our previous analyses (Todem et al., 2012), zero-inflated negative binomial (ZINB) models have been found to describe reasonably well the dental caries scores in this population. As a basic starting model, we consider a negative binomial (NB) model with mean $\lambda = \exp\{\beta^T (1, x^T)^T\}$ and overdispersion parameter κ , where $\beta = (\beta_0, \beta_1, \beta_2, \beta_3)^T$ and $x = (\text{Age}, \text{SI}, \text{Age} * \text{SI})^T$. This basic model is evaluated against its zero-inflated counterpart with the mixing probability model $g(a + \gamma^T x) = \{1 + \exp\{-a - \gamma^T x\}\}^{-1}$, $\gamma = (\gamma_1, \gamma_2, \gamma_3)^T$, assuming an unstratified population, $J = 1$. For this application, we will use the transformation $\psi = \exp\{a\}$, and use the DMFS score for y , $v = 1$, $z = x$, and $\theta = \{\beta, \kappa\}$. To perform the proposed test, the two approaches for selecting the set Γ are contrasted. Set Γ is set *a priori* to either $[-10, 10]^{\times 3}$ or $[-5, 5]^{\times 3}$, with $A^{\times 3} = A \times A \times A$, or approximated using 99% simultaneous confidence regions of γ using 1/4 or 1/3 of the total sample size. Computationally, the grid size of each dimension of γ is set to 0.4 for the two preselected sets Γ and 0.1 for the split sample approach. Fine grids of Γ are also considered. For comparison purposes, score tests with constant mixing proportions (when $\Gamma = \{(0, 0, 0)\}$)

are also performed using the usual mixture of a chi-square distribution with 1 degree of freedom and a point mass at 0, with weight 1/2 each. Likewise, the test proposed by Jansakul and Hinde (2002) that assumes the identity link function $g(a + \gamma^T x) = a + \gamma^T x$ is also performed. Under this parameterization, the homogeneity hypothesis simply reduces to $\alpha = 0$, $\gamma = 0$ and the limiting distribution of the test statistic is a chi-square with 4 degrees of freedom. We set $K = 10,000$ resamples to approximate the null distribution of the proposed test statistic. Results of this analysis are presented in Table 1.

The observed value of the test statistic appears virtually insensitive to the two pre-selected sets of γ and the grid size. In view of the resampling variation, the p-values of the tests are also insensitive to a wider set Γ , and all remain significant at 5% significance. In practice, contemplating such extreme values of γ may lead to an overly conservative test. Interestingly, in this example, the evidence of heterogeneity is rather strong, so that the supremum test rejects over a pre-specified wide region. The split sample approach appears to be sensitive to the pre-specified sample fractions used to approximate the set Γ , but yields significant p-values. Another evidence of heterogeneity in this population is provided by the test of Jansakul and Hinde (2008) which also rejects homogeneity at 5% significance level. However, the classical test that neglects covariate information under heterogeneity fails to reject the homogeneity hypothesis at 5% significance level. These data coupled with the working models provide an excellent example where the test based on covariate-varying alternatives greatly improves efficiency in detecting heterogeneity in the population. Estimates of the three models (the NB model, the ZINB model with weights depending on covariates, and the ZINB model with a constant mixing proportion) being evaluated are provided in Table 3.1. From this analysis, both the ZINB with constant mixing proportion and the homogeneous NB model have the same log-likelihood values leading to the conclusion that there is no evidence from data to reject the latter model. But the ZINB model based on covariate-varying alternatives greatly improves efficiency in detecting an age-dependent heterogeneity in these data (significant effect of age on ϕ).

It is worth noting that failure to reject homogeneity does not give evidence that the zero-inflated negative binomial model with varying mixing proportions provides the best fit for these data. Instead, such rejection only gives grounds for this model to be further evaluated against other competing models. One example of such models is the marginal mixture model that allows for negative mixing proportion to account for zero-deflation and therefore does not maintain the hierarchical representation of the mixture. A thorough investigation of this marginal model is outside the scope of this article but can be found elsewhere (see, for example, Todem et al., 2012).

3.2. Simulation study

We conduct a numerical study to evaluate the finite sample performance of the test statistic T_n assuming only continuous covariates in the mixture proportions. The performance is compared to that of the test statistic proposed by Jansakul and Hinde (2002) and to that of the test statistic that neglects parameter heterogeneity under the alternative. Because no categorical covariate is assumed in the mixture proportions, the test based on the constancy assumption is a simple version of the proposed test obtained by restricting the support set Γ

in (3) to the origin, $\Gamma = \{0\}$. In situations where γ^* the true γ is known, the performance of T_n is further compared to that of the statistic evaluated on the singleton set $\Gamma = \{\gamma^*\}$ and on a set that does not contain the true γ^* .

To keep the simulation simple, we consider a Poisson model with true mean $\lambda^* = \exp\{\theta_0^* + 0.04x_1 + 0.05x_2\}$, where x_1 and x_2 are generated independently from a uniform random variable on the interval $[0, 1]$ and a standard normal variable truncated on the interval $[-1, 1]$, respectively. The intercept parameter θ_0^* takes values in the set $\{-0.4, -0.3, -0.2\}$, and captures the separation between the two components of the mixture model. The MLE $\hat{\theta}$ of θ^* is obtained from a Poisson model with mean $\lambda = \exp\{\theta^\top(1, x^\top)^\top\}$, where $\theta = (\theta_0, \theta_1, \theta_2)^\top$ and $x = (x_1, x_2)^\top$. Throughout our simulations, we perform the test T_n assuming the mixing probability model $g(\alpha + \gamma^\top x) = \{1 + \psi^{-1} \exp\{-\gamma^\top x\}\}^{-1}$, with $\psi = \exp\{\alpha\}$ and $\gamma = (\gamma_1, \gamma_2)^\top$. We also perform the test proposed by Jansakul and Hinde (2002) assuming the identity link function $g(\alpha + \gamma^\top x) = \alpha + \gamma^\top x$. In our calculations, we set K the number of resamples to approximate the null distribution of T_n to 1,000. Finally, all simulations were replicated 1,000 times and for sample sizes 100, 200 and 400.

The proposed supremum tests have conservative type I error rates at 5% level for small sample sizes, regardless of the value of θ_0^* (see Table 3). In other words, these tests tend to reject the null hypothesis less often than anticipated. This conservativeness, however, diminishes as the sample size increases. These findings are consistent with the literature in that the supremum test can be somewhat conservative in some practical situations (see, for example, Hansen, 1996). In contrast, the test proposed by Jansakul and Hinde (2002) and the test that ignores varying heterogeneity under the alternative both maintain their size, regardless of the sample size and the value of θ_0^* .

We further investigate the empirical power of the tests under various specifications of the true mixing proportion ϕ^* . We specifically consider the situations where the true mixing probability; 1) does depend on covariates via a logistic transformation consistent with the working model; 2) does not depend on covariates; or 3) does depend on covariates via a (log-log) transformation misspecified in view of the working model (Table 4). The proposed test is conducted under the assumption that the working mixing probability is related to covariates through a logistic transformation. Overall, larger sample sizes and increasing values of θ_0^* (well separated mixture components) improve the power of detecting the alternatives under consideration for all tests considered. Under a correct specification of the working mixing probability, the power increases for tighter pre-selected sets of γ , with the highest power being achieved when Γ is restricted to the singleton $\{(-0.5, 3.5)\}$, the true γ . Tests based on varying alternatives, including the test of Jansakul and Hinde (2002), appear to be consistently more powerful than their constant-based counterparts. For our proposed test, this is particularly true whenever the pre-selected set Γ contains the true parameter. Note however that these results are reversed when pre-selected Γ does not contain the true parameter γ^* or does contain γ^* but is overwhelming wide. These are situations where sample splitting is proven to be useful in calibrating the set Γ . Indeed for moderate to large sample sizes, the split sample approach outperforms the test based on constant mixing

proportions, when the true proportion depends on covariates. Its usefulness, however, can be limited when the original sample size is small. This is expected because sample splitting reduces power by discarding observations, which has a dramatic impact on inferences when the original sample is small. Another limitation of the split sample approach is that it requires the alternative model to be fit which could be a daunting task in many practical settings, especially for small to moderate samples.

The rationale of considering tests that rely on varying alternatives builds on the behavior of such tests under model misspecification. When the true mixing proportion does not vary with the covariate profile, we have noticed a slight deterioration of power of tests that rely on varying heterogeneity. But when the true mixing proportion depends on covariates, the test based on varying heterogeneity even with a misspecified link function appears to be more powerful than its counterpart that ignores this parameter heterogeneity under alternatives.

4. Remarks

The issue of evaluating homogeneity in models that accommodate heterogeneity in the population has attracted considerable attention in the literature, but most of this debate has focused primarily on restrictive assumptions on the form of the alternative. One common restriction is that the heterogeneity parameter is often assumed constant under the maintained hypothesis. This paper extends this literature by developing a statistic that does not require fitting the alternative model to evaluate homogeneity against covariate-varying heterogeneity. Due to identifiability concerns, alternative approaches that require the models to be fitted under the alternative may not be used. This is particularly true if the heterogeneity parameter depends upon covariates, which is very typical in real applications of these models. Hall & Shen (2010) emphasized the importance of this issue by presenting a simple example of a non-identifiable zero-inflated Poisson model with covariate dependent mixing probabilities.

The testing methodology does not require the full distribution of the data to be entirely specified. Instead, a general estimating function for a finite dimensional component of the model that is of interest is assumed but other characteristics of the population which may be described by infinite dimensional parameters are left completely unspecified. In situations where the estimating function is derived from a likelihood function as it is the case in our numerical studies, the proposed test and the likelihood ratio test (LRT) are asymptotically equivalent. For a complete discussion on this issue, one should refer to the work of Ritz and Skovgaard (2005). We should emphasize however that the LRT may be more appealing computationally when a likelihood function is readily available.

The proposed test requires the analyst to specify a working model for the heterogeneity component of the model. This specification involves the selection of variables v and x which can be a daunting task in some applications. Our recommendation is to use the subject matter knowledge as a guideline for selecting these covariates. For the dental caries data, age is a natural covariate to include in the model for the probability of excess zeros. Younger children, typically under the age of 1, contribute most to the frequency of zero caries indices

due to a smaller number of teeth and a shorter exposure time for caries development compared to their older counterparts. In situations where the science is not very informative, the analyst should be prepared to consider the most elaborate model that is computationally less demanding in view of the data.

The paper did not discuss the asymptotic optimality properties of the proposed score-type tests. By asymptotic optimality, we mean tests that are uniformly most powerful among tests of asymptotic significance level $0 < \alpha < 1$, for testing the null against contiguous alternatives. Other authors have established these asymptotic properties in other contexts. For example, Andrews and Ploberger (1994) studied asymptotic optimal results for the likelihood ratio test, and more recently Song et al., (2009) established more general results using the score, Wald and likelihood ratio statistics in the context of semiparametric models. It might be possible to extend these results that have appeared in the literature to our proposed test statistics. But this endeavor merits further research.

Although the paper focuses primarily on smooth estimating functions and a $\sqrt{n}$ convergence rate, results for non-smooth functions converging at rates below $\sqrt{n}$ can be derived in the same spirit as Lee et al. (2011). And the resampling procedure adapted accordingly. The methodology may also be extended to general alternatives that do not assume a specific form for heterogeneity. Such an extension may be more robust against any misspecified alternatives, for example when an important covariate is missing from the working model under the alternative. These generalizations of the methodology are also the subject of further research.

Acknowledgments

The authors thank the Associate Editor and the referees for their helpful comments and constructive suggestions. The authors are also grateful to Amid Ismail and Woosung Sohn for their permission to use the dental caries data. This work was supported by the first author's NCI/NIH K-award, 1K01 CA131259 and its supplement from the 2009 ARRA funding mechanism.

Appendix

A1. Large sample properties

To prove Theorem 1, we first state and prove the following lemma which is of interest itself.

Lemma 1

Under the null and Conditions B1–B2, as $n \rightarrow \infty$, $\sup_{\gamma \in \Gamma} \|n^{-1} u_n(\gamma, \theta)\| = o_p(1)$.

Proof of Lemma 1

Condition B2 implies that $\mathcal{B}$ is Donsker and hence Glivenko-Cantelli (van der Vaart and Wellner, 2000a). Therefore, $\sup_{\gamma \in \Gamma, \theta \in \Theta} \|n^{-1} \sum_{i=1}^n b_i(\gamma, \theta) - E\{b_1(\gamma, \theta)\}\| \rightarrow_p 0$. From Condition B1 that requires θ to be an interior point of Θ , we immediately have $\sup_{\gamma \in \Gamma} \|n^{-1} \sum_{i=1}^n b_i(\gamma, \theta) - E\{b_1(\gamma, \theta)\}\| \rightarrow_p 0$. Moreover, by noting that $E\{b_1(\gamma, \theta)\} = 0$, for

all $\gamma \in \Gamma$ and that $n^{-1}u_n(\gamma, \theta) = n^{-1}\sum_{i=1}^n b_i(\gamma, \theta)$, we have the desired uniform consistency result.

Proof of Theorem 1

i) Uniform consistency of process $n^{-1}u_n(\gamma, \hat{\theta})$, $\gamma \in \Gamma$ —Using the Taylor expansion, we have

$$u_n(\gamma, \hat{\theta}) = u_n(\gamma, \theta) - h_n(\gamma, \check{\theta})(\hat{\theta} - \theta), \quad (4)$$

where $h_n(\gamma, \check{\theta}) = \sum_{i=1}^n d_i(\gamma, \check{\theta})$ and $\check{\theta}$ belongs to the line segment between θ and $\hat{\theta}$. That is, $\check{\theta} = \alpha^*\theta + (1 - \alpha^*)\hat{\theta}$ for some $\alpha^* \in [0, 1]$. This expansion is possible because of Condition B1 that ensures the true value of θ is an interior point of Θ . By the triangular inequality, we then have $\|n^{-1}u_n(\gamma, \hat{\theta})\| \leq \|n^{-1}u_n(\gamma, \theta)\| + \|n^{-1}h_n(\gamma, \check{\theta})(\hat{\theta} - \theta)\|$. From Lemma 1, $\sup_{\gamma \in \Gamma} \|n^{-1}u_n(\gamma, \theta)\| \rightarrow_p 0$ as $n \rightarrow \infty$. We focus on $\|n^{-1}h_n(\gamma, \check{\theta})(\hat{\theta} - \theta)\|$. Note that, $\|n^{-1}h_n(\gamma, \check{\theta})(\hat{\theta} - \theta)\| \leq \|\{n^{-1}h_n(\gamma, \check{\theta}) - h(\gamma, \check{\theta})\}(\hat{\theta} - \theta)\| + \|h(\gamma, \check{\theta})(\hat{\theta} - \theta)\|$. We define the norm of matrix M as $\|M\|_2 = \sup_{z \neq 0} \|Mz\|/\|z\|$. Hence, we have for any $z \neq 0$, $\|M\|_2 \geq \|Mz\|/\|z\|$ and subsequently $\|Mz\| \leq \|M\|_2\|z\|$. Applying this inequality to matrices $n^{-1}h_n(\gamma, \check{\theta}) - h(\gamma, \check{\theta})$ and $n^{-1}h_n(\gamma, \check{\theta})$ and assuming that $\hat{\theta} - \theta \neq 0$ w.p.1, we get,

$$\|n^{-1}h_n(\gamma, \check{\theta})(\hat{\theta} - \theta)\| \leq \{\|n^{-1}h_n(\gamma, \check{\theta}) - h(\gamma, \check{\theta})\|_2 + \|h(\gamma, \check{\theta})\|_2\} \|(\hat{\theta} - \theta)\|. \quad (5)$$

From Condition B2, $\mathcal{D}$ is Donsker and hence Glivenko-Cantelli (van der Vaart and Wellner, 2000a). Therefore, $\sup_{\gamma \in \Gamma, \theta \in \Theta} \|n^{-1}h_n(\gamma, \theta) - h(\gamma, \theta)\|_2 \rightarrow_p 0$ as $n \rightarrow \infty$ and in particular, $\sup_{\gamma \in \Gamma} \|n^{-1}h_n(\gamma, \check{\theta}) - h(\gamma, \check{\theta})\|_2 \rightarrow_p 0$. From Condition B4, it is also clear that $\sup_{\gamma \in \Gamma} \|h(\gamma, \check{\theta})\|_2 < \infty$. Condition A1 gives $\|\hat{\theta} - \theta\| \rightarrow_p 0$ which coupled with inequality (5), leads to the uniform consistency of $n^{-1}u_n(\gamma, \hat{\theta})$.

ii) Asymptotic distribution of process $n^{-1/2}u_n(\gamma, \hat{\theta})$, $\gamma \in \Gamma$ —Based on the uniform consistency of $n^{-1}u_n(\gamma, \hat{\theta})$ and applying the Taylor expansion to $u_n(\gamma, \hat{\theta})$, $u_n(\gamma, \hat{\theta}) \approx u_n(\gamma, \theta) - h_n(\gamma, \theta)(\hat{\theta} - \theta)$, where $h_n(\gamma, \theta) = \sum_{i=1}^n d_i(\gamma, \theta)$. Under Condition B1, we have $n^{1/2}(\hat{\theta} - \theta) = n^{-1/2}A^{-1}v_n(\theta) + o_p(1)$. This then leads to, $n^{-1/2}u_n(\gamma, \hat{\theta}) \approx n^{-1/2}u_n(\gamma, \theta) - n^{-1}h_n(\gamma, \theta)n^{-1/2}A^{-1}v_n(\theta)$. From Condition B2, $\mathcal{D}$ is Donsker and hence Glivenko-Cantelli (van der Vaart and Wellner, 2000a). Therefore, $\sup_{\gamma \in \Gamma, \theta \in \Theta} \|n^{-1}h_n(\gamma, \theta) - h(\gamma, \theta)\|_2 \rightarrow_p 0$ as $n \rightarrow \infty$ and in particular, $\sup_{\gamma \in \Gamma} \|n^{-1}h_n(\gamma, \theta) - h(\gamma, \theta)\|_2 \rightarrow_p 0$. Hence, $n^{-1/2}u_n(\gamma, \hat{\theta}) \approx n^{-1/2}u_n(\gamma, \theta) - h(\gamma, \theta)n^{-1/2}A^{-1}v_n(\theta)$. From the definition of $u_n(\gamma, \theta)$ and $v_n(\theta)$, the previous approximation in distribution can be rewritten,

$$n^{-1/2}u_n(\gamma, \hat{\theta}) \approx n^{-1/2}\sum_{i=1}^n \{b_i(\gamma, \theta) - h(\gamma, \theta)A^{-1}a_i(\theta)\}.$$

Under Condition B2, the function class $\mathcal{B}$ is Donsker and because of Condition B4, $h(\gamma, \theta)$ is uniformly bounded for $\gamma \in \Gamma$, the function classes $\mathcal{C} = \{b(\gamma, \theta) - h(\gamma, \theta)A^{-1}a(\theta), \gamma \in \Gamma$

Γ , $i = 1, 2, \dots, n$, are also Donsker. Therefore, $n^{-1/2}u_n(\gamma, \hat{\theta})$, random process in γ , converges in distribution to a centered Gaussian process as $n \rightarrow \infty$ with covariance kernel $\sigma_c^*(\gamma_1, \gamma_2) = E\{c_1(\gamma_1, \theta)c_1(\gamma_2, \theta)^\top\}$, $\gamma_1, \gamma_2 \in \Gamma$, where $c_1(\gamma, \theta) = b_1(\gamma, \theta) - h(\gamma, \theta)A^{-1}a_1(\theta)$, $\gamma \in \Gamma$.

A2. Asymptotic distribution of the resampling procedure

Proof of Theorem 2

To prove Theorem 2, we need to show that the conditional distribution of process $n^{-1/2}\tilde{u}_n(\gamma, \theta)$, $\gamma \in \Gamma$ given sample data $\{y_i, z_i\}_{i=1}^n$ converges asymptotically to a centered Gaussian process as $n \rightarrow \infty$ with covariance kernel $\sigma_c^*(\gamma_1, \gamma_2)$, $\gamma_1, \gamma_2 \in \Gamma$. Note that given observed data $\{y_i, z_i\}_{i=1}^n$, $n^{-1/2}\tilde{u}_n(\gamma, \theta)$ is a Gaussian process with a conditional covariance function,

$$\text{Cov}(n^{-1/2}\tilde{u}_n(\gamma_1, \theta), n^{-1/2}\tilde{u}_n(\gamma_2, \theta)) = n^{-1}\text{Cov}(\sum_{i=1}^n c_i(\gamma_1, \theta)\zeta_i, \sum_{i=1}^n c_i(\gamma_2, \theta)\zeta_i).$$

From the independence assumption we have,

$$\text{Cov}(n^{-1/2}\tilde{u}_n(\gamma_1, \theta), n^{-1/2}\tilde{u}_n(\gamma_2, \theta)) = n^{-1}\sum_{i=1}^n \text{Cov}(c_i(\gamma_1, \theta)\zeta_i, c_i(\gamma_2, \theta)\zeta_i), \gamma_1, \gamma_2 \in \Gamma.$$

Note that, $n^{-1}\sum_{i=1}^n \text{Cov}(c_i(\gamma_1, \theta)\zeta_i, c_i(\gamma_2, \theta)\zeta_i) = n^{-1}\sum_{i=1}^n \{c_i(\gamma_1, \theta)E\{\zeta_i^2\}c_i(\gamma_2, \theta)^\top\}$, with $E\{\zeta_i^2\} = 1$. Using Condition B3, the random process $n^{-1}\sum_{i=1}^n \{c_i(\gamma_1, \theta)c_i(\gamma_2, \theta)^\top\}$ indexed by γ_1, γ_2 , converges to $\sigma_c^*(\gamma_1, \gamma_2)$, uniformly over Γ . We have the desired result in view of Theorem 1.

Proof of Theorem 3

Given sample data $\{y_i, z_i\}_{i=1}^n$, we show that under H_0 and Conditions B1–B4, the conditional distribution of process $n^{-1/2}\sum_{i=1}^n c_i(\gamma, \hat{\theta})\zeta_i$, $\gamma \in \Gamma$ is asymptotically equivalent to the conditional distribution of process $n^{-1/2}\sum_{i=1}^n c_i(\gamma, \theta)\zeta_i$, $\gamma \in \Gamma$. Using the Taylor expansion, we have, $c_i(\gamma, \hat{\theta}) \approx c_i(\gamma, \theta) - d_i(\gamma, \theta)(\hat{\theta} - \theta)$. This then gives, $n^{-1/2}\sum_{i=1}^n c_i(\gamma, \hat{\theta})\zeta_i \approx n^{-1/2}\{\sum_{i=1}^n c_i(\gamma, \theta)\zeta_i - \sum_{i=1}^n d_i(\gamma, \theta)(\hat{\theta} - \theta)\zeta_i\}$. From the previous approximation, we consider the quantity, $n^{-1/2}\sum_{i=1}^n d_i(\gamma, \theta)(\hat{\theta} - \theta)\zeta_i = \{n^{-1}\sum_{i=1}^n d_i(\gamma, \theta)\zeta_i\}\{n^{-1/2}(\hat{\theta} - \theta)\}$. From Condition A1, we have $n^{1/2}(\hat{\theta} - \theta) = O_p(1)$, where $O_p(1)$ represents the boundedness in probability. Function classes $\{d_i(\gamma, \theta)\zeta_i, \gamma \in \Gamma, i = 1, 2, \dots, n\}$ are Donsker due to $\mathcal{D}$ being Donsker and $\zeta_i = O_p(1)$. For fixed data $\{y_i, z_i\}_{i=1}^n$, we then have $\sup_{\gamma \in \Gamma} \|n^{-1}\sum_{i=1}^n d_i(\gamma, \theta)\zeta_i\| \rightarrow_p 0$ as $n \rightarrow \infty$. Given sample data $\{y_i, z_i\}_{i=1}^n$, this finally gives, $n^{-1/2}\sum_{i=1}^n c_i(\gamma, \hat{\theta})\zeta_i \approx n^{-1/2}\sum_{i=1}^n c_i(\gamma, \theta)\zeta_i$, the desired result.

References

- Andrews DWK. Testing when a parameter is on the boundary of the maintained hypothesis. *Econometrica*. 2001; 69:683–734.
- Andrews DWK, Ploberger W. Optimal tests when a nuisance parameter is present only under the alternative. *Econometrica*. 1994; 62:1383–1414.
- Beg ABMRA, Silvapulle MJ, Silvapulle P. Tests against inequality constraints when some nuisance parameters are present only under the alternative: Test of ARCH in ARCH-M models. *Journal of Business & Economic Statistics*. 2001; 19:245–253.
- Besag J. Statistical analysis of non-lattice data. *The Statistician: Journal of the Institute of Statisticians*. 1975; 24:179–196.
- Chernoff H. On the distribution of the likelihood ratio. *The Annals of Mathematical Statistics*. 1954; 25:573–578.
- Davies RB. Hypothesis testing when a nuisance parameter is present only under the alternative. *Biometrika*. 1977; 64:247–254.
- Dean CB. Testing for overdispersion in Poisson and binomial regression models. *Journal of the American Statistical Association*. 1992; 87:451–457.
- Efron B, Tibshirani R. *An Introduction to the Bootstrap*. Chapman & Hall Ltd; 1993.
- 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]
- Hall DB, Shen J. Robust estimation for zero-inflated Poisson regression. *Scandinavian Journal of Statistics*. 2010; 37:237–252.
- Hansen BE. Inference when a nuisance parameter is not identified under the null hypothesis. *Econometrica*. 1996; 64:413–430.
- Heller R, Rosenbaum PR, Small DS. Split Samples and Design Sensitivity in Observational Studies. *Journal of the American Statistical Association*. 2009; 104:1090–1101.
- Jansakul N, Hinde J. Score tests for zero-inflated Poisson models. *Computational Statistics and Data Analysis*. 2002; 40:75–96.
- Jansakul N, Hinde JP. Score tests for extra-zero models in zero-inflated negative binomial models. *Communications in Statistics - Simulation and Computation*. 2008; 38:92–108.
- Lambert D. Zero-inflated Poisson regression, with an application to defects in manufacturing. *Technometrics*. 1992; 34:1–14.
- Lee S, Seo MH, Shin Y. Testing for threshold effects in regression models. *Journal of the American Statistical Association*. 2011; 106:220–231.
- 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 socioeconomic status. *Community Dentistry and Oral Epidemiology*. 2004; 32:183–189. [PubMed: 15151688]
- Lindsay BG. Composite likelihood methods. In: Prabhu NU, editor *Statistical Inference from Stochastic Processes*. American Mathematical Society; 1988. 221–239.
- Mullahy J. Specification and testing of some modified count data models. *Journal of Econometrics*. 1986; 33:341–365.
- Parzen MI, Wei LJ, Ying Z. A resampling method based on pivotal estimating functions. *Biometrika*. 1994; 81:341–350.
- Ritz C, Skovgaard IM. Likelihood ratio tests in curved exponential families with nuisance parameters present only under the alternative. *Biometrika*. 2005; 92:507–517.
- Self SG, Liang K-Y. Asymptotic properties of maximum likelihood estimators and likelihood ratio tests under nonstandard conditions. *Journal of the American Statistical Association*. 1987; 82:605–610.
- Shapiro A. Towards a unified theory of inequality constrained testing in multivariate analysis. *International Statistical Review*. 1988; 56:49–62.

- Silvapulle MJ, Silvapulle P. A score test against one-sided alternatives. *Journal of the American Statistical Association*. 1995; 90:342–349.
- Song R, Kosorok MR, Fine JP. On asymptotically optimal tests under loss of identifiability in semiparametric models. *The Annals of Statistics*. 2009; 37:2409–2444. [PubMed: 20454527]
- Sy JP, Taylor JMG. Estimation in a Cox proportional hazards cure model. *Biometrics*. 2000; 56:227–236. [PubMed: 10783800]
- 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]
- Todem D, Hsu W-W, Kim K. On the efficiency of score tests for homogeneity in two-component parametric models for discrete data. *Biometrics*. 2012; 68:975–982. [PubMed: 22348298]
- van den Broek J. A score test for zero inflation in a Poisson distribution. *Biometrics*. 1995; 51:738–743. [PubMed: 7662855]
- van der Vaart AW, Wellner JA. Preservation theorems for Glivenko-Cantelli and uniform Glivenko-Cantelli theorems. *High Dimensional Probability II*. Giné E, Mason DM, Wellner JA, editors Birkhäuser, Boston: 2000a.
- van der Vaart AW, Wellner JA. Weak convergence and empirical processes. New York: Springer; 2000b.
- Verbeke G, Molenberghs G. The use of score tests for inference on variance components. *Biometrics*. 2003; 59:254–262. [PubMed: 12926710]
- Zhu H, Zhang H. Generalized score test of homogeneity for mixed effects models. *The Annals of Statistics*. 2006; 34:1545–1569.

Table 1

Observed test statistics (p-values, 5% critical points) for the dental caries data

Full sample		Random split samples		
$\Gamma = (0, 0, 0)$	JH	$\Gamma = \Gamma_0^{(1)}$	$\Gamma = \Gamma_0^{(2)}$	$\Gamma = \Gamma_{\frac{1}{4}}$
0.0123 (0.4627, 2.71)	126.781 (<0.0001, 9.49)	170.5356 (<0.0001, 7.84)	170.5356 (<0.0001, 7.80)	135.2439 (<0.0001, 6.25)

$\Gamma_0^{(1)} = [-10, 10] \times 3, \Gamma_0^{(2)} = [-5, 5] \times 3$

$\mathcal{F}_T$ Tests performed using the usual mixture of chi-square distributions

JH: Test of Jansakul and Hinde (2008)

Table 2

Parameter estimates, standard errors (SE) and p-values for Detroit dental caries data

<i>Effects</i>	NB model (null model)			ZINB with constant ϕ			ZINB with covariate-dependent ϕ		
	Estimate	SE	p-value	Estimate	SE	p-value	Estimate	SE	p-value
λ									
Intercept	-0.323	0.190	0.089	-0.285	0.259	0.273	1.588	0.171	<0.001
Age	0.591	0.055	<0.001	0.585	0.062	<0.001	0.151	0.044	<0.001
SI	0.841	0.223	<0.001	0.840	0.222	<0.001	0.190	0.190	0.319
Age*SI	-0.213	0.056	<0.001	-0.213	0.056	<0.001	-0.047	0.048	0.325
ϕ									
Intercept (α)				-3.852	4.628	0.405	2.514	0.391	<0.001
Age (γ_1)							-1.106	0.147	<0.001
SI (γ_2)							-0.798	0.450	0.077
Age*SI (γ_3)							0.204	0.130	0.118
Dispersion $\log(\kappa)$									
Intercept	0.962	0.067	<0.001	0.923	0.195	<0.001	-0.073	0.114	0.521
Summary statistics									
-2 logL		4059.1			4059.1			3815.4	
AIC		4069.1			4071.1			3833.4	

Empirical size of test statistics setting $\phi^* = 0$ and generating data from a Poisson model with mean $\lambda^* = \exp\{\theta_0^* + 0.04x_1 + 0.05x_2\}$, at 5% significance level

Table 3

Support of γ	$\theta_0^* = -0.4$			$\theta_0^* = -0.3$			$\theta_0^* = -0.2$		
	100	200	400	100	200	400	100	200	400
$\Gamma_0^{(1)}$	0.023	0.026	0.036	0.022	0.028	0.043	0.022	0.029	0.043
$\Gamma_0^{(2)}$	0.029	0.030	0.043	0.027	0.035	0.045	0.028	0.038	0.050
$\Gamma_{\frac{1}{4}}$	0.031	0.036	0.034	0.026	0.030	0.041	0.024	0.028	0.035
$\Gamma_{\frac{1}{3}}$	0.022	0.032	0.033	0.025	0.023	0.032	0.017	0.029	0.041
$\{(0, 0)\}^{\mathcal{F}}$	0.037	0.029	0.050	0.044	0.031	0.040	0.029	0.045	0.048
JH	0.044	0.052	0.052	0.054	0.057	0.056	0.043	0.044	0.049

$\Gamma_0^{(1)} = [9 \times 2, \Gamma_0^{(2)}] \times [-4, 4] \times 2$

$\mathcal{F}$ Tests performed using the usual mixture of chi-square distributions

JH: Jansakul and Hinde's test (2002)

Table 4

Empirical power of test statistics with data generated from a zero-mixture Poisson model withmixing proportion $\phi^*(x)$, and Poisson mean $\lambda^* = \exp\{\theta_0^* + 0.04x_1 + 0.05x_2\}$, at 5%significance level

Support of γ	$\theta_0^* = -0.4$			$\theta_0^* = -0.3$			$\theta_0^* = -0.2$		
	Sample size n			Sample size n			Sample size n		
	100	200	400	100	200	400	100	200	400
$\Gamma_0^{(1)}$	$\phi^*(x) = \{1 + 0.1^{-1} \exp\{0.5x_1 - 3.5x_2\}\}^{-1}$								
	0.138	0.400	0.773	0.201	0.441	0.794	0.223	0.535	0.871
$\Gamma_0^{(2)}$	0.183	0.448	0.814	0.260	0.504	0.839	0.281	0.594	0.911
	$\{(-0.5, 3.5)\}$								
$\Gamma_0^{(3)}$	0.486	0.695	0.938	0.535	0.755	0.937	0.571	0.815	0.969
	0.122	0.237	0.393	0.136	0.245	0.418	0.164	0.261	0.418
$\Gamma_{\frac{1}{4}}$	0.112	0.240	0.577	0.119	0.276	0.663	0.145	0.377	0.734
	$\frac{1}{4}$								
$\Gamma_{\frac{3}{3}}$	0.082	0.226	0.519	0.081	0.303	0.621	0.117	0.314	0.686
	$\frac{3}{3}$								
$\{(0, 0)\}^{\neq}$	0.121	0.213	0.345	0.156	0.219	0.368	0.155	0.255	0.432
	0.137	0.207	0.368	0.130	0.245	0.463	0.161	0.276	0.576
$\Gamma_0^{(1)}$	$\phi^* = 1/6$								
	0.072	0.156	0.277	0.093	0.171	0.339	0.104	0.220	0.429
$\Gamma_0^{(2)}$	0.092	0.197	0.332	0.124	0.207	0.383	0.139	0.269	0.498
	$\frac{1}{4}$								
$\Gamma_{\frac{1}{4}}$	0.054	0.108	0.199	0.084	0.119	0.234	0.070	0.151	0.308
	$\frac{1}{4}$								

Support of γ	$\theta_0^* = -0.4$			$\theta_0^* = -0.3$			$\theta_0^* = -0.2$		
	Sample size n			Sample size n			Sample size n		
	100	200	400	100	200	400	100	200	400
$\Gamma_{\frac{1}{3}}$	0.053	0.096	0.186	0.056	0.103	0.214	0.063	0.140	0.278
$\{(0, 0)\}^{\frac{1}{2}}$	0.163	0.308	0.483	0.199	0.320	0.552	0.245	0.409	0.668
JH	0.092	0.149	0.251	0.095	0.192	0.287	0.100	0.185	0.389
$\Gamma_0^{(1)}$	$\phi^*(x) = \exp\{-\exp\{\log(10) + 0.5x_1 - 3.5x_2\}\}$								
	0.145	0.315	0.664	0.138	0.391	0.730	0.187	0.435	0.798
$\Gamma_0^{(2)}$	0.161	0.334	0.670	0.166	0.409	0.743	0.221	0.453	0.807
$\Gamma_{\frac{1}{4}}$	0.072	0.212	0.478	0.083	0.240	0.585	0.087	0.297	0.605
$\Gamma_{\frac{1}{3}}$	0.053	0.172	0.460	0.076	0.233	0.517	0.094	0.254	0.574
$\{(0, 0)\}^{\frac{1}{2}}$	0.077	0.110	0.169	0.092	0.117	0.195	0.098	0.139	0.207
JH	0.077	0.110	0.149	0.092	0.108	0.178	0.069	0.140	0.238

$\Gamma_0^{(1)} = [-6, 6] \times \Gamma_0^{(2)} = [-4, 4] \times \Gamma_0^{(3)} = [-18, -10] \times [10, 18]$

$\frac{1}{2}$ Tests performed using the usual mixture of chi-square distributions

JH: Jansakul and Hinde's test (2002)