SMMR
STATISTICAL METHODS IN MEDICAL RESEARCH

Article

Statistical Methods in Medical Research
2020, Vol. 29(8) 2179–2197
© The Author(s) 2019
Article reuse guidelines:
sagepub.com/journals-permissions
DOI: 10.1177/0962280219885985
journals.sagepub.com/home/smm
SAGE

Statistical tests for latent class in censored data due to detection limit

Hua He[1], Wan Tang[2], Tanika Kelly[1], Shengxu Li[3] and Jiang He[1]

Abstract

Measures of substance concentration in urine, serum or other biological matrices often have an assay limit of detection. When concentration levels fall below the limit, the exact measures cannot be obtained. Instead, the measures are censored as only partial information that the levels are under the limit is known. Assuming the concentration levels are from a single population with a normal distribution or follow a normal distribution after some transformation, Tobit regression models, or censored normal regression models, are the standard approach for analyzing such data. However, in practice, it is often the case that the data can exhibit more censored observations than what would be expected under the Tobit regression models. One common cause is the heterogeneity of the study population, caused by the existence of a latent group of subjects who lack the substance measured. For such subjects, the measurements will always be under the limit. If a censored normal regression model is appropriate for modeling the subjects with the substance, the whole population follows a mixture of a censored normal regression model and a degenerate distribution of the latent class. While there are some studies on such mixture models, a fundamental question about testing whether such mixture modeling is necessary, i.e. whether such a latent class exists, has not been studied yet. In this paper, three tests including Wald test, likelihood ratio test and score test are developed for testing the existence of such latent class. Simulation studies are conducted to evaluate the performance of the tests, and two real data examples are employed to illustrate the tests.

Keywords

Censored normal regression, detection limit, latent class, likelihood ratio test, mixture Tobit model, score test, Tobit model, Wald test

1 Introduction

Measures of substance concentration in urine, serum or other biological matrices that fall below the assay limit of detection are pretty common in environmental science and medical research.[1–10] When the concentrations are under the limit of detection, accurate measures cannot be obtained. Instead, their values are only partially known and left censored. For example, triclosan is a broad-spectrum antimicrobial chemical and widely used in household and health care related products. Currently the detection limit for urine triclosan concentration is 2.3 ng/ml. Only triclosan concentrations greater than or equal to 2.3 ng/ml can be detected. For triclosan concentrations

[1]Department of Epidemiology, Tulane University School of Public Health and Tropical Medicine, New Orleans, LA, USA
[2]Department of Biostatistics and Data Science, Tulane University School of Public Health and Tropical Medicine, New Orleans, LA, USA
[3]Children's Minnesota Research Institute, Children's Hospitals and Clinics of Minnesota Medicine, Minneapolis, MN, USA

Corresponding author:
Hua He, Department of Epidemiology, Tulane University School of Public Health and Tropical Medicine, 1440 Canal Street, Suite 2014, New Orleans, LA 70112-2703, USA.
Email: hhe2@tulane.edu

lower than 2.3 ng/ml, the value is censored. Instead of a precise measure of the triclosan concentration, the value is only partially known, namely that is somewhere between 0 and 2.3. Methods for handling censored data due to detection limit include deletion or substitution with 0, the detection limit, or half or one-third of the detection limit. These methods are commonly used in practice despite their inappropriateness.[1,11–15]

When data are collected from a single normal distribution with some observations under the detection limit, a Tobit regression model should be applied.[4,16–22] The Tobit regression model is widely applied in economics,[23–27] and other fields such as medical research[10,28–32] and environmental research.[6,7,33–35] In cases where the data are not from a normal distribution, data transformation such as log-transformation can be employed first, and the Tobit regression model can then be applied on the transformed data.

A Tobit regression model assumes that the underlying latent continuous measure follows a single normal distribution, but the observed outcome is subject to censoring due to the detection limit. Given a distribution and detection limit, the proportion of under detection can be approximately determined. However, when there is a subgroup of subjects who don't have the substance at all, their measures are of course under the detection limit and thus are censored. In such case, the subgroup of subjects make up a latent class as their measures are always censored, and the data can exhibit more censored observations than what would be expected according to the Tobit regression model, and makes the Tobit regression model assumption violated. This latent class issue was acknowledged in Halsey et al.[36] Some methods such as a mixture model were proposed to address this issue.[37–39] However, a fundamental question about whether there is a latent class is not studied in the literature.

In this paper, three tests including the Wald test, likelihood ratio (LR) test and score test[40] are developed for testing if a latent class exists. A brief review of the Tobit regression models and mixture Tobit regression models are given first in Section 2, and the three tests are developed in Section 3. Simulation studies to investigate and compare the performance of tests are given in Section 4, and two real data examples to illustrate the methods are given in Section 5. The paper is concluded with a discussion in Section 6.

2 Tobit model and mixture Tobit model

Consider an independent sample $(\mathbf{x}_i, y_i)$, $i = 1, \ldots, n$, where $\mathbf{x}_i = (x_{i1}, x_{i2,} \ldots, x_{ip})$ is a p-dimensional covariate, and y_i is the observed censored measurement. The observed y_i is obtained based on a latent variable y_i^*, which is assumed to have a linear relationship with $\mathbf{x}_i$ through a parameter vector β, i.e.

$$y_i^* = \mathbf{x}_i^\top \beta + \varepsilon_i, \quad \varepsilon_{\mathrm{i}} \sim \mathrm{N}(0, \sigma^2) \tag{1}$$

Let L be the lower detection limit, and a Tobit model censored at L is defined as

$$y_i = \begin{cases} y_i^* & \text{if } y_i^* \geq L \\ L & \text{if } y_i^* < L \end{cases}$$

Due to censoring, the variable y_i^* cannot be measured (or detected) if its value is below L. In such cases, its value is substituted by the threshold L. Under the assumption that the underlying variable y_i^* follows a normal distribution with mean $\mu_i = \mathbf{x}_i^T \beta$, where the design matrix includes a constant for the intercept, then the observed y_i follows a Tobit model, denoted as $y_i \sim$ Tobit (μ_i, σ^2, L), and has the following distribution

$$\Pr(Y_i = y_i) = \begin{cases} \dfrac{1}{\sqrt{2\pi}\sigma} \exp\left(-\dfrac{(y_{i-}\mu_i)^2}{2\sigma^2}\right) & \text{if } y_i > L, \\ \Phi\left(\dfrac{L - \mu_i}{\sigma}\right) & \text{if } y_i = L \end{cases}$$

The Tobit regression model can be expressed as

$$y_i | \mathbf{x}_i \sim \text{i.d. Tobit } (\mu_{\mathrm{i}}, \sigma^2, \mathrm{L}), \quad \mu_{\mathrm{i}} = \mathbf{x}_{\mathrm{i}}^{\mathrm{T}} \beta \tag{2}$$

Let r_i be an indicator indicating whether y_i is censored or not, with $r_i = 1$ for $y_i = L$ and $r_i = 0$ for $y_i > L$. The likelihood function for the ith subject is given by

$$L_i = \left[\Phi\left(\frac{L - \mu_i}{\sigma}\right)\right]^{r_i}\left[\frac{1}{\sqrt{2\pi}\sigma}\exp\left(-\frac{(y_{i} - \mu_i)^2}{2\sigma^2}\right)\right]^{(1-r_i)} \tag{3}$$

where $\Phi(\cdot)$ is the cumulative distribution function of the standard normal. Given the likelihood in equations (3) and model (2), maximum likelihood method can be applied to estimate β and σ.

For the Tobit regression model, the proportion of under the detection limit is $E\left[\Phi\left(\frac{L-\mu_i}{\sigma}\right)\right]$, and can be estimated by $\frac{\sum_{i=1}^{n} r_i}{n}$. But when there is a latent class of subjects who don't have the substance at all, their measures are of course under the detection limit and thus are censored. In such case, the data can exhibit more censored observations than what would be expected according to the Tobit regression model, i.e. the proportion of under-detection is much higher than $E\left[\Phi\left(\frac{L-\mu_i}{\sigma}\right)\right]$.

When there is a latent class such as non-exposure, the Tobit regression model for a single population such as exposure is not appropriate to model such censored data anymore, and some methods such as a mixture model were proposed to address this issue.[37–39] Let ω be the probability of the latent class. Data from a mixture of the latent class with probability ω and Tobit model with probability $(1 - \omega)$, named mixture Tobit model and denoted as mTobit $(\omega, \mu_i, \sigma^2, L)$, has likelihood function for the ith subject expressed as

$$L_i = \left[\omega + (1-\omega)\Phi\left(\frac{L - \mu_i}{\sigma}\right)\right]^{r_i}\left[(1-\omega)\frac{1}{\sqrt{2\pi}\sigma}\exp\left(-\frac{(y_{i} - \mu_i)^2}{2\sigma^2}\right)\right]^{(1-r_i)} \tag{4}$$

The mTobit $(\omega, \mu_i, \sigma^2, L)$ is a mixture of a latent class with probability ω and a Tobit model with probability $(1 - \omega)$. Thus, the proportion of under detection now becomes $\omega + (1-\omega)E\left[\Phi\left(\frac{L-\mu_i}{\sigma}\right)\right]$, which is always greater than $E\left[\Phi\left(\frac{L-\mu_i}{\sigma}\right)\right]$ for $\omega > 0$. Therefore, ω is a parameter indicating the excessive observations under the detection limit.

If the probability of the latent class depends on some covariates, say $\mathbf{u}_i$, a generalized linear model such as a logistic regression model can usually be applied to model the latent class and a linear model is used to model the Tobit component. A mixture Tobit regression model can have the following form

$$y_i|\mathbf{x}_i \sim \text{i.d. mTobit}(\omega_\text{i}, \mu_\text{i}, \sigma^2, \text{L}), \quad \text{logit}(\omega_\text{i}) = \mathbf{u}_\text{i}^\top \beta_\omega, \quad \mu_\text{i} = \mathbf{x}_\text{i}^\text{T}\beta_\mu \tag{5}$$

The covariates $\mathbf{u}_i$ and $\mathbf{x}_i$ in the two components can be the same or different. If the probability ω does not depend on any covariates, i.e. ω is a constant, then there is no need for a link function, and thus the mTobit model in equation (5) can actually be represented as a linear regression model with

$$y_i|\mathbf{x}_i \sim \text{i.d. mTobit}(\omega_\text{i}, \mu_\text{i}, \sigma^2, \text{L}), \quad \omega_\text{i} = \omega, \quad \mu_\text{i} = \mathbf{x}_\text{i}^\text{T}\beta \tag{6}$$

Given the likelihood in equations (4) and model (5) or (6), we can apply the maximum likelihood method to obtain the MLE of β_ω, β_μ or β, as well as ω and σ.

Under equation (5), the probability ω_i is always positive, and thus the models (2) and (5) are not nested, and commonly used tests such as the Wald and LR tests cannot be directly applied to test the latent class. However, if we simply assume the probability ω_i is a constant as in equation (6), the Tobit regression model (2) is now nested in the mTobit regression model (6) as it corresponds to the cases with $\omega = 0$ under equation (6). Therefore, the Wald, LR and score tests for testing whether $\omega = 0$ can be applied. When ω is a constant and no link function is necessary as in equation (6), ω can have a negative value to imply that not only there is no latent class, but also the probability of data censored is lower than what would be expected under equation (2), i.e. the data exhibits less amount of observations under detection than what would be expected under the Tobit model. Thus, $\omega = 0$ is actually an interior point and we don't have the non-standard condition issue when ω can only take non-negative values.

3 Tests for the latent class

In this section, we will develop three tests, the Wald, LR, and score tests, for testing whether there is a latent class in a Tobit model.

3.1 Wald test

The Wald test for testing $H_0 : \omega = 0$ vs. $H_A : \omega \neq 0$ is developed based on the MLE estimate of ω under the mTobit model (6). Let $\mu_i = \mathbf{x}_i^T \beta$ be the mean of the Tobit component, and the log-likelihood for the ith subject is

$$l_i = r_i \log\left[\omega + (1-\omega)\Phi\left(\frac{L-\mu_i}{\sigma}\right)\right] + (1-r_i)\left[\log(1-\omega) - \log(\sqrt{2\pi}\sigma) - \frac{(y_i - \mu_i)^2}{2\sigma^2}\right]$$

So, the log-likelihood for the whole sample is

$$l(\omega, \beta, \sigma) = \sum\nolimits_{i=1}^{n} r_i \log\left[\omega + (1-\omega)\Phi\left(\frac{L-\mu_i}{\sigma}\right)\right] + (1-r_i)\left[\log(1-\omega) - \log(\sqrt{2\pi}\sigma) - \frac{(y_i - \mu_i)^2}{2\sigma^2}\right] \tag{7}$$

Taking the first derivative of $l(\omega, \beta, \sigma)$ □ corresponding to ω, β, σ, we have

$$S_{\beta_s} = \frac{\partial l}{\partial \beta_s} = \sum_{i=1}^{n}\left\{-\frac{r_i}{\varphi_i}(1-\omega)A_i\frac{x_{is}}{\sigma} + (1-r_i)\frac{y_i - \mu_i}{\sigma^2}x_{is}\right\}, \quad s = 1, 2, \ldots, p \tag{8}$$

$$S_{\sigma} = \frac{\partial l}{\partial \sigma} = \sum_{i=1}^{n}\left\{-\frac{r_i}{\varphi_i}A_i(1-\omega)\frac{L-\mu_i}{\sigma^2} + (1-r_i)\left[-\frac{1}{\sigma} + \frac{(y_i - \mu_i)^2}{\sigma^3}\right]\right\} \tag{9}$$

$$S_{\omega} = \frac{\partial l}{\partial \omega} = \sum_{i=1}^{n}\left\{\frac{r_i - \varphi_i}{(1-\omega)\varphi_i}\right\} \tag{10}$$

where $A_i = \frac{1}{\sqrt{2\pi}}\exp\left(-\frac{(L-\mu_i)^2}{2\sigma^2}\right)$ and $\varphi_i = \omega + (1-\omega)\Phi\left(\frac{L-\mu_i}{\sigma}\right)$. The MLE of ω, β and σ can be obtained by simultaneously solving $S_{\beta_s} = 0$, $s = 1, 2, \ldots, p, S_\sigma = 0$ and $S_\omega = 0$. The asymptotic variance of the MLE of ω can further be estimated by the Fisher information matrix. Let $\hat{\omega}$ be the MLE of ω under equation (6) and $\hat{\sigma}_\omega$ the estimated variance of $\hat{\omega}$. Then, the Wald statistic is defined as $S_{Wald} = \hat{\omega}^2/\hat{\sigma}_\omega^2$. Under $H_0 : \omega = 0$, the Wald statistic asymptotically follows a chi-square distribution, i.e. $S_{Wald} = \hat{\omega}^2/\hat{\sigma}_\omega^2 \sim \chi_1^2$. Since the parameter ω is only one-dimensional, it is equivalent to the Z-statistic

$$Z_{Wald} = \hat{\omega}/\hat{\sigma}_\omega \sim N(0, 1) \tag{11}$$

For a two-sided test against the alternative $H_A : \omega \neq 0$, i.e. the amount of observations under detection is different from what would be expected by the Tobit model in either direction, with a type I error α, we reject H_0 if $|Z_{Wald}| > Z_{\alpha/2}$, where $Z_{a/2}$ is the $100(1-\alpha/2)$ percentile of the standard normal. For a one-sided test with alternative $H_A : \omega > 0$, i.e. the amount of observations under detection is more than what would be expected under a Tobit regression model, we reject H_0 if $Z_{Wald} > Z_\alpha$.

It is worth to note that the Hessian matrix of mTobit models can be singular, and thus the MLEs may not exist. For example, if the mean $\mu_i's$ are small while the detection limit L is large, in which most of the outcomes are undetectable, or the $\mu_i's$ are large while L is small, in which there are too few observations under detection, the optimization of the log-likelihood of the mTobit models (6) is likely to fail, and thus the MLEs do not exist. In such cases, the Wald test cannot be applied.

3.2 Likelihood ratio test

The LR test is based on the likelihoods of the data under both models (6) and (2). Let $\hat{\omega}$, $\hat{\beta}$ and $\hat{\sigma}$ be the MLE of ω, β, and σ under equation (6), and $\hat{\beta}'$ and $\hat{\sigma}'$ be the MLE of β and σ under equation (2). Then, the corresponding log-likelihood at the MLEs for the two models are $l(\hat{\omega}, \hat{\beta}, \hat{\sigma})$ and $l(0, \hat{\beta}', \hat{\sigma}')$. The LR statistic would then be

$$S_{LR} = 2(l(\hat{\omega}, \hat{\beta}, \hat{\sigma}) - l(0, \hat{\beta}', \hat{\sigma}')) \sim \chi_1^2 \tag{12}$$

The LR statistic S_{LR} asymptotically follows a χ_1^2 distribution as there is only one more parameter in the mTobit model than the Tobit model. For a two-sided test with type I error α, we reject H_0 if $S_{LR} > \chi^2_{1,\ 1-\alpha}$, where $\chi^2_{1,1-\alpha}$ is the $100(1-\alpha)$ percentile of a χ^2 distribution. For a one-sided test with alternative $H_A : \omega > 0$, we may combine the test statistic with the estimate $\hat{\omega}$ by restricting on the cases where $\hat{\omega} > 0$. The H_0 is rejected when $\hat{\omega} > 0$ and $S_{LR} > \chi^2_{1,1-2\alpha}$. Note that the $100(1-2\alpha)$ percentile of a χ^2 distribution is used for type I error α because, under H_0, there is about a 50% chance for $\hat{\omega}$ to be negative or positive.

The LR test requires the fit of both the Tobit and mTobit models, as well as the existence of MLEs, and thus suffers the same issue as the Wald test.

3.3 Score test

Without actually fitting a mTobit model, the score statistic can be computed under $H_0 : \omega = 0$. The score functions are the first derivative of (7) corresponding to β, σ and ω, which is given by equations (8), (9) and (10). Under the null hypothesis $H_0 : \omega = 0$, and given the MLE $\hat{\beta}$ and $\hat{\sigma}$ of β and σ, based on equations (8), (9) and (10), we have $S_{\beta_s}|\hat{\beta} = 0, s = 1, 2, \ldots, p$, $S_\sigma|\hat{\sigma} = 0$, and $S_\omega|_{\hat{\beta},\hat{\sigma},\omega=0} = \sum_{i=1}^n \frac{r_i - \hat{B}_i}{\hat{B}_i}$, where $\hat{B}_i = \Phi\left(\frac{L-\hat{\mu}_i}{\hat{\sigma}}\right)$. Therefore the score function under H_0 and MLE of β and σ becomes

$$U^T(\hat{\beta}, \hat{\sigma}, 0) = \left(0, 0, \ldots, 0, 0, \sum_{i=1}^n \frac{r_i - \hat{B}_i}{\hat{B}_i}\right) \tag{13}$$

where $U^T(\hat{\beta}, \hat{\sigma}, 0)$ is a vector with $(p+2)$ elements with the first p zeros are for the score values of $\hat{\beta}_1, \hat{\beta}_2, \ldots, \hat{\beta}_p$, and the $(p+1)^{th}$ zero is for the score value of $\hat{\sigma}$.

Let $\mathbf{J}(\beta, \sigma, \omega)$ be the expected information matrix, and $\mathbf{J}(\hat{\beta}, \hat{\sigma}, 0)$ be the estimate of $\mathbf{J}(\beta, \sigma, \omega)$ under $H_0 : \omega = 0$ and $\hat{\beta}, \hat{\sigma}$, then $\mathbf{J}(\hat{\beta}, \hat{\sigma}, 0)$ can be expressed as

$$\begin{aligned} \mathbf{J}(\hat{\beta}, \hat{\sigma}, 0) = \mathbf{J}(\beta, \sigma, \omega)|_{\hat{\beta},\hat{\sigma},\omega=0} &= \begin{bmatrix} \hat{\mathbf{J}}_{\beta\beta} & \hat{\mathbf{J}}_{\beta\sigma} & \hat{\mathbf{J}}_{\beta\omega} \\ \hat{\mathbf{J}}_{\sigma\beta} & \hat{\mathbf{J}}_{\sigma\sigma} & \hat{\mathbf{J}}_{\sigma\omega} \\ \hat{\mathbf{J}}_{\omega\beta} & \hat{\mathbf{J}}_{\omega\sigma} & \hat{\mathbf{J}}_{\omega\omega} \end{bmatrix} \\ &= \begin{bmatrix} \mathbf{E}\left(-\frac{d^2 l(\cdot)}{d\beta d\beta}\right)\Big|_{\hat{\beta},\hat{\sigma},\omega=0} & \mathbf{E}\left(-\frac{d^2 l(\cdot)}{d\beta d\sigma}\right)\Big|_{\hat{\beta},\hat{\sigma},\omega=0} & \mathbf{E}\left(-\frac{d^2 l(\cdot)}{d\beta d\omega}\right)\Big|_{\hat{\beta},\hat{\sigma},\omega=0} \\ \mathbf{E}\left(-\frac{d^2 l(\cdot)}{d\beta d\sigma}\right)^T\Big|_{\hat{\beta},\hat{\sigma},\omega=0} & \mathbf{E}\left(-\frac{d^2 l(\cdot)}{d\sigma d\sigma}\right)\Big|_{\hat{\beta},\hat{\sigma},\omega=0} & \mathbf{E}\left(-\frac{d^2 l(\cdot)}{d\sigma d\omega}\right)\Big|_{\hat{\beta},\hat{\sigma},\omega=0} \\ \mathbf{E}\left(-\frac{d^2 l(\cdot)}{d\beta d\omega}\right)^T\Big|_{\hat{\beta},\hat{\sigma},\omega=0} & \mathbf{E}\left(-\frac{d^2 l(\cdot)}{d\sigma d\omega}\right)\Big|_{\hat{\beta},\hat{\sigma},\omega=0} & \mathbf{E}\left(-\frac{d^2 l}{d\omega d\omega}\right)\Big|_{\hat{\beta},\hat{\sigma},\omega=0} \end{bmatrix} \end{aligned} \tag{14}$$

where $\hat{\mathbf{J}}_{\beta\beta}$ is a $p \times p$ matrix, $\hat{\mathbf{J}}_{\beta\sigma}$ and $\hat{\mathbf{J}}_{\beta\omega}$ are $p \times 1$ matrix, $\hat{\mathbf{J}}_{\sigma\sigma}, \hat{\mathbf{J}}_{\sigma\omega}$ and $\hat{\mathbf{J}}_{\omega\omega}$ are scalar. Other elements $\hat{\mathbf{J}}_{\sigma\beta} = \hat{\mathbf{J}}^T_{\beta\sigma}$, $\hat{\mathbf{J}}_{\omega\beta} = \hat{\mathbf{J}}^T_{\beta\omega}$ and $\hat{\mathbf{J}}_{\omega\sigma} = \hat{\mathbf{J}}^T_{\sigma\omega}$.

Given the estimate of the expected information matrix $\mathbf{J}(\hat{\beta}, \hat{\sigma}, 0)$, the score statistic for testing $\omega = 0$ can be written as

$$S_{Score} = U^T(\hat{\beta}, \hat{\sigma}, 0)[\mathbf{J}(\hat{\beta}, \hat{\sigma}, 0)]^{-1} U(\hat{\beta}, \hat{\sigma}, 0)$$
$$= \frac{\left(\sum_{i=1}^{n} \frac{r_i - \hat{B}_i}{\hat{B}_i}\right)^2}{\hat{\mathbf{J}}_{\omega\omega} - \hat{\mathbf{J}}_{\omega\beta}\hat{\mathbf{J}}_{\beta\beta}^{-1}\hat{\mathbf{J}}_{\beta\omega} - \hat{\mathbf{J}}_{\omega\beta}VWV^T\hat{\mathbf{J}}_{\beta\omega} + \hat{\mathbf{J}}_{\omega\sigma}WV^T\hat{\mathbf{J}}_{\beta\omega} + \hat{\mathbf{J}}_{\omega\beta}VW\hat{\mathbf{J}}_{\sigma\omega} - \hat{\mathbf{J}}_{\omega\sigma}W\hat{\mathbf{J}}_{\sigma\omega}} \quad (15)$$

where $V = \hat{\mathbf{J}}_{\beta\beta}^{-1}\hat{\mathbf{J}}_{\beta\sigma}$ and $W = (\hat{\mathbf{J}}_{\sigma\sigma} - \hat{\mathbf{J}}_{\sigma\beta}\hat{\mathbf{J}}_{\beta\beta}^{-1}\hat{\mathbf{J}}_{\beta\sigma})^{-1}$. Under $H_0 : \omega = 0$, the statistic $S_{Score} \sim \chi_1^2$. The technical details for developing the score test are given in the Web Appendix.

For a two-sided test with type I error α, we reject $H_0 : \omega = 0$ vs. $H_A : \omega \neq 0$ when $S_{Score} > \chi^2_{1,1-\alpha}$. For a one-sided test $H_0 : \omega = 0$ vs. $H_A : \omega > 0$, we can consider the statistic $Z_{Score} = \frac{\sum_{i=1}^{n} \frac{r_i - \hat{B}_i}{\hat{B}_i}}{s_I}$, where s_I is the square-root of the $(p+2, p+2)^{th}$ term of the inverse of the Fisher information matrix evaluated at $\omega = 0$, $\hat{\beta}$ and $\hat{\sigma}$, i.e. $s_I^2 = [\mathbf{J}(\hat{\beta}, \hat{\sigma}, 0)]^{-1}_{(p+2,p+2)}$. Under $H_0 : \omega = 0$, the statistic $Z_{Score} \sim N(0, 1)$. We reject the null hypothesis if $Z_{Score} > Z_\alpha$ for the one-sided test with type I error α.

Compared to the Wald and LR tests, the score test only requires fitting a Tobit model (2), the Wald test requires fitting an mTobit model, and the LR test requires fitting both models. When the MLE for the mTobit model doesn't exist, the Wald test and LR test cannot be applied, but the score test can be still performed. Hence, the score test is preferred to the other two tests in such situations.

4 Simulation studies

4.1 Simulation setup

We use simulation studies to examine and compare the performance of the three tests. In all the simulation studies, we consider a one-sided test to test whether there is a latent class in the data, i.e. we tested $H_0 : \omega = 0$ vs $H_A : \omega > 0$. Two sets of simulation studies are considered, one with data generated from a Tobit model, in which type I error of rejecting the null hypothesis is assessed when $H_0 : \omega = 0$ is true, and the other with data generated from a mTobit model. In this case, the power of rejecting the null hypothesis is assessed when it is not true. In all the simulation studies, the detection limit L is fixed to be -1, and the variance for the Tobit model is fixed to be 4, but with varying means, i.e. $y \sim$ Tobit $(\mu, 4, -1)$. The varying mean yields different proportions of data under the detection limit and allows us to investigate how the performance changes with different proportions of data undetected. For the Tobit model, we consider three scenarios, a constant mean and mean changing with covariates from either a bounded uniform distribution or an unbounded normal distribution. For the mTobit model, we consider five scenarios: no covariates for both the latent class and the Tobit component, no covariates for the latent class only, and covariates for both the latent class and the Tobit component. For all the scenarios, small (50 and 100), moderate (200 and 500) and large (1000) sample sizes are considered, and a Monte Carlo (MC) sample size of 1000 is used. The simulations are carried out using R.[41] For the Wald and LR tests, we used the R function "optim" to find the MLE $\hat{\omega}$ of ω, and the standard error of $\hat{\omega}$ is based on estimated Hessian matrix.

4.2 Tobit response

Under the Tobit model, the null hypothesis $H_0 : \omega = 0$ (vs. $H_1 : w > 0$) is true, and the type I error of rejecting H_0 is used to evaluate the performance of the tests. In addition to providing rejection rates across the 1000 MC replications, we also used QQ plots, a plot of p values based on the asymptotic distribution of test statistic versus the corresponding empirical type I errors, which can serve as the true nominal level, based on 1000 MC replications, for a comprehensive evaluation of the performances. For a test to have a good performance, the p-value calculated based on the asymptotic distribution of the test statistics should be close to its corresponding empirical type I error. Thus, a QQ plot close to the diagonal line indicates a good performance, while deviation of the QQ plot from the diagonal line indicates how poor the performance is. Since the nominal level is usually very small,

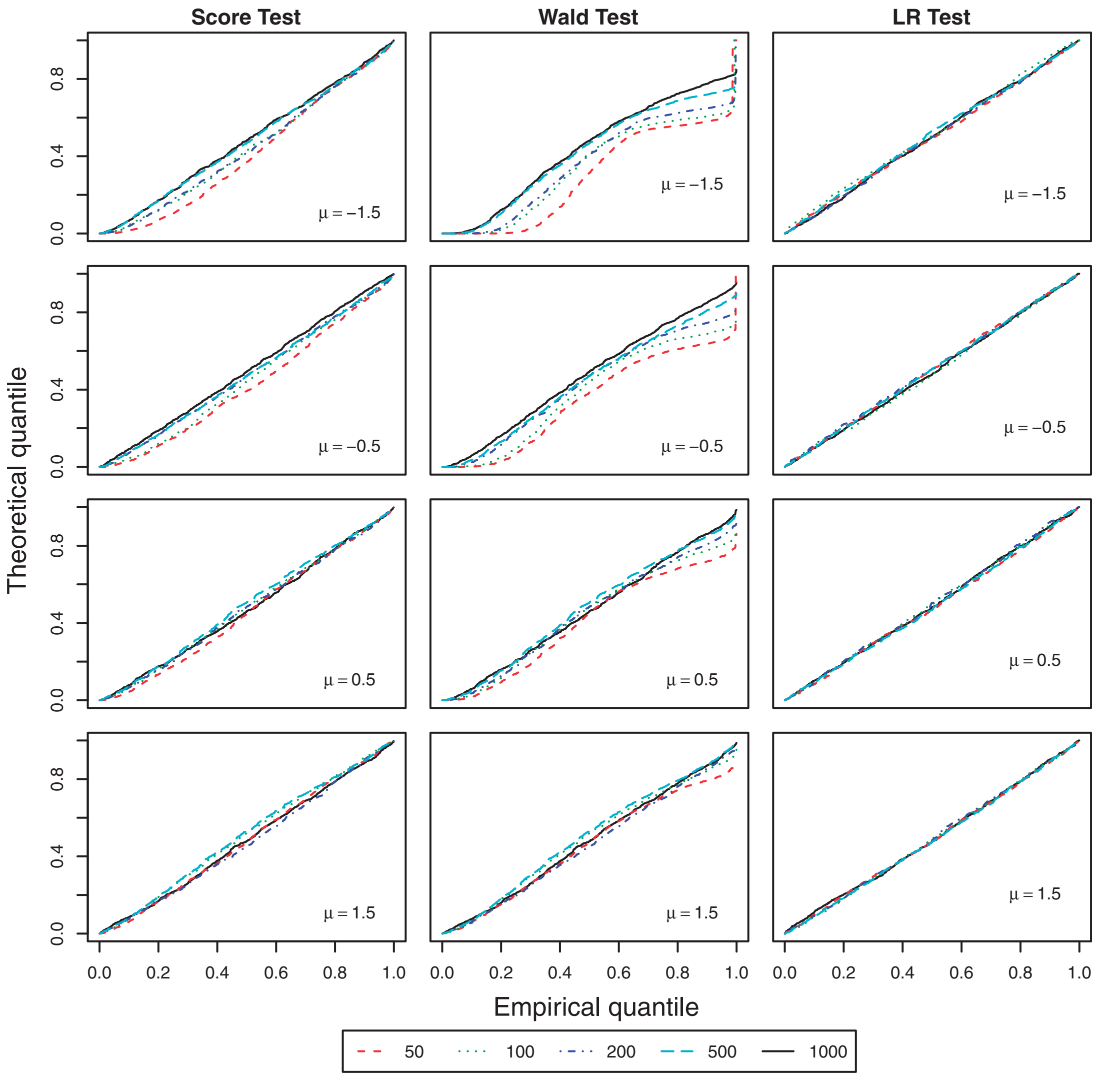

Figure 1. QQ plots of theoretical *p*-values and the corresponding empirical type I errors for the Tobit model without covariates and sample sizes 50, 100, 200, 500, and 1000.

such as 0.05, we focused on the (0, 0) end of the QQ plot. If a QQ plot lies above (under) the diagonal lines, then the tests are less (more) likely to reject the null hypothesis than it should be.

4.2.1 No covariate.

Data y_i is generated from N $(\mu, 4)$, with mean $\mu = -1.5, -0.5, 0.5$ and 1.5. y_i is replaced by -1 if its value is less than -1. The corresponding proportion of data under the detection limit is about 60%, 40%, 22% and 10%, respectively. The *p* values of rejecting the null hypothesis are summarized in the QQ plots presented in Figure 1. Overall, the LR test outperforms other two tests and performs very well in general, even when sample size is small. As expected, all the tests perform better as the sample size increases. When the sample size is 1000, the tests perform quite well except the Wald test when $\mu = -1.5$ and –0.5, which corresponds to 60% or 40% of under detection.

The performances of the Wald test and score test are clearly affected by the proportion of data under detection. When the proportion of under detection is large, such as $\mu = -1.5$ or -0.5, the score test and Wald test do not perform well, especially when the sample size is small. In these cases, the QQ plots for the Wald and score tests lie below the diagonal lines, which indicate that the two tests are more likely to falsely reject the null hypothesis. This is further confirmed by the rejection rates summarized in Table S1 as a supplementary material. The rejection rates for the Wald test and score test are uniformly higher than the nominal level 0.05. The rates for the Wald test goes as high as 33% at its worst situation when μ is small. As μ increases, i.e. less data falls below the detection limit, the p values are closer to the nominal level. When $\mu = 1.5$, i.e. about 10% of the data are under detection, the QQ lines for the score and Wald tests are very close to the diagonal line as the rejection rates are close to 0.05. It seems that the proportion of under detection does not have notable impact on the LR test, its performance is pretty stable with varying proportion of data under detection, and the rejection rates vibrate around 0.05.

As we noted in Section 3, the LR and Wald tests depend on the existence of an MLE for the mTobit model. Our simulation studies show that this can be problematic. In our study, out of the 1000 samples, when $\mu = -1.5$, MLEs cannot be obtained, and hence the Wald and LR statistics cannot be applied for 52, 19 and 4 samples for sample sizes 50, 100, and 200, respectively. For $\mu = -0.5$, there are eight samples for which the MLE cannot be obtained for sample size 50. For most of cases, there are no under-limit observations at all, thus the MLEs do not exist as the likelihood function (4) is monotone in ω. Note, however, that as a mixture of Tobit and degenerate components, equation (6) is a special case of the mixture model. Since the components of origin are known for observations over the limits, the model (6) is in general identifiable and the methods apply for large samples.

4.2.2 *Covariate* $x \sim$ Uniform (0, 1)

In this study, the covariate x is generated from a uniform distribution on [0,1], and y is simulated from Tobit$(\mu_i, 4, -1)$ with $\mu_i = \alpha - x_i$, where α is set to be $-1, 0$, 1 and 2, corresponding to about 60%, 40%, 22% and 10% of the data being under the detection level. Shown in Figure 2 are the QQ plots of the theoretical p value versus the corresponding empirical type I error of the three tests for different sample sizes and α values. The patterns and trends of QQ plots for the three tests are very similar to those in Figure 1 when there are no covariates for the Tobit model. Again, in general, the LR test still outperforms the other tests. Large deviations from the diagonal lines are observed for the Wald test, especially for data with a large proportion of under detection level and small sample sizes. The score test performs pretty well when the proportion of under detection level is around 20% or less. In most of cases, the QQ plots for the score test and Wald tests lie below the diagonal line, and the two tests are also more likely to falsely reject the null hypothesis. The rejection rates across the 1000 replications are summarized in Table S2 in the supplementary material. The Wald test and score test have rejection rates higher than the nominal level 0.05. While for the LR test, the rejection rates are pretty close to the nominal level.

There are still some samples for which the Wald and LR tests cannot be applied as the MLEs do not exist. Out of 1000 samples, when $\alpha = -1$, the Wald and LR statistics cannot be applied for 36, 13 and 3 samples for sample sizes 50, 100 and 200, respectively. For $\alpha = 0, 1$ and 2, there are 5, 13 and 112 samples which don't have MLEs. Those samples are most prevalent for a sample size of 50.

4.2.3 *Covariate* $x \sim N(0, 1)$.

The covariate x is generated from $N(0, 1)$, and y is generated from Tobit $(\mu_i, 4, -1)$ with $\mu_i = \alpha - x_i$, where α is set to be $-1.5, -0.5, 0.5$ and 1.5, which corresponds to a proportion of about 59%, 41%, 25% and 13% under detection, respectively. The QQ plots are presented in Figure 3. The patterns and trends for the Wald test and score test are pretty similar to that in Figures 1 and 2. However, the LR test performs a little worse than the previous cases based on the plots. The score test performs slightly better than the LR test with the Wald test the worst, especially when the large proportion of data under detection. The rejection rates are summarized in Table S3 as supplementary.

The number of samples that MLE does not exist are 63, 38, 96 and 265 for $\alpha = -1.5, -0.5, 0.5$ and 1.5, respectively, and most of them are for sample size 50. When sample size increases to 1000, the issue can be very minor.

4.3 mTobit Model

Next we simulate data from a mixture Tobit model to assess the performance of the tests in terms of power of detecting the latent class when the latent class does exist. We generate data with and without covariates for both the Tobit component and the latent class. We consider the covariate from either uniform or normal distribution. For the Tobit component, the data are generated similarly to the above cases, while for the latent class, we consider two scenarios, no covariates and with covariates.

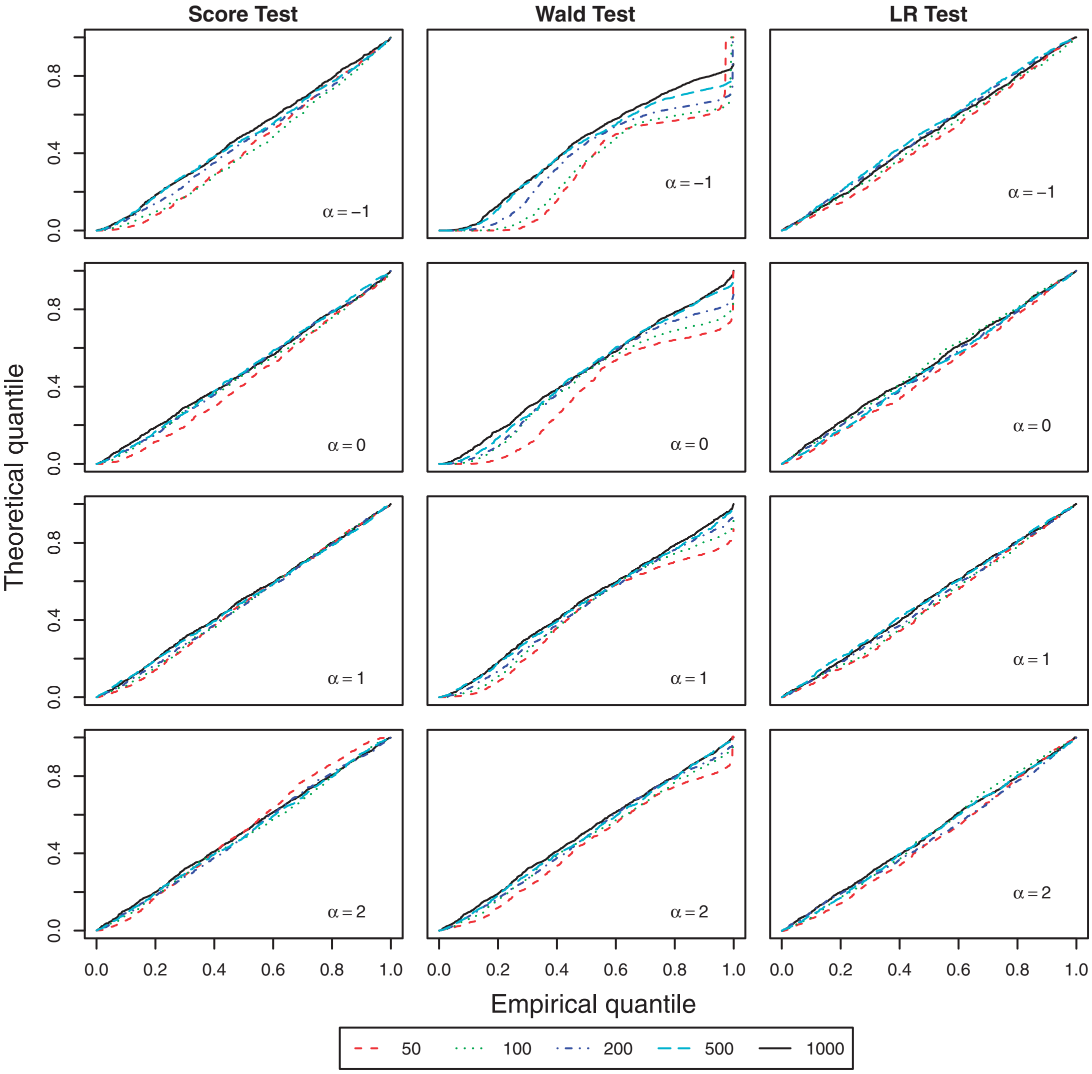

Figure 2. QQ plots of theoretical p-values and the corresponding empirical type I errors for the Tobit model with uniformly distributed predictors and sample sizes 50, 100, 200, 500, and 1000.

4.3.1 No covariate for the mTobit model.

In this case, we generated data from a mixture Tobit model $y \sim$ mTobit $(\omega, \mu, \sigma^2, L)$ with $\mu = -0.5, , 0.5, 1.5$ and 2.5, which corresponds to about 40%, 23%, 11% and 4% of data under detection for Tobit component. We let $\omega = 0.05k, k = 1, 2, \ldots, 6$ to investigate how the power changes with ω varying from very small (5%) to moderate (30%). Figures 4 shows the empirical powers of being able to detect the latent class, i.e. the proportion of rejecting the null hypothesis, with a type I error of 5%. The figures show that the power of the LR test in general is less than the Wald and score tests, with the Wald test giving the highest power. This coincides with the fact that the Wald test tends to be more likely to reject the null hypothesis. When μ is fixed, as expected, the power increases as the proportion of the latent class and sample size increases. However, the power is also significantly affected by μ. Since the detection limit is fixed, μ reflects how likely the data is censored, with higher values of μ meaning the data is less likely to be censored. The plots indicate that with fixed ω and sample size, as more data from the Tobit components are censored, the tests are less powerful in detecting the latent class.

4.3.2 Covariate $x \sim$ Unif[0, 1]

The data is simulated similarly as in Section 4.3.1 but with the Tobit component depending on a covariate x from Unif[0,1]. For the Tobit component, we use the same setting as in Section 4.2.2, i.e. $\mu = \alpha - x$ for the mean of the Tobit regression component with $\alpha = 0, 1, 2$ and 3. These α's correspond with about 40%, 23%, 11% and 4% data

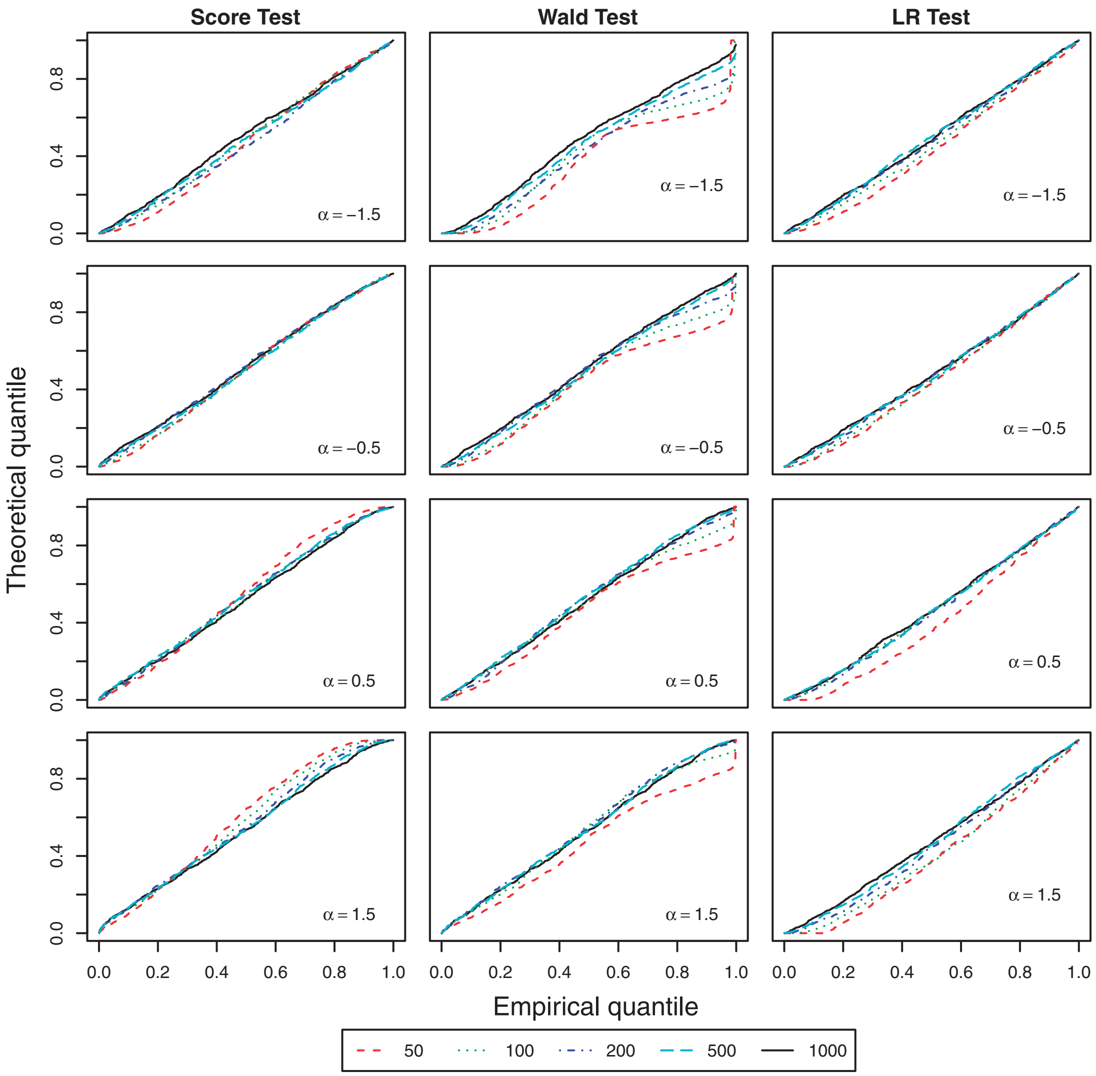

Figure 3. QQ plots of theoretical p-values and the corresponding empirical type I errors for the Tobit model with normally distributed predictors and sample sizes 50, 100, 200, 500, and 1000.

under detection for Tobit component. While for the latent class, two scenarios are considered below, one with covariate and the other without covariate for ω.

No covariate for ω: As in Section 4.3.1, we still let $\omega = 0.05k, k = 1, 2, \ldots, 6$. The mTobit model is generated as below

$$y \sim \text{mTobit}(\omega, \mu, 4, 1), \quad \mu = \alpha - x, \quad \alpha = 0, 1, 2, 3 \quad \text{and} \quad \omega = 0.05\text{k}, \quad \text{k} = 0, 1, \ldots, 6$$

Summarized in Figure 5 are the empirical powers for detecting the latent class when x follows a uniform distribution. The patterns are very similar to those in Section 4.3.1. Again the LR test has the lowest power and the Wald test has the highest power in most cases. As sample size and ω increase, the power increases as well. However, for fixed sample size and ω, more censored data in the Tobit component results in lower power as the tests are less likely to distinguish the data of the latent class from those of the Tobit component if they both are under detection.

Covariate for ω: Assume the probability of the latent class follows a generalized linear model with logit link and covariate x. The outcome y is generated according to the following model

$$y \sim \text{mTobit}(\omega, \mu, 4, -1), \quad \mu = \alpha - x, \quad \text{logit}(\omega) = -b + x \tag{16}$$

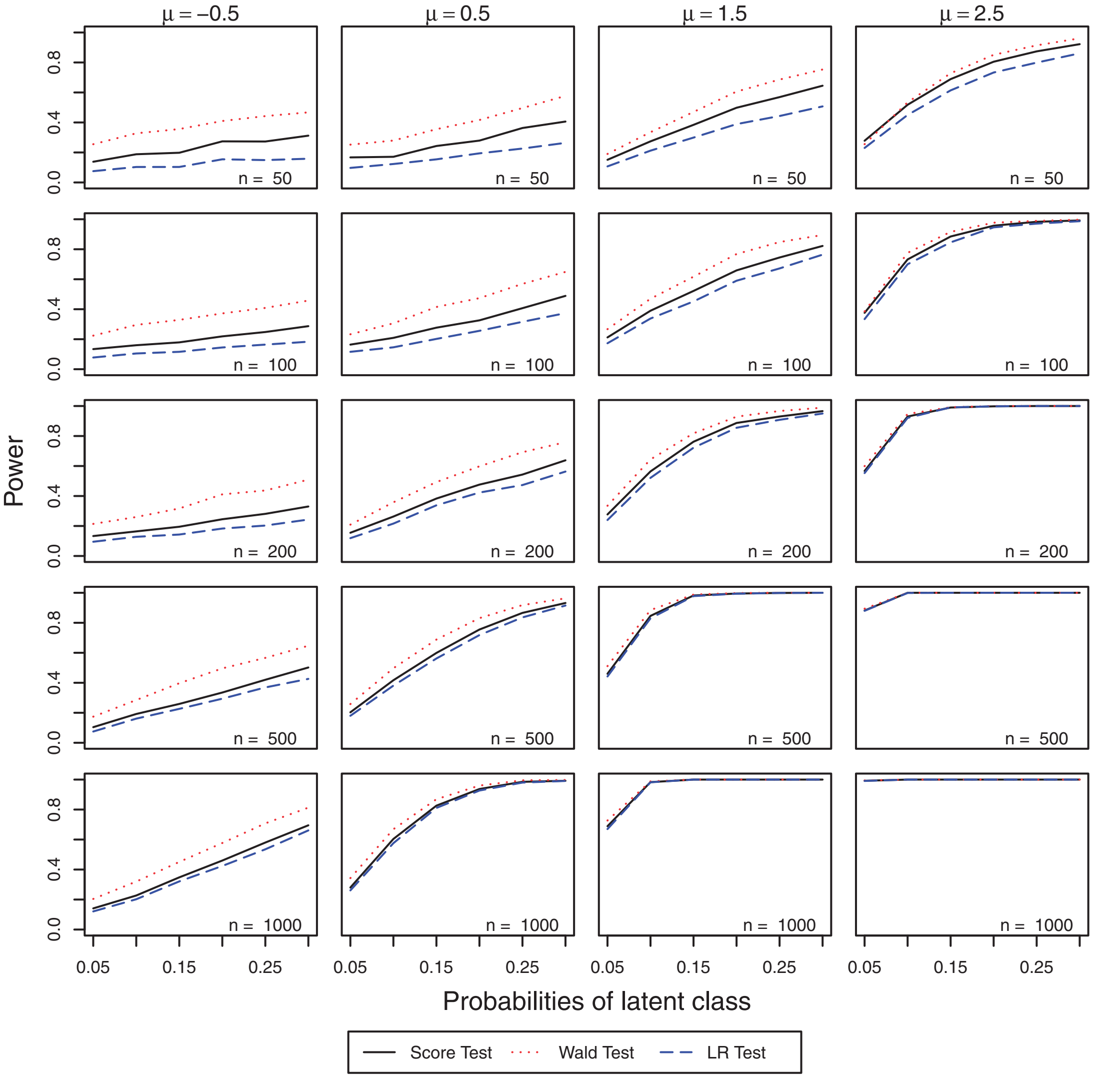

Figure 4. Power of detecting the latent class in mTobit model when there are no covariates for both Tobit model and the latent class.

where α is the same as the above, b is set to be 0.5, 1.0, 1.5, 2.0, 2.5 and 3.0, which in average corresponds to a probability of 0.32, 0.22, 0.15, 0.10, 0.06 and 0.04, respectively, for the latent class.

Shown in Figure 6 are the empirical powers for detecting the latent class when the covariate x is uniformly distributed. The patterns are very similar to those when the latent class does not depend on any covariates.

4.3.3 Covariate $x \sim$ N(0, 1).

The data is simulated similarly as in Section 4.3.2 but with the Tobit component depending on a covariate x from N(0,1), i.e. the mean of the Tobit component is defined by $\mu = \alpha - x$ with $x \sim$N(0,1). The α is set to be $-0.5, 0.5, 1.5$ and 2.5 which corresponds to about 41%, 25%, 13% and 6% of data under detection for Tobit component. We also consider two scenarios, one with covariate and the other without covariate for ω.

No covariate for ω: As in section 4.3.2, we set $\omega = 0.05k$, $k = 1, 2, \ldots, 6$ for the latent class. The powers are summarized in Figure 7. The patterns are very similar to those in Figure 5. As expected, as ω and sample size increase, the powers increase too. The power also increases as the data are less censored for fixed sample size and ω.

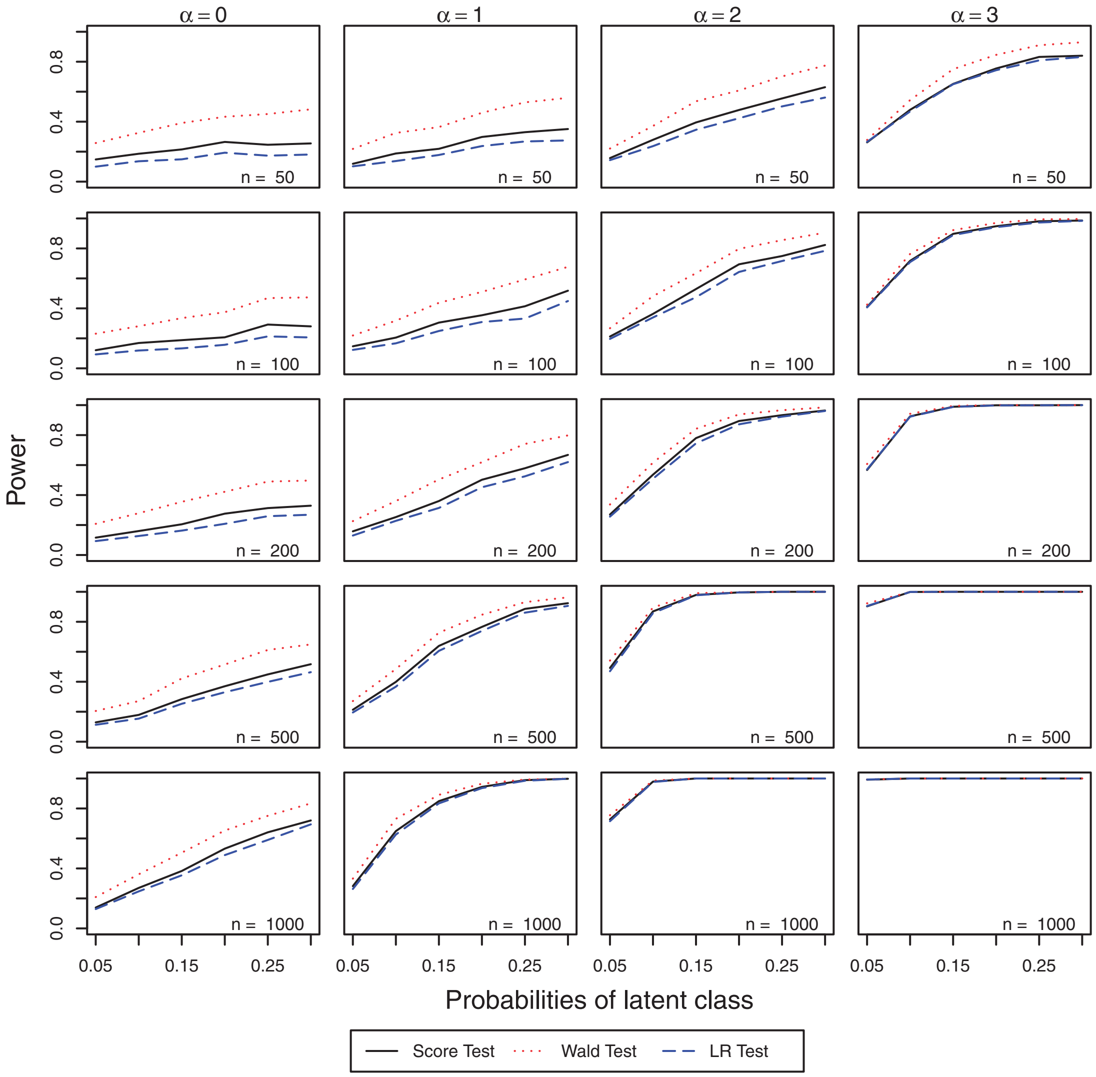

Figure 5. Power of detecting the latent class in mTobit model with uniform distributed covariate for Tobit component only.

Covariate for ω. Assume the probability of latent class depend on the covariate x through $\text{logit}(\omega) = -b + x$, the data are generated according to the following model

$$y \sim \text{mTobit}(\omega, \mu, 4, -1), \quad \mu = \alpha - x, \quad \text{logit}(\omega) = -b + x \tag{17}$$

where α is the same as the above, b is set to be 0.5, 1.0, 1.5, 2.0, 2.5 and 3.0 corresponding a probability of 0.38, 0.28, 0.19, 0.13, 0.08 and 0.05, respectively, for the latent class.

Figure 8 shows the empirical power of detecting the latent class. The results are similar to other cases. The LR test continues to have a little lower power than the other two tests, while the Wald test has the highest power as it tends to be more likely to reject the null hypothesis.

For all the above simulations, the means of the estimated ω are summarized in Table S4 as supplementary material.

To assess the performance of the tests for more covariates and possible correlated covariates, we also have conducted simulation studies using covariates in equation (19) with outcome generated as

$$y_i \sim \text{mTobit}(\omega, \mu_i, \sigma^2, L), \quad \mu_i = \beta_0 + \beta_1 Age_i + \beta_2 Gender_i + \beta_3 Race_i + \beta_4 BMI_i$$

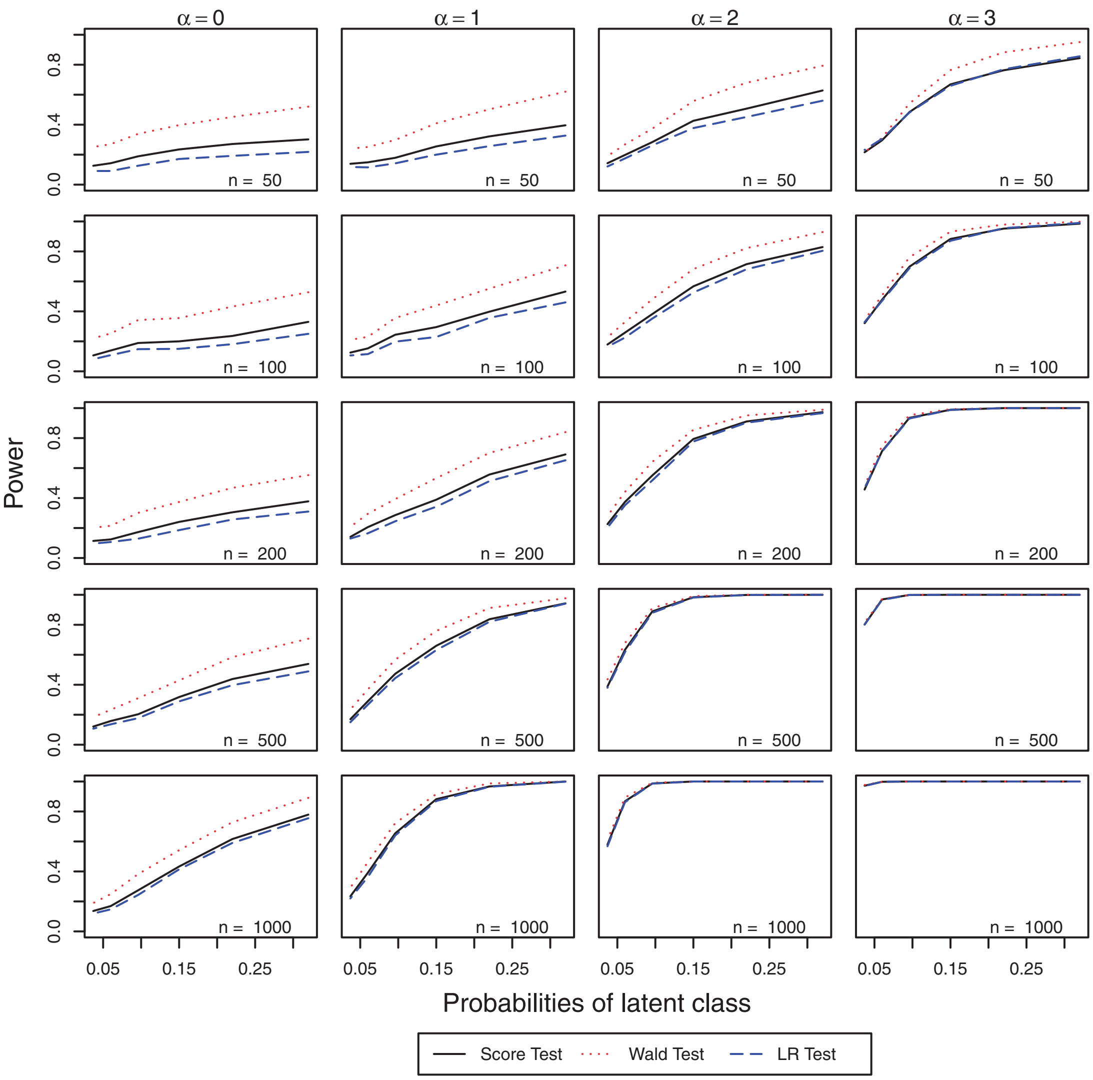

Figure 6. Power of detecting the latent class in mTobit model with uniform distributed covariate for both the Tobit component and the latent class component.

where $\beta_0 = -1, \beta_1 = 1, \beta_2 = 0.5, \beta_3 = 1$ and $\beta_4 = 2$. We consider $\omega = 0.0$, 0.1, 0.2 and 0.3 to examine the rejection rate for $\omega = 0.0$ and power for $\omega = 0.1$, 0.2 and 0.3. The rejection rates and powers are summarized in Table S5 in the Web Appendix.

5 Case studies

In this section, we use two real data examples to illustrate the tests. The first one is the NHANES 2003–2010 Study to examine if there is a latent class in the urinary triclosan concentrations, the second one is the Bogalusa Heart Study to examine if the serum metabolites have a mixture Tobit distribution.

5.1 NHANES 2003–2010 study

NHANES is a continuous program that examines a nationally representative sample of about 5000 persons each year to assess the health and nutritional status of adults and children in the general population of the USA (http://www.cdc.gov/nchs/nhanes.htm). Demographic, socioeconomic, dietary, and health-related data were collected via interviews. Blood and urine samples were also collected for laboratory testing. In four surveys conducted between 2003 and 2010, urinary triclosan concentration was measured in a random sample of survey participants aged 6 years or older. Literature shows that triclosan has potential to alter both gut microbiota and endocrine function

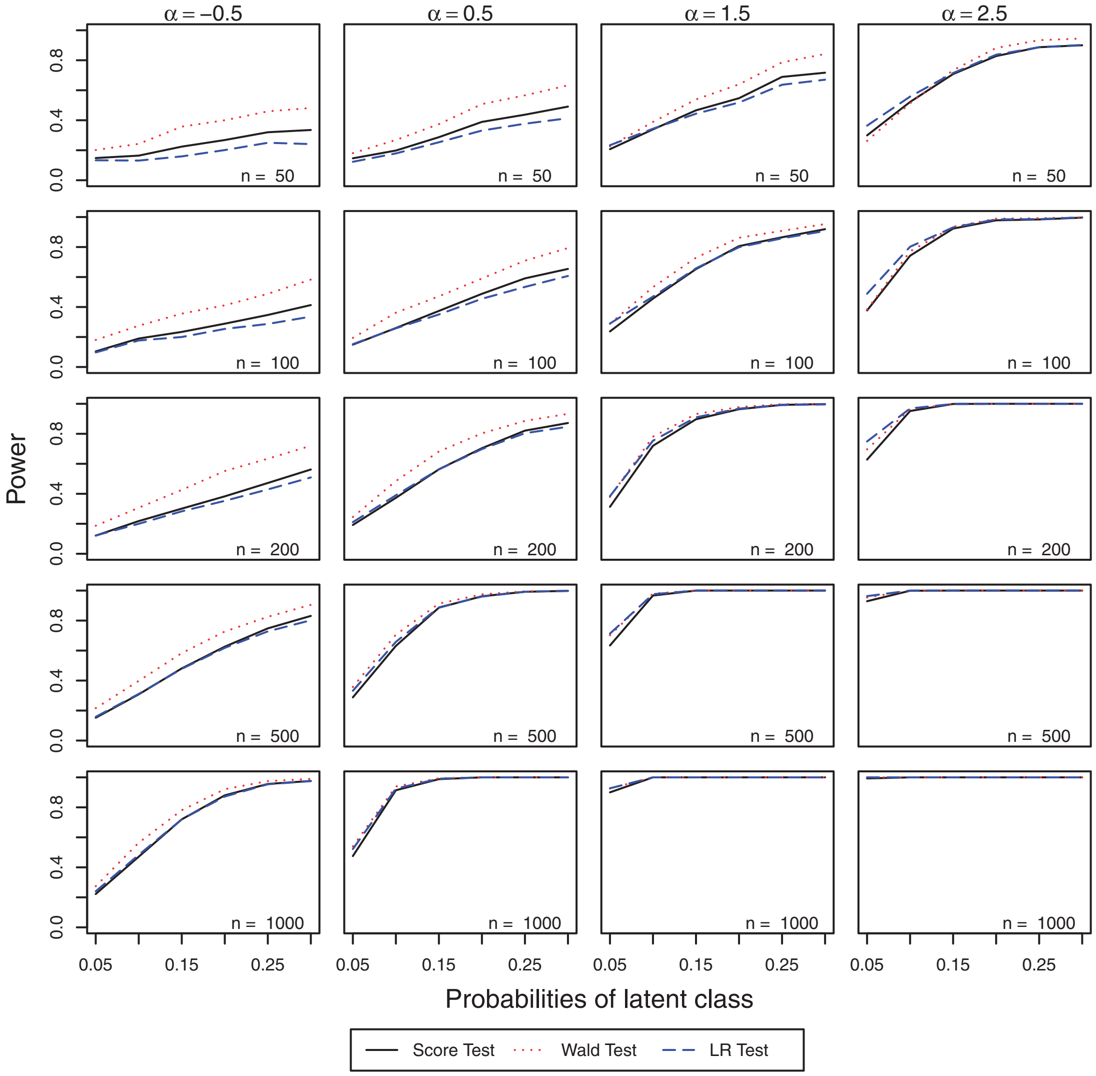

Figure 7. Power of detecting the latent class in mTobit model with normal distributed covariate for Tobit component only.

and has a negative impact on human health.[42,43] In the 2003–2010 NHANES database, urinary triclosan concentration was measured in 3659 children (6–19 years old) and 6566 adults (20 years or older). Of these, 2898 children and 5066 adults had detectable levels of urinary triclosan, which means that there are about 22% participates with their triclosan concentration undetected for a detection limit 2.3 ng/ml. Li et al.[43] treated the censored data as missing and used the complete dataset to examine the relationship between triclosan and body mass index, while Lankester et al.[42] treated the triclosan as a categorical variable and the censored data was classified as a single category. Here we apply the three tests to test if there is a latent class in the urinary triclosan concentration.

Due to the skewed distribution, a logarithm transformation is first conducted before testing the null hypothesis $H_0 : \omega = 0$ vs. $H_A : \omega > 0$. We assume a mTobit regression model with covariates age, gender, race, education, BMI, urine cotinine and creatinine for the transformed data, that is

$$\begin{aligned} y_i &\sim \text{mTobit}\Big(\omega, \mu_i, \sigma^2, L = \log(2.3)\Big), \\ \mu_i &= \beta_0 + \beta_1\, Age_i + \beta_2\, Gender_i + \beta_3\, Race_i + \beta_4\, Edu_i + \beta_5\, BMI_i \\ &\quad + \beta_6\, Cotinine_i + \beta_7\, Creatinine_i \end{aligned} \tag{18}$$

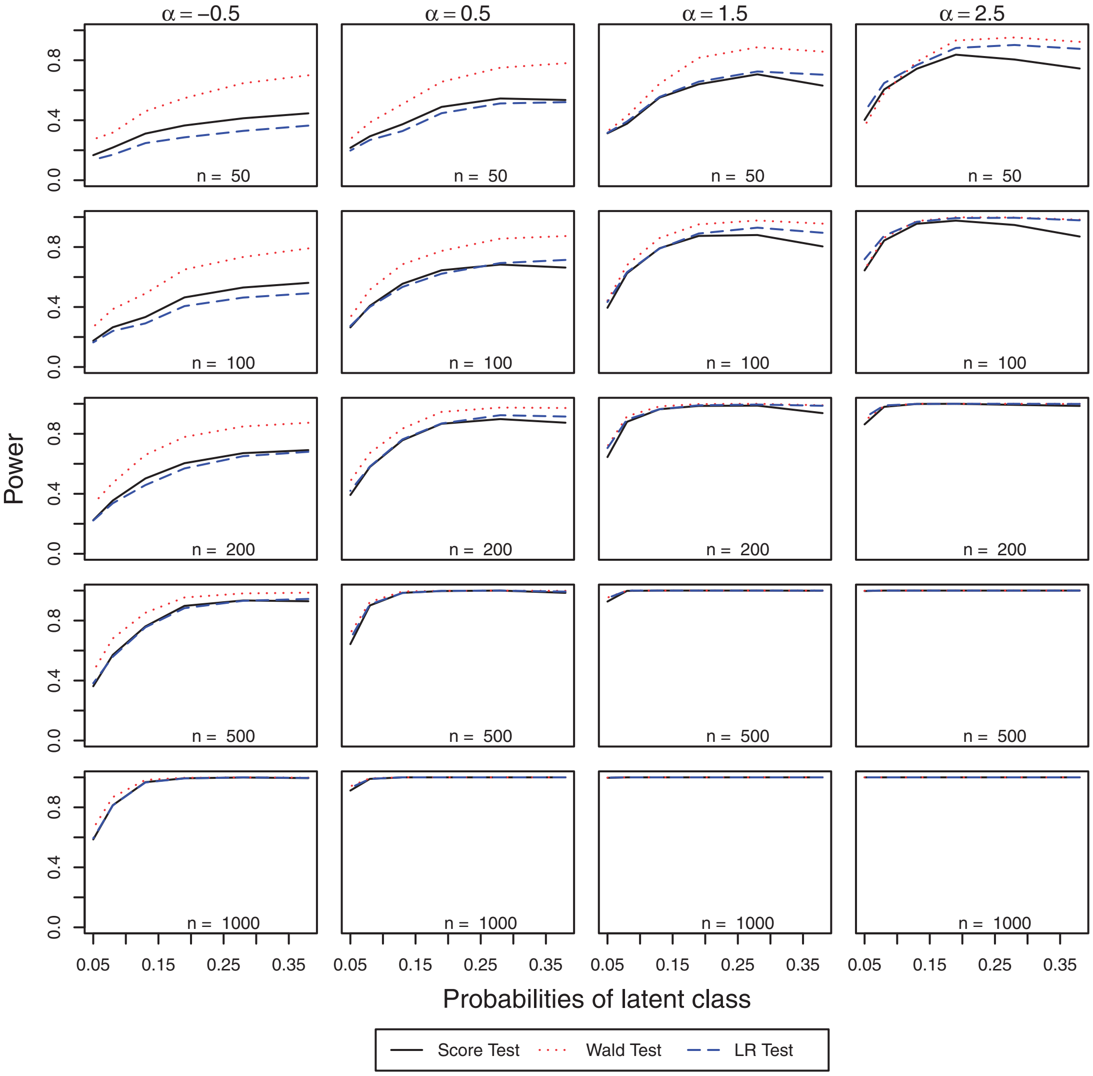

Figure 8. Power of detecting the latent class in mTobit model with normal distributed covariate for Tobit component only.

The estimated probability of the latent class is −0.000012 and the associated p-values of testing $H_0 : \omega = 0$ vs $H_A : \omega > 0$ are 0.5000, 0.5037 and 0.768 for the Wald test, LR test and score test, respectively. All the tests yield the same conclusion that there is no evidence to reject H_0, i.e. there is no latent class in the urinary triclosan concentration. The results are not surprising, because the study participants are 6 year or older, and it is very likely that everyone has been exposed to triclosan to some degree. Some of measures are not detected because of their low concentration, instead of not being exposed to triclosan at all.

5.2 Bogalusa Heart Study

The Bogalusa Heart Study, a series of long-term studies in a semirural biracial (65% white and 35% black) community in Bogalusa, Louisiana, was founded in 1973. This study focuses on the early natural history of cardiovascular disease since childhood. More details about the BHS can be found here (https://www.clersite.org/bogalusaheartstudy/). In the current BHS study, a total of 1466 metabolites were quantified from 1360 BHS blood samples. The BHS blood samples include 1261 unique BHS participants with blood samples collected during the 2013–2016 visit cycle, 64 random blind duplicate samples from a random sample of the same participants of the 2013–2016 visit cycle, and 35 replicate samples collected during the 2017–2019 visit cycle. Quality control procedures were conducted and data were cleaned. The cleaned analysis dataset includes 1202 metabolites for 1261

Table 1. Metabolits identified with a latent class based on p value less than 0.001 for at least one of the Wald test, LR test and Score test.

Metabolites	Super-pathway	Est. of Prob. (latent class)	*P* value		
			Wald	LR	Score
Y_100003260	Amino Acid	0.105	5.77×10^{-12}	1.65×10^{-8}	2.17×10^{-6}
Y_35	Amino Acid	0.021	1.75×10^{-3}	3.09×10^{-4}	6.69×10^{-3}
Y_241	Amino Acid	0.058	8.64×10^{-10}	5.85×10^{-10}	1.48×10^{-7}
Y_100002185	Amino Acid	0.227	0.00×10^{0}	4.06×10^{-11}	5.56×10^{-9}
Y_1094	Amino Acid	0.024	1.22×10^{-2}	5.77×10^{-7}	7.10×10^{-2}
Y_100015735	Lipid	0.189	0.00×10^{0}	0.00×10^{0}	0.00×10^{0}
Y_273	Lipid	0.020	5.02×10^{-7}	0.00×10^{0}	1.15×10^{-7}
Y_100019957	Lipid	0.162	6.68×10^{-11}	7.13×10^{-7}	3.83×10^{-5}
Y_100015833	Lipid	0.074	6.88×10^{-5}	4.60×10^{-4}	3.24×10^{-3}
Y_100019958	Lipid	0.136	1.54×10^{-8}	3.96×10^{-6}	1.16×10^{-4}
Y_100008934	Lipid	0.074	0.00×10^{0}	0.00×10^{0}	0.00×10^{0}
Y_100005716	Lipid	0.172	2.56×10^{-10}	1.11×10^{-5}	1.07×10^{-4}
Y_100008921	Lipid	0.042	1.30×10^{-2}	2.10×10^{-4}	1.28×10^{-2}
Y_100009154	Lipid	0.254	0.00×10^{0}	0.00×10^{0}	1.00×10^{-15}
Y_100009345	Lipid	0.035	2.47×10^{-5}	2.06×10^{-5}	8.51×10^{-5}
Y_100015666	Lipid	0.033	1.74×10^{-4}	2.05×10^{-4}	9.28×10^{-4}
Y_100015609	Lipid	0.104	0.00×10^{0}	0.00×10^{0}	0.00×10^{0}
Y_100015731	Lipid	0.271	0.00×10^{0}	1.34×10^{-9}	2.55×10^{-8}
Y_100015792	Lipid	0.022	1.06×10^{-3}	6.95×10^{-4}	2.14×10^{-3}
Y_207	Nucleotide	0.282	0.00×10^{0}	5.27×10^{-8}	4.71×10^{-7}
X_17335	Unknown	0.424	0.00×10^{0}	0.00×10^{0}	0.00×10^{0}
X_21788	Unknown	0.140	1.54×10^{-7}	2.64×10^{-5}	6.28×10^{-5}
X_21796	Unknown	0.064	6.14×10^{-11}	3.86×10^{-12}	2.51×10^{-6}
X_22776	Unknown	0.215	1.18×10^{-6}	1.30×10^{-3}	9.30×10^{-3}
X_23294	Unknown	0.417	0.00×10^{0}	1.10×10^{-14}	0.00×10^{0}
X_23637	Unknown	0.010	3.76×10^{-3}	5.41×10^{-4}	5.95×10^{-3}
X_24832	Unknown	0.022	5.63×10^{-5}	3.50×10^{-7}	1.51×10^{-3}
Y_100004601	Xenobiotics	0.398	0.00×10^{0}	0.00×10^{0}	0.00×10^{0}
Y_100001623	Xenobiotics	0.056	0.00×10^{0}	0.00×10^{0}	0.00×10^{0}
Y_100003696	Xenobiotics	0.167	6.43×10^{-11}	2.45×10^{-7}	1.04×10^{-4}
Y_100002324	Xenobiotics	0.117	1.40×10^{-14}	1.51×10^{-9}	1.42×10^{-8}

unique BHS participants from the 2013 to 2016 visit cycle. Among the 1202 metabolites, 167 have a missing rate or below detection rate >50% and 1035 have a missing rate or below detection rate ≤50%. Among the 1035 metabolites, 398, 401, 77, 56, 52 and 51 metabolites had 0%, <10%, 10–20%, 20–30%, 30–40% and 40–50% missing values or values below detection limit, respectively. The metabolites with more than 50% missing values were categorized as 1 for missing or below detection limit, the ones with greater than below detection limit but less than median were categorized as 2, and the ones with equal or greater than the median were categorized as 3.

To illustrate the tests, we consider metabolites with missing values less than 50% and apply the three tests to test whether there is a latent class in these serum metabolites. To account for individual differences, the Tobit regression component is adjusted for age, gender, race and BMI, i.e. the mTobit model is given by

$$y_i \sim \text{mTobit}(\omega, \mu_i, \sigma^2, L), \quad \mu_i = \beta_0 + \beta_1 Age_i + \beta_2 Gender_i + \beta_3 Race_i + \beta_4 BMI_i \tag{19}$$

There are a total 31 metabolites identified to have a latent class with p value less than 0.001 for at least one of the three tests. The metabolites, the estimated probability of the latent class, as well as the associated p-values for the three tests are provided in Table 1. Several of the identified metabolites belong to the xenobiotic super-pathway. This pathway includes many drug metabolites which we would expect to have a latent class.

6 Discussion

In this paper, we have developed three tests for testing if there is a latent class in a Tobit regression model. The simulation studies show that the Wald test gives inflated type I errors, especially when sample size is small and the proportion of data under detection is large. The Wald test tends to be more likely to reject the null hypothesis, and therefore it yields elevated powers compared to the other tests. The LR test controls for the type I error very well when there are no covariates or covariates from uniform distribution. The trend of score test in controlling for type I error is pretty stable regardless of covariates. When there are covariates, the score test performs slightly better than the LR test, especially when the covariates are from unbounded normal distribution.

The proportion of data under detection has impact on the type I error of the tests, especially on the Wald test. When the proportion of the data under the detection level is large, the Wald test doesn't provide valid type I error, especially for small to moderate sample sizes. The impact of the proportion seems small for the LR test. The impact of the proportion of censored data on the score test is somewhere between the LR test and the Wald test.

The proportion of under detection also has impact on the power. Even with large sample size 1000 and large probability of the latent class, the power is not high if the proportion of under detection is large. In such cases, the tests are not powerful to distinguish data from the latent class and data from Tobit model but under detection. In general, the Wald test has the highest power because it tends to be more likely rejecting the null hypothesis. In the paper, we focused on cross-sectional data where one measure is obtained from each subject. One may improve the power by obtaining repeated measurements. However, since repeated measurements from the same subjects are unavoidably correlated, it may be necessary to study the correlation structure. While in principle approaches in other latent class situations such as those studied in literature[44,45] should be adapted to the current situation, future research is needed to generalize the tests to such cases.

Since the simulated data sets with nonexistence of MLE for mTobit model were discarded in our simulation studies, the type I error and power can be affected for Wald and likelihood ratio test, but the performance of score test is not affected. Based on the performances of the tests in terms of both power and controlling type I error, the Wald test is the least favorable. The choice of LR test versus score test depends on covariates. If no covariates, the LR test would be preferred, while if the covariates are unbounded, the score test would be preferred. However, both the LR test and the Wald test require the MLE of the mTobit model which can be problematic, especially when the sample size is small and the proportion of data under detection is large, and therefore the LR test and Wald test may not be applied. The score test only needs the MLE for Tobit model, and in general it can be applied. Thus, only the score test can be used if the MLE does not exist for the mTobit model.

The tests proposed in the paper for latent class assume constant probability for the latent class, and it may depend on some covariates. The extension of the tests to this case is the next natural step.

Acknowledgement

The authors would like to thank Dr. Ye Peng for his suggestions.

Declaration of conflicting interests

The author(s) declared no potential conflicts of interest with respect to the research, authorship, and/or publication of this article.

Funding

The author(s) disclosed receipt of the following financial support for the research, authorship, and/or publication of this article: This work was supported by the NIH under grants R01GM108337 and P20GM109036.

ORCID iDs

Hua He https://orcid.org/0000-0001-8989-0297
Wan Tang https://orcid.org/0000-0003-0404-1494

Supplemental Material

Supplemental material for this article is available online.

References

1. Hornung Richard W and Reed Laurence D. Estimation of average concentration in the presence of nondetectable values. *Appl Occupation Environment Hygiene* 1990; **5**: 46–51.
2. Jamjoum LS, Bielak LF, Turner ST, et al. Relationship of blood pressure measures with coronary artery calcification. *Med Sci Monitor* 2002; **8**: CR775–CR781.
3. Reilly MP, Wolfe ML, Russell Localio A, et al. Coronary artery calcification and cardiovascular risk factors: impact of the analytic approach. *Atherosclerosis* 2004*;***173**: 69–78.
4. Lubin JH, Colt JS, Camann D, et al. Epidemiologic evaluation of measurement data in the presence of detection limits. *Epidemiology* 2005; **16**: S40.
5. Dinse GE, Jusko TA, Ho LA, et al. Accommodating measurements below a limit of detection: a novel application of cox regression. *Am J Epidemiol* 2014; **179**: 1018–1024.
6. Nassan FL, Coull BA, Gaskins AJ, et al. Personal care product use in men and urinary concentrations of select phthalate metabolites and parabens: results from the environment and reproductive health (earth) study. *Environ Health Perspect* 2017; **125**: 1–10.
7. Ferrero A, Esplugues A, Estarlich M, et al. Infants' indoor and outdoor residential exposure to benzene and respiratory health in a Spanish cohort. *Environ Pollut* 2017; **222**: 486–494.
8. Østergren PB, Kistorp C, Fode M, et al. Luteinizing hormone-releasing hormone agonists are superior to subcapsular orchiectomy in lowering testosterone levels of men with prostate cancer: results from a randomized clinical trial. *J Urol* 2017; **197**: 1441–1447.
9. Kakourou A, Vach W and Mertens B. Adapting censored regression methods to adjust for the limit of detection in the calibration of diagnostic rules for clinical mass spectrometry proteomic data. *Stat Meth Med Res* 2018; **27**: 2742–2755.
10. Kim S, Chang Y, Sung E, et al. Association between sonographically diagnosed nephrolithiasis and subclinical coronary artery calcification in adults. *Am J Kidney Dis* 2018; **71**: 35–41.
11. Olson DR. A simple method for estimation when there is a detection limit. In: *Joint statistical meeting of American Statistical Society and Biometric Society*, San Francisco, California, 8 August 1993.
12. LaFleur B, Lee W, Billhiemer D, et al. Statistical methods for assays with limits of detection: Serum bile acid as a differentiator between patients with normal colons, adenomas, and colorectal cancer. *J Carcinogenesis* 2011; **10**: 12–20.
13. Slymen DJ, de Peyster A and Donohoe RR. Hypothesis testing with values below detection limit in environmental studies. *Environ Sci Technol* 1994; **28**: 898–902.
14. Gleit A. Estimation for small normal data sets with detection limits. *Environ Sci Technol* 1985; **19**: 1201–1206.
15. Newman MC, Dixon PM, Looney BB, et al. Estimating mean and variance for environmental samples with below detection limit observations 1. *J Am Water Resources Assoc* 1989; **25**: 905–916.
16. Tobin J. Estimation of relationships for limited dependent variables. *Econometrica: J Econometric Soc* 1958; **26**: 24–36.
17. Cohen AC. Simplified estimators for the normal distribution when samples are singly censored or truncated. *Technometrics* 1959; **1**: 217–237.
18. McDonald JF and Moffitt RA. The uses of tobit analysis. *Rev Economics Stat* 1980; **62**: 318–321.
19. Amemiya T. Tobit models: a survey. *J Econometrics* 1984; **24**: 3–61.
20. Olsen RJ. Note on the uniqueness of the maximum likelihood estimator for the Tobit model. *Econometrica* 1978; **46**: 1211–1215.
21. Wang W and Griswold ME. Natural interpretations in tobit regression models using marginal estimation methods. *Stat Meth Med Res* 2017; **26**: 2622–2632.
22. Dagne GA and Huang Y. Bayesian semiparametric mixture tobit models with left censoring, skewness, and covariate measurement errors. *Stat Med* 2013; **32**: 3881–3898.
23. Zhou C, Yu NN and Losby JL. The association between local economic conditions and opioid prescriptions among disabled medicare beneficiaries. *Med Care* 2018; **56** :62–68.
24. Awalime DK, Davies-Teye BBK, Vanotoo LA, et al. Economic evaluation of 2014 cholera outbreak in Ghana: a household cost analysis. *Health Econom Rev* 2017; **7**: 45.
25. Jou R-C and Chen T-Y. The willingness to pay of parties to traffic accidents for loss of productivity and consolation compensation. *Accident Analysis Prevent* 2015; **85**: 1–12.
26. Etilé F and Sharma A. Do high consumers of sugar-sweetened beverages respond differently to price changes? A finite mixture IV-tobit approach. *Health Econom* 2015; **24**: 1147–1163.
27. Al-Hanawi MK, Alsharqi O and Vaidya K. Willingness to pay for improved public health care services in Saudi Arabia: a contingent valuation study among heads of Saudi households. *Health Econom Policy Law* 2018; **13**: 1–28.
28. Park M and Lee D. Analysis of severe injury accident rates on interstate highways using a random parameter tobit model. *Math Problems Eng* 2017; **2017**: 1–6.
29. Marriott E-R, van Hazel G, Gibbs P, et al. Mapping eortc-qlq-c30 to eq-5d-3l in patients with colorectal cancer. *J Med Econom* 2017; **20**: 193–199.
30. Chun S, Choi Y, Chang Y, et al. Sugar-sweetened carbonated beverage consumption and coronary artery calcification in asymptomatic men and women. *Am Heart J* 2016; **177**: 17–24.

31. Ko B-J, Chang Y, Jung H-S, et al. Relationship between low relative muscle mass and coronary artery calcification in healthy adults. *Arteriosclerosis, Thrombosis Vasc Biol* 2016; **36**: 1016–1021.
32. García-Esquinas E, Pérez-Gómez B, Fernández-Navarro P, et al. Lead, mercury and cadmium in umbilical cord blood and its association with parental epidemiological variables and birth factors. *BMC Public Health* 2013; **13**: 841.
33. Philippat C, Bennett D, Calafat AM, et al. Exposure to select phthalates and phenols through use of personal care products among Californian adults and their children. *Environ Res* 2015; **140**: 369–376.
34. Setty KE, Kayser GL, Bowling M, et al. Water quality, compliance, and health outcomes among utilities implementing water safety plans in France and Spain. *Int J Hygiene Environ Health* 2017; **220**: 513–530.
35. Darrow LA, Jacobson MH, Preston EV, et al. Predictors of serum polybrominated diphenyl ether (pbde) concentrations among children aged 1–5 years. *Environ Sci Technol* 2016; **51**: 645–654.
36. Halsey NA, Boulos R, Mode F, et al. Response to measles vaccine in Haitian infants 6 to 12 months old: influence of maternal antibodies, malnutrition, and concurrent illnesses. *New Engl J Med* 1985; **313**: 544–549.
37. Moulton LH and Halsey NA. A mixture model with detection limits for regression analyses of antibody response to vaccine. *Biometrics* 1995; **51**: 1570–1578.
38. Taylor DJ, Kupper LL, Rappaport SM, et al. A mixture model for occupational exposure mean testing with a limit of detection. *Biometrics* 2001; **57**: 681–688.
39. Reisetter AC, Muehlbauer MJ, Bain JR, et al. Mixture model normalization for non-targeted gas chromatography/mass spectrometry metabolomics data. *BMC Bioinform* 2017; **18**: 84.
40. Engle RF. Wald, likelihood ratio, and Lagrange multiplier tests in econometrics. *Handbook Econometrics* 1984; **2**: 775–826.
41. R Development Core Team. *R: A Language and Environment for Statistical Computing*. R Foundation for Statistical Computing, Vienna, Austria, 2011. ISBN 3-900051-07-0.
42. Lankester J, Patel C, Cullen MR, et al. Urinary triclosan is associated with elevated body mass index in nhanes. *PloS One*, 2013; **8**: e80057.
43. Li S, Zhao J, Wang G, et al. Urinary triclosan concentrations are inversely associated with body mass index and waist circumference in the US general population: experience in nhanes 2003–2010. *Int J Hygiene Environ Health*, 2015; **218**: 401–406.
44. Espeland MA and Handelman SL. Using latent class models to characterize and assess relative error in discrete measurements. *Biometrics* 1989; **86**: 587–599.
45. Albert PS and Dodd LE. A cautionary note on the robustness of latent class models for estimating diagnostic error without a gold standard. *Biometrics* 2004; **60**: 427–435.