

# Two parts are better than one: modeling marginal means of semicontinuous data

Valerie A. Smith<sup>1,2</sup> · Brian Neelon<sup>3</sup> · Matthew L. Maciejewski<sup>1,2</sup> · John S. Preisser<sup>4</sup>

Received: 30 August 2016 / Revised: 21 February 2017 / Accepted: 21 March 2017 /  
Published online: 29 March 2017  
© Springer Science+Business Media New York (outside the USA) 2017

**Abstract** In health services research, it is common to encounter semicontinuous data characterized by a point mass at zero followed by a continuous distribution with positive support. These are often analyzed using two-part mixtures that separately model the probability of use to account for the portion of the sample with zero values. Commonly, but not always, the second component models the continuous values conditional on them being positive. Prior work examining whether such two-part models are needed to appropriately draw inference from semicontinuous data compared to standard one-part regression models has found mixed results. However, prior studies have generally used only measures of model fit on a single dataset, leaving a definitive conclusion uncertain. This paper provides a detailed evaluation using simulations of the appropriateness of standard one-part generalized linear models (GLMs) compared to a recently developed marginalized two-part (MTP) model. The MTP model, unlike the one-part GLMs, explicitly accounts for the point mass at zero, yet takes the same form for the marginal mean as the commonly used GLM with log link, making the covariate effects directly comparable. We simulate data scenarios with varying sample sizes and percentages of zeros. One-part GLMs resulted in increased bias, lower than nominal coverage of confidence intervals, and inflated type I error rates, rendering them inappropriate for use with semicontinuous data. Even when

---

**Electronic supplementary material** The online version of this article (doi:[10.1007/s10742-017-0169-9](https://doi.org/10.1007/s10742-017-0169-9)) contains supplementary material, which is available to authorized users.

---

✉ Valerie A. Smith  
valerie.a.smith@duke.edu

<sup>1</sup> Center for Health Services Research in Primary Care (152), Durham VA Medical Center, 508 Fulton Street, Durham, NC 27705, USA

<sup>2</sup> Division of General Internal Medicine, Department of Medicine, Duke University Medical Center, Durham, NC, USA

<sup>3</sup> Department of Public Health Sciences, Medical University of South Carolina, Charleston, SC, USA

<sup>4</sup> Department of Biostatistics, University of North Carolina, Chapel Hill, NC, USA

distributional assumptions were violated, estimates of covariate effects and type I error rates under the MTP model remained robust.

**Keywords** Generalized gamma distribution · Health care expenditures · Log-skew-normal distribution · Marginalized models · Two-part models

## 1 Introduction

Semicontinuous data, characterized by a “clumping” of zero values combined with a positive, often skewed, distribution of continuous values, commonly arise in health services research. For example, health care expenditures, substance abuse, and inpatient length of stay can all be characterized by a portion of the sample who are non-users with zero values and another portion with a distribution of positive values. Response data with such features are often thought of as arising from two distinct stochastic processes: a binary part governing the occurrence of zeros and a continuous part determining the observed value conditional on it being a nonzero response.

To accommodate the two aspects of semicontinuous data, analysts often consider two-part models. Most commonly, the binary part is modeled via logistic regression while the log-transformed continuous component is modeled via standard linear regression, although many other specifications have been used (Mullahy 1998; Blough et al. 1999). When covariates are included in the regression models, covariates in the second, or continuous, part are interpreted conditionally upon having observed a positive outcome.

As an alternative to two-part models, analysts often will fit a one-part generalized linear model (GLM), usually with a log link to ensure that predicted values are non-negative. GLMs incorporate both the zero and positively continuous values into a single stochastic process rather than explicitly specifying a separate component to account for the point mass at zero. In doing so, they permit direct interpretation of covariate effects on the overall mean. Nonlinear least squares estimation (Mullahy 1998) and quasilielihood estimation (Buntin and Zaslavsky (2004)) have been proposed for fitting one-part models, avoiding fully parametric assumptions. Empirical standard errors, which provide asymptotically valid inference even if the variance model is misspecified, may also be incorporated; however, their finite sample performance in the presence of many zero values has not been fully evaluated.

This article addresses the common problem of estimating individual covariate effects on the overall mean of a semicontinuous outcome. To accommodate this, we recently introduced a fully parametric marginalized two-part (MTP) modeling approach that specifies the same marginal mean model as a typical one-part GLM, while simultaneously accounting for the point mass at zero (Smith et al. 2014). The MTP model parameterizes covariate effects directly on the overall mean on the untransformed original scale via the log link function, allowing parameter estimates to be interpreted as the multiplicative effect on the overall mean. This approach also has the advantage of separately providing estimates of covariate effects on the probability of incurring a positive-valued outcome as in the first part of two-part models, as well as accounting for the zero-inflated and skewed nature of many semicontinuous outcomes. While not examined in this article, random effects can be easily incorporated into MTP models to address repeated measures and clustered data (Smith et al. 2015).

GLMs rely on fewer assumptions than MTP models, so it may be natural to question whether one-part models fit to semicontinuous data perform better than MTP models in terms of bias, precision, and type I error when interest lies in marginal inferences on the overall mean. Duan et al. (1983), Diehr et al. (1999), Madden et al. (2000), and Buntin and Zaslavsky (2004) have each compared the performance of “standard” two-part models with various one-part models. In each case, the models were assessed using real datasets and performance was determined using a combination of goodness of fit criteria and predictive accuracy. Conclusions were mixed, with one-part models performing equally well or better on some datasets (Diehr et al. 1999; Buntin and Zaslavsky 2004) and two-part models exhibiting better performance on others (Duan et al. 1983; Madden et al. 2000).

In particular, Buntin and Zaslavsky (2004) compared several one- and two-part models with the goal of predicting Medicare expenditures using a sample of 10,134 elderly Medicare beneficiaries with 8.6% of individuals having zero expenditures. They assessed each model’s predictive ability via several metrics, including mean squared error and split-sample cross-validation. They concluded that excess zeros in the data pose little problem when fitting a one-part GLM and suggested one-part GLMs could be used across an array of semicontinuous outcomes. While this comparison has been frequently cited as justification of the appropriateness of one-part GLMs for use in semicontinuous data, examination of model fit on a single dataset with a comparatively small proportion of zeros does not in general answer the question of whether one-part GLMs are appropriate to use when data contain many zeros. More work is needed to assess model performance under the presence of a greater proportion of zeros as well as the ability of one- and two-part models to accurately estimate the effects of covariates.

Previously, model-estimated covariate effects across one- and two-part models were not generally comparable because standard two-part models separately specified the probability of a positive expenditure and the level of expenditure conditional on it being positive, thereby indirectly specifying a different parameterization for the overall marginal mean. The recent introduction of the MTP model, however, provides an analytic approach that explicitly accounts for excess zeros without sacrificing the direct interpretability of regression parameters as covariate effects on the overall mean. The performance of the MTP model has not been examined in comparison to one-part models, nor has a formal simulation study comparing such one- and two-part models been conducted.

With this goal in mind, we conducted a series of simulation studies to compare the performance of GLMs fit with quasilielihood and various MTP models. We evaluate bias, test size, and coverage of nominal 95% confidence intervals under varying data generating mechanisms with proportions of zeros set at 10, 20, and 40% where inferential focus is on the regression parameter estimates for the overall mean. Our simulation design was motivated in part by an analysis to assess the impact of a behavioral weight loss program on health care expenditures in the year following enrollment, presented in Smith et al. (2014). In addition to the simulation study, we reanalyze the data from this weight loss program using GLMs and MTP models and compare results.

The remainder of this paper is laid out as follows. Section 2 briefly reviews models commonly used for semicontinuous data, while Sect. 3 discusses the details of the simulations conducted. Section 4 shows the results of the simulations, and Sect. 5 presents the results of the weight loss program analysis. Section 6 provides a discussion of the implications of the results and points to areas for future research and investigation.

## 2 Models for semicontinuous data

We begin with a brief review of models commonly used for semicontinuous data.

### 2.1 Standard two-part models

Traditionally, conventional two-part models have been used to analyze data with zero-heavy outcomes. In these models, the binary part is commonly modeled via logistic regression and the continuous component via a linear model for the natural log transformed outcome conditional on the response being greater than zero. Specifically, the generic form of the conventional two-part model, presented in Cragg (1971), Manning et al. (1981), Duan et al. (1983) and elsewhere, can be written as

$$f(y_i) = (1 - \pi_i)^{1_{(y_i=0)}} \times [\pi_i g(y_i|y_i > 0)]^{1_{(y_i > 0)}}, \quad y_i \geq 0, i = 1, \dots, n, \quad (1)$$

where  $\pi_i = \Pr(Y_i > 0)$ ,  $1_{(\cdot)}$  is the indicator function, and  $g(y_i|y_i > 0)$  is any density function applicable to the positive values of  $Y_i$ . This model is commonly parameterized as

$$\text{logit}(\pi_i) = \mathbf{z}'_i \boldsymbol{\alpha} \quad \text{and} \quad (2)$$

$$\mu_i = E(\ln Y_i | Y_i > 0) = \mathbf{x}'_i \boldsymbol{\gamma}. \quad (3)$$

To conduct inference on the overall mean  $E(Y_i|\mathbf{x}_i)$  of the response  $Y_i$  conditional on covariates  $\mathbf{x}_i$ , the analyst must remove the conditioning on  $Y_i > 0$  and transform back from  $\ln(y)$  space to  $y$ -space. As the error retransformation is a function of  $\mathbf{x}_i$ , the computation of covariate effects and their interpretation must proceed with care (Mullahy 1998). A more direct, two-part generalized linear model (GLM) specifies an exponential conditional model (ECM) for the untransformed response in the second part as  $E(Y_i|Y_i > 0, \mathbf{x}_i) = \exp(\mathbf{x}'_i \boldsymbol{\beta})$ , which does not have the need for retransformation (Mullahy 1998; Blough et al. 1999; Basu and Manning 2009).

Both the two-part GLM model and the conventional two-part model have limitations for estimation of covariate effects on the overall mean. While marginal effects of covariates on the overall mean can be calculated from these conditional two-part models (Belotti et al. 2015), the effects vary heterogeneously with each combination of values observed in the other covariates in the model. In an attempt to derive a single marginal effect in the presence of this heterogeneity, analysts have averaged over the effects calculated at each combination of covariates, in a process often referred to as “recycled predictions” or “standardization”. Specifically, the two parts must be combined as  $E(Y_i|\mathbf{x}_i) = \Pr(Y_i > 0|\mathbf{x}_i)E(Y_i|Y_i > 0, \mathbf{x}_i)$  and the exact form of this overall mean will depend on the model specification and the assumed distribution. Moreover, the effect of a covariate on the overall mean varies depending on the values of the other covariates in the model and the effect of a continuous covariate further depends on a reference value (e.g., the effect of age for 40 vs. 50 years differs from the effect of age for 50 vs. 60 years). The complex relationships between marginal means and covariates inherently implies both non-linearity and heterogeneity. However, interest may lie in obtaining a single value to describe the impact of a covariate. Analysts often produce a single value of a covariate effect by averaging all individual’s specific treatment effects (often functions like contrasts or ratios of predicted marginal means) over the entire study sample in a process commonly referred to as recycled predictions. However, this averaging process does not recover the true marginal effect if the true covariate effect is homogeneous on the overall mean or

systematically varies via an interaction term relative to the overall mean (Smith et al. 2014; Neelon et al. 2016).

Two-part models, therefore, do not in general provide homogeneous covariate effects on the marginal mean on the original scale of the response variable. In addition to requiring recycled predictions, they typically require resampling or the delta method to estimate the variances of covariate effects. The problem of obtaining straightforward marginal inference may become prohibitive as the complexity of two-part GLMs increases, such as when clustering and/or heterogeneity in scale parameters are modeled (Liu et al. 2010).

## 2.2 The MTP model

For independent observations, the generic form of a MTP model for semicontinuous data is given by Eq. (1). To obtain interpretable covariate effects on the marginal mean, Smith et al. (2014) proposed the MTP model that parameterizes the covariate effects directly in terms of the marginal mean,  $v_i = E(Y_i)$ , on the original (i.e., untransformed) data scale. The MTP model specifies the linear predictors

$$\begin{aligned} \text{logit}(\pi_i) &= \mathbf{z}'_i \boldsymbol{\alpha} \text{ and} \\ E(Y_i) &= v_i = \exp(\mathbf{x}'_i \boldsymbol{\beta}). \end{aligned} \quad (4)$$

Model-predicted means and standard errors can be easily obtained under this parameterization in a single step by estimating  $\exp(\mathbf{x}'_i \boldsymbol{\beta})$  at the desired values of the covariates. SAS code (SAS Institute, Cary, NC) implementing the MTP model using PROC NL MIXED is provided in Smith et al. (2014) and Neelon et al. (2016).

Most distributions with closed forms for the overall mean can be specified for the MTP model. While commonly used for two-part models, the log-normal distribution imposes a sometimes unrealistic condition of symmetry on the log-scale. Alternative distributions such as the log-skew-normal or generalized gamma have been proposed for the continuous part in an effort to relax these assumptions (Azzalini 1985; Chai and Bailey 2008; Manning et al. 2005; Liu et al. 2010) and each take the log-normal distribution as a special or limiting case.

### 2.2.1 The log-skew-normal MTP Model

Smith et al. (2014) developed this model with  $g(y_i|y_i > 0)$  taking either the log-normal or log-skew-normal (LSN) density. The LSN density relaxes the log-normal density's assumption of log-scale normality through inclusion of a shape parameter,  $\kappa$ , allowing skewness on the log-scale, with the log-normal density taking the special case of  $\kappa = 0$  (Azzalini 1985). Smith et al. found that the LSN density displayed better properties and more appropriately accounted for skewness commonly observed in semicontinuous data than the log-normal distribution. For this reason, we utilize the LSN MTP model here and omit the log-normal MTP model.

The generic form of the two-part LSN model for independent data is given by:

$$f(y_i) = (1 - \pi_i)^{1_{(y_i=0)}} \times [\pi_i \text{LSN}(y_i; \zeta_i, \sigma, \kappa)]^{1_{(y_i > 0)}}, \quad y_i \geq 0, i = 1, \dots, n, \quad (5)$$

where  $\text{LSN}(\cdot; \zeta_i, \sigma, \kappa)$  denotes the LSN distribution with location parameter  $\zeta_i$ , scale parameter  $\sigma > 0$ , and shape parameter  $\kappa$ , all on the log scale, given by

$$g(y_i|y_i > 0) = \frac{2}{\sigma y_i} \phi\left(\frac{\ln y_i - \xi_i}{\sigma}\right) \Phi\left(\frac{\kappa}{\sigma}(\ln y_i - \xi_i)\right), \tag{6}$$

where  $\phi(\cdot)$  and  $\Phi(\cdot)$  are the probability density function and cumulative distribution function, respectively, of the standard normal density. The marginal mean of  $Y_i$  is then given by:

$$E(Y_i) = v_i = 2\pi_i \exp\left(\xi_i + \frac{\sigma^2}{2}\right) \Phi(\sigma\delta), \tag{7}$$

where  $\delta = \kappa/\sqrt{1 + \kappa^2}$ . In order to re-express the LSN likelihood as a function of  $\beta$  from the MTP model in Eq. (4), we solve Eq. (7) for  $\xi_i$  in terms of  $\beta$ :

$$\begin{aligned} \xi_i &= \ln v_i - \ln 2 - \ln \pi_i - \ln[\Phi(\sigma\delta)] - \frac{\sigma^2}{2} \\ &= \mathbf{x}'_i \beta - \ln 2 - \ln \pi_i - \ln[\Phi(\sigma\delta)] - \frac{\sigma^2}{2}. \end{aligned}$$

After plugging this expression into Eq. (5) above, parameter estimates can be obtained using standard optimization routines such as Newton–Raphson or Fisher scoring.

### 2.2.2 The generalized gamma MTP model

Extending the ideas developed in Smith et al. (2014) and Smith and Preisser (2015), we also consider the MTP model fit with a generalized gamma (GG) distribution. The generalized gamma is a flexible, three-parameter distribution that takes as special cases the standard gamma, inverse gamma, Weibull, and log-normal distributions (Manning et al. 2005; Liu et al. 2010). The GG density is given by

$$g(y_i; \kappa, \mu_i, \sigma) = \frac{\eta^\eta}{\sigma y_i \Gamma(\eta) \sqrt{\eta}} \exp[u_i \sqrt{\eta} - \eta \exp(|\kappa| u_i)], \tag{8}$$

where  $\eta = |\kappa|^{-2}$ ,  $u_i = \text{sign}(\kappa)(\log(y_i) - \mu_i)/\sigma$ ,  $\mu_i$  is the location parameter,  $\sigma > 0$  is the scale parameter, and  $\kappa$  is the shape parameter. The GG density reduces to the log-normal density as  $\kappa \rightarrow 0$ . Thus, the LSN and GG distributions both take the log-normal distribution as a special case, but in general, do not overlap.

Under the GG distribution, the marginal mean of  $Y_i$  is given by (Manning et al. 2005; Liu et al. 2010):

$$\begin{aligned} E(Y_i) &= v_i = \exp(\mathbf{x}'_i \beta) \\ &= \pi_i \exp\left\{ \mu_i + \frac{\sigma \log(\kappa^2)}{\kappa} + \log[\Gamma(1/\kappa^2 + \sigma/\kappa)] - \log[\Gamma(1/\kappa^2)] \right\}. \end{aligned}$$

Following similar steps as done above for the LSN MTP model, we solve for  $\mu_i$  in terms of  $\beta$  to obtain

$$\begin{aligned}\mu_i &= \log(v_i) - \log(\pi_i) - \frac{\sigma \log(\kappa^2)}{\kappa} - \log[\Gamma(1/\kappa^2 + \sigma/\kappa)] + \log[\Gamma(1/\kappa^2)] \\ &= \mathbf{x}'_i \boldsymbol{\beta} - \log(\pi_i) - \frac{\sigma \log(\kappa^2)}{\kappa} - \log[\Gamma(1/\kappa^2 + \sigma/\kappa)] + \log[\Gamma(1/\kappa^2)].\end{aligned}$$

By plugging this expression for  $\mu_i$  into the GG density given by Eq. (8) and plugging this density into the generic form of the two-part model likelihood given by Eq. (1), the GG MTP model involving covariates as shown in Eq. (4) can be fit by maximum likelihood using standard optimization techniques.

### 2.3 GLMs fit with quasilielihood

GLMs fit using quasilielihood require only the specification of the mean and variance, as opposed to the full distribution, making them an attractive alternative when assumptions regarding the underlying parametric distribution are questionable. Specifically, when using a log link as is most commonly specified for health care expenditures, the overall mean model is given by

$$E(Y_i) = v_i = \exp(\mathbf{x}'_i \boldsymbol{\beta}), \quad (9)$$

the same as specified in the MTP model. A commonly used family of variance functions is the power family, taking the form

$$\text{Var}(Y_i) = \rho v_i^\lambda = \rho \exp(\lambda \mathbf{x}'_i \boldsymbol{\beta}), \quad (10)$$

and methods have been proposed to assist in determining the optimal value of  $\lambda$  (Manning and Mullahy 2001; Park 1966; Basu and Rathouz 2005). Specifically, commonly used values include  $\lambda = 0$ , constant variance,  $\lambda = 1$ , variance proportional to the mean, and  $\lambda = 2$ , variance proportional to the square of the mean, or equivalently, standard deviation proportional to the mean. Empirical “sandwich” variance estimators (Royall 1986; Kauermann and Carroll 2001) are commonly paired with such GLMs, such that if the variance is misspecified, they yield valid inference under many conditions where the marginal mean model is correctly specified (Fitzmaurice et al. 2012). For this comparison, we utilize the empirical standard errors and fit GLMs with  $\lambda = 0, 1$ , and 2, or with constant variance, variance proportional to the mean, and standard deviation proportional to the mean. Such models are implementable in most standard statistical software packages.

## 3 Simulation details

Given the limitations of standard two-part models in terms of marginal inference, we focus here on comparing models specifically designed to directly model the marginal mean. In this article, we compare five models: (1) two MTP models incorporating different parametric distributions; (2) three GLMs incorporating different mean-variance relationships fit with quasilielihood. These models take the same mean structure, providing easily comparable quantities, and fit the data on the original untransformed scale, so retransformation methods are not required.

### 3.1 Motivating example

To evaluate the performance of the MTP models and the GLMs, we conducted a series of simulation studies motivated in part by the analysis of a behavioral weight loss program that has been described in Smith et al. (2014). That study evaluated the effect of a system-wide weight loss intervention (MOVE!) implemented by the Veterans Affairs (VA) health care system beginning in 2006 to address the high prevalence of obesity among VA patients (Kahwati et al. 2011). Briefly, the total expenditures in the year following enrollment of 18,214 MOVE! enrollees were compared to those of 18,214 non-enrollees who were matched to the enrollees on sex, race (white or non-white), marital status (married or non-married), copay status (exempt vs. non-exempt), veterans integrated service network (VISN) of residence, BMI, and comorbidity burden, assessed via the 2002 diagnostic cost group (DCG) score. The goal of the analysis was to assess whether MOVE! enrollment was associated with a difference in mean total health care expenditures in the following year. With 17% of the MOVE! enrollees having zero expenditures in the year, results from one-part GLMs may have been unreliable, and use of the MTP model was therefore motivated.

### 3.2 Mean structure and properties examined

Basing covariate distributions on those of the MOVE! study, all simulated data scenarios considered here were generated assuming the following marginal mean structure:

$$E(Y_i) = v_i = \exp(6 + 0.2x_{1i} - 0.01x_{2i} + 0.05x_{3i}), \quad (11)$$

where  $x_{1i} \sim \text{Bernoulli}(0.5)$ ,  $x_{2i} \sim N(0, 1)$ , and  $x_{3i} \sim \text{Pois}(1)$ . Parameter values were chosen to mimic similar distributions as those observed in the MOVE! data, while also maintaining a similar magnitude difference between treatment arms. We considered three different scenarios for the distribution of the positive values of  $Y_i$ : (1) distributed as LSN with low log-scale skewness, (2) distributed as LSN with higher log-scale skewness, and (3) distributed as generalized gamma (GG). For each of these three scenarios, we considered data with approximately 10, 20, and 40% zeros to assess the influence of the size of the discrete point mass on the performance of each model. Specifically, zeros were introduced in the  $Y_i$ 's with probability  $\pi_i$ , where  $\pi_i$  was given by  $\text{logit}(\pi_i) = 3 - 2.4x_{1i} + 1.5x_{2i} + 2x_{3i}$ ,  $\text{logit}(\pi_i) = 3 - 4x_{1i} + 3.5x_{2i} + 2.5x_{3i}$ , and  $\text{logit}(\pi_i) = 3 - 7x_{1i} + 5x_{2i} + 2x_{3i}$  to achieve approximately 10, 20, and 40% zeros, respectively. For each of these nine combinations of distributions and percentages of zeros, we evaluated datasets of sample sizes 200, 1000, 10,000, and 50,000 to assess the impact of sample size on model performance, resulting in a total of 36 simulated scenarios with 1000 datasets each. In each case, the mean model of the GLMs and MTP models fit to the data were correctly specified as  $E(Y_i) = \exp(\beta_0 + \beta_1x_{1i} + \beta_2x_{2i} + \beta_3x_{3i})$ . Empirical sandwich variance estimators were used for the GLMs.

To assess the performance of each model, we examined the bias of parameter estimates and model-predicted “total expenditures”, the simulated outcome. Total expenditure bias was calculated as the average difference in an individual’s model-predicted total expenditure and their true theoretical mean total expenditure, based on their respective combination of covariates, calculated from Eq. (11). We also examined coverage probability of nominal 95% Wald-type confidence intervals for each parameter as well as total expenditure predictions. For each of the 36 scenarios, we then re-generated data with  $\beta_1 = 0$  to



mimic a null hypothesis of no treatment effect for the binary variable  $x_{1i}$  in order to evaluate Type I error rates for each model at a nominal 0.05 significance level. The simulations presented here examine the performance of both the log-skew-normal MTP model (Smith et al. 2014) and the generalized gamma MTP model (Smith and Preisser 2015).

### 3.3 Simulation 1: log-skew-normal data

In the first set of simulations, we assumed the positive values of  $Y_i$  followed the LSN density shown in Eq. (6). Thus, in this simulation, the parametric assumptions of the LSN MTP model were met, while those of the GG MTP were not. We set the scale parameter,  $\sigma$ , at 1.2 and set the log-scale skewness parameter  $\kappa$  at 0.5 and 5.0 for the low log-scale skewness and high log-scale skewness simulations, respectively.

### 3.4 Simulation 2: generalized gamma data

In the second set of simulations, we investigated the performance of the MTP models relative to that of the one-part GLMs when the parametric distributional assumptions of the LSN MTP model were not met and those of the GG MTP model were. As in Simulation 1, we set the scale parameter,  $\sigma$ , at 1.2, and we set the shape parameter,  $\kappa$ , at 0.63 based on the analysis from Liu et al. (2010).

## 4 Simulation results

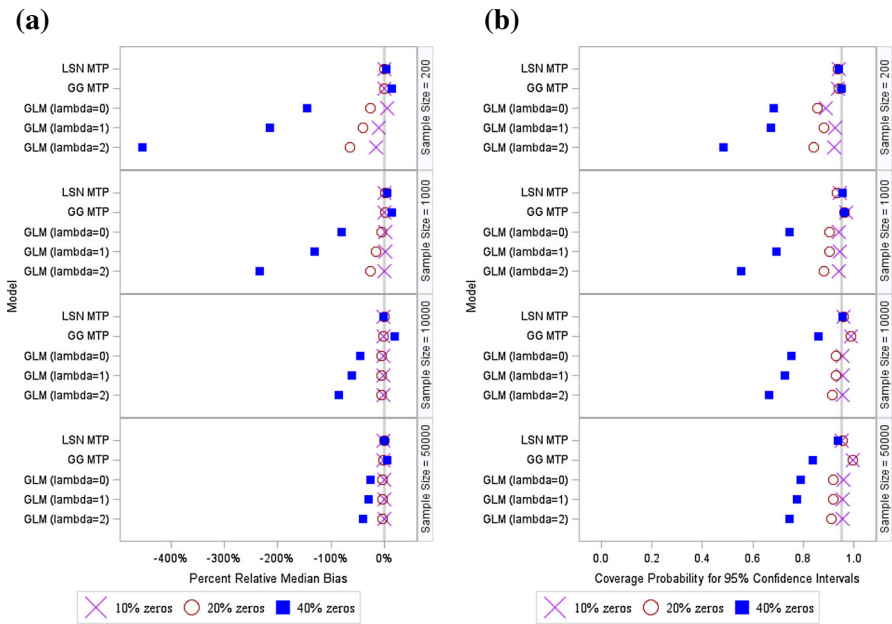
### 4.1 Log-skew-normal with low log-scale skewness results

Descriptive statistics on the 36 simulated datasets are shown in Table 1. Percent relative median bias and coverage probabilities of the 95% Wald-type confidence intervals for the main binary covariate effect of interest from the models fit on the datasets generated from the LSN distribution with low log-scale skewness are shown in Fig. 1. Across all simulations and models, results were similar for each of the covariates, so only the main binary covariate of interest is shown here. The remaining results can be found in the online supplementary material. The LSN MTP model generally provided the least biased estimates under all scenarios, which is expected given that the parametric assumptions of the model were met. The GG MTP model also incurred minimal bias, lower than that of the GLMs under almost all scenarios. Among all models, bias generally decreased with sample size and was noticeably larger among the GLMs when the data had 40% zeros as opposed to 10 or 20% zeros. In particular, estimates of  $\beta_1$ , the treatment effect of main interest, were negatively biased under the GLMs. With 40% zeros, the negative bias increased such that, for sample sizes of 200 and 1000, the GLMs were on average producing negative treatment effect estimates instead of positive ones.

The LSN MTP model maintained approximately 0.95 coverage probability for covariate effects with all percentages of zeros. The GG MTP model often provided 0.95 coverage probability, with some reduction seen with 40% zeros when the sample size increased to 10,000 or larger. Even with empirical standard errors, modest reductions in coverage probability were seen for the GLMs with 20% zeros, with coverage ranging from 0.74 to

**Table 1** Descriptive statistics on the simulated datasets

Distribution	Sample size	10% zeros			20% zeros			40% zeros		
		Percent Zeros (%)	Mean (SD)	Median (Q1–Q3)	Percent Zeros (%)	Mean (SD)	Median (Q1–Q3)	Percent Zeros (%)	Mean (SD)	Median (Q1–Q3)
LSN low log-skewness	200	10	470 (884)	227 (89–526)	21	466 (1511)	184 (41–478)	40	472 (9082)	89 (0–351)
	1000	10	471 (878)	227 (88–525)	21	470 (1804)	184 (40–476)	40	452 (6489)	88 (0–351)
	10,000	10	471 (878)	228 (89–525)	21	472 (2368)	184 (41–477)	40	466 (12301)	88 (0–351)
	50,000	10	472 (880)	228 (88–526)	21	472 (2259)	184 (41–477)	40	473 (28105)	88 (0–351)
LSN high log-skewness	200	10	470 (722)	285 (171–527)	21	420 (1083)	231 (119–452)	40	472 (7902)	166 (0–388)
	1000	10	471 (714)	285 (171–527)	21	470 (1540)	249 (124–492)	40	454 (5721)	166 (0–388)
	10,000	10	471 (715)	285 (171–527)	21	472 (2010)	250 (124–493)	40	466 (10563)	166 (0–388)
	50,000	10	472 (718)	285 (171–528)	21	472 (2868)	250 (124–493)	40	473 (23002)	167 (0–389)
GG	200	10	474 (707)	243 (73–597)	21	475 (1703)	188 (23–538)	40	498 (15784)	74 (0–390)
	1000	10	472 (702)	241 (71–593)	21	471 (1664)	186 (22–535)	40	470 (13132)	72 (0–387)
	10,000	10	472 (701)	242 (72–594)	21	473 (6344)	187 (22–537)	40	460 (7907)	72 (0–389)
	50,000	10	472 (700)	242 (72–594)	21	472 (1991)	187 (22–537)	40	473 (22821)	72 (0–389)



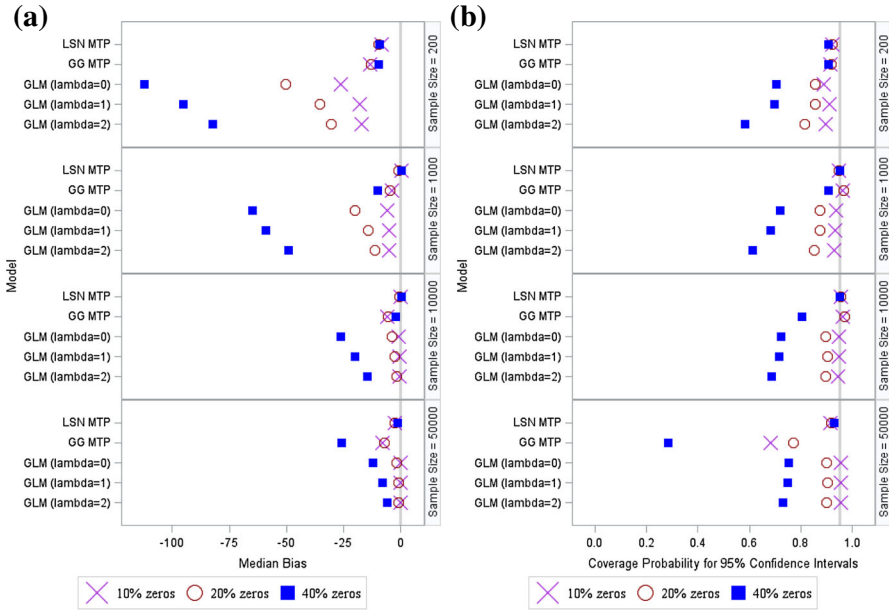
**Fig. 1** Percent relative median bias and coverage of 95% Wald-type confidence intervals for  $\beta_1$ , the binary covariate effect of interest, from low log-scale skewness LSN data. **a** Percent relative median bias, **b** coverage probability

0.93. With 40% zeros, coverage dropped significantly for the GLMs. In particular, coverage for  $\beta_1$ , the treatment effect, ranged from 0.48 to 0.79 for the GLMs with 40% zeros.

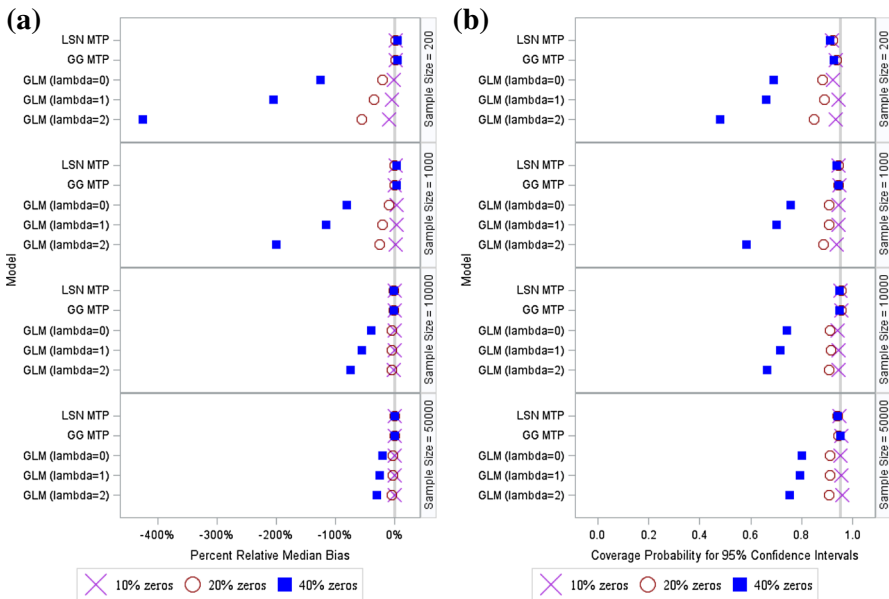
Median bias and coverage probabilities of the 95% Wald-type confidence intervals for the total expenditure prediction from the models fit on the datasets generated from the LSN distribution with low log-scale skewness are shown in Fig. 2. The GLMs generally incurred more bias than the MTP models in total expenditure prediction, particularly for the smaller sample sizes and greater proportion of zeros, with the exception of the GG MTP model when the sample size was 50,000. A similar pattern was seen with coverage probabilities, with very poor coverage seen among the GLMs when sample sizes were small or there were a greater proportion of zeros. The LSN MTP maintained coverage near 0.95; the GG MTP suffered somewhat worse coverage for prediction with larger sample sizes.

**4.2 Log-skew-normal with high log-scale skewness results**

Percent relative median bias and coverage probabilities of the 95% Wald-type confidence intervals for the covariate effect of interest from the models fit on each of the datasets generated from the LSN distribution with higher log-scale skewness are shown in Fig. 3. The LSN MTP model again generally provided the least biased estimates under all scenarios, as its parametric assumptions were met. The GG MTP model also incurred minimal bias for covariate effects. Similarly to the results above, bias generally decreased with sample size, and among the GLMs, was noticeably larger when the data had 40% zeros as opposed to 10 or 20%. Once again, estimates of  $\beta_1$ , the treatment effect of main interest, were negatively biased under almost all of the GLMs, and with 40% zeros, the GLMs were



**Fig. 2** Median bias and coverage of 95% Wald-type confidence intervals for total cost prediction from low log-scale skewness LSN data. **a** Median bias, **b** coverage probability



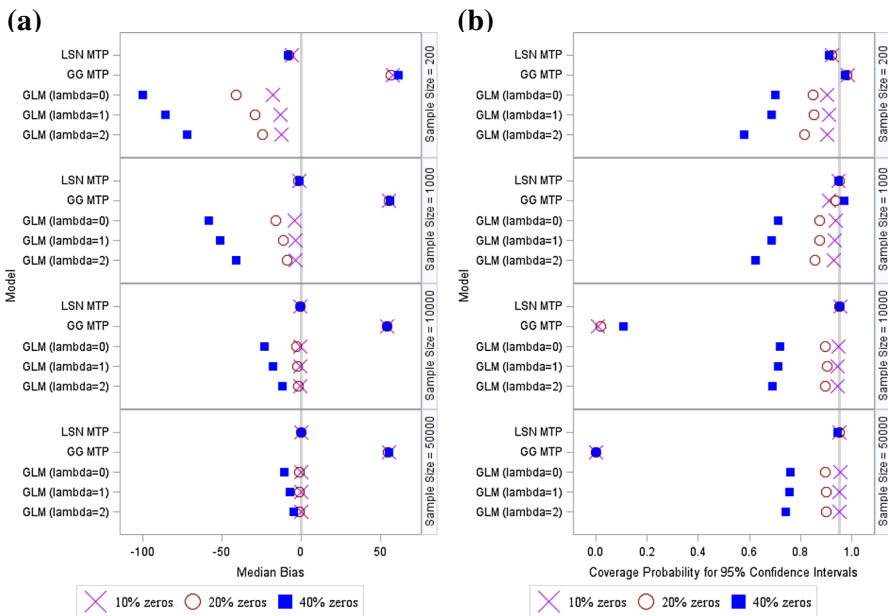
**Fig. 3** Percent relative median bias and coverage of 95% Wald-type confidence intervals for  $\beta_1$ , the binary covariate effect of interest, from high log-scale skewness LSN data. **a** Percent relative median bias, **b** coverage probability

on average producing negative treatment effect estimates instead of positive ones for sample sizes of 200 and 1000.

For coverage probabilities, results were relatively similar to those from the LSN with low log-scale skewness data. Coverage probabilities for the covariate effects remained fairly close to 0.95 under the both MTP models. The coverage for  $\beta_1$  from the LSN MTP model dropped to as low as 0.86 with a sample size of 200. Similarly to the previous results, the GLMs incurred modest reductions in coverage probability with 20% zeros, with coverage ranging from 0.73 to 0.94. With 40% zeros, coverage from the GLMs dropped more significantly, with coverage for the treatment effect ranging from 0.48 to 0.80.

Median bias and coverage probabilities of the 95% Wald-type confidence intervals for the total expenditure prediction from the models fit on the datasets generated from the LSN distribution with higher log-scale skewness are shown in Fig. 4. The LSN MTP model, as expected, has low bias and appropriate coverage probability. The GG MTP model, on the other hand, experienced increased bias for the intercept, and therefore, the total expenditure prediction.

Additionally, the GG MTP model suffered severe lack of coverage for the total expenditure prediction when the sample size was 10,000 or greater. In this scenario, coverage under the GG MTP model dropped to <0.0001 for the total expenditure prediction. This is likely due to a combination of bias and underestimated standard errors, with a larger problem seen in the high log-skewness LSN data than the low log-skewness because the low log-skewness LSN data are closer to log-normally distributed, a special case of the GG distribution. The LSN MTP model was the only model to provide sufficiently good coverage for total expenditure predictions with 40% zeros and 10,000 or greater subjects, with the next highest coverage probability being 0.76.



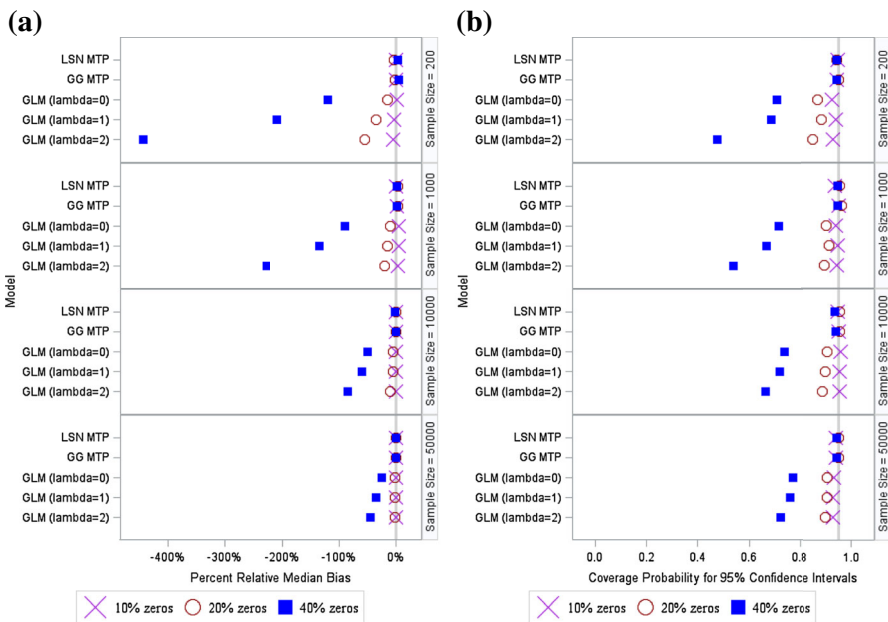
**Fig. 4** Median bias and coverage of 95% Wald-type confidence intervals for total cost prediction from high log-scale skewness LSN data. **a** Median bias, **b** coverage probability

### 4.3 Generalized gamma results

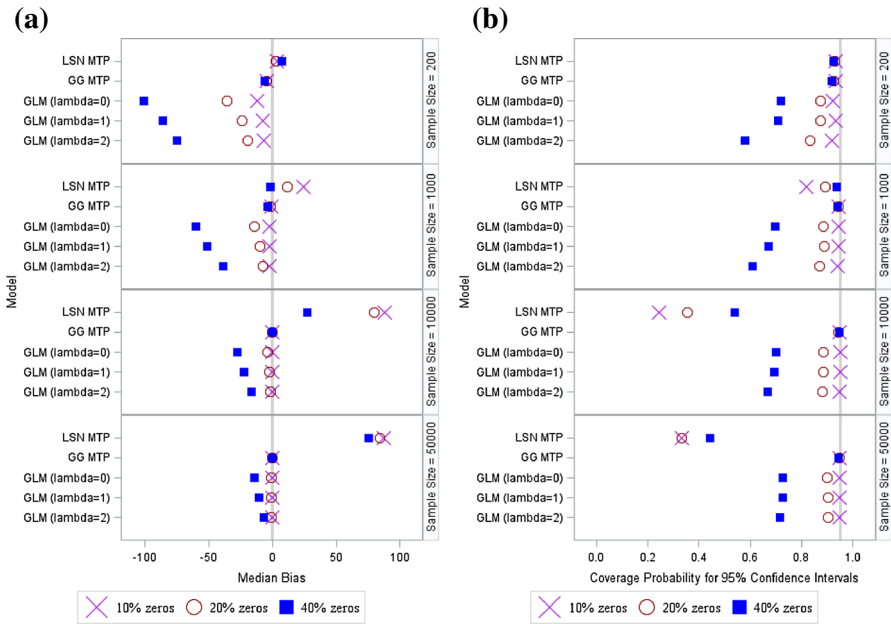
Percent relative median bias and coverage probabilities for the covariate effect of interest from the models fit on each of the datasets generated from the GG distribution is shown in Fig. 5. Under this scenario, when the parametric assumptions of the LSN MTP model were no longer met, bias in the estimation of covariate effects remained low for the LSN MTP model regardless of sample size or percentage of zeros. The GG MTP incurred almost no bias as expected given that the parametric assumptions were met. The GLMs again performed much better with 10 or 20% zeros than with 40%, and the bias incurred appeared to decrease with sample size. Even with a sample of 50,000, however, the estimate of treatment effect under the GLMs with 40% zeros was strongly negatively biased, and with the smaller sample sizes, often resulted in estimates of treatment effect that were in the wrong direction.

Similar trends were seen in the coverage probabilities. Coverage for the covariate effect parameters remained close to 0.95 under the LSN MTP model regardless of sample size or percentage of zeros. The GG MTP model maintained coverage probability near 0.95 in all cases. Similar to the results using the LSN data, the GLMs showed a modest reduction in coverage with 20% zeros, with values ranging from 0.74 to 0.94. With 40% zeros, however, coverage for the GLMs dropped significantly for all parameters. In particular, coverage for  $\beta_1$ , the treatment effect, ranged from 0.48 to 0.77 in this scenario.

Median bias and coverage probabilities of the 95% Wald-type confidence intervals for the total expenditure prediction from the models fit on the datasets generated from the GG distribution are shown in Fig. 6. When the parametric assumptions of the LSN MTP model were no longer met, the LSN MTP model incurred more bias in estimating the intercept,



**Fig. 5** Percent relative median bias and coverage of 95% Wald-type confidence intervals for  $\beta_1$ , the binary covariate effect of interest, from GG data. **a** Percent relative median bias, **b** coverage probability

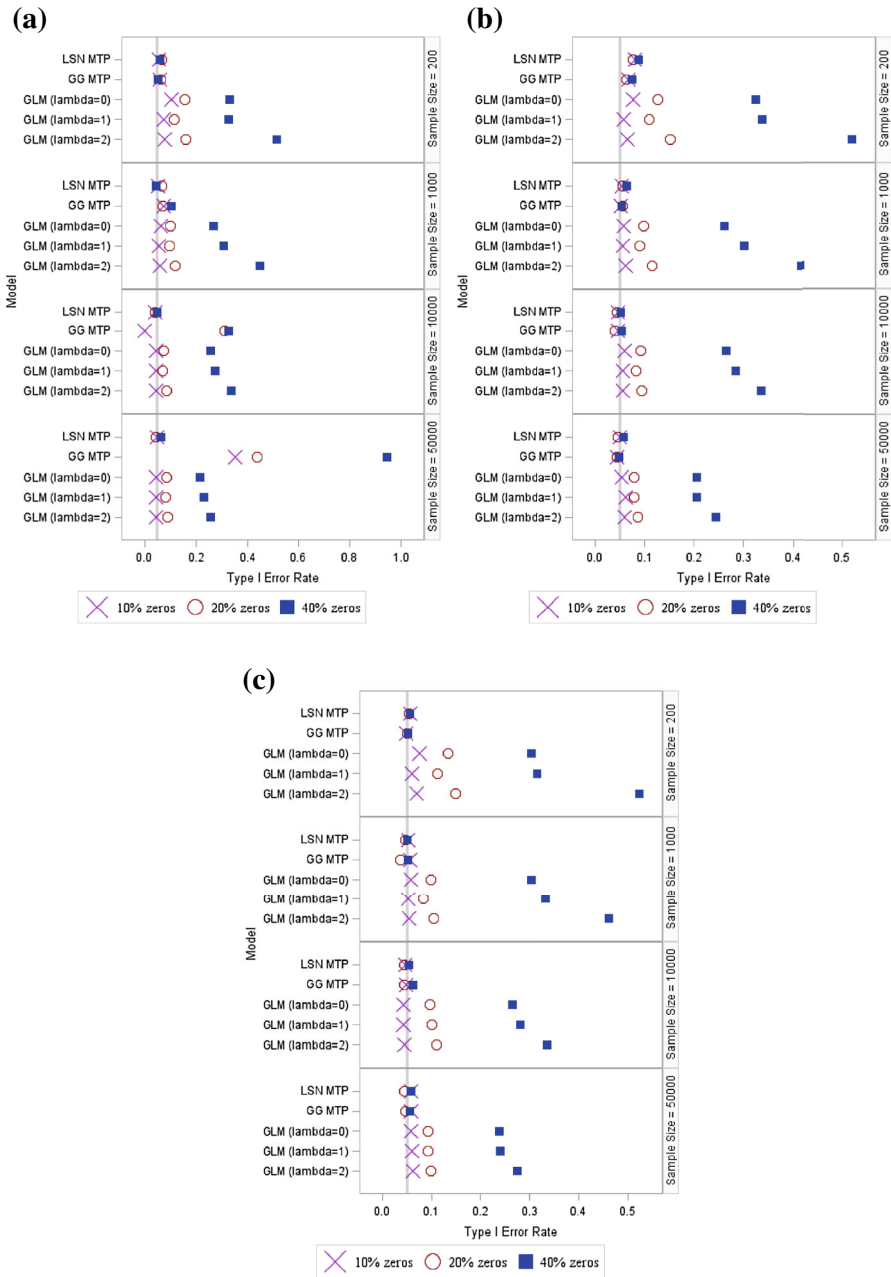


**Fig. 6** Median bias and coverage of 95% Wald-type confidence intervals for total cost prediction from GG data. **a** Median bias, **b** coverage probability

$\beta_0$ , and subsequently, in total expenditure prediction. Notably, the bias in the intercept and total expenditure prediction did not improve with increased sample size. Coverage probabilities for the total expenditure prediction under the LSN MTP model dropped to as low as 0.25. The LSN MTP outperformed the GLMs with regards to coverage for total expenditure prediction with 40% zeros when the sample size was smaller (200 or 1000). With 10,000 or 50,000 subjects, however, the LSN MTP model coverage of total expenditure prediction dropped substantially and the GLMs provided higher coverage, particularly for the lower percentage of zeros. The GG MTP model was the only model to provide sufficiently good coverage for total expenditure prediction with 40% zeros and 10,000 or greater subjects, with the next highest coverage probability being 0.73.

#### 4.4 Type I error rates

Type I error rates from each of the models re-run on data simulated with  $\beta_1 = 0$  under each of the distributions are shown in Fig. 7. Type I error rates remained close to 0.05 for the LSN MTP model under all scenarios. Type I errors for the GG MTP model remained close to 0.05 under most scenarios, but increased substantially when fit to the LSN data with low log-skewness, particularly for larger sample sizes; for the sample size of 10,000, the type I error under the GG MTP model increased to 0.32, and with a sample size of 50,000 increased to as high as 0.94. Type I errors remained at least somewhat inflated under almost all scenarios with 20 or 40% zeros for the GLMs, but were near 0.05 with only 10% zeros. When the data contained 20% zeros, the GLM type I error rates ranged from 0.07 to 0.16. With 40% zeros, they ranged from 0.21 to 0.52. Type I errors seemed to generally



**Fig. 7** Type I error rates at nominal significance level 0.05 for LSN and GG data. **a** Low log skewness LSN data, **b** high log skewness LSN data, and **c** GG data

decrease with increasing sample size for the GLMs, but particularly with 40% zeros, rates remained significantly higher than the nominal 0.05 significance level at all sample sizes examined.



### 5 MOVE! study analysis

To assess the impact of model choice in our motivating example described in Sect. 3.1, we fit the two MTP models and the three one-part GLMs to the same data, using the same mean model. In order to evaluate the effect of MOVE! enrollment on mean health care expenditures in the following year under each model, we fit the overall mean model specified as:

$$E(Y_i) = \exp(\beta_0 + \beta_1 x_{i1} + \beta_2 x_{i2} + \beta_3 x_{i3} + \beta_4 x_{i4}),$$

where  $x_{i1} = 1$  if individual  $i$  was enrolled in MOVE! and 0 otherwise, and we additionally adjusted for  $x_{i2}$ ,  $x_{i3}$ , and  $x_{i4}$ , individual  $i$ 's BMI, age, and DCG score, respectively. For the binary part of the MTP models, we included the same covariates:

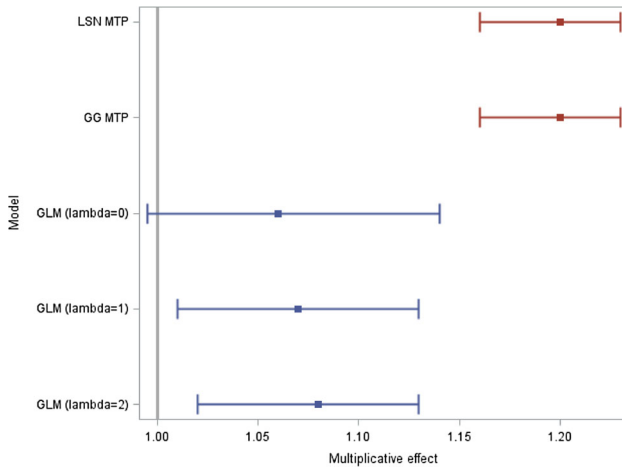
$$\text{logit}(\pi_i) = \alpha_0 + \alpha_1 x_{i1} + \alpha_2 x_{i2} + \alpha_3 x_{i3} + \alpha_4 x_{i4}.$$

To compare fit across models, we calculated the mean squared error (MSE) for each. Table 2 presents the parameter estimates, standard errors, and MSE from each of the models. Figure 8 presents the model-estimated multiplicative effects of MOVE! enrollment on mean health care expenditures, calculated by exponentiating  $\beta_1$ , with 95% confidence intervals.

Parameter estimates from the LSN and GG MTP models were quite similar. While the parameter estimates from the GLMs were similar to each other, they differed from those of the MTP models. In particular, the one-part GLMs provided estimates of the effect of MOVE! enrollment that were noticeably lower in magnitude than those estimated by the MTP models. Specifically, both MTP models estimated that MOVE! enrollment was associated with 20% higher mean health care expenditures in the year following

**Table 2** Overall mean parameter estimates (standard errors) and mean squared error from each of the five models compared

	Parameter	LSN MTP	GG MTP	GLM $\lambda = 0$	GLM $\lambda = 1$	GLM $\lambda = 2$
Intercept	$\beta_0$	9.11 (0.087)	9.11 (0.087)	9.51 (0.173)	9.42 (0.156)	9.43 (0.156)
MOVE! enrollment	$\beta_1$	0.18 (0.014)	0.18 (0.015)	0.06 (0.034)	0.07 (0.029)	0.07 (0.027)
Age	$\beta_2$	0.01 (0.002)	0.01 (0.002)	-0.003 (0.004)	0.002 (0.004)	-0.02 (0.004)
BMI	$\beta_3$	-0.02 (0.001)	-0.02 (0.001)	-0.02 (0.002)	-0.02 (0.001)	-0.02 (0.001)
DCG score	$\beta_4$	1.29 (0.044)	1.29 (0.044)	1.20 (0.082)	1.28 (0.085)	1.29 (0.098)
Scale parameter	$\sigma$	1.41 (0.018)	1.21 (0.005)	-	-	-
Shape parameter	$\kappa$	0.84 (0.048)	-0.09 (0.012)	-	-	-
Mean squared error		348,107,503	348,133,800	397,489,839	397,489,878	397,489,865



**Fig. 8** Multiplicative effects of MOVE! enrollment overall mean parameter estimates (95% confidence intervals) estimated from each of the five models compared

enrollment, while the one-part models estimated a much smaller effect, ranging from 6 to 8%. Further, the 95% confidence intervals for the effect of MOVE! enrollment from the MTP models and GLMs do not overlap. The LSN MTP model provided the lowest MSE among all five models. The GG MTP model was the next best fitting model, and all GLMs performed similarly with much higher MSE values than those of the MTP models.

## 6 Discussion

Our results suggest that, in general, one-part GLMs are not appropriate for use with data including a significant proportion of zeros. The one-part GLMs incurred increased bias, lower than nominal coverage, and increased type I error rates in all scenarios with 20 or 40% zeros, and for the sample size of  $n = 200$  with 10% zeros. Results improved with larger sample sizes but, particularly with 40% zeros, even a sample size of 50,000 was not large enough for the one-part GLMs to overcome the bias and low coverage. This conclusion differs from that provided in Buntin and Zaslavsky (2004), which was based on a dataset with 8.6% zeros and over 10,000 individuals. For datasets with >10% zeros, we advise that one-part GLMs be avoided as they provide biased and unreliable results.

Despite reliance on a parametric model, the MTP models, and in particular the LSN MTP model, appeared to be fairly robust to distributional misspecification in covariate effect estimation. Specifically, and conversely to the GLMs, the MTP models appeared particularly robust in the smaller sample sizes. At sample sizes of 200 and 1000, both MTP models maintained low bias and appropriate coverage and type I error rates for covariate effects, regardless of whether the distribution was correctly specified. In larger sample sizes, the LSN MTP continued to perform well for covariate effects, with low bias and appropriate coverage and type I error rates, regardless of the data-generating distribution. The GG MTP model, however, suffered lower coverage and high type I error rates with larger sample sizes (10,000 and 50,000) when the distribution was incorrectly specified, likely due in large part to underestimated standard errors.

With regards to prediction, results were more nuanced. The MTP models fit with the distribution incorrectly specified provided biased predictions with lower than nominal coverage probabilities, particularly for larger sample sizes. If an analyst is interested in both estimation of covariate effects and in outcome prediction, specification testing to determine the most appropriate distribution is crucial. While the semi-parametric nature of the one-part GLMs has often been viewed as an appealing alternative to avoid distributional dependence, our results suggest that this is not a reliable option for semicontinuous data with a significant proportion of zeros. Examination of skewness parameters, residual plots, and plots of observed versus predicted values can inform distributional decisions for parametric models.

Results from the MOVE! analysis showed that the GLMs estimated a much smaller effect of MOVE! enrollment on mean health care expenditures in the following year than did the MTP models. This pattern matches that seen in the simulation results, in which the GLMs provided negatively biased treatment effect estimates. The MOVE! analysis contained 17% zeros, slightly less than the lowest percentage considered in our simulation studies, and had a sample size of 36,428. While our simulation results showed improvement in the performance of the GLMs with larger sample sizes, the sample size needed to alleviate problems with bias and coverage may be quite large and may depend upon other aspects of the data, such as skewness and the heaviness of the long tail covering higher expenditures, and not only the percentage of zeros.

This study is subject to several limitations. We focus primarily on the estimation of covariate effects as opposed to prediction, and thus we consider only models that provide direct, homogeneous covariate effects on the overall mean, including both the zero and positive valued outcomes. As such, models such as the conventional two-part model (Duan et al. 1983; Manning et al. 1981) were not considered as it does not, in most cases, provide estimates of homogeneous covariate effects on the marginal mean. Rather, the covariate effects estimated from conventional two-part models typically vary based on the values specified for the other covariates in the model, making such estimates non-comparable with those of the one-part GLMs and MTP models (Smith et al. 2014). Additionally, other specifications of the GLMs could be considered. We focused on GLMs with a log link so as to fit the correct mean model in the simulation studies, but other link functions could be evaluated. Additional mean-variance relationships or link functions could also be estimated from the data, such as proposed by Basu and Rathouz (2005).

Additionally, we did not examine effects of covariate misspecification. As when fitting any model, attention must be paid to correct covariate specification, and the two parts of the MTP may take different covariates in each part. Which covariates to consider should generally be based upon a combination of subject matter expertise and standard covariate selection processes, such as forward or backward selection, comparison of AIC values or likelihood ratio tests. Asymptotically, if the model is overspecified, correct inference will still be obtained, but computational issues could arise, particularly for smaller datasets. Alternatively, when a large number of covariates are all deemed relevant, they may all be included in a shared-parameter model like that of Preisser et al. (2016). In the case of under-specification, or omission of necessary covariates, biased parameter estimates may result in any of the models examined.

Limitations notwithstanding, MTP models present a significant breakthrough in the next generation of analytic options for semicontinuous data. MTP models achieve accurate and precise analyses of covariate effects on marginal mean expenditures with interpretations based on the original scale without the need for post-modeling computations. Future work may be needed to find methods that accommodate a large proportion of zeros with less

reliance on parametric assumptions, particularly when interest in is prediction of mean or total expenditures. The MTP model could be extended to additional distributions or mixtures of distributions, and addition of empirical standard errors to the MTP model may increase coverage probabilities for predictions under questionable distributional assumptions. The MTP models could also be extended to allow heteroscedasticity by specifying the scale parameter to depend on covariates following Liu et al. (2010). Regardless of modeling approach chosen, analysts will continue to need to carefully balance trade-offs in model fit, robustness, and interpretability with their specific analytic goals in mind.

**Acknowledgements** The views expressed in this article are those of the authors and do not necessarily reflect the position or policy of the Department of Veteran Affairs, Duke University, the Medical University of South Carolina, or the University of North Carolina.

**Funding** This work was funded by the Diabetes QUERI and the National Center for Health Promotion and Disease Prevention within the Department of Veterans Affairs (VA). This work was supported by the Center of Innovation for Health Services Research in Primary Care (CIN 13-410) at the Durham VA Medical Center. Dr. Maciejewski is supported by a Research Career Scientist award (RCS 10-391) and a Grant (IIR 10-159) from the VA. The Diabetes QUERI, the National Center for Health Promotion and Disease Prevention, and the Health Services Research and Development Service, Department of Veterans Affairs had no role in the design, conduct, collection, management, analysis, or interpretation of the data; or in the preparation, review, or approval of the manuscript.

### Compliance with ethical standards

**Conflict of interest** Dr. Maciejewski owns stock in Amgen due to his spouse's employment. No other authors have conflicts of interest.

**Ethical approval** All procedures performed in studies involving human participants were in accordance with the ethical standards of the institutional and/or national research committee and with the 1964 Helsinki declaration and its later amendments or comparable ethical standards.

**Informed consent** This study was approved by the institutional review board (including waiver of informed consent) of the Durham VA Medical Center.

## References

- Azzalini, A.: A class of distributions which includes the normal ones. *Scand. J. Stat.* **12**, 171–178 (1985)
- Basu, A., Manning, W.G.: Issues for the next generation of health care cost analyses. *Med. Care* **47**, S109–S114 (2009)
- Basu, A., Rathouz, P.J.: Estimating marginal and incremental effects on health outcomes using flexible link and variance function models. *Biostatistics* **6**, 93–109 (2005)
- Belotti, F., Deb, P., Manning, W.G., Norton, E.C.: twopm: two-part models. *Stata J.* **15**, 3–20 (2015)
- Blough, D.K., Madden, C.W., Hornbrook, M.C.: Modeling risk using generalized linear models. *J. Health Econ.* **18**, 153–171 (1999)
- Buntin, M.B., Zaslavsky, A.M.: Too much ado about two-part models and transformation?: comparing methods of modeling Medicare expenditures. *J. Health Econ.* **23**, 525–542 (2004)
- Chai, H.S., Bailey, K.R.: Use of log-skew-normal distribution in analysis of continuous data with a discrete component at zero. *Stat. Med.* **27**, 3643–3655 (2008)
- Cragg, J.G.: Some statistical models for limited dependent variables with application to the demand for durable goods. *Econometrica* **39**, 829–844 (1971)
- Diehr, P., Yanez, D., Ash, A., Hornbrook, M., Lin, D.: Methods for analyzing health care utilization and costs. *Annu. Rev. Public Health* **20**, 125–144 (1999)
- Duan, N., Manning Jr., W.G., Morris, C.N., Newhouse, J.P.: A comparison of alternative models for the demand of medical care. *J. Bus. Econ. Stat.* **1**, 115–126 (1983)

- Fitzmaurice, G.M., Laird, N.M., Ware, J.H.: *Applied Longitudinal Analysis*. Wiley, New York (2012)
- Kahwati, L.C., Lance, T.X., Jones, K.R., Kinsinger, L.S.: RE-AIM evaluation of the Veterans Health Administration's MOVE! weight management program. *Transl. Behav. Med.* **1**, 551–560 (2011)
- Kauermann, G., Carroll, R.J.: A note on the efficiency of sandwich covariance matrix estimation. *J. Am. Stat. Assoc.* **96**, 1387–1396 (2001)
- Liu, L., Cowen, M.E., Strawderman, R.L., Shih, Y.-C.T.: A flexible two-part random effects model for correlated medical costs. *J. Health Econ.* **29**, 110–123 (2010)
- Madden, C.W., Mackay, B.P., Skillman, S.M., Ciol, M., Diehr, P.K.: Risk adjusting capitation: applications in employed and disabled populations. *Health Care Manag. Sci.* **3**, 101–109 (2000)
- Manning, W.G., Mullahy, J.: Estimating log models: to transform or not to transform? *J. Health Econ.* **20**, 461–494 (2001)
- Manning, W.G., Morris, C.N., Newhouse, J.P., Orr, L.L., Duan, N., Keeler, E., Leibowitz, A., Marquis, K., Marquis, M., Phelps, C.: A two-part model of the demand for medical care: preliminary results from the health insurance study. In: *Health, Economics, and Health Economics*, pp. 103–123. North-Holland, Amsterdam (1981)
- Manning, W.G., Basu, A., Mullahy, J.: Generalized modeling approaches to risk adjustment of skewed outcomes data. *J. Health Econ.* **24**, 465–488 (2005)
- Mullahy, J.: Much ado about two: reconsidering retransformation and the two-part model in health econometrics. *J. Health Econ.* **17**, 247–281 (1998)
- Neelon, B., O'Malley, A.J., Smith, V.: Modeling zero-modified count and semicontinuous data in health services research, Part 2: Case studies. *Stat. Med.* **35**, 5094–5112 (2016)
- Park, R.E.: Estimation with heteroscedastic error terms. *Econometrica* **34**, 888 (1966)
- Preisser, J.S., Das, K., Long, D.L., Divaris, K.: Marginalized zero-inflated negative binomial regression with application to dental caries. *Stat. Med.* **35**, 1722–1735 (2016)
- Royall, R.M.: Model robust confidence intervals using maximum likelihood estimators. *Int. Stat. Rev.* **54**, 221–226 (1986)
- Smith, V.A., Preisser, J.S.: Direct and flexible marginal inference for semicontinuous data. *Stat. Methods Med. Res.* (2015). doi:[10.1177/0962280215602290](https://doi.org/10.1177/0962280215602290) (published online September 1, 2015)
- Smith, V.A., Preisser, J.S., Neelon, B., Maciejewski, M.L.: A marginalized two-part model for semicontinuous data. *Stat. Med.* **33**, 4891–4903 (2014)
- Smith, V.A., Neelon, B., Preisser, J.S., Maciejewski, M.L.: A marginalized two-part model for longitudinal semicontinuous data. *Stat. Methods Med. Res.* (2015). doi:[10.1177/0962280215592908](https://doi.org/10.1177/0962280215592908) (published online July 7, 2015)