A semiparametric 2-part mixed-effects heteroscedastic transformation model for correlated right-skewed semicontinuous data

HUAZHEN LIN
Department of Mathematics, Sichuan University, Chengdu, Sichuan 610064, People Republic of China

XIAO-HUA ZHOU*
HSR&D Center of Excellence, VA Puget Sound Health Care System, Seattle, WA 98108, USA and Department of Biostatistics, University of Washington, Seattle, WA 98195, USA
azhou@u.washington.edu

SUMMARY
In longitudinal or hierarchical structure studies, we often encounter a semicontinuous variable that has a certain proportion of a single value and a continuous and skewed distribution among the rest of values. In this paper, we propose a new semiparametric 2-part mixed-effects transformation model to fit correlated skewed semicontinuous data. In our model, we allow the transformation to be nonparametric. Fitting the proposed model faces computational challenges due to intractable numerical integrations. We derive the estimates for the parameter and the transformation function based on an approximate likelihood, which has high-order accuracy but less computational burden. We also propose an estimator for the expected value of the semicontinuous outcome on the original scale. Finally, we apply the proposed methods to a clinical study on effectiveness of a collaborative care treatment on late-life depression on health care costs.

Keywords: Laplace approximation; Mixed effects; Right skewed; Semicontinuous; Semiparametric; Transformation model.

1. INTRODUCTION
This study is motivated by an analysis to examine the effectiveness of the Improving Mood-Promoting Access to Collaborative Treatment (IMPACT) program for late life depression (Unutzer and others, 2002). Intervention patients had access for up to 12 months to a depression care manager who offered education, care management, and support of antidepressant management. One primary outcome, the total inpatient cost over a half-year period, was collected at months 6, 12, 18, and 24.

*To whom correspondence should be addressed.

© The Author 2009. Published by Oxford University Press. All rights reserved. For permissions, please e-mail: journals.permissions@oxfordjournals.org.
We are interested in assessing the cost difference between intervention and control groups and how the difference changes with patient’s covariates. This problem can be considered as a special case of inference on a change in the mean cost associated with a change in one or more covariates (e.g. increase in depression, comparison of treatment groups). Statistically, we need to develop accurate regression models for the mean function \( \mu(x) = E(Y \mid X = x) \). The main challenge for such an estimation is how to deal with 3 analytic problems: correlated data, zero inpatient costs for some patients, and a highly skewed distribution of nonzero costs. Unlike estimation on regression coefficients, estimation of \( \mu(x) \) may be sensitive to how to treat the correlation and skewness (Manning 1998; Mullahy 1998; Blough and others, 1999; Manning and others, 2005).

In the literature, a continuous variable with addition zero values is also called a semicontinuous variable. For cross-sectional data, a 2-part model, which has a long history in econometrics, is most appropriate for dealing with semicontinuous data. The 2-part model assumes that a semicontinuous response results from 2 processes: 1 determining whether the response is zero and the other determining the actual level if it is nonzero (Duan and others, 1983; Manning and others, 1981; Manning 1998; Mullahy 1998). Olsen and Schafer (2001) extended the 2-part model to longitudinal data by introducing random effects into the 2-part model. Tooze and others (2002) independently developed a similar extension of the 2-part model. Albert and Shen (2005) further extended the model of Olsen and Schafer (OS) to incorporate serial correlations. All these mixed-effect 2-part models use a linear normal model to fit the actual level of nonzero observations, which may not be appropriate for highly skewed data.

Since the transformation of \( Y \) can simplify the relationship of \( Y \) and \( X \) by inducing a particular type of distribution, for example normal, homoscedastic, symmetric distribution, or remove extreme skewness so that more efficient estimators and more appropriate plotting can be obtained (Ruppert 2001), econometricians and statisticians have historically relied on logarithmic or other specific transformations of \( Y \), followed by regression of the transformed \( Y \) on \( X \) using ordinary least square (OLS) estimation to overcome problems of heteroscedasticity, severe skewness, and kurtosis (Box and Cox 1964; Duan 1983; Ruppert 2001; Manning 1998; Manning and Mullahy 2001). Since the parametric transformation in OLS is not based on any meaningful mechanism and may not be reasonable, Horowitz (1996), Cheng (2002), and Zhou and others (2009) proposed nonparametric transformation models for nonzero cost data in cross-sectional studies. In this paper, we extend OS’s parametric 2-part mixed-effects model to a semiparametric transformation 2-part mixed-effects model.

Fitting our semiparametric 2-part mixed-effects transformation model faces a computational challenge because of intractable numerical integration, which is also encountered in generalized linear random effects models and nonlinear variance component models. In this paper, by transforming the integral in the likelihood function to a “conditional expectation,” we obtain an approximation to the likelihood function that has a closed form. The simulation shows that our approximation is even more accurate than the sixth-order Laplace approximation in finite sample sizes. However, the computational requirement on our accurate approximation is minimal; we only need to evaluate first and second derivatives and maximize the 2 integrands.

This paper is organized as follows. In Section 2, we derive the estimates for the regression parameters and the transformation function based on the approximate log likelihood and a system of estimating equations. In Section 3, we present a method for calculating the unbiased estimator for the mean of the untransformed cost of a patient given the patient’s covariates. In Sections 2 and 3, we also show that under some regularity conditions that our estimators for the unknown transformation function and the mean of the untransformed scale are asymptotically normal, both with the parametric rate of \( O(n^{-1/2}) \). We report results of simulation studies on the accuracy of our approximation and the robustness and efficiency of our method in Section 4. Finally, we apply our methods to the IMPACT data in Section 5.
2. Model and Estimation

2.1 Notation and model

Let \( Y_{ij} \) denote a semicontinuous response for subject \( i \) at occasion \( j \), where \( i = 1, \ldots, n \), and \( j = 1, \ldots, n_i \). This response can be recorded as 2 different responses,

\[
U_{ij} = \begin{cases} 
1 & \text{if } Y_{ij} \neq 0 \\
0 & \text{if } Y_{ij} = 0 
\end{cases}, \quad \text{and} \quad V_{ij} = \begin{cases} 
Y_{ij} & \text{if } Y_{ij} \neq 0 \\
\text{irrelevant} & \text{if } Y_{ij} = 0 
\end{cases}.
\]

We model these 2 responses by a pair of correlated random effects models: 1 for the probability that 
\( U_{ij} = 1 \) and 1 for the continuous response \( V_{ij} \). Let \( \delta_{1i} \) and \( \delta_{2i} \) be the random effects due to subject \( i \) for the 2 parts. We allow \( \delta_{1i} \) and \( \delta_{2i} \) to be correlated, reflecting possible correlations across the 2 parts of the model. Denote \( \delta_i = (\delta_{1i}, \delta_{2i})' \) and assume that \( \delta_i \)'s are i.i.d. with the density function \( f(\delta; \psi) \). A common choice of \( f \) is a multivariate normal distribution with zero mean vector and covariance matrix,

\[
\psi = \begin{pmatrix} \psi_{11} & \psi_{12} \\
\psi_{12} & \psi_{22} \end{pmatrix},
\]

where \( \psi_{11} \) and \( \psi_{22} \) are the covariance matrices of \( \delta_{1i} \) and \( \delta_{2i} \), respectively. We assume that \( Y_{ij}'s \) are conditionally independent, given \( \delta_i \). Let \( \pi_{ij}(\delta_{1i}) = P(U_{ij} = 1|\delta_{1i}) \). The first part of the 2-part model predicts the probability of having a nonzero cost by the following mixed effects model:

\[
\eta(\pi_{ij}(\delta_{1i})) = X'_{1ij}a + Z'_{1ij}\delta_{1i}, \quad (2.1)
\]

where \( a \) is a vector of unknown parameters and \( \eta \) is a known link function. A common choice of \( \eta \) is the logistic function \( \eta(x) = \log(x/(1-x)) \), but other choices are possible.

The second part of the 2-part model predicts the continuous response by the following model:

\[
h(V_{ij}) = b_0 + X'_{2ij}\beta + Z'_{2ij}\delta_{2i} + \epsilon_{ij}, \quad (2.2)
\]

where \( h \) is a monotone increasing but unknown transformation function, satisfying \( h(0) = -\infty \), that makes the distribution of the error term \( \epsilon_{ij} \) to be the normal distribution with mean zero and variance \( \sigma^2 \); \( \beta \) is a vector of unknown parameters; \( \epsilon_{ij} \) and \( \delta_{2i} \) are independent; and \( b_0 \) is a known constant for identifiability. The requirement \( h(0) = -\infty \) ensures that \( \Phi(a + h(0)) = 0 \) for any finite \( a \), where \( \Phi \) is the distribution function of the standard normal variable.

Let \( U_i = (U_{i1}, \ldots, U_{in_i})' \), \( X_i = (X_{i11}, \ldots, X_{1in_i})' \), and \( Z_i = (Z_{i11}, \ldots, Z_{1in_i})' \). Denote \( h(V_i), V_i, X_{2i}, \) and \( Z_{2i} \) to be the vectors or matrices of all relevant values of \( h(V_{ij}), V_{ij}, X_{2ij}, \) and \( Z_{2ij} \) for subject \( i \) with \( U_{ij} = 1 \), respectively.

Let \( \Theta = (\beta, a, \sigma, \psi) \). Hence \( \Theta \) and \( h \) are the unknown parameters and function to be estimated in our 2-part mixed-effects transformation regression model, defined by (2.1) and (2.2). In the rest of the paper, we denote the \((k_1+k_2+\cdots)\)th order partial derivative of a function \( f(x_1, x_2, \ldots) \) by \( f^{(k_1,k_2,\ldots)}(x_1, x_2, \ldots) \); that is \( f^{(k_1,k_2,\ldots)}(x_1, x_2, \ldots) = \frac{d^{k_1+k_2+\cdots}}{dx_1^{k_1}dx_2^{k_2}\cdots} f(x_1, x_2, \ldots) \).

2.2 The approximate likelihood for the parameter vector \( \Theta \) given \( h \)

In this section, we propose a likelihood-based estimation method for \( \Theta \) given that \( h \) is known. Given \( h \), the marginal likelihood for the model defined by (2.1) and (2.2) can be expressed as follows:

\[
L = \prod_{i=1}^{n} \int f(U_i|\delta_{1i}) f(h(V_i)|\delta_{1i}, \delta_{2i}) f(\delta_{1i}, \delta_{2i}) d\delta_{1i} d\delta_{2i}.
\]
Clearly, \( f(h(V_i)|\delta_{1i}, \delta_{2i})f(\delta_{1i}, \delta_{2i}) \) can be identified as the joint density of \((h(V_i), \delta_{1i}, \delta_{2i})\) and can be further written as \( f(\delta_{2i}|h(V_i), \delta_{1i})f(h(V_i), \delta_{1i}) \). Since \( \int f(\delta_{2i}|h(V_i), \delta_{1i})d\delta_{2i} = 1 \), we can further write the marginal likelihood function as follows:

\[
L = \frac{1}{n} \sum_{i=1}^{n} f(U_i|\delta_{1i})f(h(V_i), \delta_{1i})d\delta_{1i} = \frac{1}{n} \sum_{i=1}^{n} f(h(V_i)) \int f(U_i|\delta_{1i})f(\delta_{1i}|h(V_i))d\delta_{1i}, \tag{2.3}
\]

where \( f(U_i|\delta_{1i}) = \exp \left\{ \sum_{j=1}^{n_i} (U_{ij}\eta(\pi_{ij}(\delta_{1i}))) + \log(1 - \pi_{ij}(\delta_{1i})) \right\} \) comes from the model that describes the probability of being a zero observation.

When the dimension of \( \delta_{1i} \) is high, ML estimation of \( \Theta \) becomes difficult because of the intractable numerical integration in (2.3). One method to avoid this intractable numerical integration is to use a Bayesian simulation method with an Monte Carlo Markov chain (MCMC) algorithm. However, with an unknown transformation function \( h \), the Bayesian MCMC algorithm may be time consuming.

The problem of intractable integration in (2.3) is closely related to that in the marginal likelihood of a generalized linear mixed model. In the literature on maximizing the marginal likelihood of a generalized linear mixed model, several authors have proposed several methods for approximating the integrands in the marginal likelihood functions, including Gauss–Hermit quadrature (Anderson and Aitkin 1985; Hedeker and Gibbons 1994) and second-order Laplace approximations (Solomon and Cox 1992; Liu and Pierce 1993; Breslow and Clayton 1993). In general, the Laplace method is easier to implement than the quadrature method, while the quadrature method is more accurate than the Laplace approximation. Recently, Raudenbush and others (2000) proposed the sixth-order Laplace approximation.

Olsen and Schafer (2001) applied the sixth-order Laplace approximation to the parametric 2-part homoscedastic mixed-effects model for the semicontinuous data. Through the simulation, Raudenbush and others (2000) found that the sixth-order Laplace approximation is as accurate as the quadrature method but with much less computational time. However, computation of the first to sixth derivatives, required by the sixth-order Laplace approximation, is also a difficult task in our case.

Based on the idea proposed by Tierney and Kadane (1986), we propose a new approximation to the integration in (2.3), which only requires evaluation of first and second derivatives. Through simulations, we find our approximation is more accurate than the sixth-order Laplace approximation in finite sample sizes. We achieve this accurate approximation by writing the integral in (2.3) as the ratio of 2 integrals,

\[
\int f(U_i|\delta_{1i})f(\delta_{1i}|h(V_i))d\delta_{1i} = \frac{\int f(U_i|\delta_{1i})f(\delta_{1i}, h(V_i))d\delta_{1i}}{\int f(\delta_{1i}, h(V_i))d\delta_{1i}}. \tag{2.4}
\]

We approximate the numerator and denominator in (2.4), respectively, by Laplace’s approximation, instead of directly approximating \( \int f(U_i|\delta_{1i})f(\delta_{1i}|h(V_i))d\delta_{1i} \), as done in a standard Laplace’s approximation. In taking the ratio of these 2 approximations, we can cancel some portion of these residual errors. As a result, we can improve the order of accuracy of the approximation for the ratio.

Next, we give a formal statement of the proposed approximation. Denote

\[
D_i = \text{diag}(\sigma_1^2, \ldots, \sigma_n^2), \quad \Sigma_i = D_i + Z_{2i}'\psi_{22}Z_{2i}, \quad B_i = \psi_{11} - \psi_{12}Z_{2i}'\Sigma_i^{-1}Z_{2i} \psi_{21},
\]

\[
\pi_i(\delta_{1i}) = (\pi_{i1}(\delta_{1i}), \ldots, \pi_{in_i}(\delta_{1i}))', \quad \Delta_i = \left( \begin{array}{c} \Sigma_i \\ \psi_{12}Z_{2i} \psi_{21} \\ \psi_{11} \end{array} \right),
\]

\[
\Pi_i(\delta_{1i}) = \text{diag}\{\Pi_{ij}(\delta_{1i}), j = 1, \ldots, n_i\}, \quad \text{and} \quad \Pi_{ij}(\delta_{1i}) = \pi_{ij}(\delta_{1i})(1 - \pi_{ij}(\delta_{1i})).
\]

Let \( \tau_i^*(\delta_{1i}) \equiv \log\{f(U_i|\delta_{1i})f(\delta_{1i}, h(V_i))\} \) and \( \tau_i(\delta_{1i}) \equiv \log\{f(\delta_{1i}, h(V_i))\} \), which correspond to the integrands of the numerator and the denominator in (2.4), respectively. Let \( \hat{\delta}_{1i} \) and \( \hat{\delta}_{1i} \) be the modes of
$\tau^*_i(\delta_{ii})$ and $\tau_i(\delta_{ii})$, respectively. We obtain $\hat{\delta}_{ii}$ by solving the equation $\frac{\partial^2 \eta(\delta_{ii})}{\partial \delta_{ii}^2} = 0$, which has an explicit solution,

$$\hat{\delta}_{ii} = \psi_{12} Z_{ii}^T \Sigma^{-1}_i (h(V_i) - b_0 I - X_{2i} \beta),$$

where $I$ is the vector with all the component to be 1. By setting $\frac{\partial^2 \tau^*_i(\delta_{ii})}{\partial \delta_{ii}^2} = 0$, we obtain $\hat{\delta}^*_i$ by iteratively solving the following equation:

$$\hat{\delta}^*_i = B_i Z_{ii}^T (U_i - \pi_i(\hat{\delta}_{ii})) + \psi_{12} Z_{ii}^T \Sigma^{-1}_i (h(V_i) - b_0 I - X_{2i} \beta).$$

From (2.4), we see that we can further write the integral in (2.3) as the following ratio of the 2 integrals:

$$\int f(U_i | \delta_{ii}) f(\delta_{ii} | h(V_i)) d\delta_{ii} = \int \frac{\exp(\tau^*_i(\delta_{ii}))}{\exp(\tau_i(\delta_{ii}))} d\delta_{ii}. \quad (2.5)$$

We first derive the second-order Laplace’s approximations to the numerator and denominator of the ratio in (2.5). Then, taking the ratio of the 2 approximations, we have a new approximation for the integral in (2.3).

$$\int f(U_i | \delta_{ii}) f(\delta_{ii} | h(V_i)) d\delta_{ii} = \left(\frac{| - \tau^{(2)}_i(\hat{\delta}_{ii})|}{| - \tau^{(2)}_i(\delta^*_i)|}\right)^{1/2} \exp[\tau_i(\hat{\delta}^*_i) - \tau_i(\hat{\delta}_{ii})] \left(1 + \frac{a_i^* - a_i}{n_i} + O(n_i^{-2})\right),$$

where $a_i = g(\tau^{(2)}_i(\hat{\delta}_{ii}), \tau^{(3)}_i(\hat{\delta}_{ii}), \tau^{(4)}_i(\hat{\delta}_{ii}))$, $a_i^* = g(\tau^{(2)}_i(\delta^*_i), \tau^{(3)}_i(\delta^*_i), \tau^{(4)}_i(\delta^*_i))$ and $g$ is a known function. For example, when $\delta_{ii}$ is 1 dimension, denote $\tau^{(k)}_i = \tau^{(k)}_i(\hat{\delta}_{ii})$, $\sigma^2_i = - (\tau^{(2)}_i)^{-1}$, we have $a_i = \frac{1}{8} \sigma^4_i \tau^{(4)}_i + \frac{5}{24} \sigma^6_i (\tau^{(3)}_i)^2$, $a_i^*$ is defined in the same way except that $\tau_i$ and $\hat{\delta}_{ii}$ are replaced by $\tau_i^*$ and $\delta^*_i$. In Appendix A, we show that our new approximation has the error of order $O(n_i^{-3/2})$.

$$\int f(U_i | \delta_{ii}) f(\delta_{ii} | h(V_i)) d\delta_{ii} = \left(\frac{| - \tau^{(2)}_i(\hat{\delta}_{ii})|}{| - \tau^{(2)}_i(\delta^*_i)|}\right)^{1/2} \exp[\tau_i(\hat{\delta}^*_i) - \tau_i(\hat{\delta}_{ii})] (1 + O(n_i^{-3/2})). \quad (2.6)$$

The simulations in Section 4 also demonstrate that our approximation is more accurate than the sixth-order Laplace’s method in finite sample sizes.

Based on (2.6), we obtain the following final approximate likelihood:

$$l(\Theta; h) = -\frac{1}{2} \sum_{i=1}^{n} \log |\Sigma_i| - \frac{1}{2} \sum_{i=1}^{n} \log |B_i|$$

$$- \frac{1}{2} \sum_{i=1}^{n} (h(V_i) - b_0 I - X_{2i} \beta)^T \Sigma_i^{-1} (h(V_i) - b_0 I - X_{2i} \beta)$$

$$- \frac{1}{2} \sum_{i=1}^{n} \log | - \tau^{(2)}_i(\hat{\delta}_{ii})| + \sum_{i=1}^{n} \sum_{j=1}^{n_i} (U_{ij} \eta(\pi_{ij}(\hat{\delta}_{ii})) + \log(1 - \pi_{ij}(\hat{\delta}_{ii})))$$

$$- \frac{1}{2} \sum_{i=1}^{n} (U_i - \pi_i(\hat{\delta}_{ii}))^T Z_{1i} B_i Z_{1i}^T (U_i - \pi_i(\hat{\delta}_{ii})). \quad (2.7)$$
We maximize the function \( l(\Theta; h) \) by Newton–Raphson iterative procedure,
\[
\Theta^{(t+1)} = \Theta^{(t)} + C^{-1}S,
\]
where \( C = -\partial^2 l(\Theta; h)/\partial \Theta \partial \Theta' \) and \( S = \partial l(\Theta; h)/\partial \Theta \) evaluated at \( \Theta = \Theta^{(t)} \). Since the second derivative of the log likelihood is difficult to calculate, the well-known identity
\[
E(\partial^2 l(\Theta; h)/\partial \Theta \partial \Theta') = -E[(\partial l(\Theta; h)/\partial \Theta)(\partial l(\Theta; h)/\partial \Theta)']
\]
suggests an approximate scoring procedure with \( C \approx \sum_{i=1}^{n}(\partial l_i(\Theta; h)/\partial \Theta)(\partial l_i(\Theta; h)/\partial \Theta)' \), where \( l_i(\Theta; h) \) is the contribution of subject \( i \) to the approximate log likelihood. Expressions for the components of the score vector can be obtained from the authors upon a request.

### 2.3 Estimation of the transformation function \( h \) given \( \Theta \)

In this section, we discuss estimation of the transformation function \( h \) given all the parameters \( \Theta \). Since
\[
Pr(V_{ij} \leq v) = Pr(h(V_{ij}) \leq h(v)) = \Phi \left( \frac{h(v) - b_0 - X'_{2ij} \beta}{\sqrt{Z'_{2ij} \psi_{22} Z_{2ij} + \sigma^2}} \right), \quad (2.8)
\]
where \( \Phi \) is the cumulative distribution function of the standard normal random variable, we obtain an estimate \( \hat{h}(v) \) for \( h(v) \) by solving the following estimating equation:
\[
\sum_{i=1}^{n} \sum_{j=1}^{m_i} \left( I(V_{ij} \leq v) - \Phi \left( \frac{h(v) - b_0 - X'_{2ij} \beta}{\sqrt{Z'_{2ij} \psi_{22} Z_{2ij} + \sigma^2}} \right) \right) = 0, \quad (2.9)
\]
where \( v \in [v_0, v_1] \), the range of the observed \( V_{ij} \).

Using the monotone increasing property of the function \( \Phi \), we obtain that the estimator \( \hat{h}(v) \) is a nondecreasing step function in \( v \in [v_0, v_1] \) with jumps only at the observed \( V_{ij} \), where \( i = 1, \ldots, n \), \( j = 1, \ldots, n_i \). Hence, let \( v_1 < \cdots < v_K \) be the set of distinct points of \( V_{ij}, i = 1, \ldots, n, j = 1, \ldots, n_i, \) then solving the system of estimating equations defined by (2.9) is equivalent to solving the following system of \( K \) equations:
\[
\sum_{i=1}^{n} \sum_{j=1}^{m_i} \left( I(V_{ij} \leq v_k) - \Phi \left( \frac{h(v_k) - b_0 - X'_{2ij} \beta}{\sqrt{Z'_{2ij} \psi_{22} Z_{2ij} + \sigma^2}} \right) \right) = 0, \quad \text{for } k = 1, \ldots, K. \quad (2.10)
\]

The Newton–Raphson algorithm can be used to solve the system of \( K \) estimating (2.10). We can see later that the discrete property of \( \hat{h} \) provides us with a large simplification to predict the mean of the original scale. In addition, unlike a traditional nonparametric approach to estimate the transformation function (Horowitz 1996), our approach does not involve nonparametric smoothing and thus does not suffer from smoothing-related problems, for example selection of a smoothing parameter.

We estimate \( \Theta \) and \( h \) iteratively based on the approximation likelihood (2.7) and the system of estimating (2.10) until 2 successive values of \( \Theta \) do not differ significantly. An initial value of \( \Theta \) is required to start the iterations, which can be obtained by fitting a generalized linear model for \( U_i \) and a transformation model for nonzero with the dependence between models and the dependence among data being ignored. For simplicity, we set the starting values for \( \psi_{11} \) and \( \psi_{22} \) to be the identity matrix.
Let \( \hat{h} \) be the estimators of \( h, d_{ij}(\Theta) = Z_{2ij}W_{22}Z_{2ij} + \sigma^2 \),

\[
S(w; \nu, \Theta) = \frac{1}{n} \sum_{i=1}^{n} \sum_{j=1}^{m_i} \left( I(V_{ij} \leq v) - \Phi \left( \frac{w - b_0 - X_{2ij}^\prime \beta}{\sqrt{d_{ij}(\Theta)}} \right) \right) ,
\]

\[
s_1(v) = \lim_{n \to \infty} S^{(00)}(h_0(v); \nu, \Theta_0), \quad s_2(v) = \lim_{n \to \infty} S^{(000)}(h_0(v); \nu, \Theta_0),
\]

where \( h_0 \) and \( \Theta_0 \) are the true values of \( h \) and \( \Theta \), respectively. Then, under the conditions given in Appendix B, we have

\[
\hat{h}(v) - h_0(v) \approx s_1^{-1}(v)[S(h_0(v); \nu, \Theta_0) + s_2(v)\hat{\Theta} - \Theta_0],
\]

where \( \hat{\Theta} \) is the estimate of \( \Theta \). Hence, if there exist independent random variables \( \xi_i \) with \( E(\xi_i) = 0 \) and \( \text{Var}(\xi_i) < \infty \), for \( i = 1, 2, \ldots, n \), such that \( \hat{\Theta} - \Theta_0 = \frac{1}{n} \sum_{i=1}^{n} \xi_i + o_p(n^{-1/2}) \), we have

\[
\hat{h}(v) - h_0(v) = \frac{1}{ns_1(v)} \sum_{i=1}^{n} \Omega_i(v) + o_p(n^{-1/2}),
\]

where \( \Omega_i(v) = \sum_{j=1}^{m_i} \left[ I(V_{ij} \leq v) - \Phi \left( \frac{h_0(v) - b_0 - X_{2ij}^\prime \beta}{\sqrt{d_{ij}(\Theta)}} \right) \right] + s_2(v)\xi_i \). This implies that the distribution of \( n^{1/2}(\hat{h}(v) - h_0(v)) \) can be approximated by a normal random variable with mean 0 and variance \( \Sigma = \frac{1}{s_1(v)}E\Omega^2_i(v) \). Hence, we can estimate the nonparametric function \( h(.) \) with a parametric convergent rate if we can estimate the parameters \( \Theta \) at a rate of \( n^{-1/2} \). The similar conclusion, regarding \( n^{-1/2} \) convergent rate of the estimated transformation function, can be also found in Horowitz (1996), Cheng (2002), Ye and Duan (1997), and Zhou and others (2009). The conclusion assures that the resulting estimator for the mean of the original scale converges to the true value at a rate of \( n^{-1/2} \).

### 3. Predicting the Mean of the Original Scale

Given the covariates \( x = (x'_1, x'_2)' \) and \( z = (z'_1, z'_2)' \), we want to estimate \( u(x, z) = E(Y|x, z) \), where \( Y \) is the response of the outcome for the patient with the covariates \( x \) and \( z \). Unbiased and consistent quantities on the transformed scale may not automatically retransform into unbiased or consistent quantities on the untransformed scale. The smearing estimate, proposed by Duan (1983), is a popular method to consistently estimate an individual’s expected response on the untransformed scale. Since the random effects, \( \delta_i \)'s, are unobservable, it is difficult to extend the smearing estimator to the 2-part model with the random effects.

In this section, we propose a numerical method to estimate \( \mu(x, z) \). Let \( \pi(\delta_1) = \eta^{-1}(x'_1\alpha + z'_1\delta_1) \) and \( \nu(\delta_2) = h^{-1}(b_0 + x'_2\beta + z'_2\delta_2 + \sigma\varepsilon) \), where \( \delta = (\delta'_1, \delta'_2)' \sim N(0, \psi), \varepsilon \sim N(0, 1), \delta \) and \( \varepsilon \) are independent. With this notation, we obtain the following expression for \( \mu(x, z) \):

\[
u(x, z) = E(E(Y|x, z, \delta))
\]

\[
= E(\pi(\delta_1)E(\nu(\delta_2)|\delta))
\]

\[
= E[\eta^{-1}(x'_1\alpha + z'_1\delta_1)E[h^{-1}(b_0 + x'_2\beta + z'_2\delta_2 + \sigma\varepsilon)|\delta]].
\]

From this expression, we see that one way to estimate \( \mu(x, z) \) is to first estimate \( E[h^{-1}(b_0 + x'_2\beta + z'_2\delta_2 + \sigma\varepsilon)|\delta] \) for any given \( \delta \), which can be achieved by the following estimator:

\[
\frac{1}{R_1} \sum_{k=1}^{R_1} \hat{h}^{-1}(b_0 + x'_2\hat{\beta} + z'_2\hat{\delta} + \hat{\sigma}\varepsilon_k),
\]
where \( \varepsilon_k \) is generated from the standard normal distribution. Then, we can obtain the following estimator for \( u(x, z) \):

\[
\hat{u}(x, z) = \frac{1}{R_1 R_2} \sum_{r=1}^{R_1} \sum_{k=1}^{R_2} \eta^{-1}(x_1' \hat{\alpha} + z_1' \delta_r) \hat{h}^{-1}(b_0 + x_2' \hat{\beta} + z_2' \delta_r + \hat{\delta} \varepsilon_k),
\]

where \( \hat{\delta} = (\delta_{1r}, \delta_{2r})' \) is generated from the multivariate normal distribution with mean vector 0 and covariance matrix \( \psi \). Next, we give an asymptotic result for \( \hat{u}(x, z) \).

Let \( \zeta_1 = x_1' \alpha + z_1' \delta_1 \) and \( \zeta_2 = b_0 + x_2' \beta + z_2' \delta_2 + \sigma \varepsilon, \pi(\zeta_1) = \eta^{-1}(\zeta_1), \nu(\zeta_2) = h^{-1}(\zeta_2), z_{01} = (\varepsilon', 0)' \), and \( z_{02} = (0', \varepsilon)' \). Suppose that \( \delta = (\delta_{1r}', \delta_{2r})' \) is the normally distributed random vector with mean 0 and covariance matrix \( \psi \) and that \( \delta \) and \( \varepsilon \) are independent. Denote

\[
\varrho_1(x, z, \Theta) = E[\pi(\zeta_1)\nu(\zeta_2)], \quad \varrho_2(x, z, \Theta) = E[\pi(\zeta_1)\nu(\zeta_2)e],
\]

\[
\varrho_3(x, z, \Theta) = E[\pi(\zeta_1)\nu(\zeta_2)c], \quad \varrho_4(x, z, \Theta) = E[\pi(\zeta_1)\nu(\zeta_2)s_1^{-1}(\nu(\zeta_2))s_2(\nu(\zeta_2))],
\]

\[
\varrho_5(x, z, \Theta) = E[\pi(\zeta_1)\nu(\zeta_2)\varrho(\zeta_1, \varrho_2(b_0 + x_2' \beta + z_2' \delta_2 + \sigma \varepsilon)\varrho), \psi(x_2' \beta + \psi^{-1} \varrho)],
\]

\[
\varrho_6(x, z, \Theta) = E[\pi(\zeta_1)\nu(\zeta_2)\varrho(b_0 + x_2' \beta + z_2' \delta_2 + \sigma \varepsilon)\varrho, \psi(x_2' \beta + \psi^{-1} \varrho)],
\]

\[
\varrho_7(x, z, \Theta)(\hat{\Theta} - \Theta) \approx z_{01}(\psi^{-1} - \psi'^{-1})\varrho_5(x, z, \Theta) + z_{02}(\psi^{-1} - \psi'^{-1})\varrho_6(x, z, \Theta),
\]

\[
\zeta(x, z, \Theta) = (\varrho_1(x, z, \Theta)x_1', \varrho_2(x, z, \Theta)x_2', \varrho_3(x, z, \Theta), 0)',
\]

\[
\varrho(x, z, \Theta) = \zeta(x, z, \Theta) - \varrho_4(x, z, \Theta) + \varrho_7(x, z, \Theta),
\]

\[
\Upsilon(x, z, x_2', z_2', \nu^*, \Theta) = E \left\{ \pi(\zeta_1)\nu(\zeta_2)s_1^{-1}(\nu(\zeta_2)) \right\}
\]

where \( d(\Theta) = z_2' \psi_{22} z_2 + \sigma^2 \). Then, under the conditions given in Appendix B, we have

\[
\hat{u}(x, z) - u(x, z) \approx \varrho(x, z, \Theta_0)(\hat{\Theta} - \Theta_0) - \frac{1}{n} \sum_{i=1}^{n} \sum_{j=1}^{m_i} \Upsilon(x, z, x_{2ij}, Z_{2ij}, V_{ij}, \Theta_0).
\]

Hence, if there exist independent random variables, \( \zeta_i, i = 1, 2, \ldots, n \), with \( E(\zeta_i) = 0, \ Var(\zeta_i) < \infty \), such that \( \hat{\Theta} - \Theta_0 = \frac{1}{n} \sum_{i=1}^{n} \zeta_i + o_p(n^{-1/2}) \), then the distribution of \( n^{1/2}(\hat{u}(x, z) - u(x, z)) \) can be approximated by a normal random variable with mean 0 and a finite covariance matrix.

4. Simulation

4.1 Performance of the approximate log likelihood

In this subsection, we investigate the accuracy of our approximate log likelihood by comparing the estimates based on our approximation with the estimates based on the sixth-order Laplace approximation, proposed by Olsen and Schafer (2001).

In our simulation study, we use the same setting as in Olsen and Schafer (2001). For each subject, \( X_i \) is the matrix of covariates related to fixed effects and is constructed with 3 columns: a constant equal to 1,
a dummy indicator for a nontime–varying covariate drawn from Bernoulli \((p = 0.5)\), and a time-varying covariate taking values 0, 1, \ldots, \(m - 1\), where \(m\) is the number of occasions. The matrix of covariates \(Z_i\), which are related to random effects, are set to be columns of 1s. The coefficients of the fixed effects are set to \(\alpha = (-1, -0.5, 0.4)'\) and \(\beta = (-0.3, 0.1, 0.4)'\). The variance parameters are set to be \(\psi_{11} = 1, \psi_{12} = 0.2,\) and \(\psi_{22} = 0.5\). The homoscedastic variance of the transformed nonzero response is set to be \(\sigma^2 = 0.5\). The transformation function is assumed to be identity. We also vary the number of subjects and the number of occasions in a 2 \(\times\) 2 design with \(n = 1000\) or 200 and \(n_i = m = 10\) or 5.

We summarize the behavior of our new estimators and the OS’s estimators of \(\alpha\) and \(\beta\) in Table 1.

For each scenario, Table 1 lists the average, standard error (SE), and the root of mean square errors (RMSE) of the estimators. Both our estimators and the OS estimators are basically unbiased. However, the standard deviations of our estimators are smaller than those of the OS estimators in all the settings considered here. As a result, our estimators have smaller RMSE than the OS estimators in all the settings and, hence, are better than the OS estimators.

Although our estimator needs only the first and second derivatives, and the OS estimator needs the first through sixth derivatives, our estimators are still more accurate than the OS estimators.

Table 1. Simulation results when semicontinuous data are generated from a 2-part homoscedastic mixed-effects model

<table>
<thead>
<tr>
<th>(n)</th>
<th>(m)</th>
<th>(\alpha_1)</th>
<th>(\alpha_2)</th>
<th>(\alpha_3)</th>
<th>(\beta_1)</th>
<th>(\beta_2)</th>
<th>(\beta_3)</th>
<th>(\alpha_1)</th>
<th>(\alpha_2)</th>
<th>(\alpha_3)</th>
<th>(\beta_1)</th>
<th>(\beta_2)</th>
<th>(\beta_3)</th>
</tr>
</thead>
<tbody>
<tr>
<td>1000</td>
<td>10</td>
<td>-1.0039</td>
<td>-0.503</td>
<td>0.3998</td>
<td>-0.2993</td>
<td>0.0994</td>
<td>0.3999</td>
<td>0.0668</td>
<td>0.0808</td>
<td>0.0097</td>
<td>0.0405</td>
<td>0.0475</td>
<td>0.0303</td>
</tr>
<tr>
<td></td>
<td></td>
<td>0.0000</td>
<td>0.0000</td>
<td>0.0000</td>
<td>0.0000</td>
<td>0.0000</td>
<td>0.0000</td>
<td>0.0669</td>
<td>0.0810</td>
<td>0.0097</td>
<td>0.0405</td>
<td>0.0475</td>
<td>0.0303</td>
</tr>
<tr>
<td>1000</td>
<td>10</td>
<td>-0.503</td>
<td>0.3998</td>
<td>-0.2993</td>
<td>0.0994</td>
<td>0.3999</td>
<td>0.0668</td>
<td>0.0808</td>
<td>0.0097</td>
<td>0.0405</td>
<td>0.0475</td>
<td>0.0303</td>
<td>0.0000</td>
</tr>
<tr>
<td>200</td>
<td>10</td>
<td>-0.9948</td>
<td>-0.501</td>
<td>0.3982</td>
<td>-0.3029</td>
<td>0.1010</td>
<td>0.4004</td>
<td>0.0768</td>
<td>0.0849</td>
<td>0.0240</td>
<td>0.0530</td>
<td>0.0601</td>
<td>0.0123</td>
</tr>
<tr>
<td>200</td>
<td>10</td>
<td>-0.501</td>
<td>0.3982</td>
<td>-0.3029</td>
<td>0.1010</td>
<td>0.4004</td>
<td>0.0768</td>
<td>0.0849</td>
<td>0.0240</td>
<td>0.0530</td>
<td>0.0601</td>
<td>0.0123</td>
<td>0.0000</td>
</tr>
<tr>
<td>200</td>
<td>5</td>
<td>-1.0144</td>
<td>-0.512</td>
<td>0.4019</td>
<td>-0.2932</td>
<td>0.0935</td>
<td>0.3996</td>
<td>0.1434</td>
<td>0.1784</td>
<td>0.0231</td>
<td>0.0890</td>
<td>0.1052</td>
<td>0.0080</td>
</tr>
<tr>
<td>200</td>
<td>5</td>
<td>-0.512</td>
<td>0.4019</td>
<td>-0.2932</td>
<td>0.0935</td>
<td>0.3996</td>
<td>0.1434</td>
<td>0.1784</td>
<td>0.0231</td>
<td>0.0890</td>
<td>0.1052</td>
<td>0.0080</td>
<td>0.0000</td>
</tr>
</tbody>
</table>

<table>
<thead>
<tr>
<th>(\psi_{11})</th>
<th>(\psi_{12})</th>
<th>(\psi_{22})</th>
<th>(\sigma^2)</th>
<th>Transformation</th>
</tr>
</thead>
<tbody>
<tr>
<td>1</td>
<td>0.2</td>
<td>0.5</td>
<td>0.5</td>
<td>Identity</td>
</tr>
</tbody>
</table>

We also vary the number of subjects and the number of occasions in a 2 \(\times\) 2 design with \(n = 1000\) or 200 and \(n_i = m = 10\) or 5.
Since our method does not require specification of a parametric form for the transformation function, we expect that the resulting estimates and inferences are more reliable than the parametric method with the misspecified transformation (MT) function, for example the OS estimators. We want to know whether the added robustness is gained at the expense of reduced efficiency. To investigate these 2 issues, we examine the performance of the proposed method in comparison with the MT method, where the transformation function is misspecified, and the correctly specified transformation (CST) method, where the transformation function is correctly specified.

Our proposed model contains key 2 components, the transformation function, $h$, and the distribution of the random effects. We would also like to know the relative effect of misspecification of the transformation function and misspecification of the random effect distribution on our inference. To investigate this issue, we want to compare the performance of the proposed method with a misspecified distribution function of the random effects, with the performance of the MT method with the correctly specified random effect distribution.

We conduct 2 simulation studies to answer the above 3 issues. In the first simulation study, we simulate data from the setting similar to the above simulation in Section 4.1 except that $\beta = (-0.3, 0.3, 0.4)^{\top}$ and the transformation function $h(v) = 3 \log(v)$. A total of 200 data sets were generated. For each simulated data set, we obtain estimates for the fixed effect and the mean of original scale $\mu(x) = \mu(x_1, x_2, x_3)$ at the combination of $x_1 = 1$, $x_2 = 0, 1$, and $x_3 = (1, 2, 3, 4)$ using the proposed approach, the CST method, and the MT method with the MT function $h(v) = v^4$.

The MT method fails to converge for 123 of the 200 samples. The results reported in Tables 2 and 3 are based on the remaining samples. Table 2 presents the average, the SE, the standardized bias (bias as a percent of the SE), and the RMSE for the fixed effect parameters. Since the transformation function is involved only in the second part of the models, the estimates of the regression parameters in the first part of the models are basically unbiased even when we misspecified the transformation function. However, misspecification of the transformation function can lead to severely biased estimates for the parameters.

### Table 2. Simulation results when semicontinuous data are generated from a 2-part mixed-effects heteroscedastic transformation model

<table>
<thead>
<tr>
<th></th>
<th>Proposed</th>
<th>CST</th>
<th>MT (OS)</th>
<th>Proposed</th>
<th>CST</th>
<th>MT (OS)</th>
</tr>
</thead>
<tbody>
<tr>
<td>$\alpha_1$</td>
<td>0.0026</td>
<td>0.0022</td>
<td>0.0043</td>
<td>$\beta_1$</td>
<td>—</td>
<td>0.0044</td>
</tr>
<tr>
<td>SE</td>
<td>0.0588</td>
<td>0.0558</td>
<td>0.0530</td>
<td>—</td>
<td>0.0352</td>
<td>0.0476</td>
</tr>
<tr>
<td>RMSE</td>
<td>0.0589</td>
<td>0.0559</td>
<td>0.0532</td>
<td>—</td>
<td>0.0355</td>
<td>0.9825</td>
</tr>
<tr>
<td>$\alpha_2$</td>
<td>0.0104</td>
<td>0.0092</td>
<td>-0.0069</td>
<td>$\beta_2$</td>
<td>-0.0117</td>
<td>-0.0010</td>
</tr>
<tr>
<td>SE</td>
<td>0.0654</td>
<td>0.0617</td>
<td>0.0632</td>
<td>0.0633</td>
<td>0.0386</td>
<td>0.0354</td>
</tr>
<tr>
<td>RMSE</td>
<td>0.0662</td>
<td>0.0624</td>
<td>0.0635</td>
<td>0.0644</td>
<td>0.0386</td>
<td>2.6544</td>
</tr>
<tr>
<td>$\alpha_3$</td>
<td>-0.0025</td>
<td>-0.0024</td>
<td>-0.0066</td>
<td>$\beta_3$</td>
<td>0.0013</td>
<td>-0.0005</td>
</tr>
<tr>
<td>SE</td>
<td>0.0160</td>
<td>0.0166</td>
<td>0.0146</td>
<td>0.0078</td>
<td>0.0086</td>
<td>0.0097</td>
</tr>
<tr>
<td>RMSE</td>
<td>0.0162</td>
<td>0.0167</td>
<td>0.0160</td>
<td>0.0079</td>
<td>0.0086</td>
<td>3.1341</td>
</tr>
<tr>
<td>$\sigma^2$</td>
<td>0.0164</td>
<td>-0.0010</td>
<td>-0.3699</td>
<td>$\psi_{11}$</td>
<td>-0.1159</td>
<td>-0.1159</td>
</tr>
<tr>
<td>SE</td>
<td>0.0534</td>
<td>0.0138</td>
<td>0.0001</td>
<td>0.0783</td>
<td>0.0789</td>
<td>0.0787</td>
</tr>
<tr>
<td>RMSE</td>
<td>0.0559</td>
<td>0.0138</td>
<td>0.3699</td>
<td>0.1399</td>
<td>0.1402</td>
<td>0.1399</td>
</tr>
<tr>
<td>$\psi_{12}$</td>
<td>-0.0048</td>
<td>-0.0014</td>
<td>-0.1958</td>
<td>$\psi_{22}$</td>
<td>0.0176</td>
<td>0.0004</td>
</tr>
<tr>
<td>SE</td>
<td>0.0347</td>
<td>0.0340</td>
<td>0.0014</td>
<td>0.0530</td>
<td>0.0287</td>
<td>0.0001</td>
</tr>
<tr>
<td>RMSE</td>
<td>0.0351</td>
<td>0.0340</td>
<td>0.1958</td>
<td>0.0558</td>
<td>0.0287</td>
<td>0.4002</td>
</tr>
</tbody>
</table>
related to the second part of the model. In contrast, our method gives estimates close to the truth value of the parameter with the reasonable variances, suggesting that our procedure is robust.

Table 3 below presents the average, SE, and RMSE for the estimated $\mu(x)$ at the MT estimate is severely biased. In contrast, the proposed approach yields an estimate with essentially no bias, once again suggesting that our method is robust.

For each simulated data set, we also obtain estimates of the transformation $H$ using the proposed approach. Figure 1 displays the averaged estimated transformation function and their 95% empirical point-wise confidence limits based on 200 simulated data sets; Figure 1 shows that our proposed estimate of the transformation function is very close to the true transformation function.

In the second simulation study, we investigate sensitivity of inferences to the random effects distribution. We generate the data, according to the same setting as in the above simulation study, and then we

<table>
<thead>
<tr>
<th>$x_2, x_3$</th>
<th>Proposed Bias</th>
<th>CST</th>
<th>MT</th>
<th>$x$</th>
<th>Proposed Bias</th>
<th>CST</th>
<th>MT</th>
<th>$x$</th>
<th>Proposed Bias</th>
<th>CST</th>
<th>MT</th>
</tr>
</thead>
<tbody>
<tr>
<td>(1,1)</td>
<td>0.0200</td>
<td>−0.0018</td>
<td>0.0646</td>
<td>(0,2)</td>
<td>0.0268</td>
<td>−0.0011</td>
<td>0.1167</td>
<td></td>
<td></td>
<td></td>
<td></td>
</tr>
<tr>
<td>SE</td>
<td>0.0274</td>
<td>0.0125</td>
<td>0.0128</td>
<td>0.0341</td>
<td>0.0147</td>
<td>0.0146</td>
<td></td>
<td></td>
<td></td>
<td></td>
<td></td>
</tr>
<tr>
<td>RMSE</td>
<td>0.0339</td>
<td>0.0126</td>
<td>0.0658</td>
<td>0.0434</td>
<td>0.0147</td>
<td>0.1176</td>
<td></td>
<td></td>
<td></td>
<td></td>
<td></td>
</tr>
<tr>
<td>(1,2)</td>
<td>0.0061</td>
<td>−0.0012</td>
<td>0.0942</td>
<td>(0,3)</td>
<td>0.0074</td>
<td>0.0000</td>
<td>0.1489</td>
<td></td>
<td></td>
<td></td>
<td></td>
</tr>
<tr>
<td>SE</td>
<td>0.0252</td>
<td>0.0156</td>
<td>0.0153</td>
<td>0.0345</td>
<td>0.0188</td>
<td>0.0181</td>
<td></td>
<td></td>
<td></td>
<td></td>
<td></td>
</tr>
<tr>
<td>RMSE</td>
<td>0.0260</td>
<td>0.0156</td>
<td>0.0955</td>
<td>0.0353</td>
<td>0.0188</td>
<td>0.1500</td>
<td></td>
<td></td>
<td></td>
<td></td>
<td></td>
</tr>
<tr>
<td>(1,3)</td>
<td>−0.0133</td>
<td>−0.0001</td>
<td>0.0966</td>
<td>(0,4)</td>
<td>−0.0191</td>
<td>0.0013</td>
<td>0.1374</td>
<td></td>
<td></td>
<td></td>
<td></td>
</tr>
<tr>
<td>SE</td>
<td>0.0334</td>
<td>0.0204</td>
<td>0.0190</td>
<td>0.0494</td>
<td>0.0247</td>
<td>0.0221</td>
<td></td>
<td></td>
<td></td>
<td></td>
<td></td>
</tr>
<tr>
<td>RMSE</td>
<td>0.0359</td>
<td>0.0204</td>
<td>0.0984</td>
<td>0.0529</td>
<td>0.0248</td>
<td>0.1391</td>
<td></td>
<td></td>
<td></td>
<td></td>
<td></td>
</tr>
</tbody>
</table>

Fig. 1. The typical estimates of transformation curve.
A semiparametric 2-part mixed-effects transformation model

Table 4. Simulation results when the random effects distribution is not normal

<table>
<thead>
<tr>
<th></th>
<th>Proposed</th>
<th>Proposed</th>
<th>Proposed</th>
<th>Proposed</th>
</tr>
</thead>
<tbody>
<tr>
<td>Bias</td>
<td>$\alpha_1$</td>
<td>0.0086</td>
<td>$\beta_1$</td>
<td>—</td>
</tr>
<tr>
<td>SE</td>
<td>0.0599</td>
<td>—</td>
<td>0.0866</td>
<td>0.0490</td>
</tr>
<tr>
<td>RMSE</td>
<td>0.0605</td>
<td>—</td>
<td>0.1005</td>
<td>0.0505</td>
</tr>
<tr>
<td>Bias</td>
<td>$\alpha_2$</td>
<td>0.0056</td>
<td>$\beta_2$</td>
<td>$-0.0080$</td>
</tr>
<tr>
<td>SE</td>
<td>0.0675</td>
<td>0.0630</td>
<td>0.0388</td>
<td></td>
</tr>
<tr>
<td>RMSE</td>
<td>0.0678</td>
<td>0.0635</td>
<td>0.0393</td>
<td></td>
</tr>
<tr>
<td>Bias</td>
<td>$\alpha_3$</td>
<td>$-0.0033$</td>
<td>$\beta_3$</td>
<td>0.0012</td>
</tr>
<tr>
<td>SE</td>
<td>0.0171</td>
<td>0.0090</td>
<td>0.0570</td>
<td></td>
</tr>
<tr>
<td>RMSE</td>
<td>0.0174</td>
<td>0.0091</td>
<td>0.1128</td>
<td></td>
</tr>
</tbody>
</table>

discretize the generated values of the random effect $\delta_1$ to $-2$, $-1$, 0, 1, 2 and the generated values of the random effect $\delta_2$ to $-1$, 0, 1. Table 4 presents the bias, SE, and RMSE for the fixed effect parameters, suggesting that our estimator basically is unbiased even when we misspecified the random effect distribution, which implies that our estimator is not sensitive to the random effects distribution. On the other hand, from Table 2, we know that the misspecification of the transformation function leads to biased estimates for the covariate effects. Hence, misspecification of the transformation function has a worse effect on estimation of covariate effects than misspecification of the random effects distribution does.

5. Example

The sample used for this study was from a clinical study, examining the effectiveness of the IMPACT collaborative care management program for late-life depression (Unutzer and others, 2002). A total of 1801 patients aged 60 years or older with major depression (17%), dysthymic disorder (30%), or both (53%) were randomly assigned to the IMPACT intervention ($n = 906$) or to usual care ($n = 895$). Intervention patients had access for up to 12 months to a depression care manager who offered education, care management, and support of antidepressant management. The primary outcome, the total inpatient cost over the previous 6-month period, was collected at months 6, 12, 18, and 24. Denote $Y_{ij}$ to be the total inpatient cost over the $j$th half year for patient $i$. The 2 independent variables are $X_{1ij}$ and $X_{2ij}$, where $X_{1ij}$ is the treatment indicator and $X_{2ij}$ is the mean score of the 20 depression items from the symptom checklist for the $j$th observation of patient $i$. With $n_i = 4$ per subject, we do not have enough information to fit a high-dimension random effects model, and hence, we fit the following random intercept model:

\[
\logistic(\pi_{ij}(\delta_{1i})) = \alpha_0 + X_{1ij}\alpha_1 + X_{2ij}\alpha_2 + \delta_{1i},
\]

and \[ h(V_{ij}) = \beta_0 + X_{1ij}\beta_1 + X_{2ij}\beta_2 + \delta_{2i} + \sigma_{ij}, \]

where $\delta_i = (\delta_{1i}, \delta_{2i})$ is a bivariate normal vector with mean 0 and covariance matrix $\psi = \begin{pmatrix} \psi_{11} & \psi_{12} \\ \psi_{12} & \psi_{22} \end{pmatrix}$ and $\epsilon_{ij}$ is a standard normal random variable. To reduce the computational time, we first apply a log transformation to the nonzero outcome variable. To compare, we also analyze the cost data by a parametric transformation model with log transformation (termed “LOG-TRAN”).

We present parameter estimates in Table 5, which shows that the effects of treatment ($X_1$) on the mean and variance are not significant and the correlations ($\psi_{12} = -0.0398$) across the 2 parts of the models are not significant. The results for $X_2$ show that the patients with higher scores of depression are associated with higher costs and larger variation in cost, although the effect on variance is not significant. Figure 2
Table 5. The estimates of the parameters for IMPACT data

<table>
<thead>
<tr>
<th>Parameter</th>
<th>Proposed</th>
<th>LOG-TRAN</th>
</tr>
</thead>
<tbody>
<tr>
<td>$\hat{\alpha}_1$</td>
<td>-2.2363 (0.0990)</td>
<td>-2.4595 (0.1398)</td>
</tr>
<tr>
<td>$\hat{\alpha}_2$</td>
<td>-0.0508 (0.0806)</td>
<td>0.0390 (0.1159)</td>
</tr>
<tr>
<td>$\hat{\alpha}_3$</td>
<td>0.1792 (0.0534)</td>
<td>0.2752 (0.0601)</td>
</tr>
<tr>
<td>$\hat{\beta}_1$</td>
<td>20(0)</td>
<td>7.2558 (0.1799)</td>
</tr>
<tr>
<td>$\hat{\beta}_2$</td>
<td>0.2087 (0.0925)</td>
<td>0.2227 (0.1566)</td>
</tr>
<tr>
<td>$\hat{\beta}_3$</td>
<td>0.1355 (0.0484)</td>
<td>0.1915 (0.0853)</td>
</tr>
<tr>
<td>$\hat{\psi}_{11}$</td>
<td>1.1556 (0.1228)</td>
<td>1.2460 (0.1679)</td>
</tr>
<tr>
<td>$\hat{\psi}_{12}$</td>
<td>-0.0398 (0.1022)</td>
<td>0.0835 (0.1870)</td>
</tr>
<tr>
<td>$\hat{\psi}_{22}$</td>
<td>0.4948 (0.3153)</td>
<td>1.0850 (0.2282)</td>
</tr>
<tr>
<td>$\hat{\sigma}^2$</td>
<td>1.2281 (0.0864)</td>
<td>2.4351 (0.2117)</td>
</tr>
</tbody>
</table>

Fig. 2. The estimated transformation curve for IMPACT data (Solid, estimated; dashed, 95% confidential limit).

Table 6. The estimates for the mean of original scale for IMPACT data

<table>
<thead>
<tr>
<th>$\lambda_1$</th>
<th>$\lambda_2$</th>
<th>$u_0$</th>
<th>SE</th>
<th>$\lambda_1$</th>
<th>$\lambda_2$</th>
<th>$u_0$</th>
<th>SE</th>
</tr>
</thead>
<tbody>
<tr>
<td>1</td>
<td>0.6</td>
<td>970.7469</td>
<td>85.9139</td>
<td>0</td>
<td>0.6</td>
<td>836.2906</td>
<td>79.3205</td>
</tr>
<tr>
<td>1</td>
<td>1.2</td>
<td>1130.0280</td>
<td>102.4616</td>
<td>0</td>
<td>1.2</td>
<td>974.7967</td>
<td>90.8485</td>
</tr>
<tr>
<td>1</td>
<td>1.8</td>
<td>1312.3440</td>
<td>134.2079</td>
<td>0</td>
<td>1.8</td>
<td>1133.4890</td>
<td>115.2832</td>
</tr>
</tbody>
</table>

presents the estimate and its 95% confidential interval for the transformation function. Using the estimates of the parameters and transformation function, we estimate the average cost of a patient with the given covariate values. Table 6 gives some average costs. For example, for a patient in the intervention group ($\lambda_1 = 1$) with a depression score of 1.2 (around the mean of $\lambda_2$), the estimated average cost and its standard deviation are $1130.028$ and $102.4616$, respectively.

Estimating the difference of the means of health medical costs between the intervention and the control patients as a function of patients' covariates is also an important target in econometrics, and hence, we
present some differences in Table 7, which suggests that the differences in cost between intervention and control patients can vary, depending on patients’ characteristics.

Our models make some assumptions that should be investigated: normality of $\delta_i$ and $\varepsilon_i$, a linear relationships between covariates, and the logit probability and linear relationships between the covariates and the transformation of respondents. For normal mixed-effects models, only a few formal diagnostics

Table 7. *The differences of the mean of cost between 2 groups*

<table>
<thead>
<tr>
<th>Group 1</th>
<th>Group 2</th>
<th>Difference of mean (SE)</th>
</tr>
</thead>
<tbody>
<tr>
<td>$X_1 = 1, X_2 = 0.6$</td>
<td>$X_1 = 0, X_2 = 0.6$</td>
<td>134.4563 (90.0963)</td>
</tr>
<tr>
<td>$X_1 = 1, X_2 = 1.2$</td>
<td>$X_1 = 0, X_2 = 1.2$</td>
<td>155.2313 (105.2384)</td>
</tr>
<tr>
<td>$X_1 = 1, X_2 = 1.8$</td>
<td>$X_1 = 0, X_2 = 1.8$</td>
<td>178.8550 (122.9108)</td>
</tr>
</tbody>
</table>

Table 8. *The frequency in each cell defined by the observed number of zero cost and expected number of zero cost among the 4 observations*

<table>
<thead>
<tr>
<th>Expected number</th>
<th>Observed number</th>
</tr>
</thead>
<tbody>
<tr>
<td>0</td>
<td>975</td>
</tr>
<tr>
<td>1</td>
<td>0</td>
</tr>
<tr>
<td>2</td>
<td>0</td>
</tr>
<tr>
<td>3</td>
<td>0</td>
</tr>
<tr>
<td>4</td>
<td>0</td>
</tr>
</tbody>
</table>

Fig. 3. Observed log amount of the cost versus expected log amount of the cost.
have been developed, and practitioners often rely on informal techniques such as normal quantile plots of the estimated random effects. Diagnostics for generalized linear mixed models are even more scarce (Olsen and Schafer 2001).

Here, we follow the similar method as used by Olsen and Schafer (2001). We detect the large discrepancies in the model fit by comparing the observed values for $U_i = \sum_{j=1}^{n_i} U_{ij}$ and $V_i = \sum_{j=1}^{n_i} \log(V_{ij})$ with their predicted values $\hat{U}_i$ and $\hat{V}_i = \sum_{j=1}^{n_i} \log(\hat{V}_{ij})$, obtained by substituting the estimates of $\Theta$, $H$, and empirical Bayes estimates of $\delta_i$. Viewing $N(0, \psi)$ as a prior distribution for $\delta_i$, empirical Bayes estimates of $\delta_i$ can be obtained by calculating a posterior mean $E(\delta_i | Y_i)$ with the unknown parameters replaced by their estimates. Since the conditional distribution $\delta_i$ given $Y_i$ does not have a closed form, we evaluate the integrals required for posterior moments by numerical techniques. In the example $\delta_i$ has 2 dimensions, and we use numerical techniques methods to evaluate the related integrals.

Table 8 gives the frequency in each cell defined by the observed and the predicted (rounded to the nearest integers) value for $U_i$. The percentage of total agreement between the observed and the predicted values is 83.08%. Figure 3 plots $V_i$ versus $\hat{V}_i$, showing no significant deviation. Table 8 and Figure 3 suggest our models are reasonable.

6. DISCUSSION

In this paper, we have developed a flexible methodology to estimate the mean of the skewed semicontinuous outcome of a patient and regression parameters in a semiparametric 2-part mixed-effects transformation model with an unknown transformation function. The current existing methods to analyzing correlated right-skewed semicontinuous data require the specification of the transformation, which is a difficult task in practice. Our paper has several new features over the existing methods. First, our method allows the arbitrary nonparametric transformation function and thus is more flexible and robust. The asymptotic distribution theory shows that our new estimators for the transformation function converge to their true values at the parametric rate $n^{-1/2}$ if the parameters are estimated at the parametric rate $n^{-1/2}$, suggesting that the extra flexibility is gained at little cost in efficiency. The simulation studies in the paper also show that the efficiency of our new estimators is comparable to the existing parametric method with the correctly specified transformation in finite sample sizes. Finally, we propose a new and more accurate approximate likelihood function to handle intractable numerical integration in the marginal likelihood, and the computational requirement of the new approximate likelihood is rather minimal.

In modeling nonzero data, we need to decide whether to put a parametric assumption on the transformation function or the distribution of the random effects. In our proposed method, we chose to impose a normal distribution assumption on random effects but leave the transformation function unknown. Our simulation study shows that the correctly specified transformation function is more important than the correctly specified distribution function of random effects in our inferences. Future research could explore the possibility of allowing both the transformation function and the distribution functions of random effects unknown.

In our proposed model, we assume that homoscedastic variance for transformed nonzero costs. In some cases, the homoscedastic variance assumption may be not met (Manning 1998; Mullahy 1998; Zhou, Gao and others, 1997; Zhou, Melfi and others, 1997; Zhou and Tu, 1999). Mullahy (1998) gave several real situations where 2-part regression models with homoscedastic variance after transforming the nonzero responses yield inconsistent inferences on $\mu(x)$. The heteroscedasticity for the nonzero data may be complicated. However, on the other hand, it may be difficult to get a good estimate of the variance if we specify a complicated heteroscedasticity. A possible method for handling heteroscedasticity of the nonzero data is to replace the second model (2.2) with

$$h(V_{ij}) = X'_{2ij} \beta + Z'_{2ij} \delta_i + g(X'_{2ij} \theta) \epsilon_{ij},$$

(6.1)
where \( g \) is a known function and \( \theta \) is a vector of unknown parameters. In the model (6.1), heteroscedasticity is modeled by the known function \( g(.) \) with a vector of unknown parameters, \( \theta \); hence, the heteroscedasticity is not linked to the mean level; and the mean and variance may be influenced by covariates in different ways. It is straightforward to extend our proposed method to model (6.1). Another possibility to model the heteroscedasticity is setting \( \theta = \beta \), as the literature in the generalized linear model, and leave the variance function unknown.

**ACKNOWLEDGMENTS**

The views expressed in this article are those of the author and do not necessarily represent the views of the Department of Veterans Affairs. We also like to thank Dr. Jurgen Unutzer for providing us with the data. **Conflict of Interest:** None declared.

**FUNDING**

National Natural Science of China (grant 10771148 to H.L.) AHRQ grant R01HS013105 and U.S. Department of Veterans Affairs, Veterans Affairs Health Administration, HSR&D grant ECI-03-206 to X.-H.Z.

**APPENDIX A**

In Appendix A, we outline a proof for our approximation (2.6). Denoting \( G_i(\delta_{li}) = \tau_{i}^{*}(\hat{\delta}_{li}) - \tau_{i}(\delta_{li}) \), we have

\[
G_i^{(2)}(\delta_{li}) = -\frac{1}{n_i} \sum_{j=1}^{n_i} \pi_{ij}(\hat{\delta}_{li})(1 - \pi_{ij}(\delta_{li}))Z_{1ij}Z_{1ij}'
\]

Note that

\[
a^*_i - a_i = g(\tau_{i}^{(s2)}(\hat{\delta}_{li}), \tau_{i}^{(s3)}(\hat{\delta}_{li}), \tau_{i}^{(s4)}(\hat{\delta}_{li})) - g(\tau_{i}^{(s2)}(\hat{\delta}_{li}), \tau_{i}^{(s3)}(\hat{\delta}_{li}), \tau_{i}^{(s4)}(\hat{\delta}_{li}))
\]

\[
+ g(\tau_{i}^{(s2)}(\hat{\delta}_{li}), \tau_{i}^{(s3)}(\hat{\delta}_{li}), \tau_{i}^{(s4)}(\hat{\delta}_{li})) - g(\tau_{i}^{(2)}(\hat{\delta}_{li}), \tau_{i}^{(3)}(\hat{\delta}_{li}), \tau_{i}^{(4)}(\hat{\delta}_{li}))
\]

\[
= O_p(\hat{\delta}^*_i - \hat{\delta}_{li}) + O_p(G^{(2)}(\hat{\delta}_{li}) + G^{(3)}(\hat{\delta}_{li}) + G^{(4)}(\hat{\delta}_{li})).
\]

From the above expression, we see that dependence of \( G^{(k)}(\delta_{li}) \), for \( k \geq 2 \) on \( \delta_{li} \) is through \( \Pi_{ij}(\delta_{li}) = \pi_{ij}(\hat{\delta}_{li})(1 - \pi_{ij}(\delta_{li})) \), and this dependence is negligible. This negligibility can be justified by using the same argument as in Bates and Watts (1980) and Pinheiro and Bates (2000) for assessing parameter effects on nonlinearity. There, they showed that the space spanned by the columns of \( \Pi_{ij}(\delta_{li}) \) depended only on the intrinsic curvature of the nonlinear model, but not on the parameter effects curvature in the tangent plane. Therefore, \( \Pi_{ij}(\delta_{li}) \) may be assumed to vary slowly with \( \delta_{li} \). This result, coupled with \( \hat{\delta}^*_i - \hat{\delta}_{li} = O_p(n_i^{-1/2}) \), gives us that \( a^*_i - a_i = O_p(n_i^{-1/2}) \). This completes the proof of the approximation (2.6).

**APPENDIX B**

To show our asymptotic results, we need the following conditions:

1. Suppose that \([v_0, v_1]\) is the domain of \( h \). In practice, this would be the range of the observed and fitted \( V_{ij}s \). Assume that \( h \) is strictly increasing and continuous for \( v \in [v_0, v_1] \).
2. There exists a sequence \( \{\hat{\Theta}\} \) such that \( \hat{\Theta} - \Theta_0 \to 0 \).
3. \((X_{1i}, X_{2i}, Z_{1i}, Z_{2i})\) has bounded support.
4. Denote $\Xi = \{(x, z): h^{-1}_0(v_0) \leq b_0 + x^2_0 \beta_0 + z^2_0 \delta_2 + \sigma \varepsilon \leq h^{-1}_0(v_1)\}$ for $\delta_2 \sim N(0, \psi_{220})$ and $\varepsilon \sim N(0, 1)$, suppose $\text{Pr}(\Xi) > 0$.
5. $n/R_1 = o(1)$ and $n/R_2 = o(1)$.
6. Suppose that $q_j(x, z, \Theta), j = 1, \ldots, 7$, are continuous functions of $\Theta$.

A.1 Proof of (2.11)

By the monotonicity and continuity of $\Phi$, for large $n$, any $\eta > 0$ and $\Theta \in \{\Theta: \|\Theta - \Theta_0\| \leq \eta\}$, uniformly in $v \in [v_0, v_1]$, there exists a unique $\hat{h}(v; \Theta)$ such that

$$S(\hat{h}(v; \Theta); v, \Theta) = 0,$$

(A.1)

where $S(w; v, \Theta)$ is defined in Section 3. Since $S(h_0(v); v, \Theta_0) \rightarrow 0$, we have $\hat{h}(v; \Theta_0) \rightarrow h_0(v)$, so that $\hat{h}(v; \Theta) \rightarrow h_0(v)$ almost surely uniformly in $v \in [v_0, v_1]$.

Now we consider the expansion of $\hat{h}(v) = \hat{h}(v; \Theta)$. Using a Taylor series expansion of $S(\hat{h}(v; \Theta); v, \Theta)$ with respect to $\hat{h}(v; \Theta)$ around $h_0(v)$, and noting that $S(\hat{h}(v; \Theta); v, \Theta) = 0$, we obtain

$$\hat{h}(v; \Theta) - h_0(v) \approx -(S'_{(100)}(h_0(v); v, \Theta))^{-1}S(h_0(v); v, \Theta).$$

Then, using a Taylor series expansion of $S(h_0(v); v, \Theta)$ with respect to $\Theta$ around $\Theta_0$, we get

$$\hat{h}(v; \Theta) - h_0(v) \approx s_1^{-1}(v)(S(h_0(v); v, \Theta_0) + s_2(v)'(\Theta - \Theta_0)).$$

(A.2)

This result, coupled with the condition 2 and the expression of $S(w; v, \Theta)$, leads to (2.11).

A.2 Proof of (3.2)

Replace $\Theta_0, h_0$ with $\Theta$ and $h$ for notational simplicity. Denote

$$u_n(x, z) = \frac{1}{R_1 R_2} \sum_{r=1}^{R_2} \sum_{k=1}^{R_1} \pi(\zeta_{1r})v(\zeta_{2rk}),$$

where $\zeta_{1r} = x_1^r \alpha + z_1^r \delta_1, \zeta_{2rk} = b_0 + x_2^r \beta + z_2^r \delta_2 + \sigma \varepsilon_k, \pi(\zeta_1) = \eta^{-1}(\zeta_1), v(\zeta_2) = h^{-1}(\zeta_2)$, and $\delta_r = (\delta_{1r}'$, $\delta_{2r}'), \sim N(0, \psi).$ Consider (3.1) and use the expansion,

$$\hat{u}(x, z) - u_n(x, z)$$

$$\approx \frac{1}{R_1 R_2} \sum_{r=1}^{R_2} \sum_{k=1}^{R_1} \pi(\zeta_{1r})v(\zeta_{2rk})x_1'(\hat{\alpha} - \alpha) + \frac{1}{R_1 R_2} \sum_{r=1}^{R_2} \sum_{k=1}^{R_1} \pi(\zeta_{1r})v(\zeta_{2rk})x_2'(\hat{\beta} - \beta)$$

$$+ \frac{1}{R_1 R_2} \sum_{r=1}^{R_2} \sum_{k=1}^{R_1} \pi(\zeta_{1r})v(\zeta_{2rk})(\hat{\sigma} - \sigma)\varepsilon_k - \frac{1}{n} \sum_{i=1}^m \sum_{j=1}^j \bar{Y}(X_{ij}, Z_{ij}, V_{ij})$$

$$- \frac{1}{R_1 R_2} \sum_{r=1}^{R_2} \sum_{k=1}^{R_1} \pi(\zeta_{1r})v(\zeta_{2rk})s_1^{-1}(v(\zeta_{2rk}))s_2(v(\zeta_{2rk}))'(\hat{\Theta} - \Theta)$$

$$\approx (\zeta(x, z, \Theta) - \tilde{\varrho}_4(x, z, \Theta))'(\hat{\Theta} - \Theta) - \frac{1}{n} \sum_{i=1}^m \sum_{j=1}^j \tilde{Y}(X_{ij}, Z_{ij}, V_{ij}),$$

(A.3)

where $\zeta$, $\tilde{\varrho}_4$, and $\tilde{Y}$ are $\zeta$, $\varrho_4$ and $Y$ defined in Section 4 but with $(\delta_{1r}', \delta_{2r}')' \sim N(0, \psi)$. 


Denote $u$ by $\tilde{u}$ if $\delta = (\delta'_1, \delta'_2)' \sim N(0, \psi)$, then

$$u_n(x, z) - \tilde{u}(x, z) = \frac{1}{R_2} \sum_{r=1}^{R_2} \left\{ \frac{1}{R_1} \sum_{k=1}^{R_1} \pi(\zeta_{1r})v(\zeta_{2rk}) - \pi(\zeta_{1r})E[v(\zeta_{2rk})|\delta_r] \right\}$$

$$+ \frac{1}{R_2} \sum_{r=1}^{R_2} \left\{ \pi(\zeta_{1r})E[v(\zeta_{2rk})|\delta_r] - E[\pi(x'_1\alpha + z'_1\delta_1)E[v(b_0 + x'_2\beta + z'_2\phi_2 + \sigma\epsilon)|\delta]] \right\}$$

$$= O(R_1^{-1/2}) + O(R_2^{-1/2}). \quad (A.4)$$

Furthermore, we have

$$\tilde{u}(x, z) - u(x, z) \approx z'_0(\psi^{1/2} - \psi^{1/2})E[\pi(1)(x'_1\alpha + z'_0\psi^{1/2}\phi)E[v(b_0 + x'_2\beta + z'_0\psi^{1/2}\phi + \sigma\epsilon)]\phi]$$

$$+ z'_0(\psi^{1/2} - \psi^{1/2})E[\pi(x'_1\alpha + z'_0\psi^{1/2}\phi)E[v(1)(b_0 + x'_2\beta + z'_0\psi^{1/2}\phi + \sigma\epsilon)]\phi], \quad (A.5)$$

where $\phi$ and $\epsilon$ are independent standard normal random vector and variables, respectively, $z_{01} = (z', 0')'$, $z_{02} = (0', z')'$. The result (3.2) follows from (A.3), (A.4), and (A.5).

REFERENCES


[Received August 25, 2008; revised April 28, 2009; accepted for publication May 20, 2009]