MTPmle: A **SAS** Macro and **Stata** Programs for Marginalized Inference in Semi-Continuous Data

Delia C. Voronca
Emmes Corporation

Mulugeta Gebregziabher
Medical University of South Carolina

Valerie Durkalski-Mauldin
Medical University
of South Carolina

Lei Liu
Northwestern University
at Chicago

Leonard E. Egede
Medical College
of Wisconsin

Abstract

We develop a SAS macro and equivalent Stata programs that provide marginalized inference for semi-continuous data using a maximum likelihood approach. These software extensions are based on recently developed methods for marginalized two-part (MTP) models. Both the SAS and Stata extensions can fit simple MTP models for cross-sectional semi-continuous data. In addition, the SAS macro can fit random intercept models for longitudinal or clustered data, whereas the Stata programs can fit MTP models that account for subject level heteroscedasticity and for a complex survey design. Differences and similarities between the two software extensions are highlighted to provide a comparative picture of the available options for estimation, inclusion of random effects, convergence diagnosis, and graphical display. We provide detailed programming syntax, simulated and real data examples to facilitate the implementation of the MTP models for both SAS and Stata software users.

Keywords: semi-continuous, marginalized two-part models, generalized gamma, log skew normal, complex survey design.

1. Introduction and background

Semi-continuous data are non-negative continuous data that are characterized by a point mass at zero and positive values that usually follow a skewed distribution (Olsen and Schafer 2001). Some common examples of semi-continuous data in biomedical research are health care expenditures, alcohol consumption, and levels of antibody concentrations in the blood.

Currently, the most popular approach to analyze these types of data are two-part models (TP; Cragg 1971). A TP model for semi-continuous data consists of two parts. The first part (binary) models the probability of being a zero versus a non-zero, and the second part (continuous) models the strictly non-zero outcomes using a continuous distribution defined on a positive domain. Conceptually, two-part models for semi-continuous outcomes are similar with hurdle models (Mullahy 1986) for zero-inflated count data, which model zero versus non-zero values in a first step and then the strictly non-zero values in a second step, using a discrete distribution truncated at zero.

TP models have been shown to be reliable when analyzing semi-continuous data (Schafer, Olsen *et al.* 1999; Duan, Manning, Morris, and Newhouse 1983). Other methods such as adding a small number to the zero value and then fitting an ordinary least squares (OLS) regression on the log transformed values, or completely discarding the zero values and fitting a model to the strictly non-zero values, would result in biased inference and/or loss of information. Note that the former approach would only shift the point mass at zero and therefore the transformed data would still fail to be normally distributed and the later approach would treat zeroes as missing values and therefore lead to loss of information.

Despite their advantages, conventional TP models do not provide marginal inference on population of health care users and nonusers and therefore often fail to answer the main research question. To address this issue, marginalized two-part models (MTP) were developed for different continuous distributions such as lognormal, gamma, Weibull, generalized gamma and log skew normal (Smith, Preisser, Neelon, and Maciejewski 2014; Voronca, Gebregziabher, Durkalski, Liu, and Egede 2015). Unlike the TP model, the MTP model provides coefficients with a straightforward marginal interpretation, i.e., the coefficients are interpreted without conditioning on observing a non-zero outcome.

Our work fills the existing gap in statistical software packages available to program these models. Moreover, since these are newly developed methods, there are no articles that compare the SAS and Stata software for MTP models in terms of performance and available options for the statistical analysis. Therefore, in this paper we 1) develop a SAS macro and Stata programs that can fit MTP models for semi-continuous data; 2) compare the proposed software extensions in terms of performance and availability of options for the statistical analysis; and 3) provide simulated and real data examples to demonstrate the application of the SAS macro and Stata programs.

2. Methods

A summary of the main concepts and formula used for the code development of the MTP models is presented below. More detailed information on the methodology can be found in Voronca *et al.* (2015) and Smith *et al.* (2014). In terms of notations, $\boldsymbol{x}_i^{(0)}$ is the vector of covariates used in the binary part specific to subject i, $\boldsymbol{x}_i$ is the vector of covariates used in the continuous part, y_i is the semi-continuous outcome of interest, f is the probability density function (pdf) of a continuous distribution defined on a positive domain, $\boldsymbol{\alpha}$ is the vector for model regression coefficients in the binary part, and $\boldsymbol{\delta}$ and $\boldsymbol{\beta}$ are vectors for regression coefficients in the continuous part of the TP and MTP model, respectively.

The pdf of a standard TP model is defined in Equations 1 and 2 below (Cragg 1971):

$$g_{\text{TP}}(y_i) = \begin{cases} 1 - \pi_i, & \text{if } y_i = 0, \\ \pi_i f\left(y_i; \boldsymbol{x}_i^\top \boldsymbol{\delta}\right), & \text{if } y_i > 0, \end{cases} \tag{1}$$

where the probability of being non-zero, π_i, is usually modeled using a logit or a probit link, and the location parameter in the continuous part, μ_i, is modeled using a log link and a continuous distribution defined on a positive domain:

$$\begin{aligned} \text{logit}(\pi_i) &= \boldsymbol{x}_i^{(0)\top} \boldsymbol{\alpha}, \\ \log(\mu_i) &= \boldsymbol{x}_i^\top \boldsymbol{\delta}. \end{aligned} \tag{2}$$

Each component of $\boldsymbol{\delta}$ represents the change in the outcome on the log scale, for one unit increase in the corresponding covariate, given that the outcome is strictly greater than zero.

The MTP models are derived for the standard TP models. The pdf of an MTP model can be defined as in Equation 3 (Smith *et al.* 2014):

$$g_{\text{MTP}}(y_i) = \begin{cases} 1 - \pi_i, & \text{if } y_i = 0, \\ \pi_i f\left(y_i; \boldsymbol{x}_i^\top \boldsymbol{\beta}\right), & \text{if } y_i > 0, \end{cases} \tag{3}$$

where the probability of being non-zero, π_i, is modeled in the same way as for the TP models, and the marginal mean is modeled as $\mathsf{E}(Y_i) = \exp(\boldsymbol{x}_i^\top \boldsymbol{\beta}) = \xi_i$. The location parameter μ_i is parameterized in terms of the marginal mean such that:

$$\mu_i = \boldsymbol{x}_i^\top \boldsymbol{\beta} - \log(\pi_i) - C(\sigma, k), \tag{4}$$

where $C(\sigma, k)$ takes different forms depending on the distribution assumed for the continuous part. Assuming the pdf used in the continuous part is the generalized gamma family of distributions with σ the scale and k the shape parameter (Manning, Basu, and Mullahy 2005), it follows that for the MTP lognormal $C_{gg}(\sigma) = \sigma^2/2$, for the MTP gamma $C_{gg} = 0$, for the MTP Weibull $C_{gg}(\sigma) = \log[\Gamma(1+\sigma)]$, and for the MTP generalized gamma $C_{gg}(\sigma, k) = \frac{\sigma \log(k^2)}{k} + \log[\Gamma(\frac{1}{k^2} + \frac{\sigma}{k})] - \log[\Gamma(\frac{1}{k^2})]$ (Voronca *et al.* 2015). Log skew normal does not belong to the generalized gamma family. The C component associated with a log skew normal (Azzalini 1985) distribution is $C_{lsn}(\omega, k) = \log(2) + \log\left[\Phi\left(\frac{k\omega}{\sqrt{1+k^2}}\right)\right] + \omega^2/2$ where ω is the scale and k is the shape parameter.

In terms of interpretation, $\exp(\beta_j)$ represents the multiplicative effect of the covariate j on the outcome and $1 - \exp(\beta_j)$ represents the % change in the outcome for one unit increase in the covariate j. The $\boldsymbol{\alpha}$ coefficients in the binary part (for both TP and MTP models) have the usual interpretation corresponding to a logistic regression.

The MTP models above can be extended to include random effects. Using the same notations, the general form of such MTP models can be written as in Equation 5 (Smith, Neelon, Preisser, and Maciejewski 2017):

$$\begin{aligned} \text{logit}(\pi_i) &= \boldsymbol{x}_i^{(0)\top} \boldsymbol{\alpha} + \boldsymbol{z}_i^{(0)\top} \boldsymbol{a}_i, \\ \mu_i &= \boldsymbol{x}_i^\top \boldsymbol{\beta} + \boldsymbol{z}_i^\top \boldsymbol{c}_i - \log(\pi_i) - C(\sigma, k), \end{aligned} \tag{5}$$

where $z_i^{(0)}$ and z_i are the vectors of random-effect covariates used in the binary part and for the marginal mean, a_i and c_i are random effects corresponding to the binary part and to the marginal mean, respectively. The random effects are assumed to follow a multivariate normal distribution:

$$\left(\begin{array}{c} a_i \\ c_i \end{array} \right) \sim N \left(\mathbf{0}, \mathbf{\Sigma} = \left[\begin{array}{cc} \mathbf{\Sigma}_{aa} & \mathbf{\Sigma}_{ac} \\ \mathbf{\Sigma}_{ac} & \mathbf{\Sigma}_{cc} \end{array} \right] \right). \tag{6}$$

The subjects-specific marginal mean (for both users and non-users) corresponding to MTP models with random effects is:

$$\mathsf{E}\left(Y_i | a_i, c_i\right) = \exp(x_i^\top \beta + z_i^\top c_i) = \nu_i. \tag{7}$$

3. **SAS** macro for customized likelihood functions

3.1. Overview

SAS and Stata are currently two of the most popular software packages used for statistical analysis. The SAS software package facilitates data management, offers a variety of statistical methods for small data, large data and data with missingness, and it operates on validated algorithms. These features make it very popular among researchers from various fields, especially for those in the medical field. The most recent version of SAS is SAS 9.4 (SAS Institute Inc. 2013) which is also used for the analysis performed in this article. A `DATA` step creates and manipulates the data set, whereas a `PROC` step performs the statistical analysis. One common procedure used for nonlinear mixed models in SAS is `PROC NLMIXED` which allows for models with both fixed and random effects. The procedures allow for both pre-specified as well as user-defined likelihood functions. When the likelihood is integrated over random effects, an integral approximation is performed using the best available methods (default is the Gauss-Hermite quadrature; Pinheiro and Bates 1995) and then different optimization techniques are used for maximization (default is the dual quasi-Newton algorithm; Wolfinger 1999). Part of the standard `PROC NLMIXED` output are the convergence status, the number of iterations to convergence, model fit information criteria, parameter estimates, and standard errors based on the second derivative matrix of the likelihood function. Currently, `PROC NLMIXED` performs only maximum likelihood estimation with random effects from a multivariate normal distribution. Like with any other procedure, the code can be easily documented by using comments and the output can be stored in a text or PDF file using the `ODS` command with different options. Of note is that other SAS procedures such as `PROC GLIMMIX` can also fit non-linear models but only for pre-specified likelihood functions, which do not include likelihood functions for semi-continuous two-part models.

3.2. The %MTPmle **SAS** macro

The proposed `%MTPmle` SAS macro is using `PROC NLMIXED` to fit the marginalized two-part models. Before fitting the macro, dummy variables need to be created for all categorical covariates. The general syntax of the macro and an explanation of the arguments are presented below:

```
%MTPmle(modeltype = , data = , outcome = , vars0 = , vars1 = , c0 = false,
  z0 = false, corr = false, initparms = 0.1, grid = 1, id = NONE,
  tp = false, plot = false)
```

- `modeltype`: is the continuous distribution used in the second part of the MTP (or TP) model; the available options are LN (lognormal), G (gamma), W (Weibull), GG (generalized gamma), LSN (log skew normal).

- `data`: is the name of the data set where the variables of interest reside.

- `outcome`: is the semi-continuous outcome of interest.

- `vars0`: represents the names of variables and interaction terms included in the binary part of the MTP (or TP) model.

- `vars1`: represents the names of variables and interaction terms included in the continuous part of the MTP (or TP) model.

- `c0`: default value `FALSE`; it can be `TRUE` or `FALSE`; when its value is changed to `TRUE`, a random intercept is included in the continuous part of the MTP (or TP) model; the distribution of the random effect is assumed normal with mean zero.

- `z0`: default value `FALSE`; can be `TRUE` or `FALSE`; when its value is changed to `TRUE`, a random intercept is included in the binary or the zero part of the MTP (or TP) model; the distribution of the random effect is assumed normal with mean zero.

- `corr`: default value `FALSE`; can be `TRUE` or `FALSE`; when its value is changed to `TRUE`, the intercept in the continuous and the binary part are assumed to be correlated; the value of this argument has an effect only when random intercepts are included in both parts of the model.

- `initparms`: default value 0.1; represents the initial values given to the model parameters corresponding to the fixed effects specified in `vars0` (`a_`) and `vars1` (`b_`).

- `grid`: default value 1; it represents the initial values given to the variances (`a_sigma2` and `b_sigma2`) of the random effects (`z0` and `c0`, respectively); the value for this argument can be also specified as a range of values using an incremental step; for example `grid` = 0 to 1 by 0.1.

- `id`: default value is `NONE`; it represents the name of the variable corresponding to the unique identifier; the unique identifier is needed only when random effects are included, in order to identify clusters or repeated measures.

- `mtp`: default value `TRUE`; can be `TRUE` or `FALSE`; when its value is `TRUE` it means that an MTP models is fitted, which is the default; when its value is changed to `FALSE`, a TP model is fitted instead.

- `plot`: default value `FALSE`; can be `TRUE` or `FALSE`; when its value is `TRUE` it graphs the kernel densities associated with the observed and predicted values.

Of note is that the macro will stop the execution and output an error message when no data set is specified ("ERROR: No dataset supplied"), no outcome is specified ("ERROR: No outcome variable supplied"), no model type is specified ("ERROR: No model type supplied") or when a `TRUE` or `FALSE` parameter (`c0`, `z0`, `mtp`, `plot`) is assigned a different string value (for example the output error could be "ERROR: 'mtp' can be either `TRUE` or `FALSE`"). In addition, for models that include random effects (`c0` or `z0` is `TRUE`) the `id` also needs to be specified to avoid an error message ("ERROR: No id variable supplied"). The macro saves the parameter estimates and other important statistics in a data set called `_out_parm`, the run time in the data set called `_time` and the predicted means in a data set called `_parm`. The data set `_parm` is used to graph observed versus predicted values. Using too many variables relative to the size of the data set may lead to non-convergence. Also, the more variables are used in the model, the longer the run time.

4. **Stata** programs for customized likelihood functions

4.1. Overview

`Stata` is a statistical package widely used by researchers, especially economists and health care investigators. Some of the most powerful features of `Stata` are quick convergence, conservative approach to declare convergence, an easy way to implement maximum likelihood estimation, a search algorithm for optimal initial values, and a variety of post-estimation commands. The software can be used for complex statistical analysis, data management, and to generate publication quality graphs. The most recent version of the software, which is also used for this article, is `Stata` 14 (StataCorp. 2015a). `Stata` commands can be stored in an ado-file which is a standard text file that can be executed interactively in `Stata` by typing:

```
. do "filename.ado"
```

The output of an ado-file can be logged, similar to logging an interactive session, by using the commands:

```
. log using "logname"
. log close
```

In addition to the built-in commands available in `Stata`, the user can develop flexible and computational efficient models that perform maximum likelihood estimation (MLE) by writing a `Stata` program. A program can be fairly general and easily extended to a variety of model specifications without changing the program syntax. The main advantage of using the `Stata` `ml` programming language is that it allows the user to implement customized likelihood functions, such as the one used for marginalized two-part models. The default optimization technique is Newton-Raphson but other options are available such as Bendt-Hall-Hall-Hausman, Davidson-Fletcher-Powell, or Broyden-Fletcher-Goldfarb-Shanno algorithms (Gould, Pitblado, and Sribney 2006). To execute an `ml` program, one must type a sequence of commands (Steenbergen 2012), some of which are optional but can prove useful in specific situations. The programs can be defined interactively or stored in an ado-file. An important note is that the `Stata` program is most straightforward to write and execute for likelihood

functions that satisfy the linear form (`lf`) restriction, i.e., the observations are (conditionally) independent. **Stata** derives the first and second (partial) derivatives numerically and uses these in the optimization algorithm. When this condition is not met, as for repeated measures or clustered data, other options are available such as `d0`, `d1`, and `d2` (StataCorp. 2015b). These options involve programming first and/or second order derivatives which complicates the program syntax but in certain situations allows extensions of simpler models to more complicated ones such as random effects models. This topic is considered for future research.

4.2. The `MTPmle` **Stata** programs

In order to maximize the likelihood function for MTP models, we develop **Stata** programs that are saved in an ado-file and can be loaded in **Stata** by using the command:

```
. do MTPmle.ado
```

Once the ado-file is loaded, the user can use the command:

```
. ml model lf model_type (outcome = vars0) (outcome = vars1) ///
> (outcome = ) (outcome = ), technique(nr)}
```

- `model_type`: identifies the continuous distribution used in the second part of the MTP model; the options available are `mtp_lognormal` (MTP lognormal), `mtp_gamma` (MTP gamma), `mtp_Weibull` (MTP Weibull), `mtp_gengamma` (MTP generalized gamma), and `mtp_logskewnormal` (MTP log skew normal).

- `outcome`: is the name of the dependent semi-continuous variable.

- `vars0`: represents an array of names for the covariates included in the binary part separated by spaces.

- `vars1`: represents the array of names for the covariates included in the continuous part separated by spaces.

- `(outcome = )` or `(outcome = ) (outcome = )`: correspond to the nuisance parameters; for example, lognormal, gamma and Weibull have only one nuisance parameter (variance for log normal, gamma scale and Weibull scale, respectively) and therefore only one set of `(outcome = )` will be used; generalized gamma has two nuisance parameters, the scale (lns2) and the shape (1/t) (Voronca *et al.* 2015), and therefore `(outcome = )` `(outcome = )` will be used; similarly, the log skew normal has two nuisance parameters, scale (omega) and shape (k) (Smith *et al.* 2014) and therefore `(outcome = )` `(outcome = )` will be used.

- `technique(nr)`: (Newton-Raphson) is the default optimization algorithm; in addition to `nr`, other available techniques are `bhh` (Bendt-Hall-Hall-Hausman), `dfp` (Davidson-Fletcher-Powell), and `bfgs` (Broyden-Fletcher-Goldfarb-Shanno).

After the `ml model` statement, a variety of commands can be used for the fitted MTP model. These commands are discussed in detail in the **Stata** documentation (StataCorp. 2015b) and

other articles (Steenbergen 2012). We recommend the following: `ml search` which performs an automated search for the best initial values for model parameters in order to optimize convergence; `ml maximize` which maximizes the specified MTP likelihood function and generates output; the `difficult` option can be used with `ml search` when a direction vector for a more difficult likelihood function is not found (suggested by the warning message "not concave"); `ml graph` which produces a graph of the iteration path useful in monitoring the convergence.

5. Simulations

To show that the proposed `%MTPmle` SAS macro and `Stata` programs are correctly estimating model parameters, we provide two simulation examples with known true parameter values from a lognormal distribution. Equations 8 and 9 describe the two models used for simulations using the same notations as in Section 2. For both simulation scenarios, we generated data for a sample size of 10000 with approximately 40% zeroes.

The first model used for simulation is a cross-sectional MTP LN:

$$\text{logit}(\pi_i) = \alpha_0 + \alpha_1 \times x_{1i} + \alpha_2 \times x_{2i},$$
$$\mu_i = \beta_0 + \beta_1 \times x_{1i} + \beta_2 \times x_{2i} - log(\pi_i) - \sigma^2/2, \tag{8}$$

where $x_{1i} \sim N(10, 2)$ and $x_{i2} \sim \text{Bernoulli}(0.5)$. The true parameter values were chosen as $\sigma^2 = 4$, $\boldsymbol{\alpha} = (5.7, -0.5, -0.5)$ and $\boldsymbol{\beta} = (7, -0.5, -0.5)$.

The second model used for simulation is a longitudinal MTP LN with uncorrelated random intercepts:

$$\text{logit}(\pi_i) = \alpha_0 + \alpha_1 \times x_{1i} + \alpha_2 \times x_{2i} + \alpha_3 \times x_{3i} + a_i$$
$$\mu_i = \beta_0 + \beta_1 \times x_{1i} + \beta_2 \times x_{2i} + \beta_3 \times x_{3ij} + c_i - log(\pi_i) - \sigma^2/2 \tag{9}$$

Cross-sectional MTP LN, 40% zeroes, $N = 10,000$			Longitudinal MTP LN, 40% zeroes, $N = 10,000$		
Model parameters	True values	`MTPmle` SAS or `Stata`	Model parameters	True values	`MTPmle` SAS
β_0	7.0	7.03	β_0	10.0	9.90
β_1	−0.5	−0.51	β_1	−0.1	−0.09
β_2	−0.5	−0.48	β_2	−0.4	−0.37
α_0	5.7	5.76	β_3	−0.6	−0.58
α_1	−0.5	−0.51	σ_c^2	2.0	1.41
α_2	−0.5	−0.51	α_0	3.8	3.33
σ^2	4.0	3.96	α_1	−0.1	−0.08
			α_2	−0.1	−0.09
			α_3	−0.2	−0.22
			σ_a^2	1.0	0.68
			σ^2	4.0	4.03

Table 1: Validation of `MTPmle` SAS macro and `Stata` programs using simulations.

where $x_{1i} \sim N(30, 4)$ and $x_{i2} \sim$ Bernoulli(0.5). The true parameter values were chosen as $\sigma^2 = 4$, $\begin{pmatrix} a_i \\ c_i \end{pmatrix} \sim N\left(\begin{bmatrix} 0 \\ 0 \end{bmatrix}, \begin{bmatrix} 1 & 0 \\ 0 & 2 \end{bmatrix} \right)$. $\boldsymbol{\alpha} = (3.8, -0.1, -0.1, -0.2)$ and $\boldsymbol{\beta} = (10, -0.1, -0.4, -0.6)$. The time variable x_{3ij}, for $j = 1, 2, 3$ represents the 3 time points at which repeated outcome measures were available for each subject.

Results are summarized in Table 1. Additional simulation scenarios are described in Voronca *et al.* (2015).

6. Data examples

6.1. Data

The first data example is simulated from a longitudinal model with 500 subjects, randomized in two treatment groups. The semi-continuous outcome (Y) is measured at 3 time points and has approximately 49% zeroes at time 3 and 45% zeroes overall. For simplicity, we assume there is no missing data. The simulation is mimicking a randomized controlled trial (RCT), such as an RCT for substance use, where subjects are randomized to an intervention or usual care group (X_2) and outcome measurements (Y) are taken at different time points (baseline, month 1 and month 2, for example). The usual aim of such RCTs is to investigate the effect of the intervention ($X_2 = 1$) in comparison to the control group ($X_2 = 0$) on lowering substance use (Y), without conditioning on observing a non-zero outcome. One approach is to look at differences in the outcome between the two treatment groups at the end of the study (Y_3) after adjusting for outcome levels at baseline (Y_1) and possible confounders (X_1). A second approach is to analyze the data longitudinally and determine if there is a difference in the rate of decline in the outcome levels over time between the two treatment groups, after adjusting for possible confounders. Summary statistics for the simulated data are presented in Table 2.

For the second example, we used publically available data provided by the Medical Expenditure Panel Survey (MEPS), household component 2011 (HC-147), to show how the Stata programs (Stata model 3) can be extended to account for a complex survey design. The aim is to determine if there are any differences in emergency room (ER) expenditures in 2011, between the adult (≥ 18 years) US noninstitutionalized non-Hispanic white and non-Hispanic Asian civilians, after adjusting for confounders and without conditioning on observing a non-zero expense. The total sample size to address this aim was 13,239 and the semi-continuous outcome of interest (ERTEXP11) had 88% zeroes. Summary statistics of the sample variables are provided in Table 3.

Treatment	Y	N	Mean	SD	Zeroes (%)
$X_2 = 0$	Time 1	252	1123.5	4951.6	41
(usual care)	Time 2	252	716.6	2855.2	45
	Time 3	252	274.9	1106.5	48
$X_2 = 1$	Time 1	248	1670.4	19317.5	39
(intervention)	Time 2	248	296.4	1961.9	46
	Time 3	248	197.7	1165.7	49

Table 2: Summary statistics for the semi-continuous outcome Y at each time point by treatment group.

Variable	Overall	Whites	Asians
	Mean ± sd	Mean ± sd	Mean ± sd
Total ER exp.	178.07 ± 931.44	196.96 ± 988.35	60.30 ± 412.93
Age (years)	48.28± 18.10	48.83±18.21	44.86± 17.03
	N (%)	*N* (%)	*N* (%)
Gender			
Male	6,335 (47.85)	5,483 (48.06)	852 (46.56)
Female	6,904 (52.15)	5,926 (51.94)	978 (53.44)
Marital status			
Married	7,406 (55.98)	6,290 (55.17)	1,116 (61.02)
Not married	3,030 (22.90)	2,537 (22.25)	493 (26.95)
Never married	2,794 (21.12)	2,574 (22.58)	220 (12.03)
Education			
< High school	1,682 (12.96)	1,450 (12.94)	232 (13.08)
High school	3,867 (29.79)	3,505 (31.27)	362 (20.41)
College	5,739 (44.21)	4,898 (43.70)	841 (47.41)
Graduate	1,694 (13.05)	1,355 (12.09)	339 (19.11)
Income			
Poor	2,242 (16.93)	1,973 (17.29)	269 (14.7)
Low	1,833 (13.85)	1,616 (14.16)	217 (11.86)
Middle	4,098 (30.95)	3,531 (30.9)	567 (30.98)
High	5,066 (38.27)	4,289 (37.59)	777 (42.46)
Insured			
Yes	10,940 (82.63)	9,471 (83.01)	1,469 (80.27)
No	2,299 (17.37)	1,938 (16.99)	361 (19.73)
Region			
Northeast	2,214 (16.72)	1,905 (16.70)	309 (16.89)
Midwest	3,465 (26.17)	3,266 (28.63)	199 (10.87)
South	4,190 (31.65)	3,825 (33.53)	365 (19.95)
West	3,370 (25.46)	2,413 (21.15)	957 (52.30)
Depression	1,071 (9.23)	972 (9.71)	99 (6.21)
Hypertension	4,354 (33.03)	3,898 (34.31)	456 (25.01)
Joint pain	4,601 (35.9)	4,252 (38.53)	349 (19.65)
Stroke	507 (3.85)	481 (4.23)	26 (1.43)
Emphysema	392 (2.97)	382 (3.36)	10 (0.55)
High Cholesterol	4,183 (31.75)	3,712 (32.69)	471 (25.86)
Arthritis	3,470 (26.32)	3,263 (28.71)	207 (11.36)
Asthma	1,191 (26.32)	1,092 (9.61)	99 (5.43)
CVD	1,990 (15.10)	1,876 (16.51)	114 (6.26)
Diabetes	1.205 (9.14)	1,048 (9.22)	157 (8.61)
Total ER exp.			
Zeroes	11,640 (87.92)	9,908 (86.8)	1,732 (94.64)
Non-zeroes	1,599 (12.08)	1,501 (13.16)	98 (5.36)

Table 3: Summary statistics of MEPS data (used for Stata model 3); MEPS = Medical Expenditure Panel Survey; ER = emergency room; CVD = cardio vascular diseases; exp. = expenditure.

6.2. Using the `%MTPmle` **SAS** macro

SAS model 1

For this example we use the simulated data described in Section 5. We assume only pre-intervention (`Y_1`, baseline) and post-intervention (`Y_3`, follow-up) outcome measurements are available for the primary analysis. The aim is to determine the magnitude and significance of the treatment effect on the outcome post-intervention after adjusting for outcome baseline levels. This type of research question is common for medical RCTs. For this scenario, the recommended approach is ANCOVA (Vickers 2001). In a real data application, the assumptions of ANCOVA (Miller and Chapman 2001) should be checked before fitting the model. Of note is that statistical methods assuming a repeated measures design, such as repeated measures ANOVA, could also be used. However, these methods answer a different research question, namely whether the mean change from pre to post is different between interventions. The SAS macro syntax used for the ANCOVA analysis is given below:

```
%MTPmle(modeltype = ln, data = ln_data_short, outcome = Y_3,
  vars0 = X1 X2 Y_1, vars1 = X1 X2 Y_1, initparms = 0,
  grid = 0.1 to 2 by 0.1, plot = true)
```

The fitted model is MTP lognormal which is specified by the `modeltype = ln`. The name of the data set used for the analysis is `ln_data_short` and the outcome is `Y_3` (semi-continuous outcome at time 3). The covariates used in the binary part (`vars0`) are the same as the ones used in the continuous part (`vars1`). These covariates are the treatment group (`X2`), the semi-continuous outcome at baseline (`Y_1`), and the continuous covariate `X1`. For a real data example, the same covariates or different covariates can be used in each part of the model, depending on what is deemed more appropriate by the statistician or the clinical investigator. To initialize model parameters, the `initparms` option was used with a value of 0 whereas the initial value for the nuisance parameter which is the variance (sigma2) corresponding to the lognormal distribution was chosen from a grid search from 0.1 to 2 by increments of 0.1 using the `grid` option. Alternative initial values or grid searches can be used when a model does not converge. However, performing multiple grid searches does not necessary lead to improved estimation or convergence, especially when more covariates are included in the model. Moreover, it increases the run time and sometimes can even terminate the execution of the macro. Therefore, the grid search is allowed only for the initial values of the nuisance parameters but not for the model parameters. The method used for optimization is the default dual quasi-Newton algorithm. The option `plot = true` displays a graph of the kernel densities for the observed versus predicted values. Fit statistics, such as Akaike information criterion (AIC; Akaike 1974) or Bayes information criterion (BIC) can be used to select among the available option for the model type (MTP lognormal, MTP gamma, MTP Weibull, MTP generalized gamma, MTP log skew normal). The lowest fit statistics across different models

Criteria	MTP ln	MTP g	MTP w	MTP gg	MTP lsn
`-2 Log`	3753.7	3863.9	3788.7	NC	3753.7
`AIC`	3771.7	3881.9	3806.7	NC	3773.7

Table 4: Comparison of model fit based on `-2 Log` and `AIC`; NC = non-convergent.

Parameter	Estimate	Standard error	p value	95% Confidence limits	
b_int	8.6635	1.1095	< 0.0001	6.4837	10.8433
b_X1	−0.08899	0.03590	0.0135	−0.1595	−0.01846
b_X2	−0.6013	0.3017	0.0468	−1.1941	−0.00844
b_Y_1	8.12E−6	7.664E−6	0.2899	−6.94E−6	0.000023
a_int	2.1809	0.7005	0.0020	0.8046	3.5572
a_X1	−0.07007	0.02269	0.0021	−0.1146	−0.02549
a_X2	−0.01768	0.1807	0.9221	−0.3727	0.3374
a_Y_1	4.93E−6	0.000012	0.6892	−0.00002	0.000029
sigma2	5.3608	0.4732	< 0.0001	4.4311	6.2905

Table 5: Parameter estimates: Selected output from `%MTPmle` SAS macro for SAS model 1 (MTP ln ANCOVA type model).
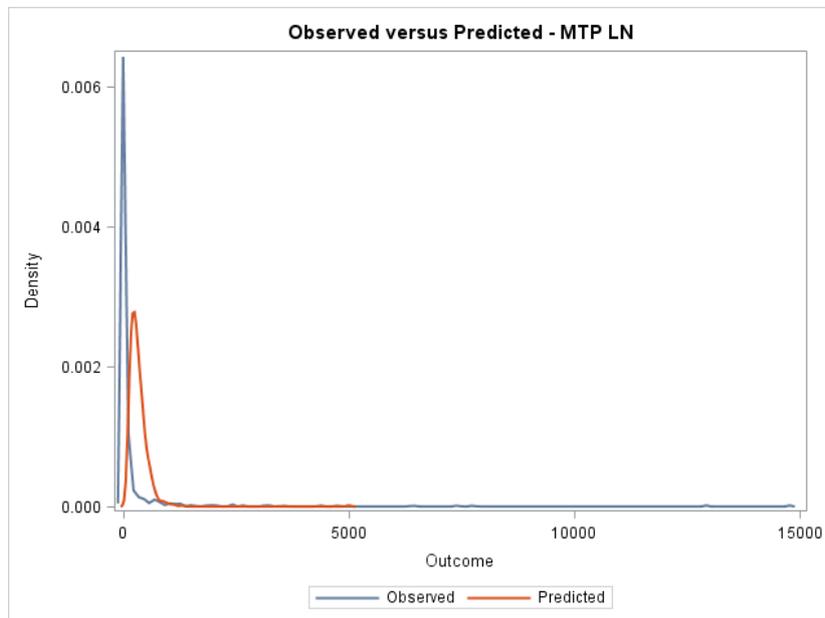

Figure 1: Graph of observed versus predicted values from `%MTPmle` SAS macro for SAS model 1 (MTP ln ANCOVA type model).

identifies the best fitting model. MTP lognormal was chosen because it had the smallest AIC (Table 4) and is more parsimonious in terms of parameters estimated comparative to other models such as MTP log skew normal.

The SAS output corresponding to the model parameter estimates, p values and the 95% confidence intervals on the log scale are presented in Table 5. Parameters that start with `b_` represent the continuous part of the MTP model, whereas parameters starting with `a_` represent the binary part of the MTP model. `sigma2` is the variance on the log scale corresponding to the lognormal distribution, which was used in the continuous part of the MTP model. The parameters in the continuous part can be interpreted as the change in the outcome on the log scale for one unit increase in the covariate. If we exponentiate the coefficients in the continuous part and subtract from 1, we get an interpretation on the original scale

as % change in the outcome. For example, `b_X2` $= -0.6$ means that on average, subjects in the intervention group had 0.6 units less in $\log(Y)$ at time 3 comparative to the usual care group. The corresponding relative risk, $\exp(\texttt{b\_X2}) = 0.55$, suggests that on average, there is a 45% reduction in the levels of outcome Y at time 3 in the intervention group comparative to the usual care group. For this model, the p value (0.047) and the 95% confidence interval $(-1.19, -0.01)$ suggest that the treatment effect is statistically significant at the 0.05 alpha level. The parameters in the binary part have the usual interpretation corresponding to a logit model. For example the odds ratio $\exp(\texttt{a\_X2}) = 0.98$ suggests that being in the intervention group decreases the chances of having a zero outcome by 2%. The graph of observed versus predicted values is presented in Figure 1. In addition, `PROC NLMIXED` outputs information about the data set and observations used, the convergence status, initial parameter values, and iteration history. The program for the above model runs in less than 1 second.

SAS model 2

An alternative analysis for the simulated data is to determine if there is a difference in the rate of decline in the outcome levels over time between the two treatment groups, after controlling for important covariates. Measurements of the outcome collected at time 1, 2 and 3 are used in a longitudinal model that controls for correlations among repeated measures through the random intercept which is included in both parts of the model. The syntax for the longitudinal analysis with two uncorrelated random effects is presented below:

```
%MTPmle(modeltype = ln, data = ln_data_long, outcome = Y,
  vars0 = X1 X2 time X2*time, vars1 = X1 X2 time X2*time,
  z0 = true, c0 = true, corr = false, id = id, plot = false)
```

The data set used for analysis (`ln_data_long`) contains repeated measures of outcome `Y` over 3 time points in the long format. The variables used in both parts of the MTP lognormal (given by `var0` and `vars1`) are time as a continuous variable (`time`), treatment group as a dichotomous variable (`X2`), and a continuous covariate (`X1`). The interaction term between treatment and time (`X2*time`) is also included in both parts of the model. The `id` is the subject unique identifier and needs to be provided in order to indicate the repeated/clustered measures. The `z0` and `c0` are set to `TRUE` and represent the random intercepts used in the binary and continuous part respectively, that follow a multivariate normal distribution. The option `corr = false` (which is also the default value) suggests that the random intercepts are not correlated. Based on model fit criteria, the smallest AIC corresponded to the model that had uncorrelated random intercepts in each part of the model. The alternative models were random intercept in the binary part only, random intercept in the continuous part only, and correlated random intercepts in both parts of the MTP lognormal model which did not converge. By adding the random intercept, a variance-covariance matrix with a compound symmetry structure is assumed in each part of the model. Other variance-covariance structures for the mixed model cannot be specified directly with the proposed macro and this topic is considered for future work.

Model parameters for the longitudinal analysis using the `%MTPmle` SAS macro are presented in Table 6. The output suggests that there is a significant effect of time (p val < 0.001) and that, on average, for every month (or unit increase in time) there is a .55 units drop (`b_time = -0.55`) in the levels of Y. Similarly, $\exp(\texttt{b\_time}) = 0.58$ suggests that there is a

Parameter	Estimate	Standard error	p value	95% Confidence limits	
`b_int`	9.6863	0.7765	< 0.0001	8.1606	11.2120
`b_X1`	-0.09221	0.02381	0.0001	-0.1390	-0.04543
`b_X2`	-0.5848	0.3978	0.1422	-1.3664	0.1968
`b_time`	-0.5522	0.1264	< 0.0001	-0.8005	-0.3038
`b_X2time`	-0.02097	0.1784	0.9065	-0.3715	0.3296
`b_sigma2`	1.9576	0.2955	< 0.0001	1.3770	2.5381
`a_int`	3.6683	0.6195	< 0.0001	2.4511	4.8855
`a_X1`	-0.09900	0.01874	< 0.0001	-0.1358	-0.06219
`a_X2`	0.1162	0.3248	0.7206	-0.5220	0.7545
`a_time`	-0.1983	0.1005	0.0491	-0.3958	-0.00076
`a_X2time`	-0.05900	0.1430	0.6800	-0.3399	0.2219
`a_sigma2`	1.1188	0.2381	< 0.0001	0.6510	1.5865
`sigma2`	3.6038	0.2442	< 0.0001	3.1240	4.0836

Table 6: Parameter estimates: Selected output from `%MTPmle` SAS macro for SAS model 2 (MTP ln with random intercepts).

42% reduction in levels of Y for each month. The interaction term (`b_X2time` $= -0.02$) is not statistically significant (p val $= 0.907$). Therefore, we conclude that the average rates of decline over time for the two treatment groups are not different. To run the above longitudinal model took 43 seconds.

In order to fit other types of MTP models, a different option can be used for the `modeltype`. For example, the syntax to fit an MTP generalized gamma is:

```
%MTPmle(modeltype = gg, data = ln_data_long, outcome = Y,
  vars0 = X1 X2 time X2*time, vars1 = X1 X2 time X2*time,
  z0 = true, c0 = true, corr = false, id = id, plot = false)
```

Of note is that all the MTP models available with the `%MTPmle` macro model the mean of a specific continuous distribution (lognormal, Weibull, gamma, generalized gamma or log skew normal) using a log link function. Therefore the interpretation of the resulting coefficients is the same as the one described for the lognormal MTP models in SAS *model 1* and *2*.

In addition to the MTP models, the SAS macro has the capability to run the equivalent TP model for comparison. This is done by setting the option `mtp` equal to `FALSE`. For example, we can run the longitudinal TP lognormal model using the syntax below:

```
%MTPmle(modeltype = ln, data= ln_data_long, outcome = Y,
  vars0 = X1 X2 time X2*time, vars1 = X1 X2 time X2*time,
  z0 = true, c0 = true, corr = false, id = id, mtp = false, plot = false)
```

Unlike the MTP models, the coefficients in the continuous part of a TP model (also starting with `b_`) are interpreted as conditioned on observing a non-zero value. The coefficients in the binary (start with `a_`) part have the interpretation of a logit model, the same as for the MTP models.

6.3. Using the `MTPmle` **Stata** programs

Stata model 1

In order to determine if there is a treatment effect on the outcome at time 3 (`y_3`), after controlling for baseline levels (`y_1`) and important covariates (`x1`), we can fit the ANCOVA type of model in Stata, after loading the `MTPmle` ado-file. The Stata commands we used to fit the MTP lognormal model, which gives equivalent results to SAS model 1, is given below:

```
. ml model lf mtp_lognormal (cont: y_3 = y_1 x1 x2) ///
> (binary: y_3 = y_1 x1 x2)(logvar: y_3 = )
. ml search
. ml maximize, difficult
. ml graph
```

The `ml model lf mtp_lognormal` performs maximum likelihood (`ml`) estimation on the MTP lognormal customized likelihood which satisfies the linear form (`lf`) condition. The first equation (`cont: y_3 = y_1 x1 x2`) represents the regression model with a log link for the continuous part of the MTP model. The second equation (`binary: y_3 = y_1 x1 x2`) represents the logistic model for the binary part. For this example the equations are the same since we use the same covariates in both parts. The third equation (`logvar: y_3 =`) represents the regression model with a log link for the variance of the lognormal distribution. If no covariates are specified for the third equations, then the nuisance parameter is assumed to be homogenous across all observations, similar to the SAS macro. Of note is that the outcome (`y_3`) is the same for all equations but the covariates in the right side of each equality can be different. The `ml search` command performs an automatic search for optimal initial values for model parameters as well as for the nuisance parameter (here the variance of the lognormal distribution). The `ml search` is not mandatory but we recommend using it to improve convergence of the model. The `ml maximize` command performs a series of iterations to maximize the likelihood function (the default Newton-Raphson algorithm is used) and generates the output below:

```
. ml maximize, difficult

initial:       log likelihood = -1917.0583
rescale:       log likelihood = -1917.0583
rescale eq:    log likelihood = -1917.0583
Iteration 0:   log likelihood = -1917.0583  (not concave)
Iteration 1:   log likelihood = -1884.9709
Iteration 2:   log likelihood = -1876.8977
Iteration 3:   log likelihood = -1876.8398
Iteration 4:   log likelihood = -1876.8398

                                        Number of obs   =        500
                                        Wald chi2(3)    =      12.00
Log likelihood = -1876.8398             Prob > chi2     =     0.0074
```

```
-------------------------------------------------------------------------------
            |      Coef.    Std. Err.      z    P>|z|     [95% Conf. Interval]
------------+------------------------------------------------------------------
cont        |
       y_1  |   8.12e-06    7.65e-06     1.06   0.289    -6.88e-06     .0000231
        x1  |  -.0889795     .035895    -2.48   0.013    -.1593323    -.0186267
        x2  |  -.6012197    .3017367    -1.99   0.046    -1.192613    -.0098267
      _cons |   8.663247    1.110148     7.80   0.000     6.487396      10.8391
------------+------------------------------------------------------------------
binary      |
       y_1  |   4.93e-06    9.06e-06     0.54   0.586    -.0000128     .0000227
        x1  |  -.0700666    .0226504    -3.09   0.002    -.1144605    -.0256727
        x2  |  -.0176199    .1807232    -0.10   0.922    -.3718309     .3365911
      _cons |   2.180806    .6989016     3.12   0.002     .8109843     3.550628
------------+------------------------------------------------------------------
logvar      |
      _cons |   1.679131    .0882371    19.03   0.000      1.50619     1.852073
-------------------------------------------------------------------------------
```

The `ml maximize` standard output displays the log likelihood value (-1876.84) that can be used for model comparison. This value can be used to get -2 log likelihood and the other information criteria such as BIC. The first equation (cont) represents the model parameters corresponding to the continuous part of the MTP lognormal model, the second equation (binary) represents the model parameters for the binary part of the MTP lognormal, and the third equation (logvar) represents the nuisance parameter, which in this situation is homogenous across all observations (the regression model is just the intercept model for `sec_3`). Note that the model parameter estimates are the same as those from the **SAS** model 1 and they have the same interpretation. The only difference is in the estimation of the nuisance parameter: the **SAS** macro estimates the variance (`sigma2` $= 5.361$) whereas the **Stata** program estimates the log of the variance (log (`sigma2`) $= 1.679$). A good convergence is suggested by a small number of iterations (12 or less) with a concave trajectory with large changes in the log likelihood between initial iterations and smaller changes between the final iterations. The warning "not concave" for iteration one suggests that **Stata** was not able to find a direction vector to update the parameter estimates. However, the warning appears early in the iterations and therefore can be ignored. If this message appeared for final iterations, the convergence would have been questionable. One way to address this is to use the `ml maximize` with the option `difficult`, as presented above, which could help determine a direction vector for difficult log likelihood functions but may increase the run time (Steenbergen 2012).

The `ml graph` displays the iteration history on a graph. This can be used to monitor the convergence of the optimization algorithm. For our simulated example, the convergence was attained in 4 steps and the trajectory is concave (Figure 2). The program for this example ran in less than 1 second.

Stata model 2

The **Stata** model 1 can be extended to account for heterogeneous variances across observations.
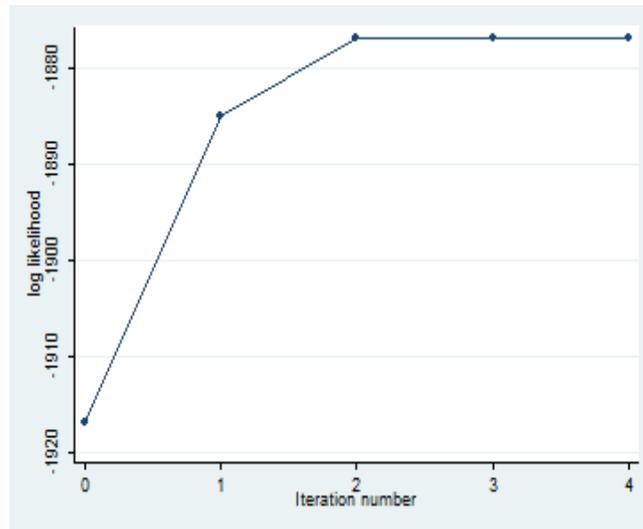
Figure 2: Graph of iterations to convergence from `MTPmle` `Stata` program for `Stata` model 1 (MTP ln ANCOVA type model).

For example, it is feasible to assume that the variability in the outcome may vary across treatment groups. To address this issue, we allow for the variance of the lognormal distribution to depend on covariate information and in this way allow for subject level heteroscedasticity. Similarly, the scale (and/or shape) parameter for Weibull and gamma, generalized gamma or log skew normal can depend on covariates (Liu, Strawderman, Johnson, and O'Quigley 2012). The syntax for the heteroscedastic MTP lognormal model is presented below:

```
. ml model lf mtp_lognormal (cont: y_3 = y_1 x1 x2 ) ///
> (binary: y_3 = y_1 x1 x2)(logvar: y_3 = x1 x2)
. ml search
. ml maximize, difficult
. ml graph
```

The syntax of this model is very similar to `Stata` model 1. The only difference is in the third equation (`logvar:  y_3 = x1 x2`) for the variance parameter of the lognormal which is allowed to depend on the treatment group `x2`, and covariate `x1`. `Stata` output is presented below:

```
. ml model lf mtp_lognormal (cont: y_3=y_1 x1 x2 ) ///
> (binary: y_3=y_1 x1 x2)(logvar: y_3= x1 x2)
. ml search
. ml maximize, difficult
. ml graph
```

The syntax of this model is very similar to `Stata` model 1. The only difference is in the third equation (`logvar:  y_3= x1 x2`) for the variance parameter of the lognormal which is allowed to depend on the treatment group `x2`, and covariate `x1`. `Stata` output is presented below:

```
. ml maximize, difficult

initial:        log likelihood = -1917.0583
rescale:        log likelihood = -1917.0583
rescale eq:     log likelihood = -1917.0583
Iteration 0:    log likelihood = -1917.0583  (not concave)
Iteration 1:    log likelihood = -1883.7531
Iteration 2:    log likelihood = -1876.5294
Iteration 3:    log likelihood = -1876.3917
Iteration 4:    log likelihood = -1876.3911
Iteration 5:    log likelihood = -1876.3911

                                       Number of obs   =         500
                                       Wald chi2(3)    =        6.46
Log likelihood = -1876.3911            Prob > chi2     =      0.0914
```

	Coef.	Std. Err.	z	P>|z|	[95% Conf. Interval]	
cont						
y_1	7.46e-06	7.27e-06	1.03	0.305	-6.79e-06	.0000217
x1	-.0809669	.0516609	-1.57	0.117	-.1822205	.0202867
x2	-1.012837	.5648123	-1.79	0.073	-2.119849	.094175
_cons	8.628227	1.617748	5.33	0.000	5.457499	11.79895
binary						
y_1	4.99e-06	9.13e-06	0.55	0.585	-.0000129	.0000229
x1	-.0702338	.0226786	-3.10	0.002	-.114683	-.0257845
x2	-.0170894	.1807306	-0.09	0.925	-.3713148	.3371361
_cons	2.185556	.6996916	3.12	0.002	.8141851	3.556926
logvar						
x1	.0053298	.017659	0.30	0.763	-.0292812	.0399408
x2	-.1573651	.1779293	-0.88	0.376	-.5061001	.1913699
_cons	1.594785	.5447993	2.93	0.003	.5269984	2.662573

The **Stata** output suggests that that the variability, on the log scale, is increasing with higher levels of covariate **x1** (0.005 units increase for each unit increase in **x1**) and is lower in the treatment group comparative to the usual care group (-0.16 units less). However, these results are not statistically significant and the log likelihood (-1876.38) is not significantly improved comparative to the homogenous MTP lognormal model. Thus, in a real data application similar to this simulated example, we would prefer the homogenous MTP model over heterogeneous MTP. To run the above **Stata** model it took less than one second.

Stata model 3

The proposed **Stata** programs can be easily extended to account for a complex survey design.

We use national data from MEPS household component 2011 (HC-147) to determine if there are any differences in emergency room (ER) expenditures in 2011 (ERTEXP11), between the adult (≥ 18 years) US noninstitutionalized non-Hispanic white and non-Hispanic Asian civilians, after adjusting for confounders and without conditioning on observing a non-zero expense. The syntax for an MTP log skew normal model that accounts for the complex survey design is given below:

```
. svyset [pweight = perwt11f], strata(varstr) psu(varpsu)
. ml model lf mtp_logskewnormal (cont: ertexp11 = asian age11x ///
> yes_married not_married insured sex lowinc middleinc highinc lesshs hs ///
> college northeast midwest south depression hypert jpain strk emph chol ///
> arth asta cvd diab) (binary: ertexp11 = asian age11x yes_married ///
> not_married insured sex lowinc middleinc highinc lesshs hs college ///
> northeast midwest south depression hypert jpain strk emph chol ///
> arth asta cvd diab) (scale:) (shape:), svy
. ml search
. ml maximize, difficult
```

The `svyset` command identifies the components of the complex survey design: the sampling weights (`perwt11f`), the stratification (`varstr`) and cluster (`varpsu`) variables. The same covariates are included in both parts of the models, based on clinical relevance (Bishu, Gebregziabher, Dismuke, and Egede 2015). The main covariate (`asian`) is a dichotomous variables that identifies non-Hispanic whites (`asian = 0`) and non-Hispanic Asians (`asian = 1`). The analysis is adjusted for important confounders, such as age, gender, marital status, education, income as percent of poverty line, insured, region, depression, cardio vascular disease, (any of myocardial infraction, angina, coronary heart disease, and other heart disease), and comorbidities such as hypertension, joint pain, stroke, emphysema, high cholesterol, arthritis, asthma, and diabetes. Of note is that dummy variables were created for each categorical covariate so that the data set can also be analyzed with the SAS macro. However, this is not necessary in Stata (one can use the `i.categorcal_covariate` to indicate a categorical covariate). Partial Stata output after running `ml maximize` is presented below:

```
Number of strata    =       165          Number of obs    =        11435
Number of PSUs      =       365          Population size   =    152242200
                                         Design df         =          200
                                         F(  25,    176)   =        10.69
                                         Prob > F          =       0.0000
```

	Coef.	Linearized Std. Err.	t	P>|t|	[95% Conf. Interval]	
cont						
asian	-1.021055	.1818156	-5.62	0.000	-1.379576	-.6625334
age11x	-.0166012	.0032848	-5.05	0.000	-.0230785	-.0101239
yes_married	-.062507	.1258404	-0.50	0.620	-.3106512	.1856371
not_married	-.090371	.1720982	-0.53	0.600	-.4297308	.2489888

```
     insured |   .5128515    .1555027     3.30    0.001      .2062164     .8194866
         sex |  -.079255     .102883     -0.77    0.442     -.2821295     .1236195
       lowinc |   .0304053    .1517947     0.20    0.841     -.2689181     .3297286
    middleinc |  -.0356091    .1314086    -0.27    0.787     -.2947333      .223515
      highinc |   .0656249    .1563192     0.42    0.675     -.2426204     .3738702
       lesshs |   .0650234    .1818582     0.36    0.721      -.293582     .4236288
          hs |   .0823075    .1695638     0.49    0.628     -.2520547     .4166698
      college |  -.042609     .1677959    -0.25    0.800      -.373485     .2882671
    northeast |  -.1738288     .154799    -1.12    0.263     -.4790763     .1314186
      midwest |   .0873369    .1516209     0.58    0.565     -.2116439     .3863176
        south |  -.0450892    .1435097    -0.31    0.754     -.3280754     .2378971
   depression |   .5103851    .1173817     4.35    0.000      .2789205     .7418496
       hypert |   .3231374    .1270143     2.54    0.012      .0726784     .5735964
        jpain |   .1238056    .1028099     1.20    0.230     -.0789249     .3265362
         strk |   .5937053    .1997078     2.97    0.003      .1999022     .9875084
         emph |   .2806056    .1992571     1.41    0.161     -.1123087     .6735199
         chol |  -.1555047     .119409    -1.30    0.194     -.3909668     .0799575
         arth |  -.0065426    .1333081    -0.05    0.961     -.2694124     .2563272
         asta |   .3231915    .1482265     2.18    0.030      .0309042     .6154788
          cvd |   .5865401    .1267053     4.63    0.000      .3366904     .8363898
         diab |   .4277702    .1406552     3.04    0.003      .1504128     .7051276
        _cons |   5.346928    .3195901    16.73    0.000       4.71673     5.977127
------------+-----------------------------------------------------------------
```

The results show that there is a significant difference in total ER expenditures for 2011 between whites and Asians living in US. More specific, after we exponentiate, the total ER cost for 2011 was 64% (RR = 0.36, 95% CI: (0.25, 0.52)) lower for non-Hispanic Asians comparative to non-Hispanic whites living in US. Similar interpretation can be given to other model coefficients. To run the above model it took 25 seconds.

7. Conclusions

The proposed `%MTPmle` SAS macro and Stata programs facilitate the implementation and comparison of different MTP models in terms of model fit as well as differences in parameter estimates and predicted values, and therefore helping researchers make a more informed decision on the final model choice. As a general rule, we recommend less complex models (such as lognormal) for smaller data sets with less than 500 observations and a more complex model (such as generalized gamma) for larger data sets. In simulations, we showed that in general, these models result in precise estimates with low bias, increased power and expected type 1 error rate, regardless of the true distribution of the data (Voronca *et al.* 2015). In specific situations depending on the distribution and sample size of the real data other models may be preferred as indicated by AIC/BIC and/or the convergence status. The use of MTP versus TP models depends on the aim of the research. If the goal is to compare the populations of subjects with zero versus non-zero outcomes and then get final inference conditional on the subjects having a non-zero outcome, the TP models are suitable. However, if final inference is needed on the whole population, including subjects with zero as well as non-zero values, then an MTP model should be used.

Functionality	`%MTPmle` (SAS)	`MTPmle` (Stata)
ML estimation	Yes	Yes
MTP models (logistic regression is used for zero versus non-zero; the distribution used for the continuous part can be lognormal, gamma, Weibull, generalized gamma, or log skew normal)	Yes	Yes
TP models (logistic regression is used for zero versus non-zero; the distribution used for the continuous part can be lognormal, gamma, Weibull, generalized gamma, or log skew-normal)	Yes	No
Random intercept from a normal (multivariate normal) distribution in the binary and/or continuous part of the MTP/TP model	Yes	No
Heteroscedastic models	No	Yes
Automated search algorithm for best initial values and a graph of iteration history to convergence	No (see `grid` and `initparms` options)	Yes
Graph of observed versus predicted values	Yes	No
MTP models for complex survey design	No	Yes

Table 7: Comparison of `MTPmle` SAS macro and `MTPmle` Stata programs.

A comparison of the functionality of the `%MTPmle` SAS macro versus the `MTPmle` Stata programs is presented in Table 7. Simple cross-sectional MTP models can be implemented by using either the proposed `%MTPmle` SAS macro or the `MTPmle` Stata programs. Both software extensions have the same types of MTP models available in terms of the distribution used in the continuous part (lognormal, Weibull, Gamma, generalized gamma, log skew normal).

Longitudinal/clustered MTP models can be implemented with the SAS macro. However, the macro can fit only random intercepts that follow a normal distribution (multivariate normal). Currently, `SAS PROC NLMIXED` does not allow for other types of distributions for the random effects and the convergence is difficult to reach when more than two random effects are included in the model. If the researcher is interested in models with more than two random effects, then other SAS procedures are more efficient in terms of convergence such as `SAS PROC MCMC` (Smith *et al.* 2017). In the proposed SAS macro, the initial values can be changed (use `initparms` or `grid` options) when convergence is not attained with default values. The run time can be longer for the SAS macro especially when using the grid search on a wide range with a small incremental step. The run time is significantly higher when random intercepts are considered in the model.

Heteroscedastic models can be implemented with the proposed Stata programs. In addition, the Stata programs can perform an automated search algorithm for best initial values that may speed up convergence and save time when running different models. A suggestive graph of the iteration history is also available. A variety of other options and commands that can be used with the proposed Stata programs can be found on the Stata website. For example,

the Stata programs can be easily extended to account for a complex survey design by using the svy option. This emphasizes the flexibility and ease of model implementation in Stata. The SAS macro is a more rigid approach and any new features must be hard coded in the body of the macro.

In summary, the advantages of using the proposed Stata programs are automated search for initial parameter values, condensed output, suggestive graphs to monitor convergence and flexible syntax that can be easily extended to fit heteroscedastic models ad complex survey designs, whereas the advantages of SAS macro are the option to fit MTP random intercept models, the option to fit the corresponding two-part model, and suggestive graphs of observed versus predicted values. In the future, it would be useful to extend the Stata capabilities to fit MTP models with random effects. Moreover, development of similar macros or functions for other popular statistical software such as SPSS (IBM Corporation 2017) and R (R Core Team 2018) would further encourage and facilitated the use of MTP models.

References

Akaike H (1974). "A New Look at the Statistical Model Identification." *IEEE Transactions on Automatic Control*, **19**(6), 716–723. doi:10.1109/tac.1974.1100705.

Azzalini A (1985). "A Class of Distributions Which Includes the Normal Ones." *Scandinavian Journal of Statistics*, **12**(2), 171–178.

Bishu KG, Gebregziabher M, Dismuke CE, Egede LE (2015). "Quantifying the Incremental and Aggregate Cost of Missed Workdays in Adults with Diabetes." *Journal of General Internal Medicine*, **30**(12), 1773–1779. doi:10.1007/s11606-015-3338-y.

Cragg JG (1971). "Some Statistical Models for Limited Dependent Variables with Application to the Demand for Durable Goods." *Econometrica*, **39**(5), 829–844. doi:10.2307/1909582.

Duan N, Manning WG, Morris CN, Newhouse JP (1983). "A Comparison of Alternative Models for the Demand for Medical Care." *Journal of Business and Economic Statistics*, **1**(2), 115–126. doi:10.2307/1391852.

Gould W, Pitblado J, Sribney W (2006). *Maximum Likelihood Estimation with Stata*. Stata Press. ISBN 1597180122.

IBM Corporation (2017). *IBM SPSS Statistics 25*. IBM Corporation, Armonk. URL http://www.ibm.com/software/analytics/spss/.

Liu L, Strawderman RL, Johnson BA, O'Quigley JM (2012). "Analyzing Repeated Measures Semi-Continuous Data, with Application to an Alcohol Dependence Study." *Statistical Methods in Medical Research*, **25**(1), 133–152. doi:10.1177/0962280212443324.

Manning WG, Basu A, Mullahy J (2005). "Generalized Modeling Approaches to Risk Adjustment of Skewed Outcomes Data." *Journal of Health Economics*, **24**(3), 465–488. doi:10.1016/j.jhealeco.2004.09.011.

Miller GA, Chapman JP (2001). "Misunderstanding Analysis of Covariance." *Journal of Abnormal Psychology*, **110**(1), 40–48. doi:10.1037//0021-843x.110.1.40.

Mullahy J (1986). "Specification and Testing of Some Modified Count Data Models." *Journal of Econometrics*, **33**(3), 341–365. doi:10.1016/0304-4076(86)90002-3.

Olsen MK, Schafer JL (2001). "A Two-Part Random-Effects Model for Semicontinuous Longitudinal Data." *Journal of the American Statistical Association*, **96**(454), 730–745. doi:10.1198/016214501753168389.

Pinheiro JC, Bates DM (1995). "Approximations to the Log-Likelihood Function in the Nonlinear Mixed-Effects Model." *Journal of Computational and Graphical Statistics*, **4**(1), 12–35. doi:10.2307/1390625.

R Core Team (2018). *R: A Language and Environment for Statistical Computing*. R Foundation for Statistical Computing, Vienna, Austria. URL https://www.R-project.org/.

SAS Institute Inc (2013). *SAS/STAT Software, Version 9.4*. Cary. URL http://www.sas.com/.

Schafer JL, Olsen MK, *et al.* (1999). "Modeling and Imputation of Semicontinuous Survey Variables." In *Proceedings of the Federal Committee on Statistical Methodology Research Conference*, pp. 565–74.

Smith VA, Neelon B, Preisser JS, Maciejewski ML (2017). "A Marginalized Two-Part Model for Longitudinal Semicontinuous Data." *Statistical Methods in Medical Research*, **26**(4), 1949–1968. doi:10.1177/0962280215592908.

Smith VA, Preisser JS, Neelon B, Maciejewski ML (2014). "A Marginalized Two-Part Model for Semicontinuous Data." *Statistics in Medicine*, **33**(28), 4891–4903. doi:10.1002/sim.6263.

StataCorp (2015a). *Stata 14 Base Reference Manual*. Stata Press, College Station. URL https://www.stata.com/.

StataCorp (2015b). *Stata Data Analysis Statistical Software: Release 14*. StataCorp LP, College Station. URL http://www.stata.com/.

Steenbergen MR (2012). "A Primer of Maximum Likelihood Programming in Stata." Technical manual.

Vickers AJ (2001). "The Use of Percentage Change from Baseline as an Outcome in a Controlled Trial Is Statistically Inefficient: A Simulation Study." *BMC Medical Research Methodology*, **1**, 6. doi:10.1186/1471-2288-1-6.

Voronca DC, Gebregziabher M, Durkalski VL, Liu L, Egede LE (2015). "Marginalized Two Part Models for Generalized Gamma Family of Distributions." arXiv:1511.05629 [stat.ME], URL https://arxiv.org/abs/1511.05629.

Wolfinger RD (1999). "Fitting Nonlinear Mixed Models with the New NLMIXED Procedure." In *Proceedings of the 24th Annual SAS Users Group International Conference (SUGI 24)*, pp. 278–284.

Affiliation:

Delia C. Voronca
Emmes Corporation
401 North Washington Street
Suite 700
Rockville, MD 20850, United States of America
E-mail: dvoronca@emmes.com