

Growth mixture modeling with non-normal distributions

Bengt Muthén*[†] and Tihomir Asparouhov

A limiting feature of previous work on growth mixture modeling is the assumption of normally distributed variables within each latent class. With strongly non-normal outcomes, this means that several latent classes are required to capture the observed variable distributions. Being able to relax the assumption of within-class normality has the advantage that a non-normal observed distribution does not necessitate using more than one class to fit the distribution. It is valuable to add parameters representing the skewness and the thickness of the tails. A new growth mixture model of this kind is proposed drawing on recent work in a series of papers using the skew-*t* distribution. The new method is illustrated using the longitudinal development of body mass index in two data sets. The first data set is from the National Longitudinal Survey of Youth covering ages 12–23 years. Here, the development is related to an antecedent measuring socioeconomic background. The second data set is from the Framingham Heart Study covering ages 25–65 years. Here, the development is related to the concurrent event of treatment for hypertension using a joint growth mixture-survival model. Copyright © 2014 John Wiley & Sons, Ltd.

Keywords: trajectory classes; skew-*t* distribution; body mass index; survival

1. Introduction

Growth mixture modeling combines the conventional Laird and Ware [1] random effects modeling with latent trajectory classes as in finite mixture modeling; see, for example, [2]. Growth mixture modeling was introduced in Verbeke and LeSaffre [3] and Muthén and Shedden [4] with related developments in Nagin and Land [5] and Roeder *et al.* [6]. Following this, many extensions and applications have been presented such as Lin *et al.* [7] considering prostate-specific antigen (PSA) biomarker trajectories with irregularly scheduled observations, Lin *et al.* [8] adding joint estimation of survival with prostate cancer, Muthén and Brown [9] considering causal inference in randomized trials of antidepressants with placebo effects, Muthén and Asparouhov [10] adding general multilevel growth mixture modeling, and Muthén *et al.* [11] modeling non-ignorable dropout in antidepressant trials. For overviews of methods with illustrations by a variety of applications, see [10, 12].

A limiting feature of the aforementioned approaches is the assumption of normally distributed variables within each latent class. With strongly non-normal outcomes, this means that several latent classes are required to capture the observed variable distributions. Consider a typical example involving body mass index (BMI) development over age. BMI is defined as kg/m^2 , where the normal range is $18 < \text{BMI} < 25$, overweight $25 < \text{BMI} < 30$, and obese > 30 . The distribution of BMI at age 15 years for men is given in Figure 1 using data from the National Longitudinal Survey of Youth (NLSY) with $n = 3194$ showing skewness of 1.5 and kurtosis of 3.1. The figure also shows the fitting of a mixture of normal distributions. The left part of Table I shows the loglikelihood and BIC values for 1–4 classes using a normal distribution. Although the four-class solution has a smaller (better) BIC than three classes, one class has less than 1%, and a three-class solution is therefore chosen. The mixture of the three classes is shown by the black curve in Figure 1 and is seen to fit the observed distribution well.

Being able to relax the assumption of within-class normality has the advantage that a non-normal observed distribution does not require using more than one class to fit the distribution. For example, with

Mplus, 3463 Stoner Ave, Los Angeles, CA 90066, U.S.A.

*Correspondence to: Bengt Muthén, 3463 Stoner Ave, Los Angeles, CA 90066, U.S.A.

[†]E-mail: bmuthen@statmodel.com

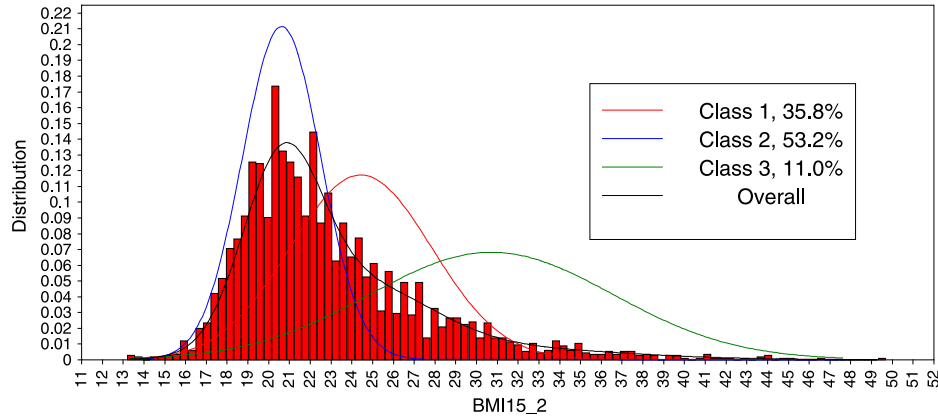


Figure 1. Observed and three-class normal fitted distribution of body mass index among 15-year-old men in the National Longitudinal Survey of Youth.

Table I. Results of fitting mixtures of normal and non-normals for BMI among 15-year old men in the NLSY.

No. classes	Normal			Skew-t		
	Loglikelihood	No. Par's	BIC	Loglikelihood	No. Par's	BIC
1	-9321	2	18,658	-8795	4	17,623
2	-8828	5	17,697	-8783	9	17,638
3	-8786	8	17,638			
4	-8774	11	17,637			

BMI, body mass index; NLSY, National Longitudinal Survey of Youth.

Table II. NLSY data on BMI for ages 12–23 years.

Accelerated longitudinal design - NLSY97												
	12	13	14	15	16	17	18	19	20	21	22	23
1997	1165	1715	1847	1868	1709	613						
1998		104	1592	1671	1727	1739	1400	106				
1999			108	1659	1625	1721	1614	1370	65			
2000				57	1553	1656	1649	1597	1390	132		
2001					66	1543	1615	1602	1582	1324	109	
2002							1614	1587	1643	1582	1324	106
2003							112	1497	1600	1582	1564	1283
Totals	1165	1819	3547	5255	6680	7272	8004	7759	6280	4620	2997	1389

NLSY, National Longitudinal Survey of Youth; BMI, body mass index.

Source: Nonnemaker *et al.* (2009). Youth BMI trajectories: Evidence from the NLSY97, obesity

a strongly skewed distribution, it is generally not of interest to interpret classes that are formed simply to match the long tail. This relates to the classic debate of whether or not the classes in Figure 1 have substantive meaning or should merely be seen as a curve-fitting device, which started with Pearson [13], continued with the Platt–Pickering hypertension debate in the 60s (McLachlan & Peel, pp. 14–17 [2]), and is still a topic of interest; see, for example, [14, 15]. From this perspective, it is valuable to have the option of fitting a model that allows within-class non-normality, adding parameters representing the skewness and the thickness of the tails. The results of fitting such a model to the BMI data are shown in the right part of Table I, where it is seen that a single-class model obtains the best BIC. Although having a somewhat lower loglikelihood value than the three-class normal model, this single-class model is more parsimonious using only four parameters instead of eight (a three-class normal model with equal variances has a worse BIC): a mean, a variance, a skew parameter, and a degrees of freedom parameter. This model is discussed in Section 2. The model draws on recent work in a series of papers that use the skew-t distribution for ‘mixtures of factor analyzers’, that is, mixture models where an exploratory factor

analysis model is applied to reduce the number of parameters in the class-specific covariance matrices; see, for example, Lin *et al.* [16] and [17]. In contrast, the skew-t growth mixture model proposed here is more closely related to confirmatory factor analysis modeling, where the class-specific covariance matrices are more parsimonious in line with conventional random effects growth modeling.

This paper illustrates the new method using the longitudinal development of BMI using two data sets. First, the NLSY data shown in Table II are used, spanning ages 12–23 years. Here, the BMI development is related to an antecedent measuring socioeconomic background. Second, data from the Framingham Heart Study are used for ages 25–65 years. Here, the BMI development is related to concurrent treatment for hypertension. For these age ranges, a quadratic growth shape has been found suitable, and Section 2 describes such a growth mixture model. The model is subsequently expanded to jointly estimate survival related to the latent trajectory classes.

2. Growth mixture modeling with non-normal random effects

Consider the quadratic random effect growth mixture model with outcome Y_{it} for individual i at time t in latent class c of the latent class variable C ,

$$Y_{it}|_{C_i=c} = \eta_{0i} + \eta_{1i} (a_t - a_0) + \eta_{2i} (a_t - a_0)^2 + \epsilon_{it}, \quad (1)$$

where a_t are age-related time scores ($t = 1, 2, \dots, T$) centered at age a_0 , the random intercepts and random slopes are expressed as

$$\eta_{ji}|_{C_i=c} = \alpha_{jc} + \gamma'_{jc} \mathbf{X}_i + \zeta_{ji}, \quad (2)$$

where $j = 0, 1, 2$, \mathbf{X}_i is a q -dimensional vector of time-invariant covariates, and the latent class probability is expressed as the multinomial logistic regression

$$P(C_i = c | \mathbf{X}_i) = \frac{\exp(a_c + \mathbf{b}'_c \mathbf{X}_i)}{\sum_s \exp(a_s + \mathbf{b}'_s \mathbf{X}_i)}. \quad (3)$$

The residuals ϵ and ζ have zero means and within-class covariance matrices to be defined later. So far, the literature on growth mixture modeling has been using a normal within-class distribution for both ϵ and ζ , a specification that has limitations as mentioned in the introduction. For reasons to be described, it is desirable with growth modeling to let the non-normality in the observed outcomes be a function of the non-normality of the random effect distribution. In this paper, a normal distribution is therefore maintained for ϵ , while a skew-t distribution is applied to $\zeta = (\zeta_0 \zeta_1 \zeta_2)'$. Using the notation for the restricted multivariate skew-t distribution given in [17],

$$\zeta_{C_i=c} \sim rMST(\mathbf{0}, \mathbf{\Psi}_c, \boldsymbol{\delta}_c, \nu_c), \quad (4)$$

where in this application $\mathbf{0}$ is the 3×1 vector of zero means for ζ , $\mathbf{\Psi}_c$ is the 3×3 within-class covariance matrix for ζ , $\boldsymbol{\delta}_c$ is the 3×1 vector of skew parameters for latent class c , and ν_c is a degree of freedom parameter. The restricted multivariate skew-t distribution can be characterized as follows.

Considering a p -dimensional vector \mathbf{V} , the restricted multivariate skew t-distribution $rMST(\boldsymbol{\mu}, \boldsymbol{\Sigma}, \boldsymbol{\delta}, \nu)$ has the stochastic representation

$$\mathbf{V} = \boldsymbol{\mu} + \boldsymbol{\delta} |U_0| + \mathbf{U}_1, \quad (5)$$

where \mathbf{U}_1 is p -dimensional vector with a multivariate t-distribution with zero mean, covariance matrix $\boldsymbol{\Sigma}$, and degree of freedom parameter ν . Here, U_0 is a one-dimensional variable with a standard t-distribution with mean 0, variance parameter 1, and degrees of freedom parameter ν , where $|U_0|$ gives rise to a half-t distribution. The term $\boldsymbol{\delta} |U_0|$ can be thought of as a univariate skewness factor with factor loadings represented by the skew parameters of $\boldsymbol{\delta}$ where the skewness is identified as that part of the \mathbf{V} distribution not captured by the symmetric part \mathbf{U}_1 . The mean and variance of \mathbf{V} for the skew t-distribution $rMST(\boldsymbol{\mu}, \boldsymbol{\Sigma}, \boldsymbol{\delta}, \nu)$ can be computed as follows:

$$E(\mathbf{V}) = \boldsymbol{\mu} + \boldsymbol{\delta} \frac{\Gamma\left(\frac{\nu-1}{2}\right)}{\Gamma\left(\frac{\nu}{2}\right)} \sqrt{\frac{\nu}{\pi}}, \quad (6)$$

$$\text{Var}(V) = \frac{\nu}{\nu - 2}(\Sigma + \delta\delta') - \left(\frac{\Gamma\left(\frac{\nu-1}{2}\right)}{\Gamma\left(\frac{\nu}{2}\right)} \right)^2 \frac{\nu}{\pi} \delta\delta'. \quad (7)$$

The univariate skewness for a single V variable can be computed as follows:

$$\text{Skew}(V) = \nu^{-3/2} \delta \sqrt{\frac{\nu}{\pi}} \left((2\delta^2 + 3\sigma) \frac{\nu}{\nu - 2} \frac{\Gamma\left(\frac{\nu-3}{2}\right)}{\Gamma\left(\frac{\nu-2}{2}\right)} - \delta^2 \frac{\nu}{\pi} \left(\frac{\Gamma\left(\frac{\nu-1}{2}\right)}{\Gamma\left(\frac{\nu}{2}\right)} \right)^3 - 3 \frac{\Gamma\left(\frac{\nu-1}{2}\right)}{\Gamma\left(\frac{\nu}{2}\right)} \nu \right), \quad (8)$$

where $\nu = \text{Var}(V)$ is given in the previous formula and the σ parameter is the diagonal element of Σ corresponding to the univariate variable. These formulas show that the δ and ν parameters affect all three quantities: the mean, the variance, and the skew. The parameter μ affects only the mean, and the σ parameters affect the variance covariance and the skew.

It can be shown (see Asparouhov & Muthén [18]) that this specification gives the distribution of the vector of observations for individual i at the T time points $Y_i = (Y_{i1}, \dots, Y_{it}, \dots, Y_{iT})'$ conditional on X and latent class c

$$Y|X \sim rMST(\mu_c, \Sigma_c, \delta_{Yc}, \nu_c), \quad (9)$$

where

$$\mu_c = \Lambda (\alpha_c + \Gamma_c X), \quad (10)$$

$$\Sigma_c = \Lambda \Psi_c \Lambda' + \Theta_c, \quad (11)$$

$$\delta_{Yc} = \Lambda \delta_c, \quad (12)$$

where for the quadratic growth model of (1)

$$\Lambda = \begin{pmatrix} 1 & a_1 - a_0 & (a_1 - a_0)^2 \\ 1 & a_2 - a_0 & (a_2 - a_0)^2 \\ \vdots & \vdots & \vdots \\ 1 & a_{T-1} - a_0 & (a_{T-1} - a_0)^2 \end{pmatrix}, \quad (13)$$

the elements of α_c and Γ_c are shown in (2) and Θ_c is the within-class covariance matrix for $\epsilon = (\epsilon_1, \epsilon_2, \dots, \epsilon_T)'$, typically specified as diagonal. This is referred to as a wide format, single-level approach to growth modeling, where elements of Λ can also be estimated to provide a flexible growth function. With this skew-t growth mixture model specification, the non-normality of the outcomes Y is generated by the non-normality of the random effects. The outcome means are a function of the means of the random effects, which as shown in (6) involves the skew and degrees of freedom parameters.

An alternative specification of a non-normal growth mixture model is to let ϵ assume a skew-t distribution while keeping ζ normal (skew-t for both is not identified). This type of model was recently proposed in Lu and Huang [19]. In this case, however, the skew parameters of δ would have to be held equal across time because otherwise, the means for Y would not follow the structure imposed by the random effect means but vary also as a function of the skew and degrees of freedom parameters for ϵ . This time invariance of the skewness is specified in the analysis of Lu and Huang [19]. Time-invariant skewness in the residuals is a special case of the model proposed earlier as it is the same as applying the skewness only to the random intercept. In the BMI data to be analyzed here, however, it is necessary to allow skewness also in the random slopes to capture increasing skewness in BMI over time.

The skew-t distribution encompasses several special cases. Fixing the δ skewness parameters to zero gives the multivariate t-distribution, fixing the ν degrees of freedom parameters at a large value such as 10,000 gives a skew-normal distribution, see [17], fixing the skewness parameters to zero and the degrees of freedom parameters to a large value gives the normal distribution.

A survival component can be added to the growth mixture model as follows. Using the standard Cox proportional hazard model, the hazard function in class c is given by

$$h_c(t) = h(t)Exp(\alpha_c), \quad (14)$$

where $h(t)$ is the baseline hazard function and $Exp(\alpha_c)$ represents the level of proportionality for the hazard functions between the classes and also captures the effect of the latent class variable C on the survival variable. The larger the coefficient α_c is, the larger the hazard and the worse the survival is. The baseline hazard function $h(t)$ is a non-parametric function as in the standard Cox regression model. The baseline hazard function is invariant across the classes and thus the class effect is captured entirely by the coefficients α_c . For identification purposes, $\alpha_c = 0$ in the last class.

2.1. Estimation

The models are estimated by maximum likelihood. Consider the estimation for a single class model. The log-likelihood can be written explicitly and maximized directly. Using equation (9), the likelihood for a particular observation is given by

$$L = 2t_{p,v}(\mathbf{Y}, \boldsymbol{\mu}, \boldsymbol{\Omega})T_1(y_1/\lambda, \nu + p), \quad (15)$$

where

$$\boldsymbol{\Omega} = \boldsymbol{\Sigma} + \boldsymbol{\delta}\boldsymbol{\delta}^T, \quad (16)$$

$$d(\mathbf{Y}) = (\mathbf{Y} - \boldsymbol{\mu})^T \boldsymbol{\Omega}^{-1} (\mathbf{Y} - \boldsymbol{\mu}), \quad (17)$$

$$q = \boldsymbol{\delta}^T \boldsymbol{\Omega}^{-1} (\mathbf{Y} - \boldsymbol{\mu}), \quad (18)$$

$$y_1 = q \sqrt{\frac{\nu + p}{\nu + d(\mathbf{Y})}}, \quad (19)$$

$$\lambda^2 = 1 - \boldsymbol{\delta}^T \boldsymbol{\Omega}^{-1} \boldsymbol{\delta}, \quad (20)$$

and $t_{p,v}(\mathbf{Y}, \boldsymbol{\mu}, \boldsymbol{\Omega})$ is the multivariate t-distribution density function given by

$$t_{p,v}(\mathbf{Y}, \boldsymbol{\mu}, \boldsymbol{\Omega}) = \frac{\Gamma\left(\frac{\nu+p}{2}\right) |\boldsymbol{\Omega}|^{-1}}{(\pi\nu)^{p/2} \Gamma\left(\frac{\nu}{2}\right) [1 + d(\mathbf{Y})/\nu]^{(\nu+p)/2}} \quad (21)$$

and $T_1(z, \nu)$ is the standard univariate t-distribution function with ν degrees of freedom.

The derivatives of the log-likelihood can be computed and thus the log-likelihood can be maximized with a general maximization algorithm such as the quasi-Newton optimization method. The only derivative that is more complex is the partial derivative of $T_1(x, \nu)$ with respect to ν . For this derivative, the method developed in Boik and Robison–Cox [20] is used. Direct maximization appears to work well and is relatively fast. The standard error (SE) estimates are based on the inverse of the information matrix as usual with the maximum likelihood (ML) estimator and robust SE can also be computed using the sandwich estimator.

For the estimation of the general mixture model, the standard EM-algorithm is used as in [4], which reduces the optimization of the mixture log-likelihood to the maximization of the log-likelihood of the single-class model at each expectation-maximization (EM) iteration.

Adding the survival component to the growth mixture model, the estimation issues are as follows. The cumulative hazard $H_c(t)$ function at time t represents the total hazard an individual is exposed to up to time t , given that the individual is in class c

$$H_c(t) = \int_0^t h_c(x)dx = Exp(\alpha_c)H(t), \quad (22)$$

where $H(t)$ is the cumulative baseline hazard

$$H(t) = \int_0^t h(x)dx. \quad (23)$$

The survival function in class c is the probability that the survival variable T is greater than t , given that the individual belongs to class c

$$S(t|c) = P(T > t|C = c) = \text{Exp}(-H_c(t)) = \text{Exp}(-\text{Exp}(\alpha_c)H(t)). \quad (24)$$

The likelihood for the survival variable T is

$$P(T|C = c) = (h(T)\text{Exp}(\alpha_c))^{(1-\delta)}S(T), \quad (25)$$

where $\delta = 1$ if the variable T is censored and zero otherwise.

The likelihood for the joint model for Y and T can be expressed as

$$P(Y, T|C) = \sum_c P(C = c)P(Y|C = c)P(T|C = c), \quad (26)$$

where $P(Y|C = c)$ is obtained from the growth mixture model. The estimation of this joint model is based on the profile likelihood approach; see [21–23]. With this approach, the baseline hazard is fully saturated, and the baseline hazard parameters are treated as auxiliary parameters. The profile likelihood is treated as a regular likelihood and is explicitly maximized.

All computations in this paper are carried out by Mplus Version 7.2 (Muthén & Muthén [24]). The Mplus software implementation is quite general, also in the sense that every parameter can be fixed, free, held equal to other parameters, or constrained in specific ways.[‡]

3. Growth mixture modeling of body mass index from the National Longitudinal Survey of Youth

This section applies the skew-t quadratic growth mixture model to the BMI development over the ages 12–23 years using the NLSY data shown in Table II. As discussed in, for example, Nonnemaker *et al.* [25], large differences in BMI are observed across gender and ethnicity subgroups, and the current analysis is restricted to black women ($n = 1160$). Several different within-class distributions are used in the modeling: normal, t, skew-normal, skew-t with equal skewnesses, and skew-t with unequal skewnesses. The latter two models differ in letting the skewness parameter be applied to the intercept random effect only or to both intercept and slopes. The parameterization is as follows for the two key distributions, the normal and the skew-t with unequal skews. For each class, the normal model uses three random effect means, six random effects variances/covariances, and one residual variance held equal over time points. With two classes, this results in 21 parameters, including one class probability parameter. The skew-t unequal model uses in addition three random effect skewness parameters for each class and one degrees of freedom parameter for each class. With two classes, this results in 29 parameters, including one class probability parameter.

Table III shows the results of fitting the five different within-class distributions for varying number of latent classes. It is seen from the BIC values that four classes are needed with the normal distribution, three are needed with the t-distribution and the skew-normal distribution, and two are needed with the two skew-t distributions. The BIC for the t-distribution improves on the BIC for the normal. The skew-normal BIC does not improve on the t-distribution, while the BIC of the skew-t distributions do. Among the two skew-t models, allowing for skewness not only for the intercept but also for the two slopes is preferable according to BIC. It is noteworthy that the four-class normal model has considerably more parameters and a worse BIC than both of the two-class skew-t models.

The estimated mean growth curves for the two classes of the preferred skew-t solution are shown at the top of in Figure 2. While starting at the same BMI level at age 12 years, one class of 54% shows a normal

[‡]Complex survey features of stratification, weights, and clustering are also handled in Mplus.

Table III. Results of fitting normal, t, skew-normal, and skew-t growth mixture models for BMI in the NLSY ages 12–23 years for black women.

No. classes	Normal			T		
	Loglikelihood	No. Par's	BIC	Loglikelihood	No. Par's	BIC
1	-17,049	10	34,168	-15,846	11	31,770
2	-15,768	21	31,684	-15,610	23	31,382
3	-15,580	32	31,386	-15,527	35	31,302
4	-15,505	43	31,314	-15,499	47	31,329
5	-15,479	54	31,338	-15,484	59	31,384
	Skew-normal			Skew-t, equal skew		
No. classes	Loglikelihood	No. Par's	BIC	Loglikelihood	No. Par's	BIC
1	-16,874	13	33,840	-15,664	12	31,412
2	-15,708	27	31,606	-15,535	25	31,247
3	-15,608	41	31,506	-15,526	38	31,320
4	-15,751	55	31,889			
	Skew-t, unequal skew					
No. classes	Loglikelihood	No. Par's	BIC			
1	-15,637	14	31,372			
2	-15,510	29	31,225			
3	-15,480	44	31,270			

BMI, body mass index; NLSY, National Longitudinal Survey of Youth.

development, whereas the other class of 46% shows an escalating development into the overweight and obese range. The four-class normal solution at the bottom shows the need for extra classes to capture the strong skewness when within-class normality is specified. The top two classes sum to 40%, which is similar to the 46% of the escalating class of the two-class skew-t solution. Combining the top two and bottom two classes, however, does not produce two classes with the same starting point at age 12 years as is seen for the skew-t solution.

Because of the skewness of the observed BMI distribution, it is of interest to present not only the estimated mean at each age but also different estimated percentiles. Figure 3 uses the percentiles of the estimated skew-t density to show that the skewness increases with age.

Figure 4 shows the estimated random intercept distribution for the normally developing class (see top figure) and the escalating class (bottom figure) of the skew-t solution. Given the choice of time scores, the random intercept corresponds to the systematic part of the development at age 17 years. It is seen that the intercept distribution for the escalating class is characterized as an approximate half-t distribution with a low frequency for BMI values less than 22. Recalling the stochastic representation in (5), this is achieved by the symmetric part of the random intercept distribution having zero variance. In contrast, the normal class has a large number with BMI less than 22, while still showing a long right tail.

The estimated skew and degrees of freedom parameters for the intercept (i), linear slope (s), and quadratic slope (q) are shown in Table IV. It is seen that all three random effects have significant skew parameters in each of the classes. The bottom of the table also gives the estimated skewness in each of the classes for these three random effects using (8).

3.1. Adding a covariate and comparing with the normal solution

To gain further understanding of the growth mixture solution, covariates that predict class membership can be added in line with (3). Such covariates may also have a direct influence on the random effects as in (2). Mother's education is an especially powerful predictor of BMI, presumably reflecting both economic circumstances and eating habits (mother's education is scored as 1: none, 2: GED, 3: high school diploma, 4: associate/junior college, 5: bachelor's degree, 6: master's degree, 7: PhD, 8: professional degree). Adding this covariate to the growth mixture model, two classes are again preferred by BIC. Appendix A shows the Mplus input for this model. It is also of interest to compare the two-class skew-t solution with a normal solution for which four classes are again preferred by BIC.

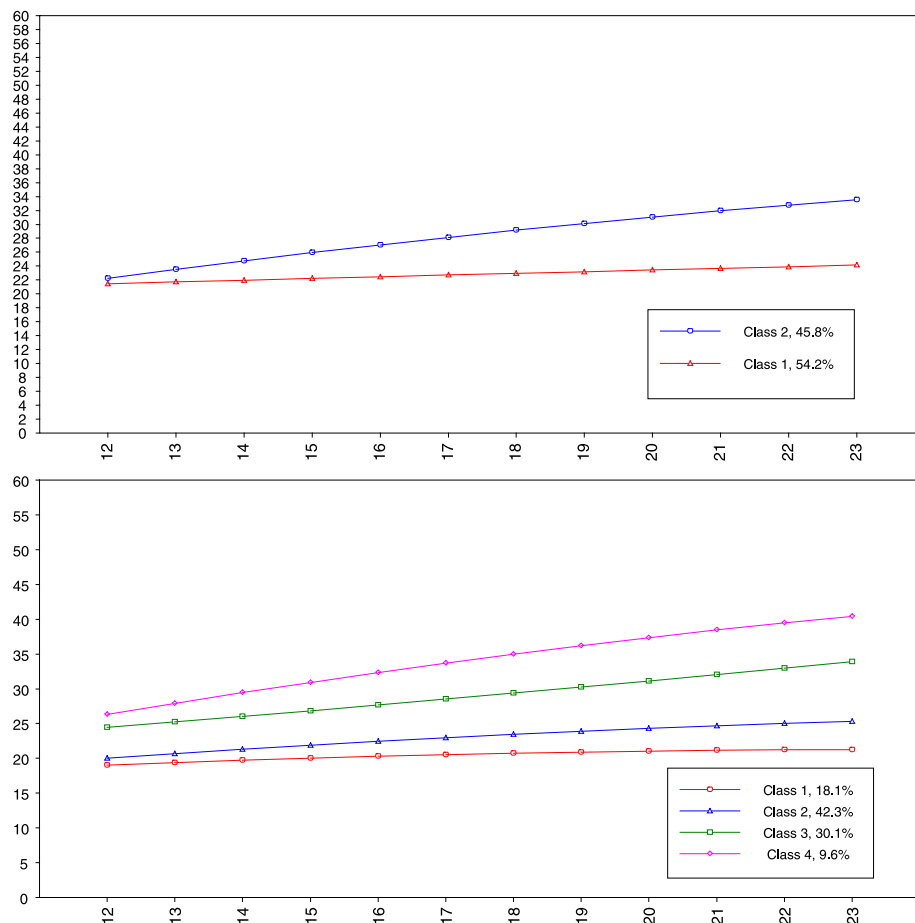


Figure 2. Estimated mean curves for two-class skew-t (top) and four-class normal (bottom) growth mixture modeling of body mass index in the National Longitudinal Survey of Youth ages 12–23 years for black women.

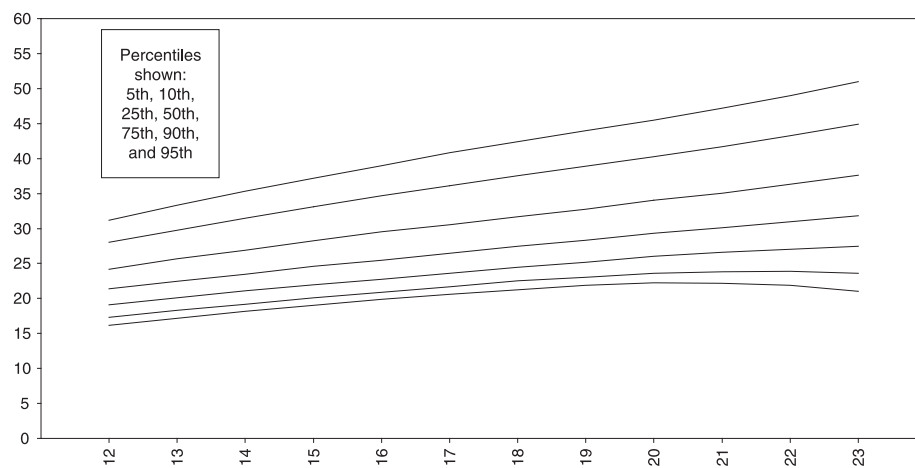


Figure 3. Estimated percentiles for the escalating class of body mass index in the National Longitudinal Survey of Youth ages 12–23 years for black women

Figure 5 shows the estimated multinomial logistic regression curves for the two-class skew-t solution (top) and the four-class normal solution (bottom). The corresponding mean curves for the two solutions were given in Figure 2. For the two-class skew-t solution, the probability of membership in the escalating class is strongly decreased by increasing mother’s education. The four-class normal solution shows two classes giving such a decrease while one class shows no relationship with mother’s education (in a

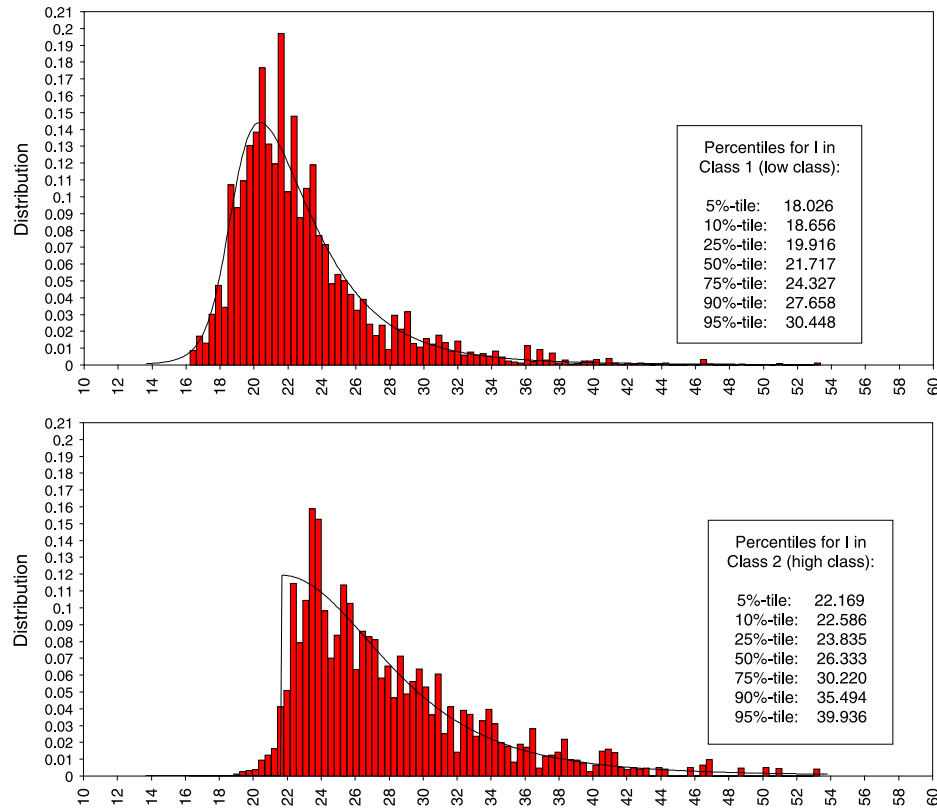


Figure 4. Random effect distributions for the normal and escalating trajectory classes using a skew-t growth mixture model of body mass index in the National Longitudinal Survey of Youth ages 12–23 years for black women.

three-class normal solution, the class with no relationship disappears while the curves for the other three classes remain the same). It is the highest class at the bottom of Figure 2 for which Figure 5 shows the less steep decrease in membership probability with increasing mother’s education. Further studies using additional information would be needed to understand if this class is merely a function of the strong skewness or has a substantive interpretation, for example, in terms of reflecting genetic susceptibility where mother’s education is a weaker predictor. The two-class skew-t solution is, however, clearly the more parsimonious model.

4. Growth mixture modeling of body mass index in the Framingham data

The classic Framingham Heart Study (Dawber *et al.* [26]) provides a second longitudinal data set with which to explore growth mixture modeling of BMI. With its focus on cardiovascular disease, the Framingham data contain information on treatment for hypertension, and it is of interest to relate this to the BMI trajectories. The current analyses use a subset of the data for women ages 25–65 years, resulting in a sample size of $n = 854$. Four repeated measures are available with individually varying ages of observation.

To capture the individually varying ages of observation, (1) in Section 2 model is modified as

$$Y_{it}|_{C_i=c} = \eta_{0i} + \eta_1 (a_{it} - a_0) + \eta_2 (a_{it} - a_0)^2 + \epsilon_{it}, \quad (27)$$

where the random effects specification in (2) is applied to the random intercept η_{0i} , but for simplicity zero, within-class variation is specified for η_1 and η_2 . In this way, the time scores $a_{it} - a_0$ and $(a_{it} - a_0)^2$ are no longer treated as fixed parameters of Λ but as variables.

Table IV. Estimated skew and degrees of freedom parameters for the i , s , and q random effects of the two-class skew- t growth mixture model for BMI in the NLSY ages 12–23 years for black women.

Skew and Df parameters			
Normal class			
	Estimate	SE	Est/SE
i	4.020	0.279	14.408
s	−0.875	0.381	−2.296
q	3.399	1.281	2.653
df	3.855	0.562	6.859
Escalating class			
	Estimate	SE	Est/SE
i	6.236	0.343	18.175
s	3.361	0.542	6.204
q	−2.746	1.399	−1.963
df	3.516	0.403	8.732
Estimated skewness for the random effects			
	i	s	q
Normal class	3.967	−0.989	1.422
Escalating class	6.653	3.437	−1.588

BMI, body mass index; NLSY, National Longitudinal Survey of Youth.

The top of Table V shows the results of fitting both normal and skew- t growth mixture models. For the normal case, BIC does not provide a guide in selecting the number of classes, but decreases for 1–5 classes. For skew- t , BIC points to three classes.

The estimated mean curves for the three-class skew- t growth mixture model are shown in Figure 6. In this age range, two escalating classes are found with 13% and 33% of the subjects, respectively.

4.1. Framingham trajectory classes related to hypertension treatment: joint growth mixture and survival analysis

Elevated BMI is associated with increased risk of developing heart disease, high blood pressure, stroke, and diabetes. Framingham data contain data on blood pressure treatment at each measurement occasion. A survival component for the first treatment can be added to the growth mixture model with survival as a function of trajectory class. A continuous-time survival approach is used here based on Cox regression where the latent class variable C is used as a predictor for the survival variable; see [22, 27].

Define the time variable T as the age when blood pressure treatment is administered for the first time. If no such treatment is observed by the end of the fourth wave of the survey, the variable T is considered censored at the time the last observation is recorded. The latent class variable C explains the correlation between the BMI developmental trajectory and the blood pressure treatment variable and can be used to evaluate the effect of the BMI trajectory on the risk of developing high blood pressure.

The bottom of Table V shows the fitting of normal and skew- t models for the joint growth mixture-survival analysis. For the skew- t model, three classes are preferred by BIC as for the growth mixture model alone, whereas for the normal model, BIC again does not give guidance on the number of classes.

The three-class skew- t model for the joint growth mixture-survival analysis gives almost the same class percentages as in the previous growth mixture analysis: 12%, 35%, and 53% for the high, middle, and low BMI trajectory class, respectively. The estimated values of α_c in the survival model of (14)

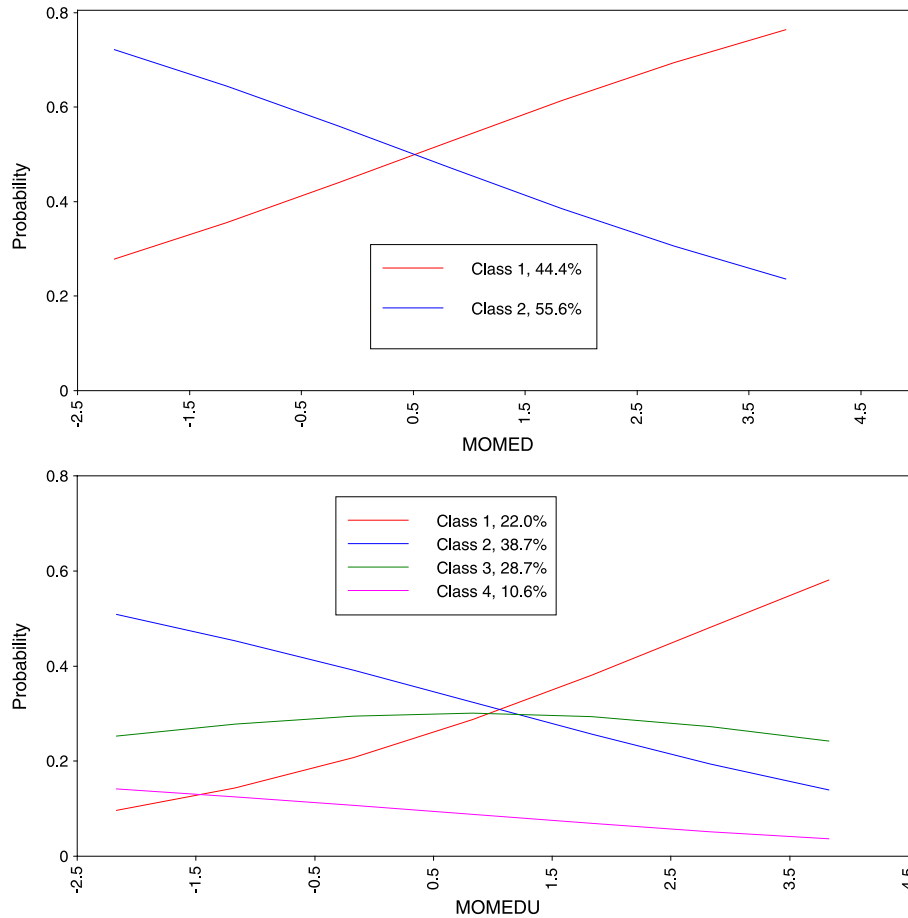


Figure 5. Relating latent class membership to mother’s education for two-class skew-t (top) and four-class normal (bottom) solutions for body mass index in the National Longitudinal Survey of Youth ages 12–23 years for black women.

Table V. Results of fitting normal and skew-t growth mixture and survival models for BMI in the Framingham data.

Growth only						
No. classes	Normal			Skew-t		
	Loglikelihood	No. Par's	BIC	Loglikelihood	No. Par's	BIC
1	-8252	8	16,557	-7771	10	15,611
2	-7954	13	15,995	-7606	17	15,327
3	-7875	18	15,871	-7567	24	15,296
4	-7787	23	15,730	-7547	31	15,304
5	-7743	28	15,674			
Growth and survival						
No. classes	Normal			Skew-t		
	Loglikelihood	No. Par's	BIC	Loglikelihood	No. Par's	BIC
1	-10,373	8	20,801	-9867	10	19,803
2	-10,067	14	20,110	-9638	18	19,401
3	-9912	20	19,961	-9596	26	19,370
4	-9820	26	19,820	-9569	34	19,372
5	-9797	32	19,814			

BMI, body mass index.

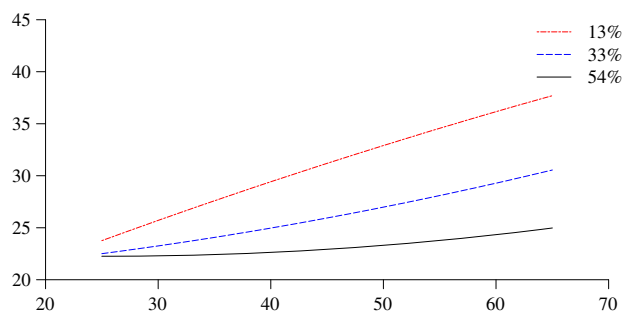


Figure 6. Estimated mean curves for the three-class skew-t growth mixture model for body mass index in the Framingham data.

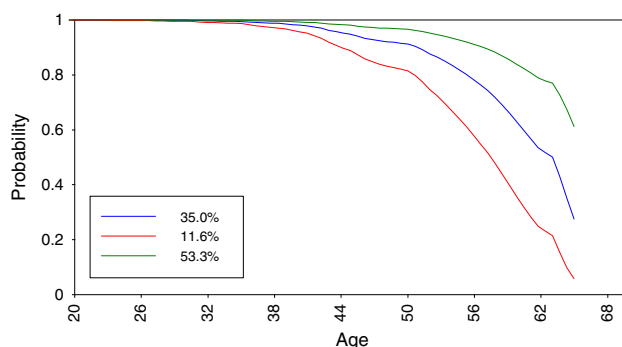


Figure 7. Estimated hypertension survival curves for the joint three-class skew-t growth mixture-survival model for BMI in the Framingham data.

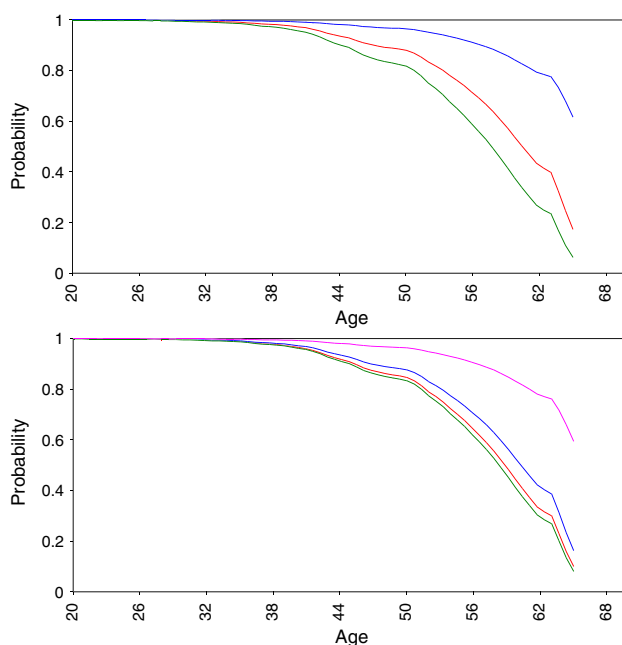


Figure 8. Estimated hypertension survival curves for the joint three- and four-class normal growth mixture-survival models for BMI in the Framingham data.

are significantly larger for the two highest BMI trajectory classes than the zero value of the reference class of normal development, and the α_c estimates of the two highest classes are significantly different from each other. The estimated survival curves for the three BMI trajectory classes are plotted in Figure 7 as a function of age ranging from 25 to 65 years. It is seen that the three survival curves

are ordered in the same way as the BMI trajectory classes, with lowest hypertension treatment survival rate for the highest BMI trajectory class. The age of median survival is estimated as about 6 years lower for subjects in the middle BMI trajectory class as compared to subjects in the lowest BMI trajectory class.

For the normal distribution version of the growth mixture-survival analysis, BIC does not give guidance on the number of classes, but BIC keeps decreasing up to the five classes attempted. The survival curves for the three-class and four-class models are shown in Figure 8. Going from the lowest to the highest curve for the three-class model, the class percentages are 6%, 33%, and 61%. The two bottom curves, corresponding to the two highest BMI trajectories, do not, however, have significantly different α_c estimates as is the case for the skew-t curves. Considering four classes, the class percentages are 4%, 6%, 26%, and 64% going from the lowest to the highest curve. The three bottom curves of the four-class model are not significantly different from each other. A conjecture is that the tendency of the normal model to add small classes to fit the tail of strongly skewed outcomes impedes its ability to fit substantively meaningful classes in this example.

5. Simulations

To further illustrate the proposed skew-t growth mixture modeling, a small Monte Carlo simulation study is carried out. Data are generated by a two-class skew-t growth mixture model and analyzed in several different ways: Using a normal distribution, using a skew-t distribution with equal skews, and using a skew-t distribution with unequal skews. The focus is on the choice of number of classes using BIC as well as bias in the parameter estimates.

Data are generated by a growth mixture model with linear growth and eight time points using parameter values given in the Mplus Monte Carlo simulation input in Appendix C. Table VI shows the estimates for key parameters using $n = 2000$ when analyzed by the same model that generated the data and allowing class-varying skew parameters for both the intercept and slope random effects, labeled i and s , respectively. Five hundred replications are used, and the results shown are based on the 498 replications that converged. Biases are small, standard deviations and average SE agree well, and the 95% coverage is good. Results for $n = 1000$, not shown here, are also good.

Table VII refers to analyses based on the same two-class skew-t growth mixture model as aforementioned. The top of Table VII shows the percentage of the replications where a certain number of classes is favored by BIC. Three different distributions are used for 1–3 latent classes. BIC for the normal distribution overestimates the number of classes, pointing to three classes 88% of the time. For the skew-t distribution assuming zero skew for the slope, that is, equal skews for the outcomes over time, BIC incorrectly points to three classes 23% of the time. Using the correct model allowing skew for both the intercept and slope, BIC points to the correct number of classes 100% of the time.

Table VI. Monte Carlo simulation results for key estimates of a two-class skew-t growth mixture model.

Parameter	Population value	Estimate	SD	SE	95% coverage
Class 1					
Mean i	4.00	4.00	0.14	0.15	0.95
Mean s	1.00	1.02	0.15	0.15	0.97
Skew i	2.00	1.98	0.18	0.19	0.96
Skew s	2.00	1.96	0.21	0.22	0.96
d.f.	5.00	5.08	0.50	0.53	0.94
Class 2					
Mean i	0.00	0.01	0.15	0.15	0.94
Mean s	0.00	0.00	0.14	0.15	0.95
Skew i	2.00	1.99	0.36	0.37	0.92
Skew s	2.00	2.01	0.23	0.25	0.94
d.f.	5.00	5.05	0.50	0.52	0.94
Logit($c=1$)	0.00	0.00	0.16	0.17	0.93

SD, standard deviations; SE, standard errors.

Table VII. Monte Carlo simulation results when analyzing data generated by a two-class skew-t growth mixture model and analyzed with different distributions and number of classes.

Distribution	Percent preferred by BIC		
	Number of classes		
	1	2	3
Normal	0	12	88
Skew-equal	0	77	23
Skew-unequal	0	100	0

Distribution	Estimated logit ($c = 1$)				
	Population value	Estimate	SD	SE	95% Coverage
Normal, 2c	0.00	-1.61 ¹	1.01	0.37	0.00
Skew-equal, 2c	0.00	0.46 ²	0.42	0.18	0.09
Skew-unequal, 2c	0.00	0.01	0.16	0.17	0.93

¹ $P(c = 1) = 0.17$

² $P(c = 1) = 0.61$

The bottom of Table VII shows estimation results when two classes are chosen by the three different distributions. The focus is on the key parameter of the logit corresponding to the probability of class 1. It is noteworthy that assuming equal skews over time in the skew-t model gives a large overestimation, where the population probability of 0.50 is estimated as 0.61. As expected, this misclassification also leads to strong biases in other parameters.

6. Conclusions

The skew-t growth mixture model has several advantages over normal growth mixture modeling. It can fit the data considerably better than normal mixtures. It can use a more parsimonious model. It can reduce the risk of extracting latent classes that are merely due to non-normality of the outcomes. It can check the stability/reproducibility of a normal mixture solution. It can describe the percentiles of skewed distributions.

There are, however, several disadvantages with skew-t growth mixture modeling. It provides much slower computations than normal mixtures especially for large sample sizes, given that computations need to handle raw data in every step as opposed to using sufficient statistics[§]. It needs larger samples, where small class sizes can create problems, although successful analyses can be carried out at $n = 100 - 200$. It needs more random starts than normal mixtures to replicate the best loglikelihood given a typically less smooth likelihood function. It often leads to classification with lower entropy. Furthermore, it needs continuous variables to provide enough information for the skew and degrees of freedom parameters.

Apart from computational efficiency, other advantages of normal mixture modeling should not be overlooked. Normal mixtures can carve out smaller subgroups in the sample that non-normal mixture modeling might miss, such as tail subgroups with different behaviors. It can handle smaller class sizes, it may be robust to mild degrees of non-normality, and it can be used as a starting point for non-normal mixture modeling.

In this paper, the skew-t growth mixture model was extended to continuous-time survival analysis. Because of the general implementation in the Mplus software, other extensions are available as well. For example, a survival part can be used to model non-ignorable dropout as in [11], and categorical and count variables can be included in the model with parameters varying as a function of the latent trajectory classes for the continuous repeated measures.

[§]The National Longitudinal Survey of Youth four-class normal growth mixture model (GMM) used 6% of the time of the two-class skew-t GMM, and the Framingham three-class normal GMM used 2% of the time of the three-class skew-t GMM. The skew-t runs took 57 and 13 min, respectively, for the two data analyses. In all cases, 400 randomly perturbed starting values were used.

Appendix A

Mplus input for the two-class skew-t growth mixture model for BMI using the NLSY data is shown in Table A1. In the second line of the MODEL command, the growth model uses i , s , and q to represent the random effects of intercept, linear slope, and quadratic slope. The random effects influence the BMI measures at ages 12–23 years with time centered at age 17 years. This is referred to as a wide format, single-level approach to growth modeling. The latent class variable is referred to as c , and c , i , s , and q are regressed on the mother's education covariate. The i , s , and q covariance matrix is specified as different in the two classes. The residual variances for the BMI outcomes are held equal across the ages but are specified as different across the classes.

Table A1. Mplus input for the 2-class skew-t growth mixture model for BMI related to mother's education using the NLSY data.

```

DATA:      FILE = nlsy97_clean.dat;
VARIABLE:  NAMES = id gender age_1996 age_1997 race1 bmi12 bmi13 bmi14
           bmi15 bmi16 bmi17 bmi18 bmi19 bmi20 bmi21 bmi22 bmi23 black
           hisp mixed c1 c2 c3 c1_wom c2_wom c3_wom momed par_bmi
           bio1_bmi bio2_bmi bmi_par currsmkr97 bingedrnk97 mjuse97
           cent_msa liv2prnts adopted income hhsz97;
           USEVARIABLES = bmi12-bmi23 momed;
           USEOBSERVATIONS = gender EQ -1 AND black EQ 1;
           MISSING ARE ALL (9999);
           CLASSES = c(2);
ANALYSIS:  TYPE = MIXTURE;
           STARTS = 400 80;
           DISTRIBUTION = SKEWT;
           ESTIMATOR = MLF;
MODEL:     %OVERALL%
           i s q | bmi12@-.5 bmi13@-.4 bmi14@-.3 bmi15@-.2 bmi16@-.1
           bmi17@0 bmi18@.1 bmi19@.2 bmi20@.3 bmi21@.4
           bmi22@.5 bmi23@.6;
           c ON momed;
           i-q ON momed;
           %c#1%
           i-q;
           i-q WITH i-q;
           bmi12-bmi23 (1);
           %c#2%
           i-q;
           i-q WITH i-q;
           bmi12-bmi23 (2);
OUTPUT:    TECH1 TECH4 TECH8 RESIDUAL;
PLOT:      TYPE = PLOT3;
           SERIES = bmi12-bmi23(s);

```

BMI, body mass index; NLSY, National Longitudinal Survey of Youth.

Appendix B

Tables B1 and B2 show the Mplus input for the three-class skew-t joint growth mixture-survival modeling of the Framingham data. The survival variable is referred to as t . The DEFINE command creates the age variables on the right-hand side of (27). The MODEL command specifies the influence of the random effect i on the BMI outcomes and how BMI is related to the age variables. The class-specific statements in Table B2 specify class-varying i variance and means as well as class-varying effects of the age variables on the BMI outcomes in line with (27). The coefficient α_c in (14) is class varying as the default.

Table B1. Mplus input for the three-class skew-t joint growth mixture-survival modeling of the Framingham data.

```

DATA:      FILE = survskew.dat;
VARIABLE:  NAMES = bmi1 bmi3 bmi5 bmi7 htnrx1 htnrx3 htnrx5 htnrx7 age1
           age3 age5 age7 T0 ind t;
           MISSING = all(999);
           USEVARIABLES = bmi1 bmi3 bmi5 bmi7 T ind cage1 cage3 cage5
           cage7 cage1q cage3q cage5q cage7q;
           SURVIVAL = t;
           TIMECENSORED = IND (1 = NOT 0 = RIGHT);
           CLASSES = c(3);
DEFINE:    cage1 = (age1 - 50)/10;
           cage3 = (age3 - 50)/10;
           cage5 = (age5 - 50)/10;
           cage7 = (age7 - 50)/10;
           cage1q = cage1*cage1;
           cage3q = cage3*cage3;
           cage5q = cage5*cage5;
           cage7q = cage7*cage7;
ANALYSIS:  TYPE = MIXTURE;
           DISTRIBUTION = SKEWT;
           STARTS = 100 40;
MODEL:     %OVERALL%
           i BY bmi1-bmi7@1;
           bmi1 ON cage1;
           bmi3 ON cage3;
           bmi5 ON cage5;
           bmi7 ON cage7;
           bmi1 ON cage1q;
           bmi3 ON cage3q;
           bmi5 ON cage5q;
           bmi7 ON cage7q;
           [bmi1-bmi7@0];
    
```

Table B2. Mplus input for the three-class skew-t joint growth mixture-survival modeling of the Framingham data, continued.

```

           %c#1%
           i;
           [i];
           bmi1 ON cage1 (1);
           bmi3 ON cage3 (1);
           bmi5 ON cage5 (1);
           bmi7 ON cage7 (1);
           bmi1 ON cage1q (2);
           bmi3 ON cage3q (2);
           bmi5 ON cage5q (2);
           bmi7 ON cage7q (2);
           %c#2%
           i;
           [i];
           bmi1 ON cage1 (3);
           bmi3 ON cage3 (3);
           bmi5 ON cage5 (3);
           bmi7 ON cage7 (3);
           bmi1 ON cage1q (4);
           bmi3 ON cage3q (4);
           bmi5 ON cage5q (4);
           bmi7 ON cage7q (4);
           %c#3%
    
```


Table B2. *Continued.*

i;
[i];
bmi1 ON cage1 (5);
bmi3 ON cage3 (5);
bmi5 ON cage5 (5);
bmi7 ON cage7 (5);
bmi1 ON cage1q (6);
bmi3 ON cage3q (6);
bmi5 ON cage5q (6);
bmi7 ON cage7q (6);

Appendix C

Table C1 shows the simulation input uses eight time points and $n = 2000$ with 500 replications, generating and analyzing with two latent classes using the skew-t distribution. The MODEL POPULATION command specifies the data-generating linear growth model with intercept i and slope s . The residual variances for the outcomes are 0.5 at all time points. The i , s means are 4, 1 and 0, 0 in classes 1 and 2, respectively. The variances for i and s are 1 and 0.7, respectively, for both classes, and the covariance is zero for both classes. These parameter values refer to the symmetric part U_1 of (5). The skew param-

Table C1. Monte Carlo simulation generating and analyzing data using a two-class skew-t growth mixture model.

```
TITLE:                2c skew-t
MONTECARLO:          NAMES = y1-y8;
                    NOBSERVATIONS = 2000;
                    NREPS = 500;
                    GENCLASSES = c(2);
                    CLASSES = c(2);
ANALYSIS:            TYPE = MIXTURE;
                    ESTIMATOR = MLR;
                    DISTRIBUTION = SKEWT;
MODEL POPULATION:    %OVERALL%
                    [c#1*0];
                    i s | y1@0 y2@.1 y3@.2 y4@.3 y5@.4 y6@0.5 y7@0.6 y8@0.7;
                    y1-y8*.5;
                    [i*4 s*1];
                    i*1 s*0.7;
                    i-s WITH i-s*0;
                    {i*2};
                    {s*2};
                    {df*5};
                    %c#2%
                    [i*0 s*0];
MODEL:                %OVERALL%
                    [c#1*0];
                    i s | y1@0 y2@.1 y3@.2 y4@.3 y5@.4 y6@0.5 y7@0.6 y8@0.7;
                    y1-y8*.5;
                    [i*4 s*1];
                    i*1 s*0.7;
                    i-s WITH i-s*0;
                    {i*2};
                    {s*2};
                    {df*5};
                    %c#2%
                    [i*0 s*0];
                    i*1 s*0.7;
                    i-s WITH i-s*0;
```

eters are 2 for i and s in both classes, whereas the degrees of freedom parameter is 5 in both classes. The MODEL command specifies that the analysis model is the same as the data-generating model with starting values equal to the population values.

References

1. Laird NM, Ware JH. Random-effects models for longitudinal data. *Biometrics* 1982; **38**:963–974.
2. McLachlan GJ, Peel D. *Finite Mixture Models*. Wiley: New York.
3. Verbeke G, Lesaffre E. A linear mixed-effects model with heterogeneity in the random effects population. *Journal of the American Statistical Association* 1996; **91**:217–221.
4. Muthén B, Shedden K. Finite mixture modeling with mixture outcomes using the EM algorithm. *Biometrics* 1999; **55**:463–469.
5. Nagin DS, Land KC. Age, criminal careers, and population heterogeneity: specification and estimation of a nonparametric, mixed Poisson model. *Criminology* 1993; **31**:327–362.
6. Roeder K, Lynch KG, Nagin DS. Modeling uncertainty in latent class membership: a case study in criminology. *Journal of the American Statistical Association* 1999; **94**:766–776.
7. Lin H, McCulloch CE, Turnbull BW, Slate E, Clark LC. A latent class mixed model for analysing biomarker trajectories with irregularly scheduled observations. *Statistics in Medicine* 2000; **19**(10):1303–1318.
8. Lin H, Turnbull BW, McCulloch CE, Slate E. Latent class models for joint analysis of longitudinal biomarker and event process data: application to longitudinal prostate-specific antigen readings and prostate cancer. *Journal of the American Statistical Association* 2002; **97**:53–65.
9. Muthén B, Brown H. Estimating drug effects in the presence of placebo response: causal inference using growth mixture modeling. *Statistics in Medicine* 2009; **28**:3363–3385.
10. Muthén B, Asparouhov T. Growth mixture modeling: analysis with non-Gaussian random effects. In *Longitudinal Data Analysis*, Fitzmaurice G, Davidian M, Verbeke G, Molenberghs G (eds). Chapman & Hall/CRC Press: Boca Raton, 2009; 143–165.
11. Muthén B, Asparouhov T, Hunter A, Leuchter A. Growth modeling with non-ignorable dropout: alternative analyses of the STAR*D antidepressant trial. *Psychological Methods* 2011; **16**:17–33.
12. Erosheva EA, Matsueda RL, Telesca D. Breaking bad: two decades of life-course data analysis in criminology, developmental psychology, and beyond. *Annual Review of Statistics and its Applications* 2014; **1**:301–332.
13. Pearson K. Contributions to the theory of mathematical evolution. *Philosophical Transactions of the Royal Society of London A* 1895; **186**:343–414.
14. Schork NJ, Schork MA. Skewness and mixtures of normal distributions. *Communications Statistics Theory Methods* 1988; **17**:3951–3969.
15. Bauer DJ, Curran PJ. Distributional assumptions of growth mixture models: implications for overextraction of latent trajectory classes. *Psychological Methods* 2003; **8**:338–363.
16. Lin T, Wu PH, McLachlan GJ, Lee SX. The skew-t factor analysis model. Submitted. Available from: <http://arxiv.org/abs/1310.5336> [accessed on 4 December 2014].
17. Lee S, McLachlan GJ. Finite mixtures of multivariate skew t-distributions: some recent and new results. *Journal of Statistics and Computing* 2014; **24**:181–202.
18. Asparouhov T, Muthén B. Structural equation models and mixture models with continuous non-normal skewed distributions. Accepted for publication in *Structural Equation Modeling* 2014.
19. Lu X, Huang Y. Bayesian analysis of nonlinear mixed-effects mixture models for longitudinal data with heterogeneity and skewness. *Statistics in Medicine* 2014; **33**:2701–2880.
20. Boik RJ, Robison-Cox JF. Derivatives of the incomplete beta function. *Journal of Statistical Software* 1998; **3**:1–19.
21. Murphy SA, van der Vaart AW. On profile likelihood. *Journal of the American Statistical Association* 2000; **95**:449–465.
22. Larsen K. Joint analysis of time-to-event and multiple binary indicators of latent classes. *Biometrics* 2004; **60**(1):85–92.
23. Asparouhov T, Masyn K, Muthén B. Continuous time survival in latent variable models. *Proceedings of the Joint Statistical Meeting in Seattle, August 2006. ASA section on Biometrics*, Seattle, Washington, 2006, 180–187.
24. Muthén L, Muthén B. *Mplus User's Guide* Seventh Edition. Muthén & Muthén: Los Angeles, CA, 2012.
25. Nonemaker JM, Morgan-Lopez AA, Pais JM, Finkelstein EA. Youth BMI trajectories: evidence from the NLSY97. *Obesity* 2009; **17**:1274–1280.
26. Dawber TR, Kannel WB, Lyell LP. An approach to longitudinal studies in a community: the Framingham study. *Annals of the New York Academy of Sciences* 1963; **107**:539–556.
27. Muthén B, Asparouhov T, Boye M, Hackshaw M, Naegeli A. Applications of continuous-time survival in latent variable models for the analysis of oncology randomized clinical trial data using Mplus, Technical Report 2009. Available from: www.statmodel.com [accessed on 4 December 2014].