

Growth Mixture Modeling with Non-Normal Distributions

Bengt Muthén & Tihomir Asparouhov

Mplus

www.statmodel.com

June 11, 2014

Abstract

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 would be 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; see, for example, [1]. The new method is illustrated using the longitudinal development of BMI in two data sets. The first data set is from the National Longitudinal Survey of Youth covering ages 12 to 23. Here, the development is related to an antecedent measuring socioeconomic background. The second data set is from the Framingham Heart Study covering ages 25 to 65. Here, the development is related to the concurrent event of treatment for hypertension using a joint growth mixture-survival model.

1 Introduction

Growth mixture modeling combines the conventional [2] random effects modeling with latent trajectory classes as in finite mixture modeling (see, e.g. [3]). Growth mixture modeling was introduced in [4] and [5] with related developments in [6] and [7]. Following this, many extensions and applications have been presented such as [8] considering PSA biomarker trajectories with irregularly scheduled observations, [9] adding joint estimation of survival with prostate cancer, [10] considering causal inference in randomized trials of antidepressants with placebo effects, [11] adding general multilevel growth mixture modeling, and [12] modeling non-ignorable dropout in antidepressant trials. For overviews of methods with illustrations by a variety of applications, see [11] and [13].

A limiting feature of the above 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 $18 < BMI < 25$; overweight $25 < BMI < 30$; and obese > 30 . The distribution of BMI at age 15 for males 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 1 shows the loglikelihood and Bayesian Information Criterion (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.

[Figure 1 about here.]

[Table 1 about here.]

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 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 [14], continued with the Platt-Pickering hypertension debate in the 60s ([3]), and is still a topic of interest (see, e.g., [15] and [16]). 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 1, 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 [17] and [1]. 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 2 is used, spanning ages 12 to 23. Here, the BMI development is related to an antecedent measuring socioeconomic background. Second, data from the Framingham Heart Study is used for ages 25 to 65. 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.

[Table 2 about here.]

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

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

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 $\boldsymbol{\zeta} = (\zeta_0 \ \zeta_1 \ \zeta_2)'$. Using the notation for the restricted multivariate skew-t distribution given in [1],

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

where in this application $\mathbf{0}$ is the 3×1 vector of zero means for $\boldsymbol{\zeta}$, $\boldsymbol{\Psi}_c$ is the 3×3 within-class covariance matrix for $\boldsymbol{\zeta}$, $\boldsymbol{\delta}_c$ 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

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

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

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

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

where $v = Var(V)$ is given in the previous formula and the σ parameter is the diagonal element of $\boldsymbol{\Sigma}$ 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 [18]) that this specification gives the distribution of the vector of observations for individual i at the T time points $\mathbf{Y}_i = (Y_{i1}, \dots, Y_{iT})'$ conditional on \mathbf{X} and latent class c

$$\mathbf{Y}|\mathbf{X} \sim rMST(\boldsymbol{\mu}_c, \boldsymbol{\Sigma}_c, \boldsymbol{\delta}_{Yc}, \nu_c), \quad (9)$$

where

$$\boldsymbol{\mu}_c = \boldsymbol{\Lambda} (\boldsymbol{\alpha}_c + \boldsymbol{\Gamma}_c \mathbf{X}), \quad (10)$$

$$\boldsymbol{\Sigma}_c = \boldsymbol{\Lambda} \boldsymbol{\Psi}_c \boldsymbol{\Lambda}' + \boldsymbol{\Theta}_c, \quad (11)$$

$$\boldsymbol{\delta}_{Yc} = \boldsymbol{\Lambda} \boldsymbol{\delta}_c, \quad (12)$$

where for the quadratic growth model of (1)

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

the elements of $\boldsymbol{\alpha}_c$ and $\boldsymbol{\Gamma}_c$ are shown in (2), and $\boldsymbol{\Theta}_c$ is the within-class covariance matrix for $\boldsymbol{\epsilon} = (\epsilon_1 \ \epsilon_2, \dots, \epsilon_T)'$, typically specified as diagonal. With this skew-t growth mixture model specification the non-normality of the outcomes \mathbf{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 [19]. In this case, however, the skew parameters of δ would have to be held equal across time because otherwise the means for \mathbf{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 [19]. Time-invariant skewness in the residuals is a special case of the model proposed above 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.

2.1 Estimation

The models are estimated by maximum likelihood. As shown in [18], the log-likelihood can be written explicitly and maximized with a general maximization algorithm such as the Quasi-Newton optimization method as long as the derivatives of the log-likelihood can be computed. The only derivative that is more complex is $T_1(x, \nu)$ with respect to ν , where T_1 is the standard t-distribution function. For this derivative the method developed in [20] is used. Direct maximization appears to work well and is relatively fast. The standard error estimates are based on the inverse of the information matrix as usual with the ML estimator and robust standard errors can also be computed using the sandwich

estimator. All computation in this paper are carried out by Mplus Version 7.2 ([21]).¹

3 Growth mixture modeling of BMI from the NLSY

This section applies the skew-t growth mixture model to the BMI development over the ages 12 to 23 using the NLSY data shown in Table 2. As discussed in e.g. [22], large differences in BMI are observed across gender and ethnicity subgroups and the current analysis is restricted to black females ($n = 1160$). Table 3 shows the results of fitting both normal and skew-t growth mixtures. It is seen from the BIC values that four classes are needed in the normal case, whereas only two classes are needed using skew-t. The best skew-t BIC value is considerably better than the best normal BIC value.

[Table 3 about here.]

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

¹The 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. Complex survey features of stratification, weights, and clustering are also handled in Mplus.

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 as is seen for the skew-t solution.

[Figure 2 about here.]

Due to 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 3 about here.]

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. 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.

[Figure 4 about here.]

The estimated skew and degrees of freedom parameters for the intercept (i), linear slope (s), and quadratic slope (q) are shown in Table 4. 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).

[Table 4 about here.]

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: Ph.D., 8: professional degree). Adding this covariate to the growth mixture model, it is also of interest to compare the two-class skew-t solution with a normal solution using the four classes considered earlier. 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 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.

[Figure 5 about here.]

4 Growth mixture modeling of BMI in the Framingham data

The classic Framingham Heart Study ([23]) provides a second longitudinal data set with which to explore growth mixture modeling of BMI. With its focus on cardiovascular disease, the Framingham data contains information on treatment for hypertension and it is of interest to relate this to the BMI trajectories. The current analyses uses a subset of the data for females ages 25-65, 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 the Section 2 model is modified as

$$Y_{it}|C_i=c = \eta_{0i} + \eta_1 (a_{it} - a_0) + \eta_2 (a_{it} - a_0)^2 + \epsilon_{it}, \quad (14)$$

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 $\mathbf{\Lambda}$ but as variables.

Table 5 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.

[Table 5 about here.]

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.

[Figure 6 about here.]

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 contains 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 [24] and [25].

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 forth wave of the survey the variable T is considered censored at the time the last observation is recorded.

Using the standard Cox proportional hazard model, the hazard function in class c is given by

$$h_c(t) = h(t)Exp(\alpha_c), \quad (15)$$

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.

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

$$H_c(t) = \int_0^t h_c(x)dx = Exp(\alpha_c)H(t), \quad (16)$$

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

$$H(t) = \int_0^t h(x)dx. \quad (17)$$

The survival function in class c is the probability that the survival variable T

is greater than t , given the individual belongs to class c ,

$$S(t|c) = P(T > t|C = c) = \text{Exp}(-H_c(t)) = \text{Exp}(-\text{Exp}(\alpha_c)H(t)). \quad (18)$$

The likelihood for the survival variable T is

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

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

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

where equation (14) provides the model for $P(Y|C = c)$. 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 skew-t growth mixture model including the survival variable 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 (15) 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. 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 six years lower for subjects in the middle BMI trajectory class as compared to subjects in the lowest BMI trajectory class.

[Figure 7 about here.]

5 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 done 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 leads to classification with lower entropy. Furthermore, it needs continuous variables to provide enough information for the skew and degrees of freedom parameters.

In this paper the skew-t growth mixture model was extended to continuous-time survival analysis. Due to the general implementation in the Mplus software,

other extensions are possible as well. For example a survival part can be used to model non-ignorable dropout as in [12] 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.

References

1. Lee S. & McLachlan G. J. (2014). Finite mixtures of multivariate skew t-distributions: some recent and new results. *Journal of Statistics and Computing*, 24, 181-202.
2. Laird, N. M. & Ware, J. H. (1982). Random-effects models for longitudinal data. *Biometrics* 38, 963-974.
3. McLachlan, G. J. & Peel, D. (2000). *Finite Mixture Models*. New York: Wiley.
4. Verbeke, G. & Lesaffre, E. (1996). A linear mixed-effects model with heterogeneity in the random effects population. *Journal of the American Statistical Association* 91, 217-221.
5. Muthén, B. & Shedden, K. (1999). Finite mixture modeling with mixture outcomes using the EM algorithm. *Biometrics*, 55, 463-469.
6. Nagin, D. S. & Land, K. C. (1993). Age, criminal careers, and population heterogeneity: Specification and estimation of a nonparametric, mixed Poisson model. *Criminology* 31, 327-362.
7. Roeder, K., Lynch, K. G., & Nagin, D. S. (1999). Modeling uncertainty in latent class membership: A case study in criminology. *Journal of the American Statistical Association* 94, 766-776.
8. Lin, H., McCulloch, C. E., Turnbull, B. W., Slate, E. & Clark, L.C. (2000). A latent class mixed model for analysing biomarker trajectories with irregularly scheduled observations. *Statistics in Medicine*, 19, 1303-1318.
9. Lin, H., Turnbull, B. W., McCulloch, C. E., & Slate, E. (2002). 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*, 97, 53-65.
10. Muthén, B. & Brown, H. (2009). Estimating drug effects in the presence of placebo response: Causal inference using growth mixture modeling. *Statistics in Medicine*, 28, 3363-3385.
 11. Muthén, B. & Asparouhov, T. (2009). Growth mixture modeling: Analysis with non-Gaussian random effects. In Fitzmaurice, G., Davidian, M., Verbeke, G. & Molenberghs, G. (eds.), *Longitudinal Data Analysis*, pp. 143-165. Boca Raton: Chapman & Hall/CRC Press.
 12. Muthén, B., Asparouhov, T., Hunter, A. & Leuchter, A. (2011). Growth modeling with non-ignorable dropout: Alternative analyses of the STAR*D antidepressant trial. *Psychological Methods*, 16, 17-33.
 13. Erosheva, E.A., Matsueda, R. L. & Telesca, D. (2014). Breaking bad: Two decades of life-course data analysis in criminology, developmental psychology, and beyond. *Annual Review of Statistics and its Applications*, 1, 301-332.
 14. Pearson, K. (1895). Contributions to the theory of mathematical evolution. *Philosophical Transactions of the Royal Society of London A*, 186, 343-414.
 15. Schork, N.J. & Schork, M.A. (1988). Skewness and mixtures of normal distributions. *Communications Statistics Theory Methods* 17, 3951-3969.
 16. Bauer, D. J., & Curran, P. J. (2003). Distributional assumptions of growth mixture models: Implications for overextraction of latent trajectory classes. *Psychological Methods*, 8, 338-363.
 17. Lin T., Wu P. H., McLachlan G. J., Lee S. X. (2013). The skew-t factor analysis model. Submitted. <http://arxiv.org/abs/1310.5336>
 18. Asparouhov, T. & Muthén, B. (2014). Structural equation models and

- mixture models with continuous non-normal skewed distributions. Accepted for publication in Structural Equation Modeling.
19. Lu, X. & Huang, Y. (2014). Bayesian analysis of nonlinear mixed-effects mixture models for longitudinal data with heterogeneity and skewness. *Statistics in Medicine*, 33, 2701-2880.
 20. Boik R. J. & Robison-Cox, J. F. (1998). Derivatives of the incomplete Beta function. *Journal of Statistical Software*, 3, 1-19.
 21. Muthén, B. and Muthén, L. (1998-2012). *Mplus User's Guide*. Seventh Edition. Los Angeles, CA: Muthén and Muthén.
 22. Nonnemaker JM, Morgan-Lopez AA, Pais JM, & Finkelstein EA. (2009). Youth BMI trajectories: evidence from the NLSY97. *Obesity*, 17, 1274-1280.
 23. Dawber, T.R., Kannel, W.B., & Lyell, L.P. (1963). An approach to longitudinal studies in a community: the Framingham Study. *Ann N Y Acad Sci*, 107, 539-556.
 24. Larsen, K. (2004). Joint analysis of time-to-event and multiple binary indicators of latent classes. *Biometrics*, 60, 85-92.
 25. Muthén, B., Asparouhov, T., Boye, M., Hackshaw, M. & Naegeli, A. (2009). Applications of continuous-time survival in latent variable models for the analysis of oncology randomized clinical trial data using Mplus. Technical Report. Retrieved from www.statmodel.com.

List of Figures

1	Observed and 3-class normal fitted distribution of BMI among 15-year old males in the NLSY	23
2	Estimated mean curves for two-class skew-t (top) and four-class normal (bottom) growth mixture modeling of BMI in the NLSY ages 12 to 23 for black females	24
3	Estimated percentiles for the escalating class of BMI in the NLSY ages 12 to 23 for black females	25
4	Random effect distributions for the normal and escalating trajectory classes using a skew-t growth mixture model of BMI in the NLSY ages 12 to 23 for black females	26
5	Relating latent class membership to mother's education for two-class skew-t (top) and four-class normal (bottom) solutions for BMI in the NLSY ages 12 to 23 for black females	27
6	Estimated mean curves for the three-class skew-t growth mixture model for BMI in the Framingham data	28
7	Estimated hypertension survival curves for the joint three-class skew-t growth mixture-survival model for BMI in the Framingham data	29

Figure 1: Observed and 3-class normal fitted distribution of BMI among 15-year old males in the NLSY

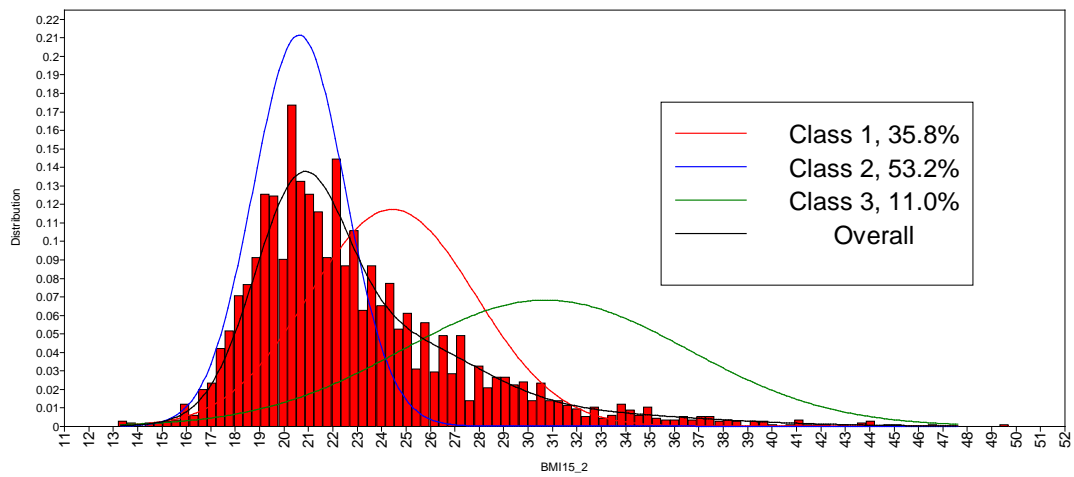


Figure 2: Estimated mean curves for two-class skew-t (top) and four-class normal (bottom) growth mixture modeling of BMI in the NLSY ages 12 to 23 for black females

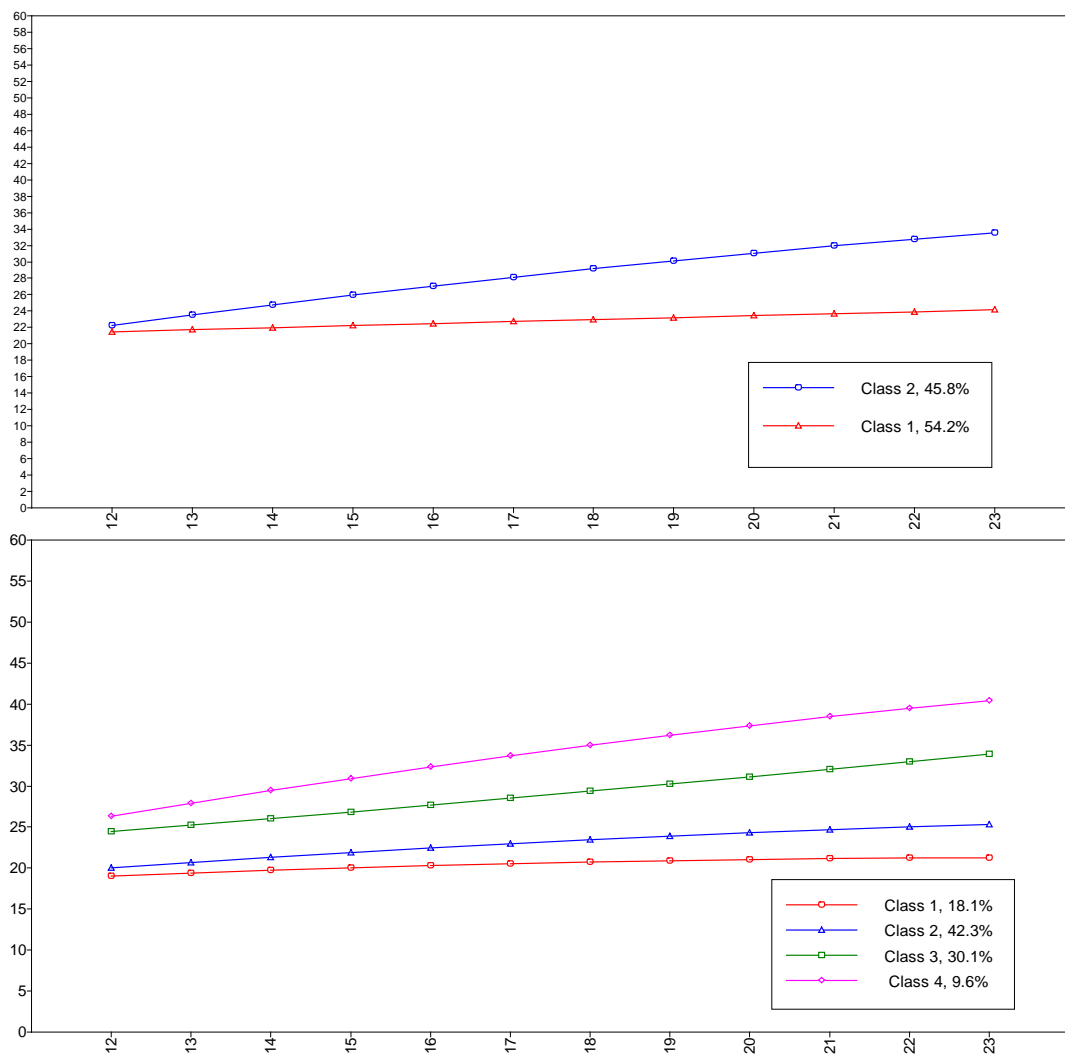


Figure 3: Estimated percentiles for the escalating class of BMI in the NLSY ages 12 to 23 for black females

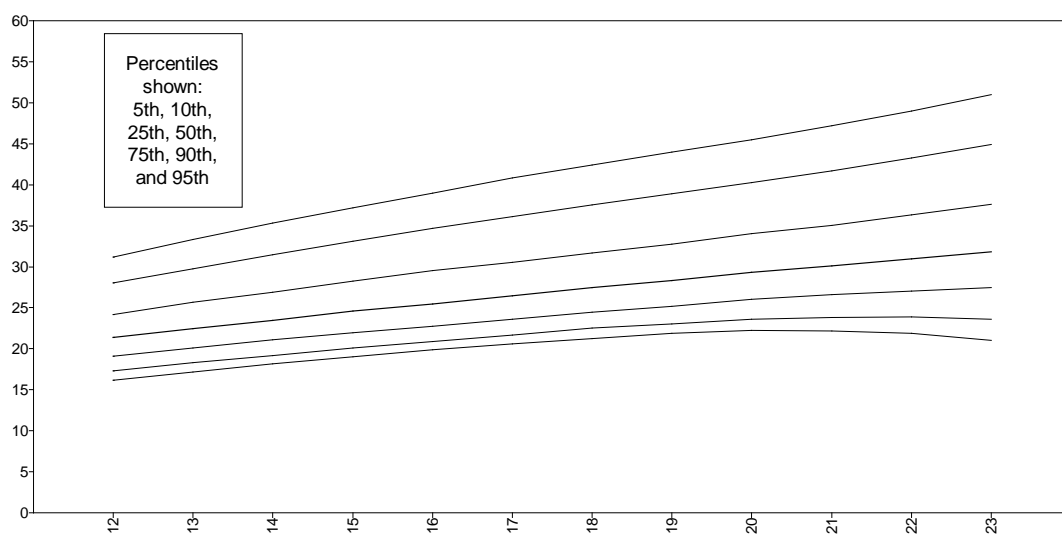


Figure 4: Random effect distributions for the normal and escalating trajectory classes using a skew-t growth mixture model of BMI in the NLSY ages 12 to 23 for black females

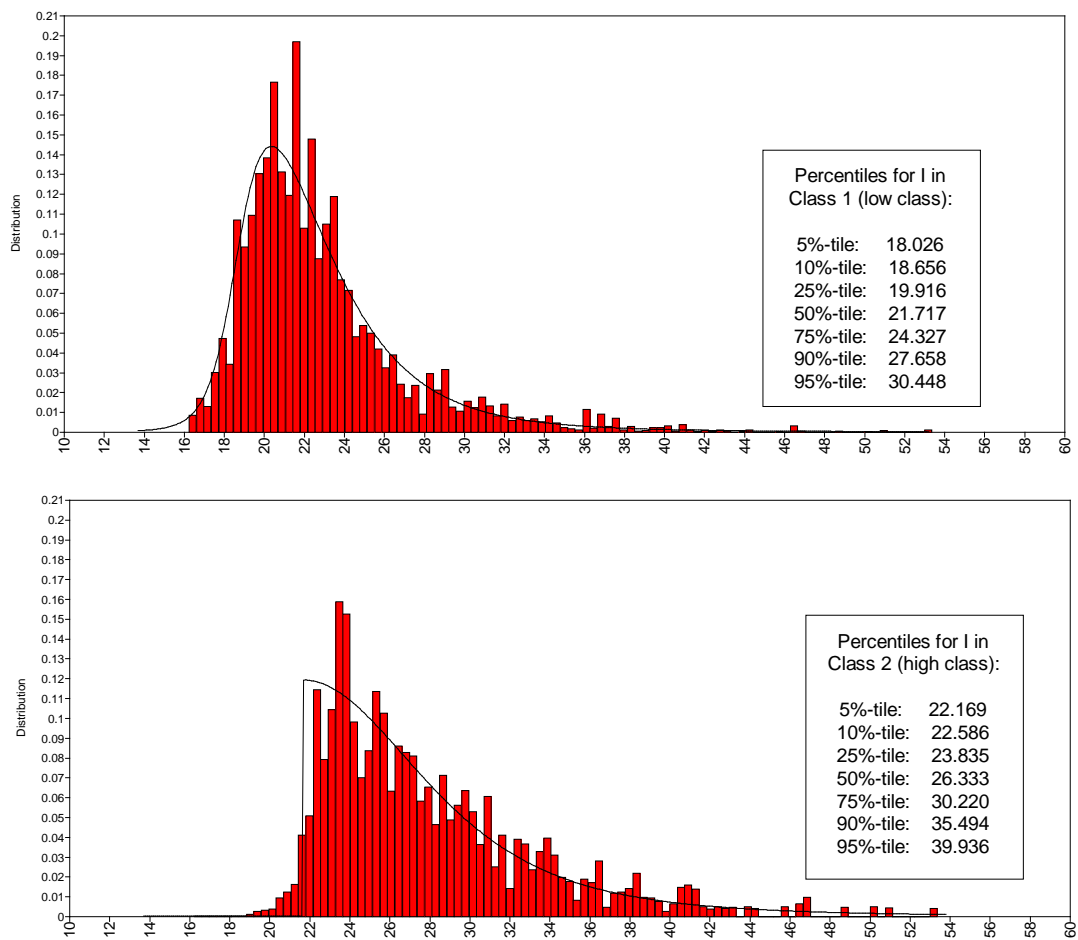


Figure 5: Relating latent class membership to mother's education for two-class skew-t (top) and four-class normal (bottom) solutions for BMI in the NLSY ages 12 to 23 for black females

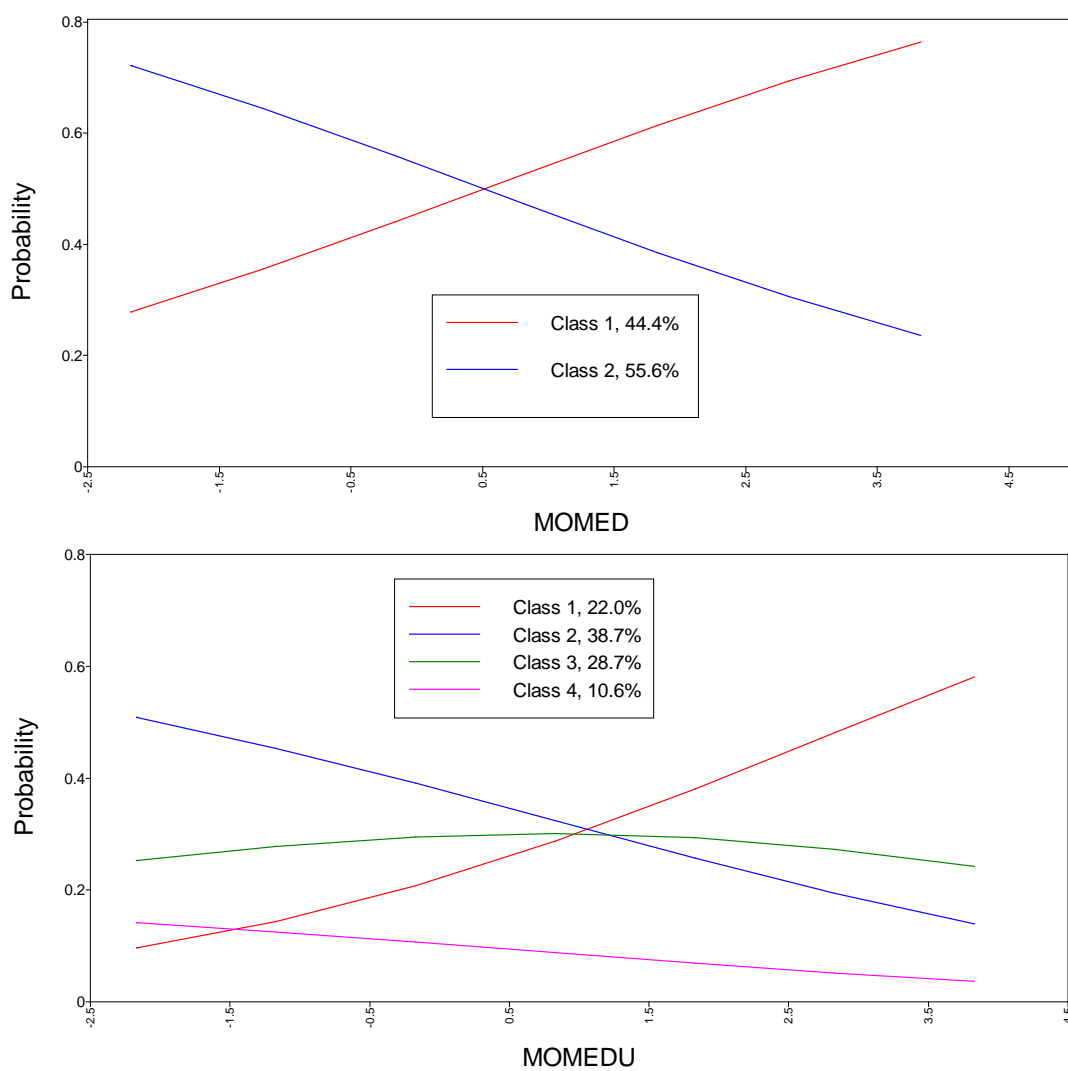


Figure 6: Estimated mean curves for the three-class skew-t growth mixture model for BMI in the Framingham data

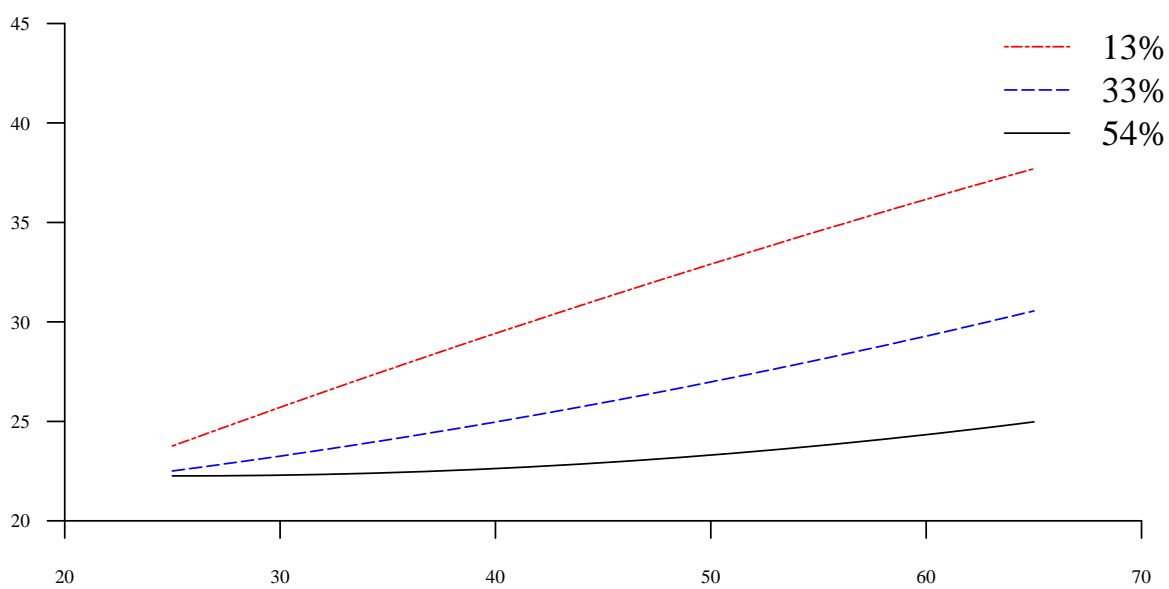
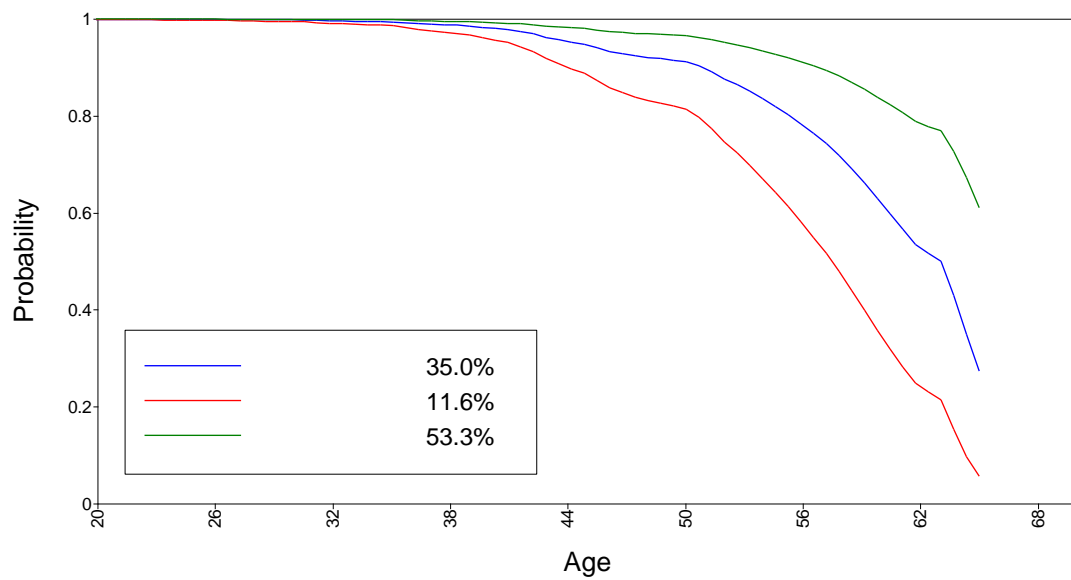


Figure 7: Estimated hypertension survival curves for the joint three-class skew-t growth mixture-survival model for BMI in the Framingham data



List of Tables

1	Results of fitting mixtures of normal and non-normals for BMI among 15-year old males in the NLSY	31
2	NLSY data on BMI for ages 12 to 23	32
3	Results of fitting normal and skew-t growth mixture models for BMI in the NLSY ages 12 to 23 for black females	33
4	Estimated skew and degrees of freedom parameters for the i , s , and q random effects of the 2-class skew-t growth mixture model for BMI in the NLSY ages 12 to 23 for black females	34
5	Results of fitting normal and skew-t growth mixture models for BMI in the Framingham data	35

Table 1: Results of fitting mixtures of normal and non-normals for BMI among 15-year old males 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			

Table 2: NLSY data on BMI for ages 12 to 23

Accelerated longitudinal design - NLSY97

	12	13	14	15	16	17	18	19	20	21	22	23
1997	1,165	1,715	1,847	1,868	1,709	613						
1998		104	1,592	1,671	1,727	1,739	1,400	106				
1999			108	1,659	1,625	1,721	1,614	1,370	65			
2000				57	1,553	1,656	1,649	1,597	1,390	132		
2001					66	1,543	1,615	1,602	1,582	1,324	109	
2002							1,614	1,587	1,643	1,582	1,324	106
2003							112	1,497	1,600	1,582	1,564	1,283
Totals	1,165	1,819	3,547	5,255	6,680	7,272	8,004	7,759	6,280	4,620	2,997	1,389

NLSY, National Longitudinal Survey of Youth

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

Table 3: Results of fitting normal and skew-t growth mixture models for BMI in the NLSY ages 12 to 23 for black females

No. Classes	Normal			Skew-t		
	Loglikelihood	No. Par's	BIC	Loglikelihood	No. Par's	BIC
1	-17,049	10	34,168	-15,617	25	31,411
2	-15,768	21	31,684	-15,510	29	31,225
3	-15,580	32	31,386	-15,480	44	31,270
4	-15,505	43	31,314			
5	-15,479	54	31,338			

Table 4: Estimated skew and degrees of freedom parameters for the i , s , and q random effects of the 2-class skew- t growth mixture model for BMI in the NLSY ages 12 to 23 for black females

Skew and Df Parameters			
Normal Class			
	Estimate	S.E.	Est/S.E.
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	S.E.	Est/S.E.
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

Table 5: Results of fitting normal and skew-t growth mixture models for BMI in the Framingham data

No. Classes	Normal			Skew-t		
	Loglikelihood	No. Par's	BIC	Loglikelihood	No. Par's	BIC
1	-8,252	8	16,557	-7,771	10	15,611
2	-7,954	13	15,995	-7,606	17	15,327
3	-7,875	18	15,871	-7,567	24	15,296
4	-7,787	23	15,730	-7,547	31	15,304
5	-7,743	28	15,674			