Technical Appendix: Methods and Results of Growth Mixture Modelling

(Supplement to: Trajectories of change in depression severity during treatment with antidepressants)

Rudolf Uher, Bengt Muthén, Daniel Souery, Ole Mors, Jan Jaracz, Anna Placentino, Ana Petrovic, Astrid Zobel, Neven Henigsberg, Marcella Rietschel, Katherine J. Aitchison, Anne Farmer, Peter McGuffin

Recent advances in statistical modelling based on mixture extension of the latent growth model make it possible to categorise subjects based on temporal patterns of change with latent variable methods such as growth mixture modelling (GMM) that can provide unbiased estimates of trajectories of change in the presence of missing data (Muthen & Asparouhov 2008; Beunckens *et al.* 2008). This appendix describes in details the application of GMM to data from the the Genome-based Therapeutic Drugs for Depression (GENDEP) project, a twelve-weeks part-randomized open-label study of depression treatment comparing two active antidepressant drugs.

Methods

Growth mixture modelling

The time course of change in depressive symptoms during treatment was modelled in 807 individuals with at least one observed post-baseline data point on the originally allocated medication. Dependent variables were MADRS total scores at baseline and one or more of the 12 weekly follow-up assessments. As it was our aim to externally test the classification, information on drug or genotypes was not included in the model. To establish the best classification based on the longitudinal pattern of change, we applied a series of growth mixture models. Growth mixture modelling (GMM) is a generalisation of repeated-measure mixed effect regression. In addition to random effects, GMM can account for subject heterogeneity in temporal patterns of change by latent classes corresponding to qualitatively distinct trajectories (Muthen & Asparouhov 2008; Beunckens *et al.* 2008). Stability of these models has been assessed through extensive simulations (Nylund *et al.* 2007). To identify the most parsimonious and interpretable solutions, we have explored several classes of GMM. We first fitted standard growth models with intercept and linear effect of time. Then we added quadratic polynomials of the time variable that allow curved trajectories. Finally, we fitted piece-wise models that allow modelling specific time-periods as distinct growth curves separated by transition points. Piece-wise models allow sharp bends at transition points in addition to straight or curved trajectories between the transition points (Duncan *et al.* 2006; Li *et al.* 2001; Muthen *et al.* 2008). Two-piece and three-piece models with several different transition points were explored. All models were fitted in the M*plus* program, version 5.1, using maximum likelihood estimation (Muthen & Muthen 2008). To ensure that the best solution corresponds to global optimum rather than a local maximum likelihood solution, we repeated the fitting procedure at least 300 times with different sets of random starting values and 50 final optimizations. Only solutions that were replicated with different starting values were accepted.

Model selection

We applied the Bayesian Information Criterion (BIC) to choose the best fitting a most parsimonious model (Schwartz, 1978). BIC, calculated from the maximised likelihood with a correction for number of parameters estimated in the model, is commonly used to select the best model, with smaller BIC indicating a better model and differences of 10 or more considered as evidence favouring one model over another (Raftery, 1995). We used the Lo-Mendell-Rubin likelihood ratio test of model fit to quantify the likelihood that the data can be described by a model with one-less class and a p value smaller than 0.05 indicating that the additional class significantly improves fit over a model with fewer classes (Lo *et al.* 2001). For the decision on the number of classes in the final model, the Lo-Mendell-Rubin test was confirmed by the more accurate and computationally demanding parametric bootstrapped likelihood ratio test with 100 draws (McLachlan & Peel 2000; Nylund *et al.* 2007). Finally, the entropy value was calculated for models with more than one class, to quantify the uncertainty of classification of subjects into latent classes. Entropy values range from 0 to 1, with 0 corresponding to randomness and 1 to a perfect classification (Celeux & Soromenho 1996). In addition to these formal criteria, the class sizes and interpretability were considered to select a model which would be applicable to the analyses of treatment trials. Specifically, we aimed for models with few large classes, as a latent class that includes only a small proportion (<10%) of subjects would have little utility in the analyses of treatment studies, unless it represents a qualitatively distinct subgroup of major interest.

Missing data

As expected in longitudinal studies with human participants, the GENDEP sample contained a significant proportion of missing data. Most of the missing data were monotone, due to 35% of participants terminating the study prematurely. Otherwise, the data were 93% complete. Maximum likelihood estimation is suitable for datasets containing missing values (as the present data do), as it allows using all available data without the need to impute missing values and provides unbiased estimates under the relatively unrestrictive "missing at random" (MAR) assumption (Little & Rubin 1987). In a previous exploration of missing data patterns in the GENDEP sample, it was concluded that the MAR assumption is reasonable (Uher *et al.* 2009).

Sensitivity analysis to assess effect of non-random allocation

GENDEP was a part-randomized study and 42% subjects were allocated to drug non-randomly. This could have introduced a bias into drug comparison. To assess if the GMM and drug comparisons are affected by any such bias, we have repeated the analyses, including GMM in the restricted sample of randomly allocated subjects.

Results

Latent trajectories of response to antidepressants

Standard growth models with linear and quadratic effects of time provided good representation of initial changes but tended to depart markedly from observed data in the second half of the twelve-week period. Addition of cubic effect of time or separation of the growth into two serial components in a piecewise growth model improved the fit. Of the simple mean curve linear mixed effect models, the model with a cubic growth factor (Model 3.1) provided the best fit. Overall, the best fit (lowest BIC) was achieved with piecewise growth mixture models separately modelling the change in weeks 0-2 and in weeks 3-12 with linear effect of time for the first period and linear and quadratic effects of time for the second period and allowing more than one class (**Table S1**). The M*plus* code for fitting this model is provided in a Text Box at the end of this Appendix. Lo-

Mendell-Rubin and bootstrapped likelihood-ratio tests showed a marked advantage of a mixture model with two latent classes over a simple linear mixed effect model (p<0.0001) and a marginal advantage for a three-class model over model with two classes (p=0.0445, **Table S1**). These two models provided comparable quality of classification of individuals into latent trajectory classes with entropy of 0.72.

The two-class piecewise growth mixture model (Model 4.2) separated the subjects into a larger class of subjects with relatively slow and approximately linear improvement over the 12-weeks (Class 1, Gradual improvement; 75% subjects) and a smaller class of subjects with rapid improvement over the first three weeks, followed by a more gradual improvement over the rest of the study period (Class 2, Rapid improvement; 25% subjects; **Table S2** and **Figure 1** (main manuscript)). The model estimates were stable across different sets of random starting values and all 50 final optimisations converged to the same solution. In the three-class piecewise growth mixture model (Model 4.3), a small third class separated with extremely rapid improvement and a floor effect after week three (6% subjects). The results of the three-class model are provided in **Table S3** and **Figure S5**. As it was our aim to find an alternative to dichotomous outcome measures, the advantage of three-class model was marginal, and the additional class included only a small proportion of subjects, we primarily explored the usefulness of the classification based on the two-class piecewise growth mixture Model 4.2. As there was some minor misspecification in the final three weeks with observed values being slightly lower than model estimates (Table S and Figures S5 and S6), we explored several three-piece models that would allow a different growth curve for the final 3-5 weeks. However, these models were less stable and did not improve model fit. The misspecifications and problems with fitting three-piece models were partly due to a larger proportion of missing values in the last four weeks.

Results of sensitivity analysis of randomly allocated subjects

To exclude an effect of selection bias, the GMM was repeated in a restricted sample of 460 subjects, who were randomly allocated to escitalopram (n=231) and nortriptyline (n=229). The results of GMM in this sample were similar to those based on the entire sample, differing on average by only 0.7 of a point on the MADRS scale (absolute average difference), corresponding to 0.1 of a standard deviation (see **Table S4** and **Figure S6** for details of the final model estimated in randomly allocated subjects. We conclude that GMM estimation was not unduly influenced by the inclusion of non-randomly allocated subjects.

References

Beunckens C, Molenberghs G, Verbeke G & Mallinckrodt C (2008). A latent-class mixture model for incomplete longitudinal Gaussian data. *Biometrics* **64**, 96-105.

Celeux G & Soromenho G (1996). An entropy criterion for assessing the number of clusters in a mixture model. *Journal of Classification* **13**, 195-212.

Duncan TE, Duncan SC & Strycker LA (2006). Piecewise and pooled interrupted time series LGMs. In *An introduction to latent variable growth curve modeling,* pp. 151-164. Lawrence Erlbaum Associates, Publishers: Mahwah, London.

Li F, Duncan TE, Duncan SC & Hops H (2001). Piecewise Growth Mixture Modeling of Adolescent Alcohol Use Data. *Structural Equation Modeling: A Multidisciplinary Journal* **8**, 175-204.

Little T & Rubin D. (1987). *Analysis with missing data.* John Wiley & Sons: New York.

Lo Y, Mendell NR & Rubin DB (2001). Testing the number of components in a normal mixture. *Biometrika* **88**, 778.

McLachlan GJ & Peel D. (2000). *Finite mixture models.* John Wiley & Sons: New York.

Muthen B & Asparouhov T (2008). Growth mixture modeling: Analysis with non-Gaussian random effects. In *Longitudinal Data Analysis*, (eds. G. Fitzmaurice, M. Davidian, G. Verbeke and G. Molenberghs), pp. 143-165. Chapman & Hall/CRC Press: Boca Raton.

Muthen B, Brown H, Leuchter A & Hunter A (2008). General approaches to analysis of course: Applying growth mixture modeling to randomized trials of depression medication. In *Causality and Psychopathology: Finding the Determinants of Disorders and their Cures*, (ed. P. E. Shrout), American Psychiatric Publishing: Washington, D.C.

Muthen B & Muthen L. (2008). *Mplus user's guide.* Muthen & Muthen: Los Angeles, CA.

Nylund KL, Asparouhov T & Muthen BO (2007). Deciding on the Number of Classes in Latent Class Analysis and Growth Mixture Modeling: A Monte Carlo Simulation Study. *Structural Equation Modeling: A Multidisciplinary Journal* **14**, 535-569.

Raftery AE (1995). Bayesian model selection in social research. *Sociological Methodology* **25**, 111-163.

Schwartz G (1978). Estimating the dimension of a model. *The Annals of Statistics* **5**, 461-464.

Uher R, Maier W, Hauser J, Marusic A, Schmael C, Mors O, Henigsberg N, Souery D, Placentino A, Rietschel M, Zobel A, Dmitrzak-Weglarz M, Petrovic A, Jorgensen L, Kalember P, Govannini C, Barreto M, Elkin A, Landau S, Farmer A, Aitchison KJ & McGuffin P (2009). Differential efficacy of escitalopram and nortriptyline on dimensional measures of depression in the GENDEP project. *British Journal of Psychiatry* **194**, 252-259.

Supplementary Table S1: Fit of growth mixture models. For each growth mixture model (GMM), the number of latent trajectory classes, maximum likelihood values and Bayesian Information Criterion (BIC) are given. For models with more than one latent class, entropy quantifies the quality of classification, Lo-Mendell-Rubin likelihood ratio tests whether the model is significantly improved compared to a simpler model with one class less, and proportion of individuals in each class is given.

Model	Classes	Likelihood	BIC	Entropy	Lo-Mendell-Rubin		Proportion of individuals in class			
					2LL	p	1	2	3	4
Linear GMM										
Model 1.1	1	-9809.2	19738.9				1.00			
Model 1.2	2	-9793.5	19727.5	0.65	31.4	0.0928	0.14	0.86		
Model 1.3	3	-9785.6	19731.9	0.66	15.8	0.0228	0.63	0.03	0.34	
Model 1.4	4	-9780.0	19740.6	0.73	11.3	0.2536	0.04	0.61	0.01	0.34
Quadratic GMM										
Model 2.1	1	-9262.8	18672.8				1.00			
Model 2.2	2	-9221.2	18616.5	0.75	80.1	0.0283	0.13	0.87		
Model 2.3	3	-9193.8	18588.4	0.81	52.8	0.0619	0.01	0.16	0.84	
Model 2.4	4	-9177.2	18581.9	0.71	32.1	0.2950	0.61	0.06	0.32	0.01
Cubic GMM										
Model 3.1	1	-9042.7	18266.2				1.00			
Model 3.2	2	-9004.3	18216.1	0.66	74.0	0.0919	0.20	0.80		
Model 3.3	3	-8984.0	18202.3	0.66	40.7	0.0157	0.30	0.64	0.06	
Model 3.4	4	-8972.2	18205.4	0.65	23.6	0.2600	0.07	0.54	0.31	0.08
Piecewise GMM (3+9)										
Model 4.1	1	-9127.1	18441.6				1.00			
Model 4.2	2	-8973.7	18175.1	0.71	306.7	0.0000	0.75	0.25		
Model 4.3	3	-8932.2	18132.1	0.71	83.1	0.0445	0.56	0.37	0.06	
Model 4.4	4	-8909.5	18126.9	0.74	45.4	0.2911	0.05	0.05	0.35	0.55

Supplementary Table S2: Estimated and observed means of the final two-class model. Estimated and observed mean values of MADRS scores and observed mean percentage improvement in MADRS scores are given for both latent classes.

	Class 1: gradual improvement			**Class 2: rapid initial improvement**		
	MADRS scores			MADRS scores		
Study week	Estimated	Observed	% improvement	Estimated	Observed	% improvement
0	28.55	28.72		28.55	28.94	
1	27.15	26.90	6.35	21.69	20.96	27.59
2	25.76	25.96	9.61	14.83	15.24	47.33
3	23.94	24.08	16.16	12.85	13.16	54.54
4	22.31	22.19	22.75	12.08	12.01	58.50
5	20.82	20.67	28.02	11.37	11.16	61.45
6	19.46	19.12	33.41	10.73	10.72	62.96
7	18.23	17.53	38.97	10.15	9.72	66.43
8	17.13	16.54	42.39	9.65	9.67	66.59
9	16.16	15.17	47.17	9.21	8.67	70.04
10	15.33	14.16	50.71	8.84	8.06	72.15
11	14.63	13.52	52.91	8.54	7.42	74.38
12	14.06	12.71	55.75	8.30	6.91	76.14

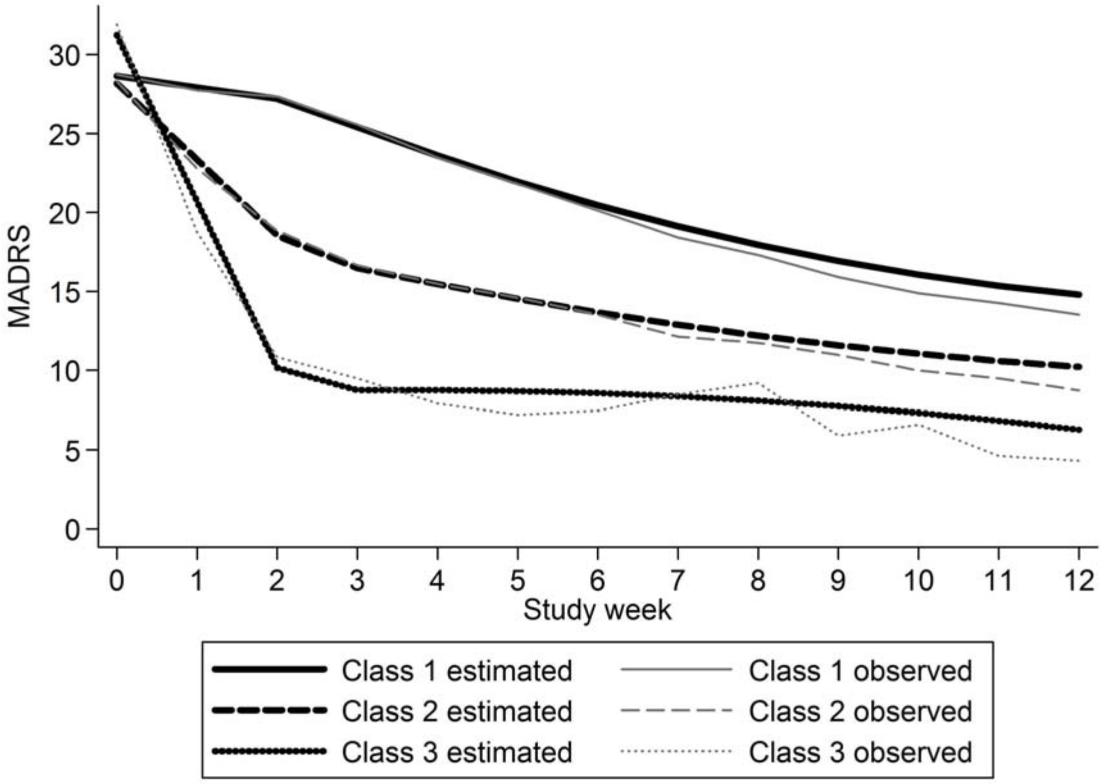
Supplementary Table S3: Three-class model. Estimated and observed mean values of MADRS scores and observed mean percentage improvement in MADRS scores are given for three latent classes in the two-piece three-class model (Model 4.3 in Table S1). The numbers of subjects in classes 1, 2 and 3 are 456, 301 and 60 respectively.

	Class 1: gradual improvement			Class 2: marked initial improvement			Class 3: very rapid initial improvement		
	MADRS scores			MADRS scores			MADRS scores		
Study week	Estimated	Observed	% improvement	Estimated	Observed	% improvement	Estimated	Observed	% improvement
0	28.64	28.72		28.17	28.32		31.23	31.91	
1	27.92	27.74	3.40	23.35	22.78	19.56	20.70	18.76	41.19
2	27.21	27.34	4.80	18.54	18.85	33.46	10.16	10.84	66.04
3	25.40	25.56	11.01	16.48	16.65	41.22	8.75	9.49	70.27
4	23.60	23.47	18.29	15.46	15.48	45.34	8.76	7.92	75.16
5	21.95	21.82	24.03	14.52	14.60	48.44	8.70	7.21	77.40
6	20.45	20.12	29.94	13.67	13.52	52.27	8.57	7.47	76.59
7	19.12	18.41	35.90	12.89	12.13	57.16	8.37	8.49	73.38
8	17.94	17.30	39.75	12.19	11.74	58.55	8.09	9.20	71.16
9	16.92	15.91	44.62	11.58	10.97	61.28	7.75	5.90	81.50
10	16.06	14.88	48.19	11.05	9.99	64.71	7.33	6.59	79.33
11	15.34	14.27	50.31	10.59	9.49	66.51	6.84	4.62	85.52
12	14.79	13.54	52.87	10.22	8.74	69.14	6.27	4.32	86.46

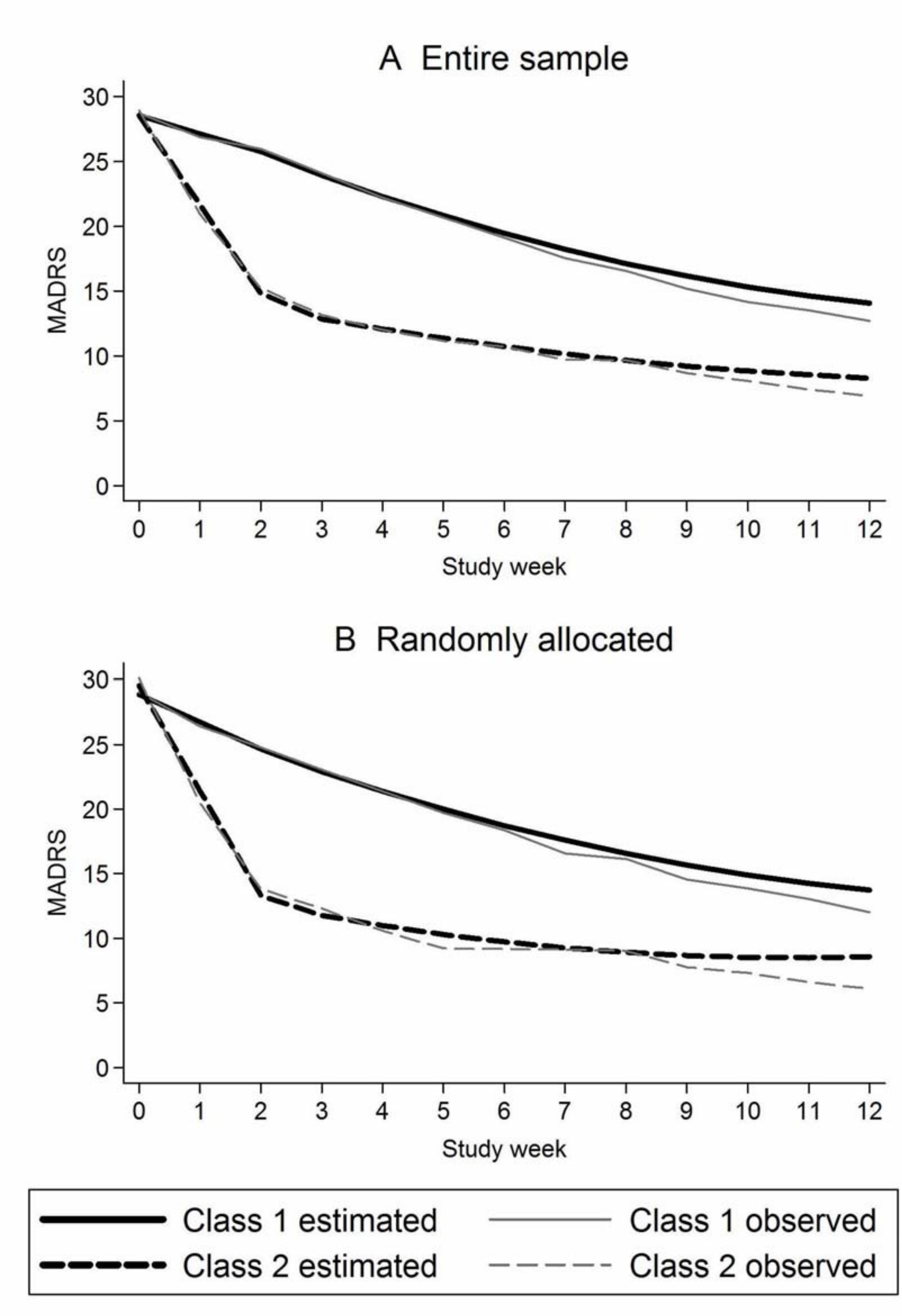
Supplementary Table S4: Sensitivity analysis of randomly allocated subjects. Estimated and observed means of the final two-class model estimated in 460 randomly allocated subjects. Estimated and observed mean values of MADRS scores and observed mean percentage improvement in MADRS scores are given for both latent classes.

	Class 1: gradual improvement			**Class 2: rapid initial improvement**		
	MADRS scores			MADRS scores		
Study week	Estimated	Observed	% improvement	Estimated	Observed	% improvement
0	28.83	29.03		29.55	30.08	
1	26.75	26.41	9.01	21.43	20.51	31.82
2	24.66	24.78	14.62	13.31	13.82	54.05
3	22.89	23.07	20.52	11.78	12.32	59.06
4	21.38	21.38	26.36	10.99	10.58	64.83
5	19.99	19.71	32.11	10.31	9.21	69.37
6	18.72	18.38	36.67	9.74	9.18	69.48
7	17.58	16.59	42.84	9.27	9.12	69.68
8	16.57	16.12	44.46	8.92	9.03	69.98
9	15.67	14.56	49.84	8.68	7.78	74.13
10	14.90	13.86	52.24	8.54	7.33	75.62
11	14.24	13.04	55.07	8.51	6.62	77.99
12	13.72	12.00	58.67	8.59	6.10	79.73

Supplementary Figure S5: The two-piece three-class model. Model estimates are printed as bold black lines. Observed average values based on most likely class are printed in fine gray lines. Note that mismatch between estimated and observed values in the later weeks is partly due to missing values.



Supplementary Figure S6: Comparison of the final two-class model estimated from the entire sample (A) and from a reduced sample of 460 randomly allocated subjects (B). Model estimates are printed as bold black lines. Observed average values based on most likely class are printed in fine gray lines. Note that mismatch between estimated and observed values in the later weeks is partly due to missing values.



Text Box: The M*plus* code for fitting the final piece-wise growth mixture model with two classes (Model 4.2).

```
TITLE:      TWO-PIECE GROWTH MIXTURE MODEL
            MADRS SCORE WEEKS 0-12 WITH TWO CLASSES

DATA:       FILE IS "C:\GJ\m.dat";

VARIABLE:   NAMES ARE d0 d1 d2 d3 d4 d5 d6 d7 d8 d9
            d10 d11 d12;
            MISSING IS .;
            CLASSES = c(2)

ANALYSIS:   ESTIMATOR = MLR;
            TYPE IS MIXTURE;
            STARTS = 300 50;
            K-1STARTS = 40 10;
            PROCESS = 2 (STARTS);

MODEL:      %OVERALL%
            i1 s1 | d0@0 d1@0.1 d2@0.2;
            i2 s2 q2 | d3@-0.9 d4@-0.8 d5@-0.7 d6@-0.6 d7@-0.5
             d8@-0.4 d9@-0.3 d10@-0.2 d11@-0.1 d12@0;
            s1@0;

OUTPUT:     RESIDUAL CINTERVAL TECH1 TECH7 TECH8 TECH11 TECH14;

PLOT:       TYPE IS PLOT1 PLOT2 PLOT3;
            SERIES = d0 d1 d2 d3 d4 d5 d6 d7 d8 d9 d10 d11
            d12(*);

SAVEDATA:   FILE IS "C:\GJ\md_3+9_2cl_cprob.dat";
            SAMPLE IS "C:\GJ\md_3+9_2cl_samp.dat";
            SAVE = CPROBABILITIES;
```