Technical Appendix

The Behavior of Growth Mixture Models Under Nonnormality: A Monte Carlo Analysis

Daniel J. Bauer & Patrick J. Curran

10/11/2002

These results are presented as a companion to the manuscript:

Bauer, D. J. & Curran, P. J. (in press). Distributional Assumptions of Growth Mixture Models: Implications for Over-Extraction of Latent Trajectory Classes. Forthcoming in *Psychological Methods*.

Further detail on the derivation of the hypotheses, design of the Monte Carlo, and all references may be found in this manuscript.

Table of Contents

Hypotheses and Method

Hypotheses:

(1) Obtaining convergence and a proper solution for a two class growth mixture model would be more difficult if data were drawn from a single group multivariate normal distribution than if they were drawn from a single group multivariate nonnormal distribution.

(2) Conventional model fit statistics would support the estimation of two (or more) trajectory classes if the data were drawn from a single group multivariate nonnormal distribution but not if they were drawn from a single group multivariate normal distribution.

(3) Estimating latent classes that do not correspond to true groups in the population could obscure the role of significant predictors of individual change, or identify spurious effects.

Method:

Data were generated to be consistent with the single group model in Figure 1. Five hundred samples at each of two sample sizes, N=200 and N=600, were generated for three distributional conditions. In the first condition, the data were generated to be normally distributed (i.e., with univariate skew 0 and kurtosis 0). The other two conditions involved transformations of the repeated measures data using Fleishman's (1978) method for generating nonnormal random variables, as extended by Vale and Maurelli (1983). Specifically, in these conditions the repeated measures data were transformed to have univariate skew 1 and kurtosis 1, and skew 1.5 and kurtosis 6, respectively. (In conditional models the covariate was generated from a normal distribution in all conditions). All models were estimated in Mplus 2.01, employing the EM estimator with the MLR option to obtain robust standard errors (Muthén & Muthén, 1998). A modified version of the RUNALL utility was used to compile the results (Nguyen, Muthén & Muthén, 2001).

Finite normal mixture models are known to have poorly behaved likelihood functions (McLachlan & Peel, 2001). For this reason, two class models were estimated both with and without across-class equality constraints on the variance components (e.g., $\Psi_k=\Psi$ and $\Theta_k=\Theta$) though these constraints are often not optimal from the standpoint of substantive theory. Second, to avoid obtaining local solutions, all two class models were estimated with six sets of start values. One set of start values was derived from the parameter estimates obtained from single group models (per Muthén & Muthén, 1998, p. 132). The single group population parameter estimates were used as start values for all of the parameters except the growth factor means, which were set higher in one group than the other for both growth factors (μ_α=1.50 and μ_β=1.60 for Class 1 and μ_α=.00 and μ_β=.00 for Class 2). The other five sets of start values were generated randomly by taking for each parameter a random draw from a normal distribution with mean equal to the single-group population value for the parameter and a standard deviation set to provide broad coverage of the surrounding parameter space. Our use of random start values is consistent with other simulation studies on finite normal mixtures (e.g., Biernacki, Celeux & Govaert, 1999; McLachlan & Peel, 2000, p. 217).

The model was allowed 1000 iterations to converge. We adopted the following algorithm for selecting solutions for analysis:

(1) When a given replication failed to converge with any of the six sets of start values, the solution was labeled nonconvergent.

(2) When more than one set of start values lead to convergence for a given replication, the solution with the maximum (best) log-likelihood was selected. This again follows standard practice in studies of finite normal mixtures (Biernacki, Celeux & Govaert, 1999; Everitt & Hand, 1981; McLachlan & Peel, 2000, p. 217).

(3) The solution selected from Step 2 was considered "improper" if any of the parameter estimates fell outside of their permissible boundaries (i.e., negative variances, or correlations greater than one).

Unless convergence was of explicit interest, nonconverged and improper solutions were excluded from the analyses since such solutions are rarely interpreted in practice (Chen et al., 2001). Additional analyses including improper solutions did not show meaningful differences from the results reported here.

(Further detail and all references may be obtained from the original manuscript)

Population Model

Figure 1.
Path diagram of a single group latent trajectory model. Displayed numbers are the population values of the parameters used in the simulation study.

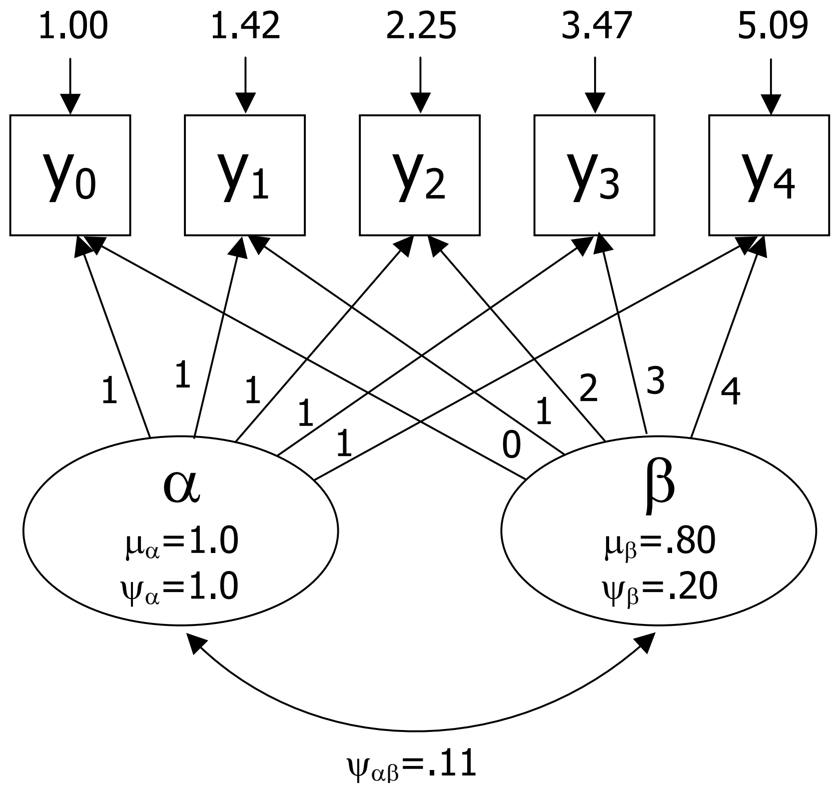

Table 1.
Convergence of 1 Class Unconditional Model

	Distribution			Converged	
N	Skew	Kurtosis	Failed to Converge	Improper Solution	Proper Solution
200	0	0	0	5 (1%)	495 (99%)
200	1	1	0	4 (1%)	496 (99%)
200	1.5	6	0	8 (2%)	492 (98%)
600	0	0	0	0	500 (100%)
600	1	1	0	0	500 (100%)
600	1.5	6	0	0	500 (100%)

Table 2.
Convergence of 2 Class Unconditional Model with Class-Invariant Variance & Covariance Parameters

	Distribution			Converged	
N	Skew	Kurtosis	Failed to Converge	Improper Solution	Proper Solution
200	0	0	6 (1%)	193 (39%)	301 (60%)
200	1	1	0	232 (46%)	268 (54%)
200	1.5	6	0	150 (30%)	350 (70%)
600	0	0	23 (5%)	108 (22%)	369 (74%)
600	1	1	0	211 (42%)	289 (58%)
600	1.5	6	0	74 (15%)	426 (85%)

Table 3.
Convergence of 2 Class Unconditional Model with Class-Varying Variance & Covariance Parameters

	Distribution			Converged	
N	Skew	Kurtosis	Failed to Converge	Improper Solution	Proper Solution
200	0	0	4 (1%)	450 (90%)	46 (9%)
200	1	1	0	170 (34%)	330 (66%)
200	1.5	6	0	162 (32%)	338 (68%)
600	0	0	31 (6%)	380 (76%)	89 (18%)
600	1	1	0	29 (6%)	471 (94%)
600	1.5	6	0	15 (3%)	485 (97%)

Table 4.
Likelihood Ratio Test of Invariance Constraints on Variance and Covariance Parameters in 2 Class Unconditional Model (Proper Solutions Only)

	Distribution				Likelihood Ratio Test [a]	
N	Skew	Kurtosis	Pairs available	Mean χ^2	$p < .05$	$p \geq .05$
200	0	0	36	14.27	14 (38.89%)	22 (61.11%)
200	1	1	185	123.50	185 (100%)	0
200	1.5	6	241	170.25	241 (100%)	0
600	0	0	76	13.70	29 (38.16%)	51 (66.23%)
600	1	1	279	334.56	279 (100%)	0
600	1.5	6	416	501.34	416 (100%)	0

[a] *Likelihood Ratio χ^2 test has 8 df.*

Table 5.

Relative Fit of 1-Class v. 2-Class Unconditional Model With Class-Invariant Variance & Covariance Parametes: Proper Solutions Only (of 500 samples) at N=200.

Fit Statistic	% of time favors 2-class model	Mean Difference[a]	Mean % Change in Fit Stat[a]
Skew 0, Kurtosis 0 (301 of 500 Samples)			
AIC	25.58%	-1.32	-0.03%
CAIC	.33%	-14.21	-0.35%
BIC	.66%	-11.21	-0.27%
Sample Size Adjusted BIC	21.59%	-1.71	-0.04%
CLC	5.98%	-97.93	-2.42%
NEC	5.98%	-36.65	-3665.09%
ICL-BIC	0%	-113.82	-2.77%
Skew 1, Kurtosis 1 (265 of 500 Samples)			
AIC	70.57%	24.06	.58%
CAIC	62.26%	11.17	.26%
BIC	64.15%	14.17	.34%
Sample Size Adjusted BIC	69.43%	23.67	.57%
CLC	36.98%	-13.60	-.34%
NEC	62.64%	-.58	-58.26%
ICL-BIC	24.91%	-29.49	-.73%
Skew 1.5, Kurtosis 6 (343 of 500 Samples)			
AIC	70.85%	37.39	.90%
CAIC	65.60%	24.50	.58%
BIC	67.06%	27.50	.65%
Sample Size Adjusted BIC	70.55%	37.00	.89%
CLC	67.06%	28.79	. 69%
NEC	95.04%	.92	91.66%
ICL-BIC	60.06%	12.89	.29%

[a] Mean difference calculated as Fit1-Fit2 where Fit1 and Fit2 are the values of the statistic for the 1- and 2-class models. Percent change calculated as (1-Fit2/Fit1)*100. Positive values indicate that the fit statistic decreased (e.g., improved) by moving to the 2-class model. Negative values indicate worse fit of the 2-class model relative to the 1-class model.

Table 6.
Relative Fit of 1-Class v. 2-Class Unconditional Model With Class-Invariant Variance & Covariance Parametes: Proper Solutions Only (of 500 samples) at N=600.

Fit Statistic	% of time favors 2-class model	Mean Difference[a]	Mean % Change in Fit Stat[a]
Skew 0, Kurtosis 0 (369 of 500 Samples)			
AIC	25.93%	-1.47	-0.01%
CAIC	0%	-17.66	-0.14%
BIC	0%	-14.66	-0.12%
Sample Size Adjusted BIC	4.88%	-5.14	-0.04%
CLC	1.90%	-402.10	-3.30%
NEC	1.90%	-174.83	-17482.70%
ICL-BIC	0%	-421.29	-3.44%
Skew 1, Kurtosis 1 (289 of 500 Samples)			
AIC	82.01%	70.58	.58%
CAIC	75.78%	54.38	.44%
BIC	77.16%	57.38	.47%
Sample Size Adjusted BIC	80.97%	66.91	.55%
CLC	16.96%	-83.52	-.69%
NEC	34.26%	1.91	190.83%
ICL-BIC	10.73%	-102.72	-.84%
Skew 1.5, Kurtosis 6 (426 of 500 Samples)			
AIC	75.59%	99.13	.81%
CAIC	73.00%	82.94	.67%
BIC	73.94%	85.94	.70%
Sample Size Adjusted BIC	75.12%	95.47	.78%
CLC	68.54%	56.45	.46%
NEC	92.02%	1.92	191.74%
ICL-BIC	63.85%	37.25	.30%

[a] Mean difference calculated as Fit1-Fit2 where Fit1 and Fit2 are the values of the statistic for the 1- and 2-class models. Percent change calculated as (1-Fit2/Fit1)*100. Positive values indicate that the fit statistic decreased (e.g., improved) by moving to the 2-class model. Negative values indicate worse fit of the 2-class model relative to the 1-class model.

Table 7.
Relative Fit of 1-Class v. 2-Class Unconditional Model With Class-Varying Variance & Covariance Parametes: Proper solutions Only (of 500 samples) at N=200.

Fit Statistic	% of time favors 2-class model	Mean Difference[a]	Mean % Change in Fit Stat[a]
Skew 0, Kurtosis 0 (46 of 500 Samples)			
AIC	32.61%	-2.04	-0.05%
CAIC	0%	-49.33	-1.2%
BIC	0%	-38.33	-0.94%
Sample Size Adjusted BIC	26.09%	-3.48	-0.09%
CLC	0%	-118.23	-2.93%
NEC	0%	-6.93	-692.75%
ICL-BIC	0%	-176.51	-4.31%
Skew 1, Kurtosis 1 (329 of 500 Samples)			
AIC	100%	133.54	3.28%
CAIC	99.70%	86.25	2.09%
BIC	99.70%	97.25	2.37%
Sample Size Adjusted BIC	100%	132.10	3.24%
CLC	98.48%	83.39	2.06%
NEC	98.48%	.50	50.42%
ICL-BIC	69.60%	25.11	.61%
Skew 1.5, Kurtosis 6 (334 of 500 Samples)			
AIC	100%	191.03	4.70%
CAIC	100%	143.74	3.49%
BIC	100%	154.74	3.77%
Sample Size Adjusted BIC	100%	189.59	4.66%
CLC	99.10%	142.75	3.52%
NEC	99.10%	.63	63.49%
ICL-BIC	92.51%	84.47	2.04%

[a] Mean difference calculated as Fit1-Fit2 where Fit1 and Fit2 are the values of the statistic for the 1- and 2-class models. Percent change calculated as (1-Fit2/Fit1)*100. Positive values indicate that the fit statistic decreased (e.g., improved) by moving to the 2-class model. Negative values indicate worse fit of the 2-class model relative to the 1-class model.

Table 8.
Relative Fit of 1-Class v. 2-Class Unconditional Model With Class-Varying Variance & Covariance Parametes: Proper solutions Only (of 500 samples) at N=600.

Fit Statistic	% of time favors 2-class model	Mean Difference[a]	Mean % Change in Fit Stat[a]
Skew 0, Kurtosis 0 (89 of 500 Samples)			
AIC	21.35%	-3.63	-0.03%
CAIC	0%	-63.00	-0.51%
BIC	0%	-52.00	-0.42%
Sample Size Adjusted BIC	0%	-17.08	-0.14%
CLC	0%	-445.39	-3.66%
NEC	0%	-26.43	-2643.08%
ICL-BIC	0%	-515.76	-4.21%
Skew 1, Kurtosis 1 (471 of 500 Samples)			
AIC	100%	390.64	3.20%
CAIC	100%	331.27	2.70%
BIC	100%	342.27	2.79%
Sample Size Adjusted BIC	100%	377.19	3.09%
CLC	98.73%	173.34	1.42%
NEC	98.73%	.41	40.54%
ICL-BIC	91.08%	102.97	0.84%
Skew 1.5, Kurtosis 6 (485 of 500 Samples)			
AIC	100%	585.62	4.80%
CAIC	100%	526.25	4.29%
BIC	100%	537.25	4.39%
Sample Size Adjusted BIC	100%	572.18	4.68%
CLC	100%	389.14	3.19%
NEC	100%	.62	62.34%
ICL-BIC	99.18%	318.77	2.60%

[a] Mean difference calculated as Fit1-Fit2 where Fit1 and Fit2 are the values of the statistic for the 1- and 2-class models. Percent change calculated as (1-Fit2/Fit1)*100. Positive values indicate that the fit statistic decreased (e.g., improved) by moving to the 2-class model. Negative values indicate worse fit of the 2-class model relative to the 1-class model.

Table 9.
Expected Value for Parameter Estimates Compared With the Mean Value of the Model Parameter Estimates (Empirical SE, Mean Estimated SE): Proper Solutions Only (of 500 Samples) at N=200.

			2 Class Model[a]	
Parameter	Population	1 Class Model	Class 1	Class 2
Skew 0, Kurtosis 0		(495 Samples)	(46 Samples)	
μ_α	1.00	1.00 (.09, .09)	1.14 (.40, 51)	.70 (.38, 34)
μ_β	.80	.80 (.05, .05)	.97 (.21, 21)	.62 (.22, 19)
ψ_α	1.00	1.00 (.20, .20)	.94 (.57, .61)	.81 (.39, .49)
ψ_β	.20	.20 (.05, .05)	.18 (.13, .13)	.15 (.09, .13)
$\psi_{\alpha\beta}$	.11	.11 (.08, .07)	-.07 (.20, .23)	-.09 (.48, .16)
$CORR_{\alpha\beta}$	.25	.27	-.09	.23
% Cases	100%	100%	48.3%	51.7%
Skew 1, Kurtosis 1		(496 Samples)	(330 Samples)	
μ_α	1.00	1.00 (.09, .09)	1.48 (.25, .19)	.22 (.23, .19)
μ_β	.80	.79 (.05, .05)	.98 (.16, .10)	.53 (.15, .10)
ψ_α	1.00	.99 (.22, .22)	.79 (.35, .36)	.24 (.15, .16)
ψ_β	.20	.20 (.05, .05)	.20 (.09, .09)	.05 (.03, .04)
$\psi_{\alpha\beta}$	.11	.11 (.08, .08)	-.06 (.13, .14)	-.03 (.04, .06)
$CORR_{\alpha\beta}$	.25	.28	-.08	-.23
% Cases	100%	100%	58.8%	41.2%
Skew 1.5, Kurtosis 6		(492 Samples)	(338 Samples)	
μ_α	1.00	1.00 (.09, .09)	1.85 (.47, .38)	.63 (.10, .12)
μ_β	.80	.79 (.05, .05)	1.07 (.24, .19)	.75 (.20, .06)
ψ_α	1.00	.99 (.28, .26)	1.22 (.81, .83)	.39 (.14, .17)
ψ_β	.20	.20 (.06, .06)	.35 (.21, .24)	.09 (.03, .04)
$\psi_{\alpha\beta}$	.11	.10 (.09, .09)	-.16 (.31, .35)	.01 (.05, .06)
$CORR_{\alpha\beta}$	.25	.28	-.19	.13
% Cases	100%	100%	26.7%	73.3%

[a] Estimated With Class-Varying Variance and Covariance Parameters.

Table 10.
Expected Value for Parameter Estimates Compared With the Mean Value of the Model Parameter Estimates (Empirical SE, Mean Estimated SE): Proper Solutions Only (of 500 Samples) at N=600.

			2 Class Model[a]	
Parameter	Population	1 Class Model	Class 1	Class 2
Skew 0, Kurtosis 0		(500 Samples)	(89 Samples)	
μ_α	1.00	1.00 (.05, .05)	1.14 (.41, .42)	.74 (.42, .42)
μ_β	.80	.80 (.02, .03)	.92 (.23, 28)	.68 (.21, .19)
ψ_α	1.00	1.00 (.11, .12)	.86 (.46, .51)	.85 (.38, .52)
ψ_β	.20	.20 (.03, .03)	.19 (.09, .12)	.18 (.10, .12)
$\psi_{\alpha\beta}$	.11	.11 (.04, .04)	.05 (.18, .19)	.06 (.15, .16)
$CORR_{\alpha\beta}$	.25	.26	.17	.18
% Cases	100%	100%	48.6%	51.4%
Skew 1, Kurtosis 1		(500 Samples)	(471 Samples)	
μ_α	1.00	1.00 (.05, .05)	1.50 (.12, .12)	.17 (.13, .13)
μ_β	.80	.80 (.03, .03)	.99 (.06, .06)	.51 (.06, .07)
ψ_α	1.00	.99 (.12, .13)	.79 (.21, .21)	.19 (.08, .09)
ψ_β	.20	.20 (.03, .03)	.21 (.05, .06)	.05 (.02, .02)
$\psi_{\alpha\beta}$	.11	.11 (.05, .05)	-.06 (.08, .08)	-.01 (.03, .03)
$CORR_{\alpha\beta}$	.25	.26	-.11	-.07
% Cases	100%	100%	60.7%	39.3%
Skew 1.5, Kurtosis 6		(500 Samples)	(485 Samples)	
μ_α	1.00	1.00 (.05, .05)	1.99 (.25, .22)	.65 (.06, .06)
μ_β	.80	.80 (.03, .03)	1.16 (.12, .11)	.69 (.09, .03)
ψ_α	1.00	.98 (.16, .15)	1.19 (.48, .50)	.40 (.08, .08)
ψ_β	.20	.20 (.04, .04)	.36 (.13, .13)	.09 (.02, .02)
$\psi_{\alpha\beta}$	.11	.11 (.05, .05)	-.15 (.21, .20)	.02 (.03, .03)
$CORR_{\alpha\beta}$	.25	.27	-.19	.13
% Cases	100%	100%	25.3%	74.7%

[a] Estimated With Class-Varying Variance and Covariance Parameters.

Table 11.

Population Values of Model Parameters Relating a Predictor to the Intercept and Slope Factors Compared With the Mean Value of the Parameter Estimates (Empirical SE, Mean Estimated SE) Obtained From 1- and 2-Class Models: Proper Solutions Only (of 500 Samples) at N=200.

Parameter	Population	1 Class Model	2 Class Model With Equality Constraints[a]	2 Class Model Without Constraints[b] Class 1	Class 2
Skew 1, Kurtosis 1		(473 samples)	(281 Samples)	(265 Samples)	
γ_1	.125	.122	.097	.140	.074
		(.029, .028)	(.027, .027)	(.043, .045)	(.034, .046)
γ_2	-.030	-.030	-.031	-.042	-.023
		(.015, .015)	(.012, .014)	(.023, .025)	(.018, .020)
% Cases	100%	100%	58.6% / 41.4%	59.6%	40.4%
Skew 1.5, Kurtosis 6		(465 Samples)	(291 Samples)	(267 Samples)	
γ_1	.125	.122	.096	.181	.085
		(.029, .029)	(.024, .024)	(.086, .086)	(.027, .028)
γ_2	-.030	-.030	-.028	-.054	-.024
		(.015, .014)	(.013, .012)	(.049, .048)	(.014, .014)
% Cases	100%	100%	26.7% / 73.3%	27.4%	72.6%

[a] Parameters γ_1 and γ_2 constrained to be equal across classes, variance and covariance parameters permitted to vary over classes.

[a] All parameters permitted to vary over classes.

Table 12.

Population Values of Model Parameters Relating a Predictor to the Intercept and Slope Factors Compared With the Mean Value of the Parameter Estimates (Empirical SE, Mean Estimated SE) Obtained From 1- and 2-Class Models: Proper Solutions Only (of 500 Samples) at N=600.

Parameter	Population	1 Class Model	2 Class Model With Equality Constraints[a]	2 Class Model Without Constraints[b] Class 1	Class 2
Skew 1, Kurtosis 1		(499 samples)	(453 Samples)	(437 Samples)	
γ_1	.125	.123	.096	.139	.070
		(.017, .016)	(.015, .015)	(.024, .025)	(.019, .021)
γ_2	-.030	-.030	-.031	-.044	-.021
		(.009, .008)	(.007, .007)	(.014, .013)	(.010, .011)
% Cases	100%	100%	59.5% / 40.5%	60.0%	40.0%
Skew 1.5, Kurtosis 6		(499 Samples)	(473 Samples)	(458 Samples)	
γ_1	.125	.123	.098	.189	.087
		(.017, .017)	(.015, .014)	(.049, .049)	(.016, .016)
γ_2	-.030	-.030	-.029	-.061	-.025
		(.009, .008)	(.007, .007)	(.029, .028)	(.008, .008)
% Cases	100%	100%	25.1% / 74.9%	25.5%	74.5%

[a] Parameters γ_1 and γ_2 constrained to be equal across classes, variance and covariance parameters permitted to vary over classes.

[a] All parameters permitted to vary over classes.

Table 13.

Evaluation of the effect of the covariate when treated as a within-class predictor of individual variability in intercepts and slopes: Table gives the percent of replications converging on a proper solution where the effect of the covariate on individual intercepts (γ_1) and slopes (γ_2) was significant at $p < .05$ (and in the same direction as the effect in the population).

	N=200		N=600	
	γ_1	γ_2	γ_1	γ_2
One-Class Model				
Skew 1, Kurtosis 1	99%	55%	100%	94%
Skew 1.5, Kurtosis 6	99%	54%	100%	93%
Two-Class Model				
Skew 1, Kurtosis 1				
Class 1	88%	49%	100%	90%
Class 2	57%	27%	93%	55%
Skew 1.5, Kurtosis 6				
Class 1	59%	23%	97%	61%
Class 2	87%	42%	100%	89%

Table 14.

Evaluation of the effect of the covariate when treated as a class predictor in a two class model.

	Distribution		Proper Solutions		Mean	% of Replications
N	Skew	Kurtosis	(of 500 Samples)	Mean Logit[a]	Odds-Ratio	Effect was NS
200	1	1	333 (67%)	.11 (.083, .080)	1.12	70.0%
200	1.5	6	330 (66%)	.11 (.077, .081)	1.12	73.9%
600	1	1	475 (95%)	.11 (.044, .044)	1.12	24.6%
600	1.5	6	489 (98%)	.10 (.044, .044)	1.11	32.9%

[a] Numbers in parentheses correspond to the empirical standard error and average estimated standard error of the logit.