Power Estimation of Slope Growth Factors in the NZAVS using Monte Carlo Simulation

Sibley, C. G., & Milojev, P. (2014). Power Estimation of Slope Growth Factors in the NZAVS using Monte Carlo Simulation. *NZAVS Technical Documents, e19.*

Available at: http://www.psych.auckland.ac.nz/uoa/NZAVS

Document last updated 8 December 2014

Funding statement

This research was supported by a Templeton World Charity Foundation Grant (ID: 0077).

Overview

This technical report describes results from a series of models we have developed to estimate the statistical power of latent growth parameters in the NZAVS using Monte Carlo simulation. The models are tailored specifically for the NZAVS, and the simulated data approximate the actual NZAVS data structure. Specifically, our simulations approximate the NZAVS datasets in terms of sample size for previous waves (Waves 1-5), projected sample size for future waves (Waves 6-10), and the structure of missing data introduced by both sample attrition and booster sampling occurring at various points during the first five years of the study.

Sample size in these models is thus known, or can be predicted with reasonable accuracy, as can missing data (with some exceptions). These parameters are thus held constant in our models. This allows us to estimate the statistical power that the NZAVS can be expected to achieve under the following conditions:

(a) Depending on the effect size for the latent growth parameter of interest,

(b) given other reasonable default parameter settings in latent growth models of NZAVS data,

(c) the complexity of the latent growth model (e.g., number of time-invariant covariates, estimation of second-order polynomial curves, etc), and

(d) depending on whether the model is tested using the currently available (as of 2014) first five waves of the NZAVS, or whether one waits for a given number of subsequent waves of data to be collected.

It is this last parameter (d) that is of critical interest to us, because holding other parameters constant, our simulations can thus provide projections about how many waves of data are needed to test latent growth curve models of varying complexity and still achieve power of .80 (a typical default for acceptable power) to detect the parameter of interest given the expected size of that parameter.

Power Estimation Using Monte Carlo Simulation in Mplus: A Brief Overview

We adapted the procedures outlined by Muthén and Muthén (2002) to estimate statistical power. Muthén and Muthén (2002, p. 8) provide the following summary of power estimation in Mplus, and we repeat it here:

Power is evaluated using the information in the column "labelled % Sig Coeff. This column gives the proportion of replications for which the null hypothesis that a parameter is equal to zero is rejected for each parameter at the .05 level (two-tailed test with a critical value of 1.96). The statistical test is the ratio of the parameter estimate to its standard error, an approximately normally distributed quantity (z-score) in large samples. For parameters with population values different from zero, this value is an estimate of power, that is, the probability of rejecting the null hypothesis when it is false. For parameters with population values equal to zero, this value is an estimate of Type I error, that is, the probability of rejecting the null hypothesis when it is true."

We followed the general recommendations outlined by Muthén and Muthén (2002) for determining adequate models when estimating desired sample size. Muthén and Muthén (2002, p. 8) summarized these as follows, and we apply them to our estimates of power when varying different model parameters (rather than to estimate sample size): "The first criterion is that parameter and standard error biases do not exceed 10 percent for any parameter in the model. The second criterion is that the standard error bias for the parameter for which power is being assessed does not exceed 5 percent. The third criterion is that coverage remains between 0.91 and 0.98."

Known and Projected NZAVS Data Characteristics

Because the first 5 waves of the NZAVS have been collected, we know quite a lot already about the structure of the data and thus aimed to approximate the actual NZAVS data structure in our models. As of the end of Wave 5, the NZAVS contains ~23,000 unique participants. There are currently about 3500 'loyal stayers' in the NZAVS, that is, people who have participated since Wave 1, and have completed all of the first 5 Waves. Many more people have completed at least 4 of the first 5 waves.

In 2009, at Wave I, the NZAVS randomly sampled a total of 6,518 registered voters from the New Zealand electoral roll.

In 2010, the NZAVS sampled 4,423 people retained from Wave I (retention rate from Wave I = 68%).

In 2011, the NZAVS sampled 6,884 New Zealanders, with 3,916 people retained from Wave I and a booster sample of 2,961 new participants (retention rate from Wave I = 60%, wave-to-wave retention from previous year = 80%).

In 2012, the NZAVS sampled 12,182 New Zealanders, with 4,054 people retained from Wave I and a booster sample of 5,377 new participants (retention rate from Wave I = 62%, wave-to-wave retention from previous year = 84%).

In 2013, the NZAVS sampled 18,211 New Zealanders, with 3,941 people retained from Wave I and a booster sample of 7,639 new participants (retention rate from Wave I = 61%, wave-to-wave retention from previous year = 81%).

Sibley (2014, NZAVS Tech Docs, e18) derived Eq 1.0 predicting sample retention from Wave 6 onwards:

$$y_w = 22764e^{-0.223w} \qquad \text{(Eq. 1.0)}$$

This simple model summarizes a rate of decay in the number of people sampled holding constant a wave-by-wave retention rate of 80%, where: y = the predicted number of people retained in the NZAVS at a given wave (w). For simplicity, Wave 1 in this model represents Time 5 as this was the current wave of the sample from which this model was derived. So setting w = 2 estimates the projected number of people who should complete Wave 6 (or the 2014 phase), w = 3 estimates the projected number of people who should complete Wave 7 (or the 2015 phase), and so forth. Sample projections are thus as follows:

Table 1. Projected sample sizes (from Sibley, 2014).

wave	year	w	sample projection
Wave 5	2013	1	18214
Wave 6	2014	2	14573
Wave 7	2015	3	11660
Wave 8	2016	4	9329
Wave 9	2017	5	7465
Wave 10	2018	6	5973

Data Simulation

All models were estimated using Mplus 7.3. Given the known and projected sample details outlined above, we generated simulation data with the following parameters:

- N = 23,000
- Variables are normally distributed.
- No missing data for 13% of cases. This ensures that the minimum possible number of loyal stayers is 3000. This is a clear underestimate for models only using data from Wave 1 – 5. Given random data generated for the remaining 87% of the sample, simulation data for Wave 1 – 5 tended to give a number of loyal stayers of ~3400-3500. Setting this value of 3000 also allows for a decrease in loyal stayers in subsequent waves.
- Data were generated with Missing at Random (MAR) for the remaining 87% of the total N. To allow for variation in sample size across time, we specified specific estimates of the proportion of missing data in this 87% at each wave. Note that this did not simply apply the decay function equation outlined in Eq. 1 because the data include independent booster samples conducted during Waves 3, 4 and 5.

- Table 2 presents the proportions of data missing completely at random that was specified for each wave for the 87% for which missing data were allowed, and the 13% which (by definition) did not have missing data. For example, at Wave 1, we calculated the proportion of missing data for the 87% for which missing data were allowed using Eq. 2.0:

  $$1 - (\text{Wave N} - \text{Loyal Stayer N}) / \text{Grand N}, \text{ or} \qquad \text{(Eq. 2.0)}$$

  $$1 - (6518 - 3000) / 23000$$

- For Wave 1 – 5, Wave N was a known value. For Waves 6 -10, Wave N was an estimate derived from Eq 1.0

Table 2. Proportions of missing data estimated for each wave (T1 – T5 based on known values, T6-10 based on sample projections).

Grand N	23000.000									
Loyal Stayers	3000.000									
Prop Loyal Stayers	0.130									
	T1	T2	T3	T4	T5	T6	T7	T8	T9	T10
Wave N	6518.000	4442.000	6884.000	12182.000	18211.000	14573.000	11660.000	9329.000	7465.000	5973.000
Wave N / Grand N	0.283	0.193	0.299	0.530	0.792	0.634	0.507	0.406	0.325	0.260
Wave N - Loyal Stayers N	0.153	0.063	0.169	0.399	0.661	0.503	0.377	0.275	0.194	0.129
Prop Loyal Stayers	0.130	0.130	0.130	0.130	0.130	0.130	0.130	0.130	0.130	0.130
Prop Missing from Grand N (varying component)	0.847	0.937	0.831	0.601	0.339	0.497	0.623	0.725	0.806	0.871
Prop Missing from Grand N (loyal stayer component)	0.870	0.870	0.870	0.870	0.870	0.870	0.870	0.870	0.870	0.870

- It is important to note that we simulated data as Missing at Random (MAR). While this mimics the obtained and projected sample sizes for each wave, our simulation did not include specific decay functions for the booster samples collected during Waves 3 -5.
- The NZAVS and simulated data have a comparable number of people who completed only one wave to those in the simulated data. However, in the NZAVS, these people completed only the wave in which they were initially booster sampled (and are thus most likely to be in Waves 1 or 5). In the simulated data, they are distributed across all waves. We thus tend to as a consequence under-estimate the N for unique cases in Wave 5 specifically.

The Mplus syntax for estimating missing data patterns in our simulations for the first 10 waves of the NZAVS was thus:

```
PATMISS = y1(.847) y2(.937) y3(.831) y4(.601) y5(.339) y6(.497) y7(.623) y8(.725) y9(.806) y10(.871)
          | y1(0) y2(0) y3(0) y4(0) y5(0) y6(0) y7(0) y8(0 y9(0) y10(0);
PATPROB = .870|.130;
```

Where for PATMISS, y1 – y10 represent Waves 1 – 10 of the study, and the values in parentheses represent the proportion of missing data.

PATPROB then extends this to specify that the first section of proportions should be estimated for 87% of simulated cases and the latter section of proportions (those with no missing values) should be simulated for 13% of cases.

Note that when estimating less than 10 waves of data, for e.g., 6 waves, the only the proportions for y1 – y6 would be included.

Table 3. Summary of missing data patterns for a simulated run using Waves 1 – 5 and without covariates (x = non-missing data).

Y1	Y2	Y3	Y4	Y5	Sim. N
				X	5853
			X	X	4039
X	X	X	X	X	3402
			X		1946
		X		X	1243
X				X	1100
		X	X	X	817
X			X	X	713
		X			598
X					542
	X			X	406
		X	X		395
X			X		383
	X		X	X	277
	X				220
X		X		X	203
X		X	X	X	153
	X		X		118
X		X			113
X		X	X		81
X	X			X	73
	X	X		X	63
X	X		X	X	60
	X	X			44
	X	X	X	X	41
X	X				33
	X	X	X		31
X	X		X		22
X	X	X		X	15
X	X	X			11
X	X	X	X		5

Model Specification: Overview

We specified a series of Latent Growth models of varying complexity and estimated the statistical power achieved by each model when varying the size of key parameters of interest when analysed using the first 5 waves, first 6 waves, and so on up to Wave 10.

All models were estimated using 100,000 replications and assuming normally distributed continuous indicators. Missing data were generated using MAR (as outlined in the section on missing data estimation above). We used Maximum Likelihood estimation. Latent Growth Models were estimated assuming a time-invariant and equally dispersed series. The means of indicators in our latent growth model were set at 0, with variances = .50. The intercept of the latent growth models was fixed at 4.0 with variance = 1.0. These default values were chosen because they generally approximate those observed in previous LGMs using NZAVS data. Syntax for a sample example of each model simulating 10 waves of data collection for each model type is included in the Appendix.

The models we estimated were:

- Model 1: LGM without covariates, varying the random effect of slope.
- Model 2: LGM without covariates, varying the mean of slope.
- Model 3: LGM with time-invariant covariate, varying the conditional effect of covariate on slope.
- Model 4: LGM with six time-invariant covariates, varying the conditional effect of focal covariate on slope.
- Model 5: LGM with second-order polynomial, varying the random effect of the quadratic component of the slope.
- Model 6: LGM with second-order polynomial, varying the mean of the quadratic component of the slope.

- Model 7: LGM with six covariates, varying the conditional effect of focal covariate on quadratic component of slope.
- Model 8: LGM where the effect of time-invariant covariate on the slope is moderated by the intercept of the growth factor.
- Model 9: Parallel process LGM, varying the regression of the intercept from one growth factor on the slope of the second growth factor.

Model 1: LGM without covariates, varying the random effect of slope

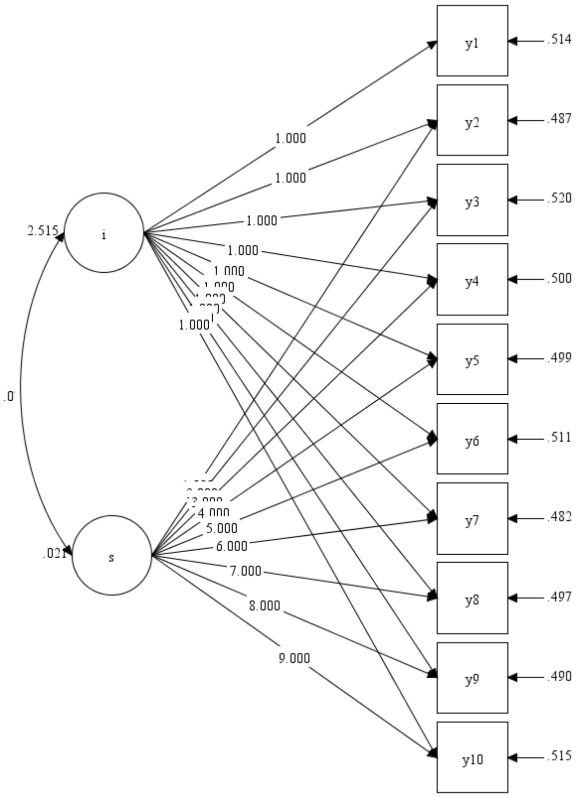

Figure 1. Example of Model 1 Simulating Waves 1 – 10.

Summary of model parameters

- Equally dispersed fixed periods between each wave (0,1,2, etc)
- Models ranged from those simulating <u>Waves 1-5 up to Waves 1-10</u>
- Mean of y indicators set to 0, residual variance of y indicators set to .50
- Intercept mean set to 4, variance set to 1
- Slope mean set to 0, <u>variance ranged from .001 to .015</u>
- Covariance of slope and intercept set to 0
  (Refer to Appendix for Syntax)

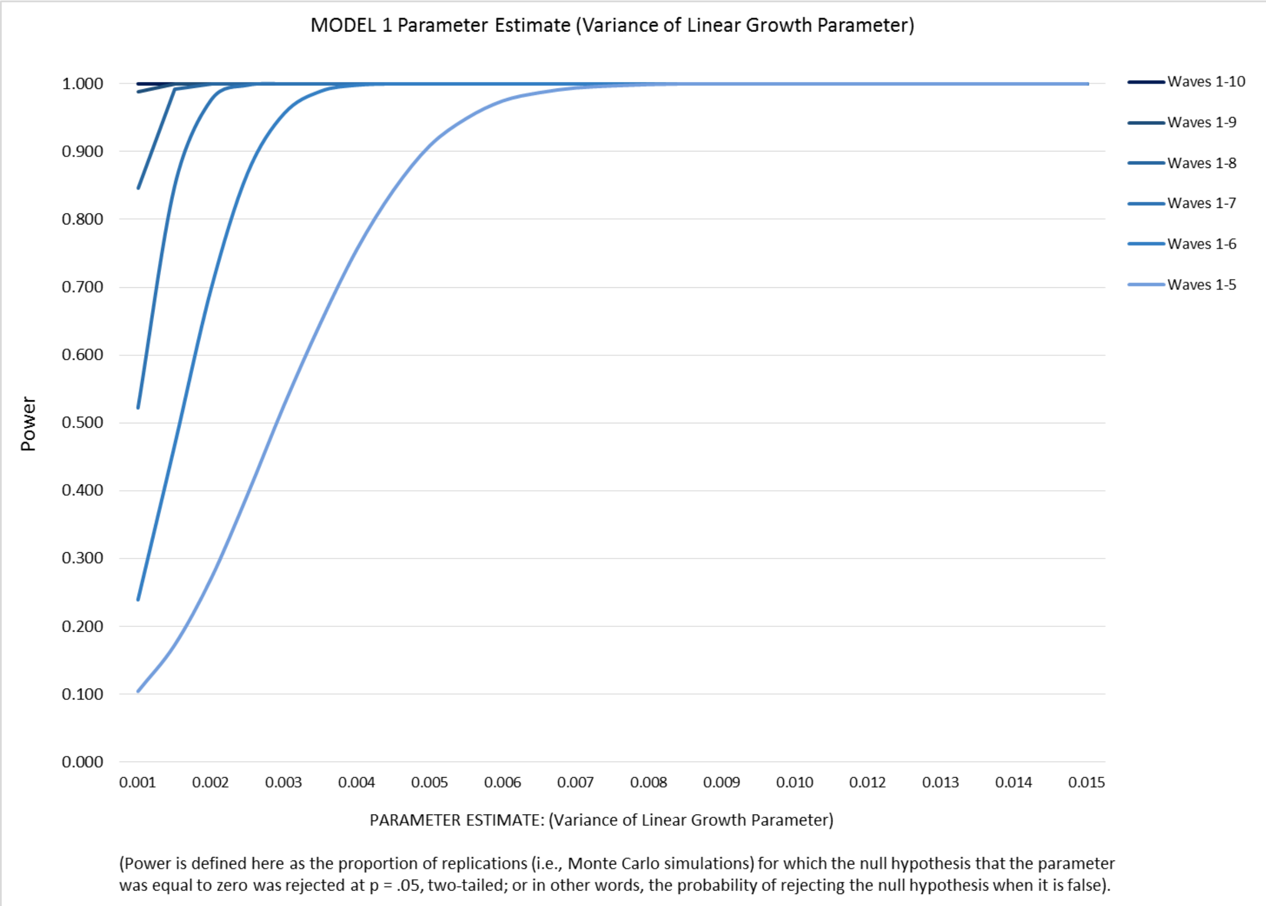

Figure 3. Power estimates for Model 1.

Table 5. Tabulated power estimates for Model 1.

Effect Size	Waves 1-5	Waves 1-6	Waves 1-7	Waves 1-8	Waves 1-9	Waves 1-10
.0010	0.104	0.239	0.522	0.846	0.988	1.000
.0015	0.172	0.466	0.848	0.992	1.000	1.000
.0020	0.270	0.697	0.976	1.000	1.000	1.000
.0025	0.394	0.868	0.998	1.000	1.000	1.000
.0030	0.526	0.956	1.000	1.000	1.000	1.000
.0035	0.647	0.989	1.000	1.000	1.000	1.000
.0040	0.756	0.998	1.000	1.000	1.000	1.000
.0045	0.843	1.000	1.000	1.000	1.000	1.000
.0050	0.909	1.000	1.000	1.000	1.000	1.000
.0055	0.949	1.000	1.000	1.000	1.000	1.000
.0060	0.975	1.000	1.000	1.000	1.000	1.000
.0065	0.987	1.000	1.000	1.000	1.000	1.000
.0070	0.994	1.000	1.000	1.000	1.000	1.000
.0075	0.997	1.000	1.000	1.000	1.000	1.000
.0080	0.999	1.000	1.000	1.000	1.000	1.000
.0085	1.000	1.000	1.000	1.000	1.000	1.000
.0090	1.000	1.000	1.000	1.000	1.000	1.000
.0095	1.000	1.000	1.000	1.000	1.000	1.000
.0100	1.000	1.000	1.000	1.000	1.000	1.000
.0105	1.000	1.000	1.000	1.000	1.000	1.000
.0120	1.000	1.000	1.000	1.000	1.000	1.000
.0125	1.000	1.000	1.000	1.000	1.000	1.000
.0130	1.000	1.000	1.000	1.000	1.000	1.000
.0135	1.000	1.000	1.000	1.000	1.000	1.000
.0140	1.000	1.000	1.000	1.000	1.000	1.000
.0145	1.000	1.000	1.000	1.000	1.000	1.000
.0150	1.000	1.000	1.000	1.000	1.000	1.000

Model 2: LGM without covariates, varying the mean of slope

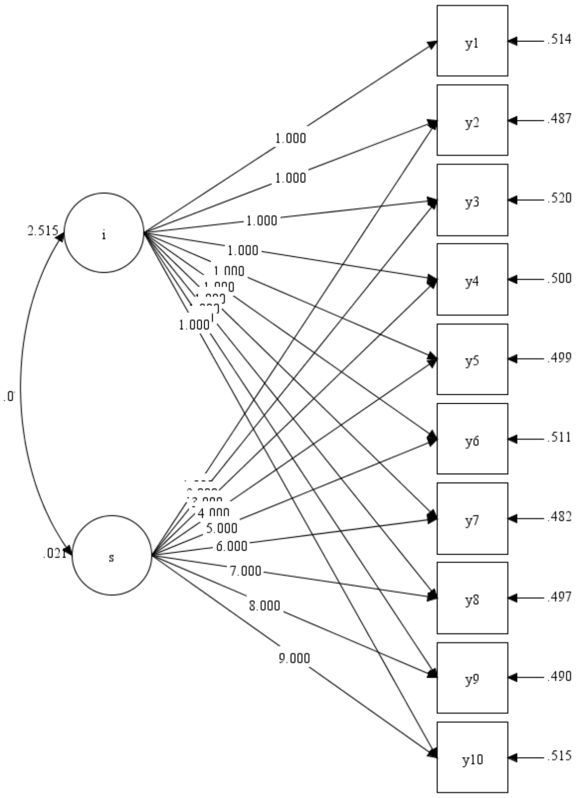

Figure 4. Example Model 2 Simulating Waves 1 – 10.

Summary of model parameters

- Equally dispersed fixed periods between each wave (0,1,2, etc)
- Models ranged from those simulating Waves 1-5 up to Waves 1-10
- Mean of y indicators set to 0, residual variance of y indicators set to .50
- Intercept mean set to 4, intercept variance set to 1
- Slope mean set to 0, slope variance ranged from .001 to .015
- Covariance of slope and intercept set to 0
  (Refer to Appendix for Syntax)

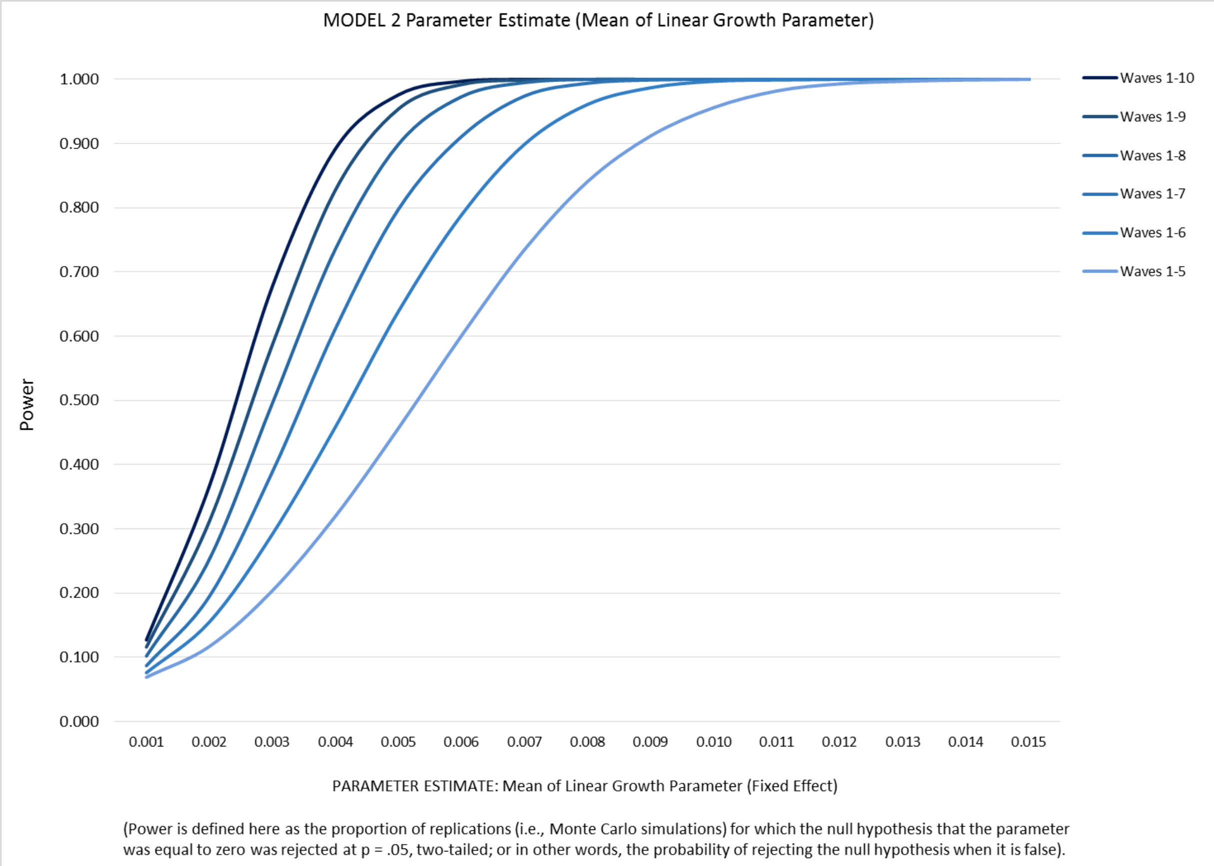

Figure 5. Power estimates for Model 2.

Table 6. Tabulated power estimates for Model 2.

Effect Size	Waves 1-5	Waves 1-6	Waves 1-7	Waves 1-8	Waves 1-9	Waves 1-10
0.001	0.069	0.076	0.087	0.102	0.116	0.127
0.002	0.117	0.155	0.195	0.253	0.312	0.367
0.003	0.204	0.292	0.389	0.496	0.587	0.678
0.004	0.320	0.459	0.612	0.736	0.828	0.892
0.005	0.457	0.639	0.798	0.899	0.954	0.976
0.006	0.601	0.789	0.911	0.973	0.992	0.997
0.007	0.735	0.899	0.974	0.995	0.999	1.000
0.008	0.841	0.961	0.994	1.000	1.000	1.000
0.009	0.912	0.987	0.999	1.000	1.000	1.000
0.010	0.956	0.997	1.000	1.000	1.000	1.000
0.011	0.982	0.999	1.000	1.000	1.000	1.000
0.012	0.993	1.000	1.000	1.000	1.000	1.000
0.013	0.997	1.000	1.000	1.000	1.000	1.000
0.014	0.999	1.000	1.000	1.000	1.000	1.000
0.015	1.000	1.000	1.000	1.000	1.000	1.000

Model 3: Model 3: LGM with time-invariant covariate, varying the conditional effect of covariate on slope

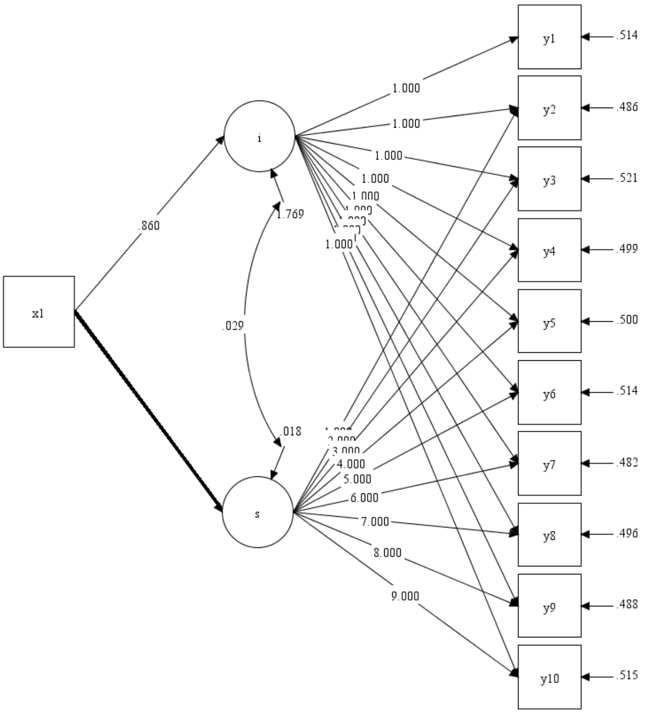

Figure 6. Example Model 3 Simulating Waves 1 – 10.

Summary of model parameters

- Equally dispersed fixed periods between each wave (0,1,2, etc)
- Models ranged from those simulating <u>Waves 1-5 up to Waves 1-10</u>
- Mean of y indicators set to 0, residual variance of y indicators set to .50
- Means of x1 covariate set to 0, variance of x indicators set to 1.
- Intercept mean set to 4, intercept variance set to 1
- Slope mean set to .01, slope variance set to .01
- Covariance of slope and intercept set to -.05
- Regression of intercept on time-invariant covariate set to .30
- <u>Regression of slope on time-invariant covariate ranged from .001 to .015</u>
  (Refer to Appendix for Syntax)

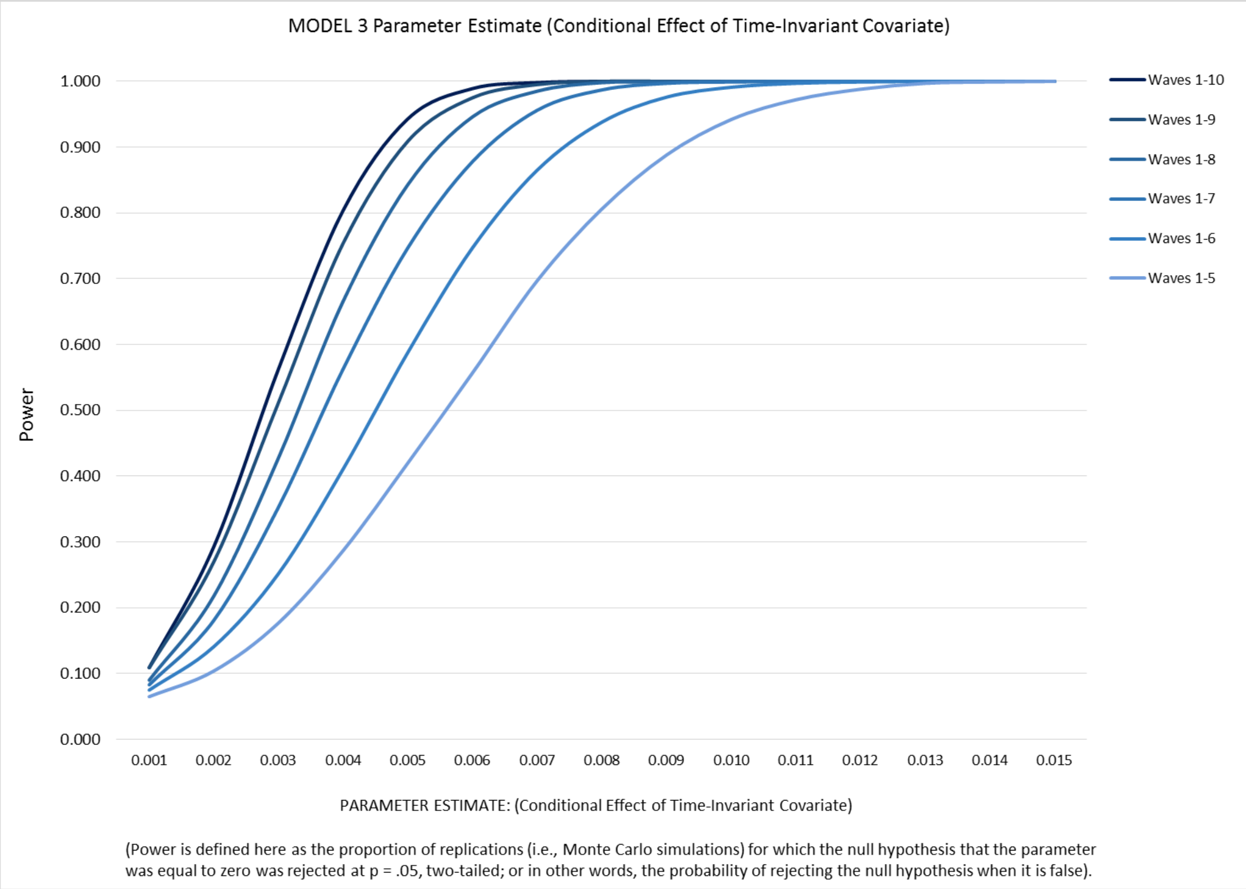

Figure 7. Power estimates for Model 3.

Table 7. Tabulated power estimates for Model 3.

Effect Size	Waves 1-5	Waves 1-6	Waves 1-7	Waves 1-8	Waves 1-9	Waves 1-10
0.001	0.065	0.075	0.083	0.090	0.109	0.109
0.002	0.104	0.141	0.181	0.218	0.270	0.293
0.003	0.177	0.252	0.353	0.428	0.512	0.563
0.004	0.287	0.411	0.563	0.666	0.754	0.804
0.005	0.420	0.588	0.747	0.843	0.909	0.943
0.006	0.557	0.747	0.878	0.946	0.975	0.989
0.007	0.697	0.865	0.956	0.985	0.995	0.998
0.008	0.806	0.938	0.987	0.998	1.000	1.000
0.009	0.888	0.976	0.997	0.999	1.000	1.000
0.010	0.942	0.991	0.999	1.000	1.000	1.000
0.011	0.972	0.997	1.000	1.000	1.000	1.000
0.012	0.988	0.999	1.000	1.000	1.000	1.000
0.013	0.997	1.000	1.000	1.000	1.000	1.000
0.014	0.999	1.000	1.000	1.000	1.000	1.000
0.015	1.000	1.000	1.000	1.000	1.000	1.000

Model 4: LGM with six time-invariant covariates, varying the conditional effect of focal covariate on slope

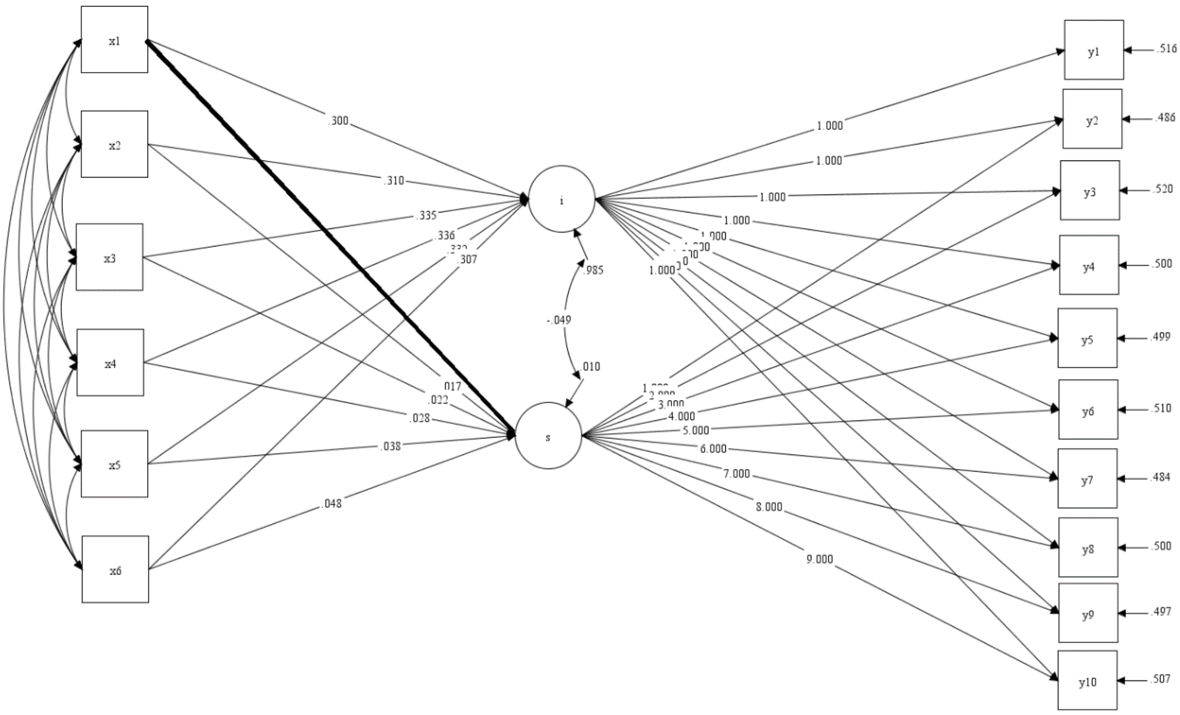

Figure 8. Example Model 4 Simulating Waves 1 – 10.

Summary of model parameters

- Equally dispersed fixed periods between each wave (0,1,2, etc)
- Models ranged from those simulating Waves 1-5 up to Waves 1-10
- Mean of y indicators set to 0, residual variance of y indicators set to .50
- Means of x1-x6 set to 0, variance of x indicators set to 1.
- One x indicator specified as categorical.
- Intercept mean set to 4, slope variance set to 1
- Slope mean set to .01, slope variance set to .01
- Covariance of slope and intercept set to -.05
- Population covariances between x1-x6 time-invariant covariates set to .40
- Regression of intercept on time-invariant focal covariate set to .30
- Regression of intercept on time-invariant other covariates set to .31, .32, .33, .34, .35
- Regression of slope on time-invariant other covariates set to .05, .04, .03, .02, .01
- Regression of slope on time-invariant focal covariate ranged from .001 to .015
  (Refer to Appendix for Syntax)

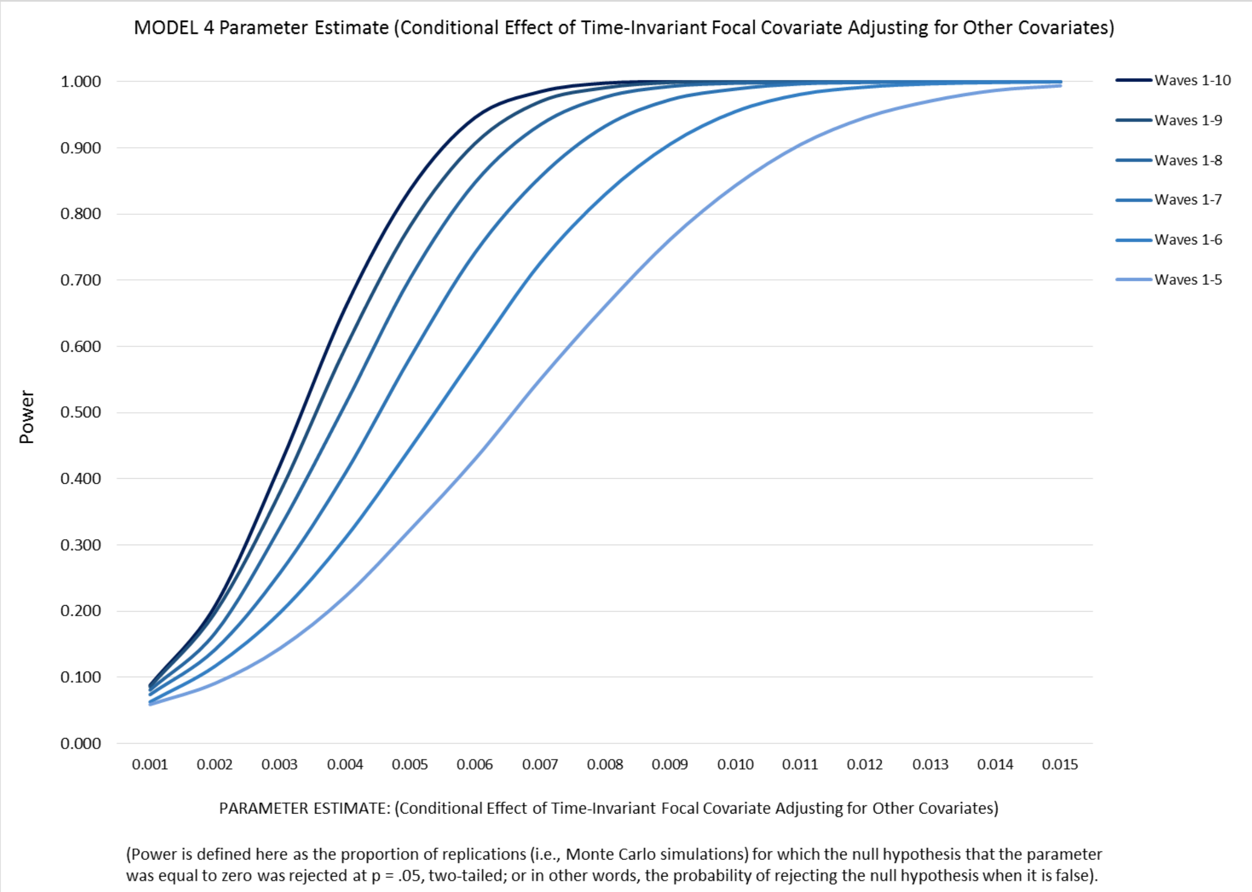

Figure 9. Power estimates for Model 4.

Table 8. Tabulated power estimates for Model 4.

Effect Size	Waves 1-5	Waves 1-6	Waves 1-7	Waves 1-8	Waves 1-9	Waves 1-10
0.001	0.059	0.063	0.074	0.081	0.086	0.088
0.002	0.091	0.117	0.142	0.167	0.198	0.208
0.003	0.144	0.198	0.258	0.327	0.380	0.421
0.004	0.222	0.310	0.408	0.512	0.597	0.658
0.005	0.323	0.446	0.583	0.703	0.782	0.837
0.006	0.430	0.588	0.742	0.848	0.907	0.946
0.007	0.550	0.726	0.856	0.935	0.970	0.985
0.008	0.661	0.830	0.933	0.977	0.991	0.998
0.009	0.762	0.906	0.973	0.993	0.999	1.000
0.010	0.843	0.955	0.989	0.998	1.000	1.000
0.011	0.905	0.981	0.997	0.999	1.000	1.000
0.012	0.946	0.992	0.999	1.000	1.000	1.000
0.013	0.971	0.997	1.000	1.000	1.000	1.000
0.014	0.987	0.999	1.000	1.000	1.000	1.000
0.015	0.994	1.000	1.000	1.000	1.000	1.000

Model 5: LGM with second-order polynomial, varying the random effect of the quadratic component of the slope.

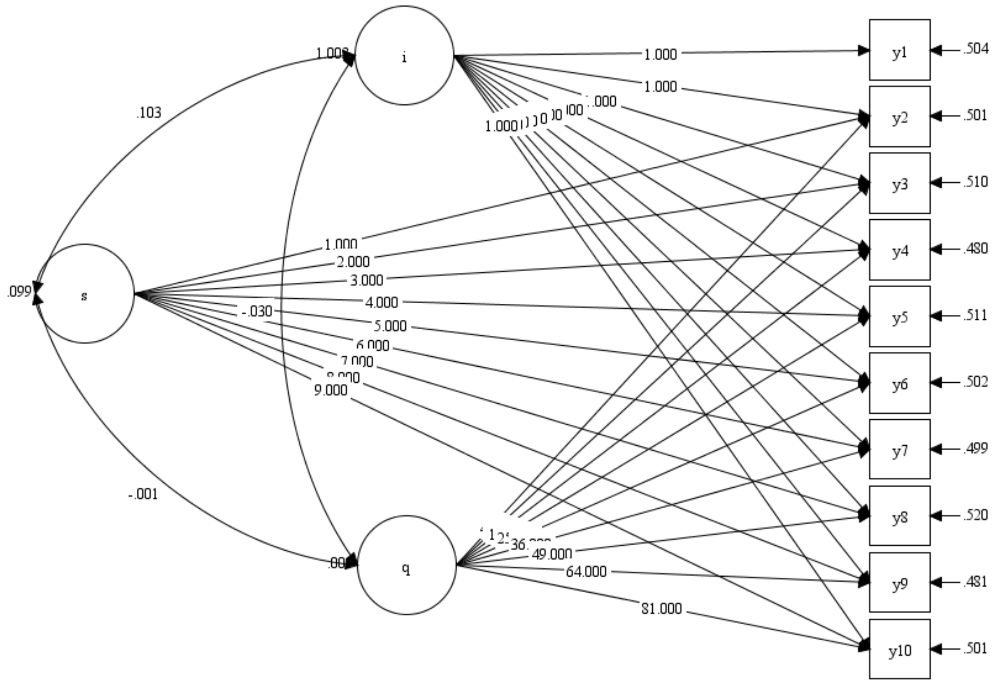

Figure 10. Example Model 5 Simulating Waves 1 – 10.

Summary of model parameters

- Equally dispersed fixed periods between each wave (0,1,2, etc)
- Models ranged from those simulating Waves 1-5 up to Waves 1-10
- Mean of y indicators set to 0, residual variance of y indicators set to .50
- Intercept mean set to 4, variance set to 1
- Slope of linear effect mean set to .20, variance set to .10
- Slope of quadratic effect mean set to -.03, variance ranged from .001 to .015
- Covariance of intercept and linear effect set to .10
- Covariance of intercept and quadratic effect set to -.03
- Covariance of linear and quadratic effect set to -.001
  (Refer to Appendix for Syntax)

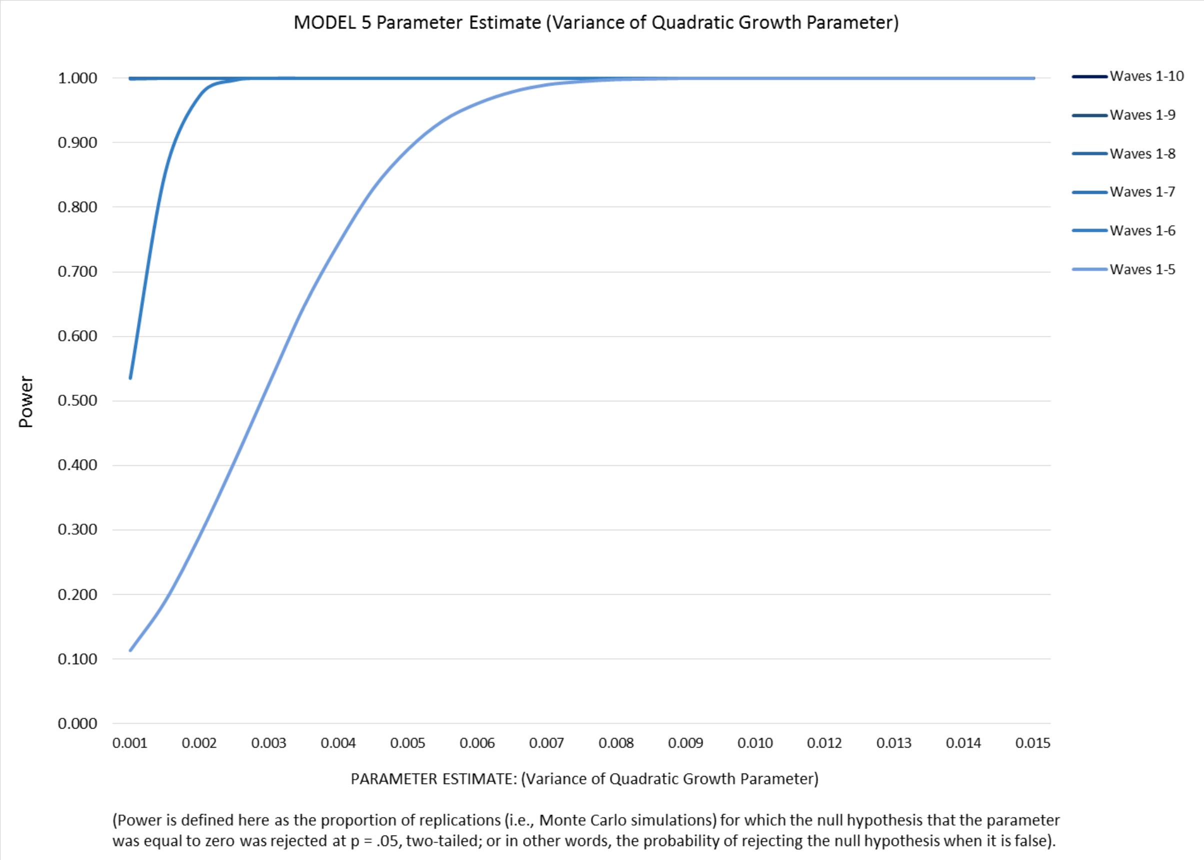

Figure 11. Power estimates for Model 5.

Table 9. Tabulated power estimates for Model 5.

Effect Size	Waves 1-5	Waves 1-6	Waves 1-7	Waves 1-8	Waves 1-9	Waves 1-10
.0010	0.113	0.535	0.999	1.000	1.000	1.000
.0015	0.189	0.852	1.000	1.000	1.000	1.000
.0020	0.291	0.973	1.000	1.000	1.000	1.000
.0025	0.406	0.997	1.000	1.000	1.000	1.000
.0030	0.527	1.000	1.000	1.000	1.000	1.000
.0035	0.646	1.000	1.000	1.000	1.000	1.000
.0040	0.744	1.000	1.000	1.000	1.000	1.000
.0045	0.829	1.000	1.000	1.000	1.000	1.000
.0050	0.890	1.000	1.000	1.000	1.000	1.000
.0055	0.934	1.000	1.000	1.000	1.000	1.000
.0060	0.961	1.000	1.000	1.000	1.000	1.000
.0065	0.979	1.000	1.000	1.000	1.000	1.000
.0070	0.990	1.000	1.000	1.000	1.000	1.000
.0075	0.995	1.000	1.000	1.000	1.000	1.000
.0080	0.998	1.000	1.000	1.000	1.000	1.000
.0085	0.999	1.000	1.000	1.000	1.000	1.000
.0090	1.000	1.000	1.000	1.000	1.000	1.000
.0095	1.000	1.000	1.000	1.000	1.000	1.000
.0100	1.000	1.000	1.000	1.000	1.000	1.000
.0105	1.000	1.000	1.000	1.000	1.000	1.000
.0120	1.000	1.000	1.000	1.000	1.000	1.000
.0125	1.000	1.000	1.000	1.000	1.000	1.000
.0130	1.000	1.000	1.000	1.000	1.000	1.000
.0135	1.000	1.000	1.000	1.000	1.000	1.000
.0140	1.000	1.000	1.000	1.000	1.000	1.000
.0145	1.000	1.000	1.000	1.000	1.000	1.000
.0150	1.000	1.000	1.000	1.000	1.000	1.000

Model 6: LGM with second-order polynomial, varying the mean of the quadratic component of the slope.

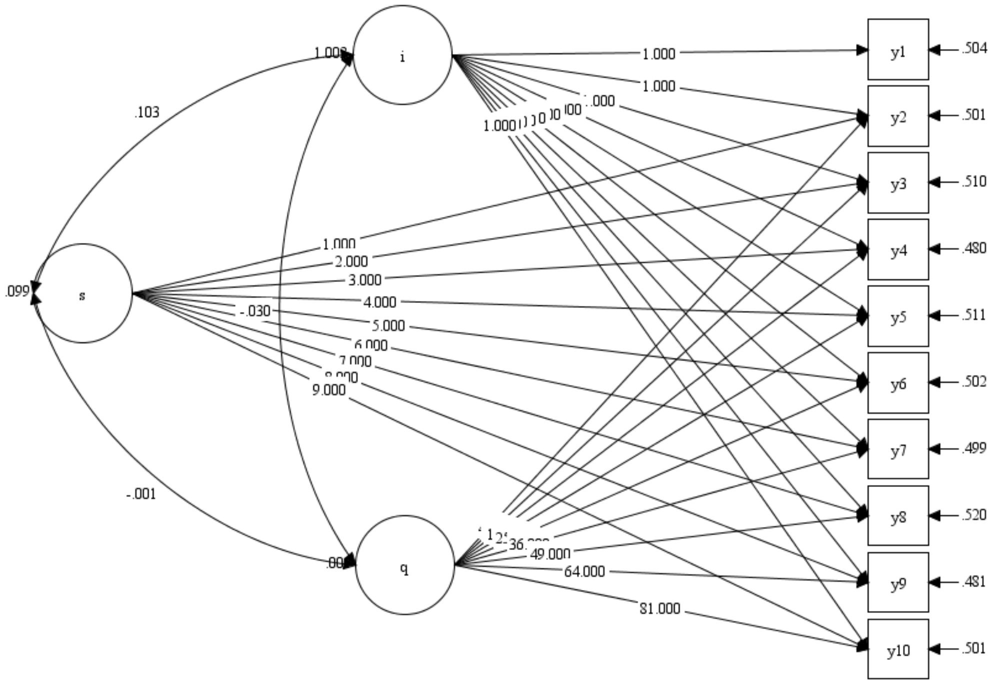

Figure 12. Example Model 6 Simulating Waves 1 – 10.

Summary of model parameters

- Equally dispersed fixed periods between each wave (0,1,2, etc)
- Models ranged from those simulating <u>Waves 1-5 up to Waves 1-10</u>
- Mean of y indicators set to 0, residual variance of y indicators set to .50
- Intercept mean set to 4, variance set to 1
- Slope of linear effect mean set to .20, variance set to .10
- Slope of quadratic effect mean <u>ranged from -.001 to -.015</u>, variance set to .01
- Covariance of intercept and linear effect set to .10
- Covariance of intercept and quadratic effect set to -.03
- Covariance of linear and quadratic effect set to -.001
  (Refer to Appendix for Syntax)

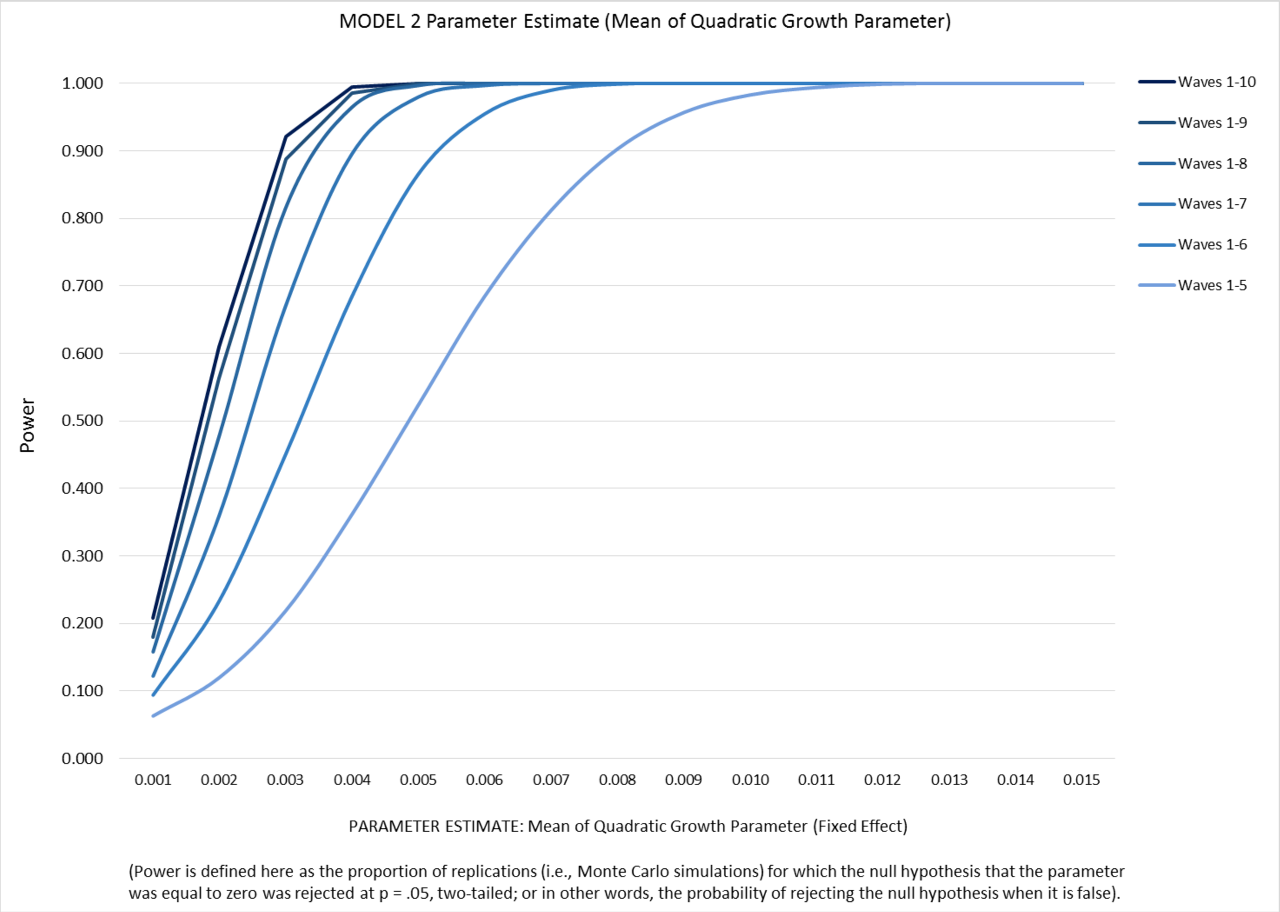

Figure 13. Power estimates for Model 6.

Table 10. Tabulated power estimates for Model 6.

Effect Size	Waves 1-5	Waves 1-6	Waves 1-7	Waves 1-8	Waves 1-9	Waves 1-10
0.001	0.063	0.094	0.122	0.158	0.180	0.208
0.002	0.120	0.234	0.361	0.479	0.564	0.610
0.003	0.219	0.451	0.671	0.816	0.888	0.921
0.004	0.362	0.685	0.896	0.965	0.986	0.995
0.005	0.524	0.866	0.980	0.997	1.000	1.000
0.006	0.685	0.955	0.997	1.000	1.000	1.000
0.007	0.812	0.990	1.000	1.000	1.000	1.000
0.008	0.903	0.999	1.000	1.000	1.000	1.000
0.009	0.957	1.000	1.000	1.000	1.000	1.000
0.010	0.983	1.000	1.000	1.000	1.000	1.000
0.011	0.994	1.000	1.000	1.000	1.000	1.000
0.012	0.999	1.000	1.000	1.000	1.000	1.000
0.013	1.000	1.000	1.000	1.000	1.000	1.000
0.014	1.000	1.000	1.000	1.000	1.000	1.000
0.015	1.000	1.000	1.000	1.000	1.000	1.000

Model 7: LGM with six covariates, varying the conditional effect of focal covariate on quadratic component of slope.

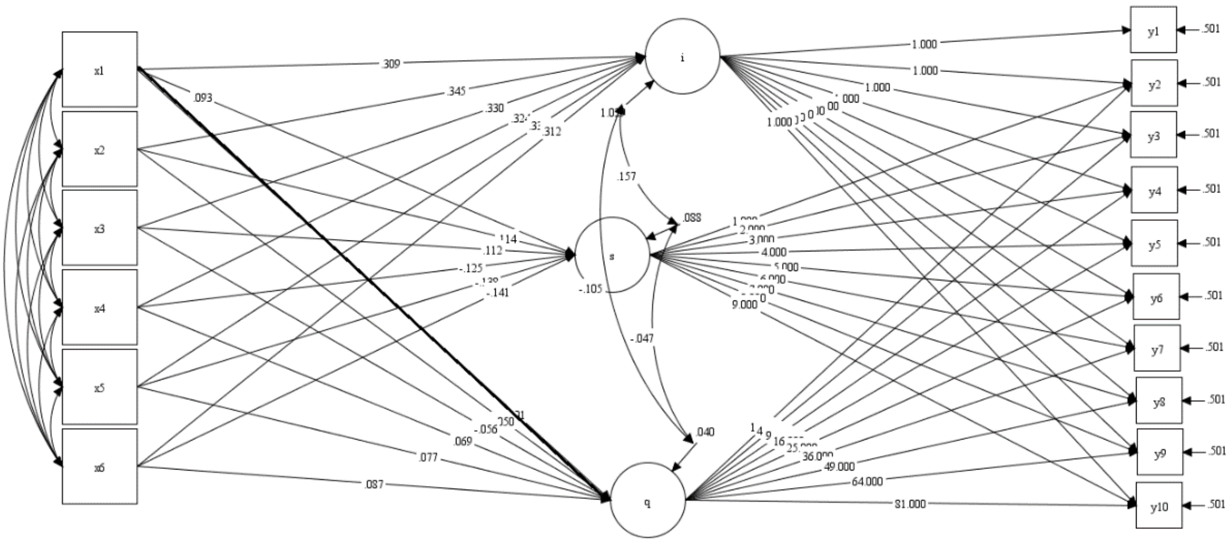

Figure 14. Example Model 7 Simulating Waves 1 – 10.

Summary of model parameters

- Equally dispersed fixed periods between each wave (0,1,2, etc)
- Models ranged from those simulating Waves 1-5 up to Waves 1-10
- Mean of y indicators set to 0, residual variance of y indicators set to .50
- Means of x1-x6 set to 0, variance of x indicators set to 1.
- One x indicator specified as categorical.
- Intercept mean set to 4, slope variance set to 1
- Slope of linear effect mean set to .20, variance set to .10
- Slope of quadratic effect mean set to -.03, variance set to .04
- Covariance of intercept and linear effect set to .15
- Covariance of intercept and quadratic effect set to -.10
- Covariance of linear and quadratic effect set to -.05
- Population covariances between x1-x6 time-invariant covariates set to .40
- Regression of intercept on time-invariant focal covariate set to .30
- Regression of linear slope on time-invariant focal covariate set to .10
- Regression of intercept on time-invariant covariates set to .35, .34, .33, .32, .31
- Regression of linear slope on time-invariant covariates set to .11, .12, -.13, -.14, -.15
- Regression of quad slope on time-invariant covariates set to -.05, -.06, .07, .08, .09
- Regression of quadratic slope on time-invariant focal covariate ranged from -.001 to --.015
  (Refer to Appendix for Syntax)

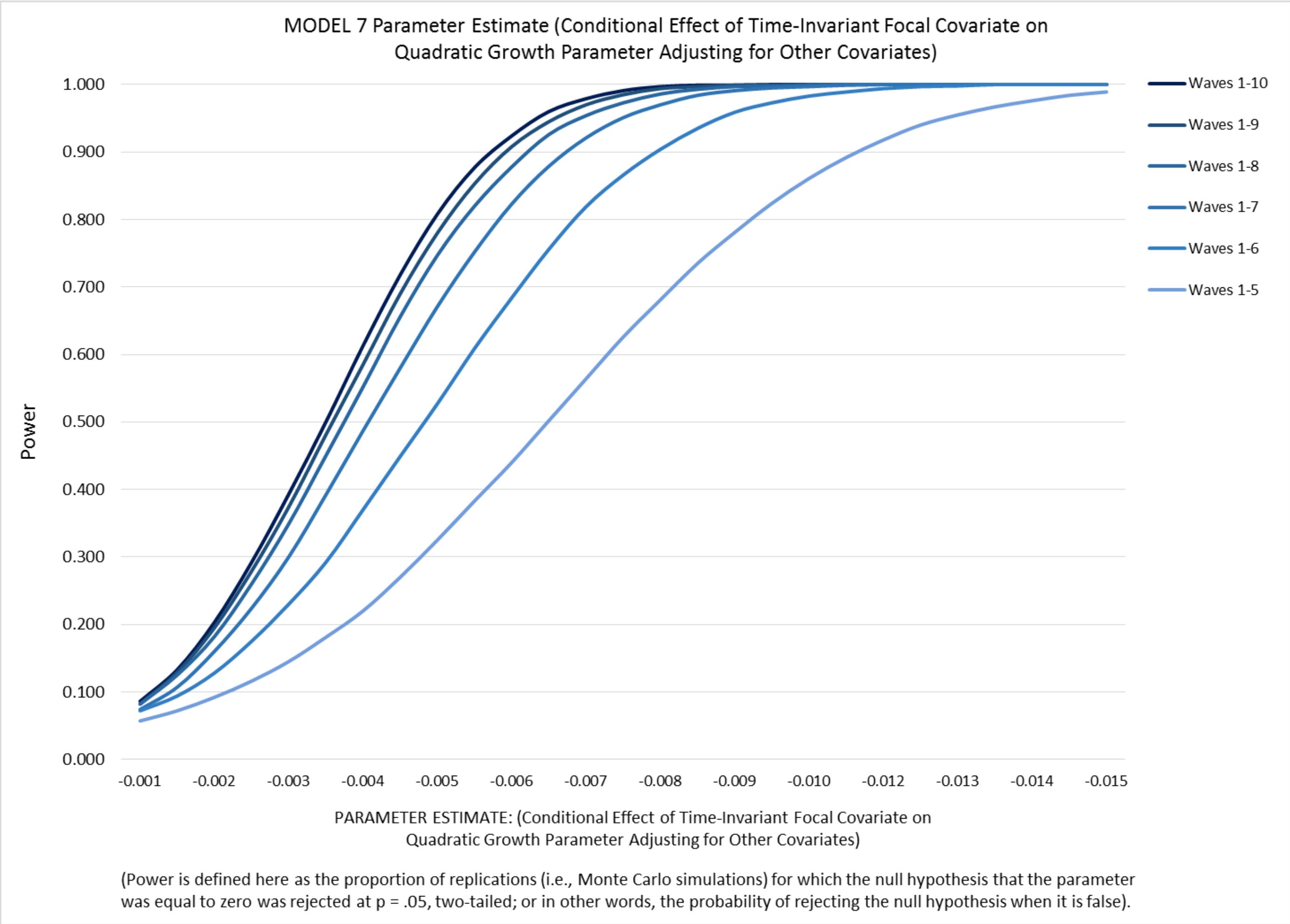

Figure 15. Example Model 7 Simulating Waves 1 – 10.

Table 11. Tabulated power estimates for Model 7.

Effect Size	Waves 1-5	Waves 1-6	Waves 1-7	Waves 1-8	Waves 1-9	Waves 1-10
-.0010	0.057	0.072	0.074	0.082	0.082	0.086
-.0015	0.072	0.094	0.107	0.125	0.129	0.133
-.0020	0.092	0.128	0.160	0.182	0.195	0.203
-.0025	0.116	0.175	0.224	0.260	0.279	0.292
-.0030	0.145	0.230	0.300	0.349	0.374	0.393
-.0035	0.181	0.292	0.392	0.450	0.481	0.500
-.0040	0.220	0.370	0.487	0.552	0.585	0.613
-.0045	0.270	0.449	0.580	0.656	0.690	0.718
-.0050	0.325	0.527	0.671	0.747	0.780	0.808
-.0055	0.383	0.609	0.752	0.820	0.853	0.877
-.0060	0.440	0.684	0.823	0.878	0.908	0.924
-.0065	0.502	0.756	0.879	0.926	0.945	0.960
-.0070	0.564	0.819	0.921	0.954	0.970	0.979
-.0075	0.626	0.866	0.951	0.973	0.985	0.991
-.0080	0.681	0.904	0.970	0.986	0.994	0.997
-.0085	0.735	0.935	0.984	0.993	0.997	0.999
-.0090	0.781	0.959	0.991	0.997	0.999	0.999
-.0095	0.824	0.973	0.995	0.998	0.999	1.000
-.0100	0.861	0.983	0.997	0.999	1.000	1.000
-.0105	0.892	0.989	0.999	1.000	1.000	1.000
-.0120	0.918	0.994	1.000	1.000	1.000	1.000
-.0125	0.940	0.997	1.000	1.000	1.000	1.000
-.0130	0.955	0.998	1.000	1.000	1.000	1.000
-.0135	0.967	1.000	1.000	1.000	1.000	1.000
-.0140	0.976	1.000	1.000	1.000	1.000	1.000
-.0145	0.984	1.000	1.000	1.000	1.000	1.000
-.0150	0.989	1.000	1.000	1.000	1.000	1.000

References

Muthen, L. K., & Muthen, B. (2002). How to use a Monte Carlo study to decide on sample size and determine power. *Structural Equation Modeling*, *4*, 599-620.

Sibley, C. G. (2014). NZAVS Sample Projections. NZAVS Technical Documents, e18. Available from: http://www.psych.auckland.ac.nz/en/about/our-research/research-groups/new-zealand-attitudes-and-values-study/nzavs-tech-docs.html

Appendix

MODEL 1

```
TITLE:

Monte Carlo Simulations used to estimate the statistical
power of latent growth parameters in the NZAVS.

Chris Sibley and Petar Milojev, 06-11-2014

MONTECARLO:

NAMES ARE y1-y10;
NOBSERVATIONS = 23000;  !Total N ever sampled
NREPS = 10000;          !Number of replications
SEED = 65474;           !Random Seed
CLASSES = C(1);         !Used for extending to non-normal models
GENCLASSES = C(1);

PATMISS = y1(.847) y2(.937) y3(.831) y4(.601) y5(.339) y6(.497)
        y7(.623) y8(.725) y9(.806) y10(.871)|
        !.870 of sample have missing data according to these proportions
        y1(0) y2(0) y3(0) y4(0) y5(0) y6(0)
        y7(0) y8(0) y9(0) y10(0);
        !.130 have 0 missing data

!y1(.847) y2(.937) y3(.831) y4(.601) y5(.339) y6(.497) y7(.623) y8(.725) y9(.806) y10(.871)
|

PATPROB = .870|.130;

ANALYSIS:

TYPE = MIXTURE;
ESTIMATOR = ML;

MODEL MONTECARLO:

%OVERALL%

i s | y1@0 y2@1 y3@2 y4@3 y5@4 y6@5 y7@6 y8@7 y9@8 y10@9;

[y1-y10@0]; y1-y10*.5;

[i*4]; i*1.0;
[s*0]; s*.001;  !Values we change

i WITH s*0;
```

```
%C#1%

[i*4];
[s*0];

MODEL:

%OVERALL%

i s | y1@0 y2@1 y3@2 y4@3 y5@4 y6@5 y7@6 y8@7 y9@8 y10@9;

[y1-y10@0]; y1-y10*.5;

[i*4]; i*1.0;
[s*0]; s*.001;  !Values we change

i WITH s*0;

%C#1%

[i*4];
[s*0];
```

MODEL 2

```
TITLE:

Monte Carlo Simulations used to estimate the statistical
power of latent growth parameters in the NZAVS.

Chris Sibley and Petar Milojev, 06-11-2014

MONTECARLO:

NAMES ARE y1-y10;
NOBSERVATIONS = 23000;  !Total N ever sampled
NREPS = 10000;          !Number of replications
SEED = 65474;           !Random Seed
CLASSES = C(1);         !Used for extending to non-normal models
GENCLASSES = C(1);

PATMISS = y1(.847) y2(.937) y3(.831) y4(.601) y5(.339)
        y6(.497) y7(.623) y8(.725) y9(.806) y10(.871) |
        !.870 of sample have missing data according to these proportions
        y1(0) y2(0) y3(0) y4(0) y5(0)
        y6(0) y7(0) y8(0) y9(0) y10(0);
        !.130 have 0 missing data

PATPROB = .870|.130;
```

```
ANALYSIS:

TYPE = MIXTURE;
ESTIMATOR = ML;

MODEL MONTECARLO:

%OVERALL%

i s | y1@0 y2@1 y3@2 y4@3 y5@4 y6@5 y7@6 y8@7 y9@8 y10@9;

[y1-y10@0]; y1-y10*.5;

[i*4]; i*1.0;
[s*.001]; s*.005;  !Values we change

i WITH s*0;

%C#1%

[i*4];
[s*.001];          !Values we change

MODEL:

%OVERALL%

i s | y1@0 y2@1 y3@2 y4@3 y5@4 y6@5 y7@6 y8@7 y9@8 y10@9;

[y1-y10@0]; y1-y10*.5;

[i*4]; i*1.0;
[s*.001]; s*.005;  !Values we change

i WITH s*0;

%C#1%

[i*4];
[s*.001];          !Values we change
```

MODEL 3

```
TITLE:

Monte Carlo Simulations used to estimate the statistical
power of latent growth parameters in the NZAVS.

Chris Sibley and Petar Milojev, 06-11-2014
```

```
MONTECARLO:

NAMES ARE y1-y10 x1;
!CUTPOINTS = x1 (0);    !Specifies Categorical x var
NOBSERVATIONS = 23000;  !Total N ever sampled
NREPS = 10000;          !Number of replications
SEED = 35479;           !Random Seed
CLASSES = C(1);         !Used for extending to non-normal models
GENCLASSES = C(1);

PATMISS = y1(.847) y2(.937) y3(.831) y4(.601) y5(.339)
        y6(.497) y7(.623) y8(.725) y9(.806) y10(.871) |
        y1(0) y2(0) y3(0) y4(0) y5(0) y6(0) y7(0) y8(0) y9(0) y10(0);

PATPROB = .870|.130;
!SAVE = data out.txt;

ANALYSIS:

TYPE = MIXTURE;
ESTIMATOR = ML;

MODEL MONTECARLO:

%OVERALL%

[x1@0]; x1@1;

i s | y1@0 y2@1 y3@2 y4@3 y5@4 y6@5 y7@6 y8@7 y9@8 y10@9;

[y1-y10@0]; y1-y10*.5;

[i*4]; i*1.0;
[s*.01]; s*.01;

i WITH s*-.05;

i ON x1*.30;
s ON x1*.001;  !Varying this parameter

%C#1%
[i*4];
[s*.01];

MODEL:

%OVERALL%

i s | y1@0 y2@1 y3@2 y4@3 y5@4 y6@5 y7@6 y8@7 y9@8 y10@9;
```

```
[y1-y10@0]; y1-y10*.5;

[i*4]; i*1.0;
[s*.01]; s*.01;

i WITH s*-.05;

i ON x1*.30;
s ON x1*.001;  !Varying this parameter

%C#1%
[i*4];
[s*.01];
```

MODEL 4

```
TITLE:

Monte Carlo Simulations used to estimate the statistical
power of latent growth parameters in the NZAVS.

Chris Sibley and Petar Milojev, 06-11-2014

MONTECARLO:

NAMES ARE y1-y10 x1-x6;
CUTPOINTS = x2 (0);            !Specifies Categorical x var
NOBSERVATIONS = 23000;         !Total N
NREPS = 10000;                 !Number of replications
SEED = 66449;                  !Random Seed
CLASSES = C(1);
GENCLASSES = C(1);

PATMISS = y1(.847) y2(.937) y3(.831) y4(.601) y5(.339)
        y6(.497) y7(.623) y8(.725) y9(.806) y10(.871) |
        y1(0) y2(0) y3(0) y4(0) y5(0) y6(0) y7(0) y8(0) y9(0) y10(0);

PATPROB = .870|.130;
!SAVE = data out.txt;

ANALYSIS:

TYPE = MIXTURE;
ESTIMATOR = ML;

MODEL MONTECARLO:

%OVERALL%

[x1-x6@0]; x1-x6*1;
```

```
x1 WITH x2*.4; x1 WITH x3*.4; x1 WITH x4*.4;
x1 WITH x5*.4; x1 WITH x6*.4; x2 WITH x3*.4;
x2 WITH x4*.4; x2 WITH x5*.4; x2 WITH x6*.4;
x3 WITH x4*.4; x3 WITH x5*.4; x3 WITH x6*.4;
x4 WITH x5*.4; x4 WITH x6*.4; x5 WITH x6*.4;

i s | y1@0 y2@1 y3@2 y4@3 y5@4 y6@5 y7@6 y8@7 y9@8 y10@9;

[y1-y10@0]; y1-y10*.5;

[i*4]; i*1.0; [s*.01]; s*.01;

i WITH s*-.05;

i ON x1*.30; s ON x1*.001;  !Varying this parameter
i ON x2*.35; s ON x2*.01;
i ON x3*.34; s ON x3*.02;
i ON x4*.33; s ON x4*.03;
i ON x5*.32; s ON x5*.04;
i ON x6*.31; s ON x6*.05;

%C#1%
[i*4]; [s*.01];

MODEL:

%OVERALL%

i s | y1@0 y2@1 y3@2 y4@3 y5@4 y6@5 y7@6 y8@7 y9@8 y10@9;

[y1-y10@0]; y1-y10*.5;

[i*4]; i*1.0; [s*.01]; s*.01;

i WITH s*-.05;

i ON x1*.30; s ON x1*.001;  !Varying this parameter
i ON x2*.35; s ON x2*.01;
i ON x3*.34; s ON x3*.02;
i ON x4*.33; s ON x4*.03;
i ON x5*.32; s ON x5*.04;
i ON x6*.31; s ON x6*.05;

%C#1%
[i*4]; [s*.01];
```

MODEL 5

```
TITLE:

Monte Carlo Simulations used to estimate the statistical
power of latent growth parameters in the NZAVS.

Chris Sibley and Petar Milojev, 06-11-2014

MONTECARLO:

NAMES ARE y1-y10;
NOBSERVATIONS = 23000;  !Total N ever sampled
NREPS = 10000;          !Number of replications
SEED = 45348;           !Random Seed
CLASSES = C(1);         !Used for extending to non-normal models
GENCLASSES = C(1);

PATMISS = y1(.847) y2(.937) y3(.831) y4(.601) y5(.339) y6(.497)
        y7(.623) y8(.725) y9(.806) y10(.871)|
        !.870 of sample have missing data according to these proportions
        y1(0) y2(0) y3(0) y4(0) y5(0) y6(0)
        y7(0) y8(0) y9(0) y10(0);
        !.130 have 0 missing data

PATPROB = .870|.130;

ANALYSIS:

TYPE = MIXTURE;
ESTIMATOR = ML;

MODEL MONTECARLO:

%OVERALL%

i s q | y1@0 y2@1 y3@2 y4@3 y5@4 y6@5 y7@6 y8@7 y9@8 y10@9;

[y1-y10@0]; y1-y10*.5;

[i*4]; i*1.0;
[s*.20]; s*.10;
[q*-.03]; q*.001;  !Values we change

i WITH s*.10;
i WITH q*-.03;
s WITH q*-.001;

MODEL:

%OVERALL%
```

i s q | y1@0 y2@1 y3@2 y4@3 y5@4 y6@5 y7@6 y8@7 y9@8 y10@9;

[y1-y10@0]; y1-y10*.5;

[i*4]; i*1.0;
[s*.20]; s*.10;
[q*-.03]; q*.001; !Values we change

i WITH s*.10;
i WITH q*-.03;
s WITH q*-.001;

MODEL 6

TITLE:

Monte Carlo Simulations used to estimate the statistical
power of latent growth parameters in the NZAVS.

Chris Sibley and Petar Milojev, 06-11-2014

MONTECARLO:

NAMES ARE y1-y10;
NOBSERVATIONS = 23000; !Total N ever sampled
NREPS = 10000;   !Number of replications
SEED = 65474;   !Random Seed
CLASSES = C(1);   !Used for extending to non-normal models
GENCLASSES = C(1);

PATMISS = y1(.847) y2(.937) y3(.831) y4(.601) y5(.339) y6(.497)
    y7(.623) y8(.725) y9(.806) y10(.871)|
    !.870 of sample have missing data according to these proportions
    y1(0) y2(0) y3(0) y4(0) y5(0) y6(0)
    y7(0) y8(0) y9(0) y10(0);
    !.130 have 0 missing data

PATPROB = .870|.130;

ANALYSIS:

TYPE = MIXTURE;
ESTIMATOR = ML;

MODEL MONTECARLO:

%OVERALL%

i s q | y1@0 y2@1 y3@2 y4@3 y5@4 y6@5 y7@6 y8@7 y9@8 y10@9;

```
[y1-y10@0]; y1-y10*.5;

[i*4]; i*1.0;
[s*.20]; s*.10;
[q*-.001]; q*.01;  !Values we change

i WITH s*.10;
i WITH q*-.03;
s WITH q*-.001;

MODEL:

%OVERALL%

i s q | y1@0 y2@1 y3@2 y4@3 y5@4 y6@5 y7@6 y8@7 y9@8 y10@9;

[y1-y10@0]; y1-y10*.5;

[i*4]; i*1.0;
[s*.20]; s*.10;
[q*-.001]; q*.01;  !Values we change

i WITH s*.10;
i WITH q*-.03;
s WITH q*-.001;
```

MODEL 7

```
TITLE:

Monte Carlo Simulations used to estimate the statistical
power of latent growth parameters in the NZAVS.

Chris Sibley and Petar Milojev, 06-11-2014

MONTECARLO:

NAMES ARE y1-y10 x1-x6;
CUTPOINTS = x2 (0);     !Specifies Categorical x var
NOBSERVATIONS = 23000;  !Total N ever sampled
NREPS = 10000;          !Number of replications
SEED = 198327;          !Random Seed
CLASSES = C(1);         !Used for extending to non-normal models
GENCLASSES = C(1);

PATMISS = y1(.847) y2(.937) y3(.831) y4(.601) y5(.339) y6(.497)
        y7(.623) y8(.725) y9(.806) y10(.871)|
        !.870 of sample have missing data according to these proportions
        y1(0) y2(0) y3(0) y4(0) y5(0) y6(0)
```

```
          y7(0) y8(0) y9(0) y10(0);
          !.130 have 0 missing data

PATPROB = .870|.130;
!SAVE = data out.txt;

ANALYSIS:

TYPE = MIXTURE;
ESTIMATOR = ML;

MODEL MONTECARLO:

%OVERALL%

[x1-x6@0]; x1-x6*1;

x1 WITH x2*.4; x1 WITH x3*.4; x1 WITH x4*.4;
x1 WITH x5*.4; x1 WITH x6*.4; x2 WITH x3*.4;
x2 WITH x4*.4; x2 WITH x5*.4; x2 WITH x6*.4;
x3 WITH x4*.4; x3 WITH x5*.4; x3 WITH x6*.4;
x4 WITH x5*.4; x4 WITH x6*.4; x5 WITH x6*.4;

i s q | y1@0 y2@1 y3@2 y4@3 y5@4 y6@5 y7@6 y8@7 y9@8 y10@9;

[y1-y10@0]; y1-y10*.5;

[i*4]; i*1.0;
[s*.20]; s*.10;
[q*-.03]; q*.04;

i WITH s*.15;
i WITH q*-.10;
s WITH q*-.05;

i ON x1*.30; s ON x1*.10; q ON x1*-.001; !Varying this parameter
i ON x2*.35; s ON x2*.11; q ON x2*-.05;
i ON x3*.34; s ON x3*.12; q ON x3*-.06;
i ON x4*.33; s ON x4*-.13; q ON x4*.07;
i ON x5*.32; s ON x5*-.14; q ON x5*.08;
i ON x6*.31; s ON x6*-.15; q ON x6*.09;

MODEL:

%OVERALL%

i s q | y1@0 y2@1 y3@2 y4@3 y5@4 y6@5 y7@6 y8@7 y9@8 y10@9;

[y1-y10@0]; y1-y10*.5;
```

```
[i*4]; i*1.0;
[s*.20]; s*.10;
[q*-.03]; q*.04;

i WITH s*.15;
i WITH q*-.10;
s WITH q*-.05;

i ON x1*.30; s ON x1*.10; q ON x1*-.001; !Varying this parameter
i ON x2*.35; s ON x2*.11; q ON x2*-.05;
i ON x3*.34; s ON x3*.12; q ON x3*-.06;
i ON x4*.33; s ON x4*-.13; q ON x4*.07;
i ON x5*.32; s ON x5*-.14; q ON x5*.08;
i ON x6*.31; s ON x6*-.15; q ON x6*.09;
```