The Concept of Change:

Difference scores or statistical control? What should I use to predict change over two time points?

Jason T. Newsom
Professor
Department of Psychology
Portland State University

Overview

- Purpose is to introduce a few basic concepts that may help guide researchers or those advising them on which analysis approach to use when
- No attempt to resolve or add to the long-standing debate about which approach is “best”
- Many complexities and many red herrings that do not really effect the essential questions addressed by the two approaches
- Despite what many if not most authors seem concerned with, there is not a “statistical winner”

Two Waves for Investigating Change

- Often what is available or affordable
- May be only change of interest, e.g., pretest-posttest
- Improvement over cross-sectional data

Definitions of Stability

- Absolute or exact sense: variable is stable to the extent that a) mean values for y_2 and y_1 are equal over time, or b) individual values of y_{2i} and y_{1i} are the equal over time
  - $\bar{y}_2 - \bar{y}_1 = 0$ or $y_{i2\text{-}1} = y_{i2} - y_{i1} = 0$
- Relative sense: variable is stable to the extent that y_{i2} is correlated with y_{i1}. Mean or individual values may increase or decrease over time, but correlation may be largely unchanged unless relative positioning changes
  - $r_{12} = 1$

Definitions of Change

Difference scores (aka: gain scores, changes scores, true change)

- Absolute or exact changes in value of y_t
  - $y_{2\text{-}1} > 0$
- Captures increase or decrease in the mean values or individual values
- Correlation between y_1 and y_2 can be anywhere between 0 or 1 and differences may be small or large
- Ex: If r_{12} is 1.0 and $y_1 = y_2$, add 5 points to y_2, and r_{12} will still be equal to 1.0.

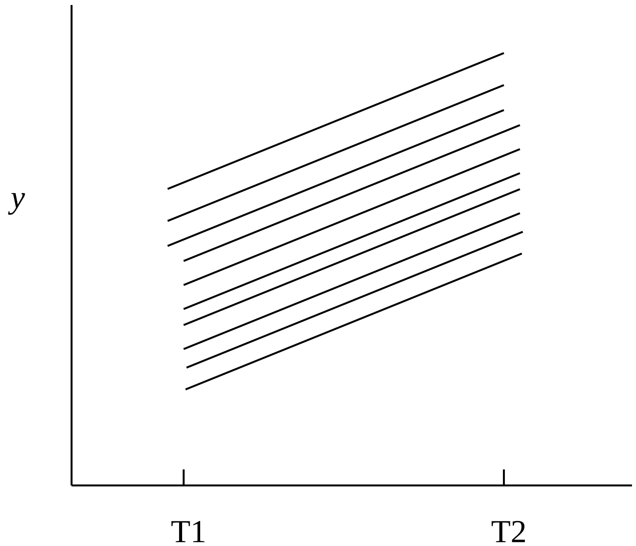
Definitions of Change

Relative change

- Lower correlation indicates greater change, but change is relative to other cases in the data set not absolute.
  - $r_{12} < 1$
- Ex. If r_{12} = .45, add 5 points to all scores, and r_{12} = .45. All scores and the mean of y change over time in the absolute value sense, but the degree of change (lack of stability) is unaffected.

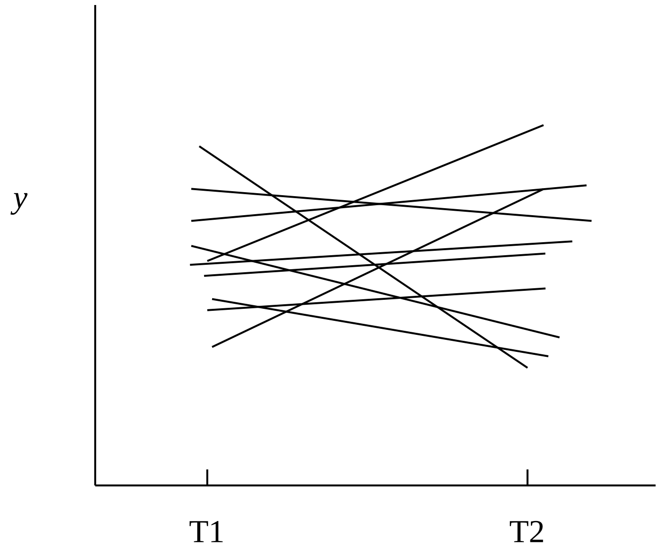
Definitions of Change

High absolute change, but no relative change



Relative position does not change at all

No absolute change of average, but high relative change



Relative position changes for many cases

Additional Approaches

- Residualized change scores (Dubois, 1957)
- True score change estimates (Lord, 1956; McNemar, 1958; see Maassen, 2001 for more)
- Highly correlated with other approaches and/or add unnecessary complexity for interpretations
- Residualized change may omit covariates otherwise included
- True score change may make inaccurate assumptions about reliability
- "One practice that should be discouraged is that of correlating residualized change with other measures. If residualized change scores are desired, it is almost always better instead to employ statistical equating by using multiple regression…" (Campbell & Kenny, 1999, p. 97).

Note on Difference Score Computation

- Always use raw scores when computing difference scores, pre-standardizing variables discards important variance information

Test of Average Difference, Paired *t*-test, Repeated Measures ANOVA

- Comparison of average difference to 0 is equal to paired (dependent) *t*-test comparing two dependent means

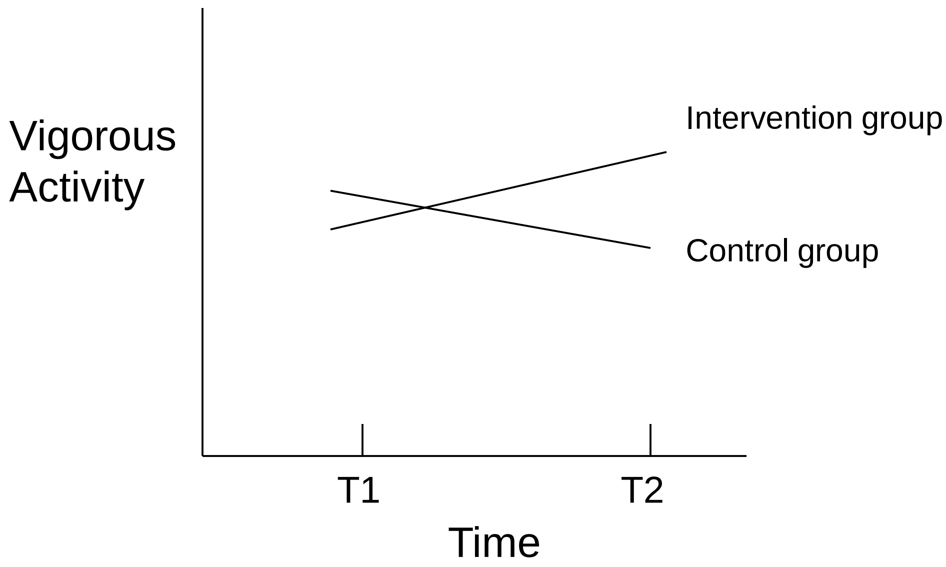
$$\bar{y}_{2-1} = \left[\sum\left(y_2 - y_1\right)\right] / N$$

$$t = \bar{y}_{2-1} / SE_{\bar{y}_{2-1}}$$

- Which is equivalent to repeated measures ANOVA

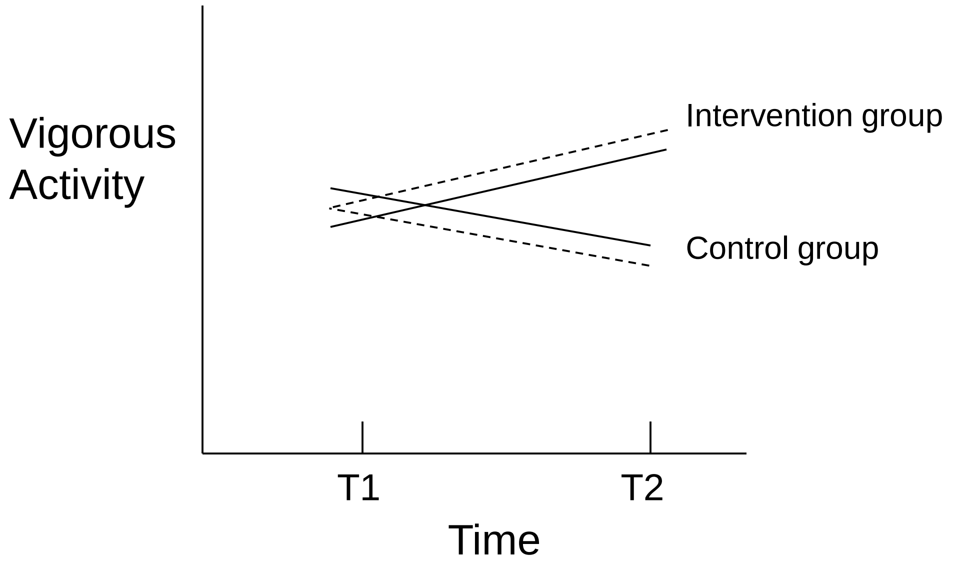
$$t^2 = F$$

Mixed Between x Repeated Measures ANOVA



- Test of interaction assesses whether change (difference score) was the same among the two groups
- Equivalent to a between-subjects t-test of difference scores, where $F = t^2$

ANCOVA



- Test of difference at Time 2, controlling for (or equating on) y at Time 1

Regression Models

- Difference score prediction (aka: change score prediction, unconditional change, gain score prediction)
  - Generalization of repeated measures ANOVA

- Lagged regression (aka: conditional change, static score model, ANCOVA)
  - Generalization of ANCOVA

Predicting Difference Scores

- Predictor x_1 (e.g., self-efficacy or intervention) predicts difference score (e.g., physical activity), where $y_{2\text{-}1} = y_2 - y_1$,

$$y_{2-1} = b_0 + b_1 x_1 + e$$

b_1 is the increment in the difference score $y_{2\text{-}1}$ for each unit change in x_1. If x_1 is binary, then this model is equivalent to the interaction in the mixed ANOVA analysis.

Lagged Regression

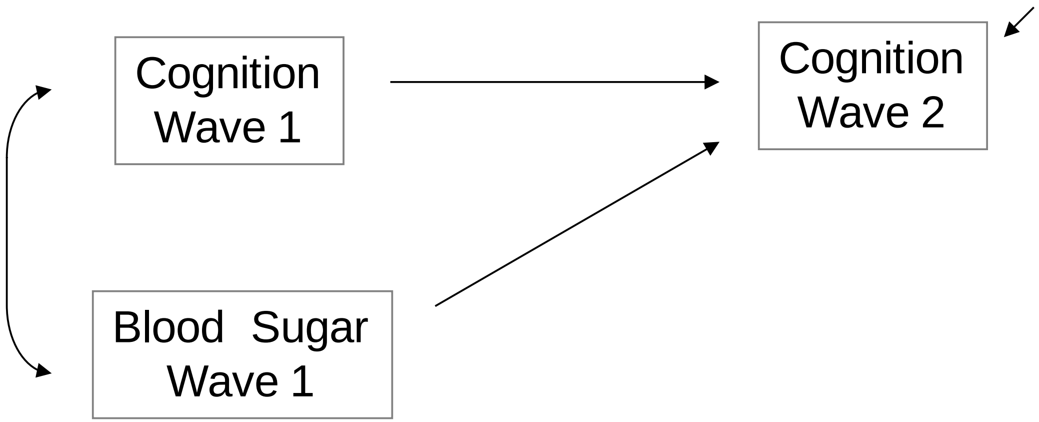
- Lagged regression model predicts change through a model in which x_1 (e.g., self-efficacy or intervention) predicts y_2 (e.g., physical activity) controlling for y_1

$$y_2 = b_0 + b_1 x_1 + b_{y1} y_1 + e$$

- Residuals represent change in the sense that any variance not accounted for by x_1 is changing
- But also holds constant y_1 (or equates across values of x_1)
- If x_1 is binary, then this model is equivalent to the ANCOVA analysis.

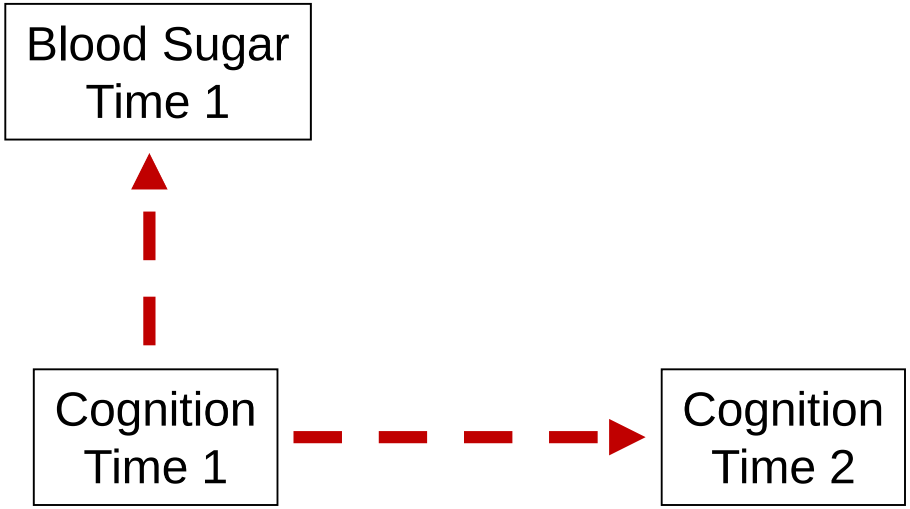
Lagged Regression

- Takes into account the prior (time 1) relationship between x_1 (e.g., blood sugar) and y_1 (e.g., cognition)



Lagged Regression

- Attempts to address initial differences in cognition due to blood sugar (e.g., participants with poorer cognitive functioning may be at greater risk for diabetes, because their dietary habits are affected by memory problems)



Lord’s Paradox

- Results from the difference score prediction and lagged regression approaches may differ from one another, referred to as “Lord’s paradox” after Frederic Lord (Lord, 1967) who noted this difference

- Estimates of the effect of the predictor on change in the dependent variable that are not equivalent

- Although the results of the two approaches may lead to the same conclusion in some instances, they may lead to different conclusions in other instances

Lord's Paradox

Example vigorous activity predicting change in BMI using the Health and Retirement Study (HRS)

Predicting Difference Scores

	B	β	p
Vigorous Activity, T1	**.046**	**.024**	**.108**

Lagged Regression

	B	β	p
Vigorous Activity, T1	**-.001**	**.000**	**.964**
BMI, T1	.926	.888	.000

Although we would not draw any different statistical conclusions from these two models in this case, the results from the two approaches may have different statistical conclusions or suggest opposite relationships between predictor and change.

Lord's Paradox

The two approaches imply different statistical models

- The difference score prediction model is

$$y_2 - y_1 = b_0 + b_1 x_1 + e$$

implies

$$y_2 = b_0 + b_1 x_1 + (1) y_1 + e$$

- which replaces the autoregressive coefficient b_{y1} with the value of 1 in the lagged regression, $y_2 = b_0 + b_{y1} x_1 + b_{y1} y_1 + e$

- The closer the autoregression effect is to 1, the more likely the results from the two approaches will lead to congruent conclusions.

Portland State University

Lord's Paradox

- Can be stated in terms of regression toward the mean (Campbell & Kenny, 1999), where extreme scores move toward the mean over time and scores near the mean move toward the extremes over time.

Lord's Paradox

- If x is a binary variable representing non-equivalent groups that have pre-existing differences on y, regression toward the mean will lead to a negative association between x_1 and y_{2-1} (assumes equal variance over time)

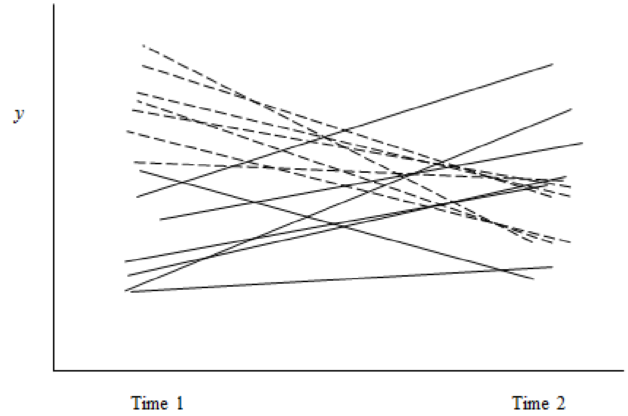
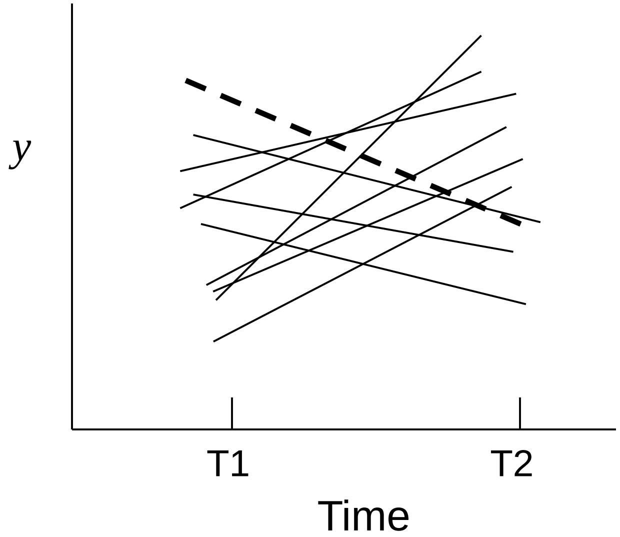


Figure 3.6 Changes in y for two groups represented by dotted lines and solid lines.

- This will happen to some degree whenever the autocorrelation is less than 1

Lord's Paradox

- More generally, the correlation between y_1 and $y_{2\text{-}1}$ will tend to be negative as long as there is a positive relationship between y_1 and y_2.



Lord's Paradox

- So an association between x_1 and y_1 implies an association between x_1 and $y_{2\text{-}1}$ as long as there is regression toward the mean (and as long as x_1 and y_1 are related)

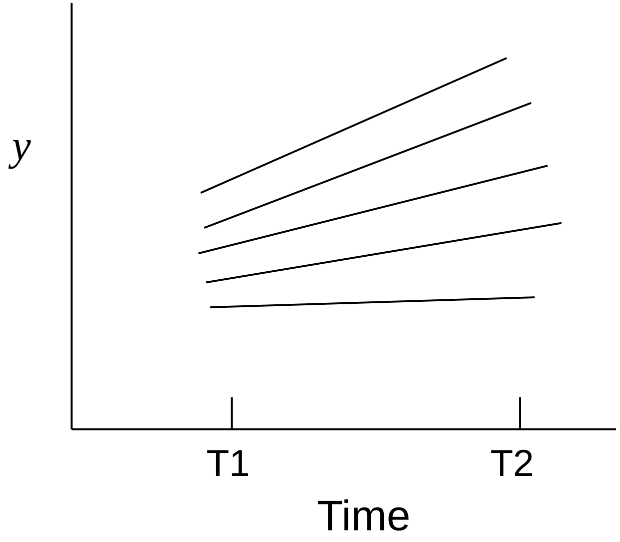
- Although we are subtracting out y_1 when computing the difference score, $y_{2\text{-}1}$, there still remains a correlation between x_1 and y_1 that is not being removed when we predict $y_{2\text{-}1}$, with x_1.

- Therefore, the possibility that the relationship between x_1 (e.g., diabetes) and y_2 (e.g., cognition) is because y_1 (e.g., cognition) causes x (e.g. diabetes) is not addressed.

Lord's Paradox

- Not always the case that there will be a positive association between y_1 and y_2, but this is rare

- Variances of y_1 and y_2 will also impact these relationships—negative correlation between y_1 and $y_{2\text{-}1}$ will occur unless variance increases (Campbell & Kenny, 1999).



Caution: Don't Do Both

Why?
If we take the standard lagged regression model

$$y_2 = b_0 + b_1 x_1 + b_2 y_1 + e$$

And then subtract y_1 from both sides,

$$y_2 - y_1 = b_0 + b_1 x_1 + b_2 y_1 - y_1 + e$$
$$y_{2-1} = b_0 + b_1 x_1 + (b_2 - 1) y_1 + e$$

We don't change the b_1 effect at all, we simply modify the autoregressive coefficient, by subtracting 1.

Caution: Don't Do Both

Reanalyzing the HRS example, with BMI2-BMI1 regressed on vigorous activity, we get

Lagged Regression
(BMI_2 as outcome)

	B	β	p
Vigorous Activity, T1	**-.001**	**.000**	**.964**
BMI, T1	.926	.888	.000

Both Difference Score and Lagged Regression
(BMI_2 – BMI_1 as outcome)

	B	β	p
Vigorous Activity, T1	**-.001**	**.000**	**.964**
BMI, T1	-.074	-.152	.000

Note that the value for the BMI effect in the both model is -.074 = .926 – 1, or $b_2 - 1$.

Conceptual Distinction in the Questions Asked

- Difference score prediction: "Whose score is most likely to increase or decrease over time?"
  - does not address pre-existing differences due to x (i.e., direction of change in y arbitrary)
  - does not address regression toward the mean

- Lagged regression: "Is x a likely cause of y?"
  - does not quantify the amount of change or variability in change
  - does not describe who is most likely to change

Conceptual Distinction in the Questions Asked

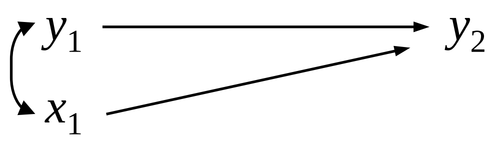
$$x_1 \longrightarrow y_{2\text{-}1}$$

is statistically equivalent to

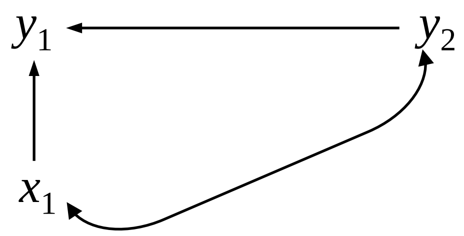
$$x_1 \longrightarrow y_{1\text{-}2}$$

Conceptual Distinction in the Questions Asked

but



is *not* statistically equivalent to



Several Limitations of Both Approaches

- Measurement error

Effects of Measurement Error Effects on Lagged Regression

- Measurement error in x_1 attenuates relationship of x_1 to y_1 and y_2 in lagged regression, thus
  - Autoregressive effect (y_1 to y_2) may be underestimated, leading to over estimated of x_1 effect
  - Underestimate of relationship between x_1 and y_1 means x_1 is not fully controlled.
  - In ANCOVA context, measurement error on y_1 and y_2 leads to over- or underestimate of group difference, depending on direction of difference at pretest and direction of the difference at posttest (see Campbell & Kenny, 1999, for more detailed analysis)

Effects of Measurement Error Effects on Difference Score Prediction

- Measurement error increases the variability of $y_{2\text{-}1}$
- Measurement error in x_1 or y attenuates relationship between x_1 and the difference score, $y_{2\text{-}1}$
- Because measurement error increases the variance of $y_{2\text{-}1}$, part of inter-individual variability in change is not true variability or explainable

Effects of Measurement Error Effects on Difference Score Prediction

- ***Does not*** impact on average difference, where $y_t = T_t + e_t$

$$E\left(y_t - y_{t-1}\right) = E\left[\left(T_t + e_t\right) - \left(T_{t-1} + e_{t-1}\right)\right]$$
$$= \left[E(T_t) + E(e_t)\right] - \left[E(T_{t-1}) + E(e_{t-1})\right]$$
$$= \left[E(T_t) + 0\right] - \left[E(T_{t-1}) + 0\right]$$
$$= E(T_t) - E(T_{t-1})$$
$$= E(T_t - T_{t-1})$$

Effects of Measurement Error Effects on Difference Score Prediction

- *Does* impact variance of differences

$$Var\left(y_t - y_{t-1}\right) = Var\left[\left(T_t + E_t\right) - \left(T_{t-1} + E_{t-1}\right)\right]$$
$$= Var\left(T_t\right) + Var\left(E_t\right) + Var\left(T_{t-1}\right) + Var\left(E_{t-1}\right) - 2Cov\left(T_t, T_{t-1}\right)$$

Effects of Measurement Error Effects on Difference Score Prediction

- True that reliability of the difference score is usually considerably below the reliability of the individual scores

$$\rho_{y_{2-1}} = \frac{\rho_y - r_{12}}{1 - r_{12}}$$

$\rho_{y_{2-1}}$ is the reliability of the difference score ρ_y is the reliability of either the T1 or T2 measure (assuming they are equal for simplicity here) and r_{12} is the correlation between y_1 and y_2.

Effects of Measurement Error Effects on Difference Score Prediction

- As the autocorrelation between in y_2 and y_1 approaches 1, the reliability of the difference score decreases

	Average reliability of two tests					
Correlation between tests	.50	.60	.70	.80	.90	.95
.00	.50	.60	.70	.80	.90	.95
.40	.17	.33	.50	.67	.83	.92
.50	.00	.20	.40	.60	.80	.90
.60		.00	.25	.50	.75	.88
.70			.00	.33	.67	.83
.80				.00	.50	.75
.90					.00	.50
.95						.00

Table from Pituch and Stevens (2016) adapted from Thorndike and Hagen (1977)

Effects of Measurement Error Effects on Difference Score Prediction

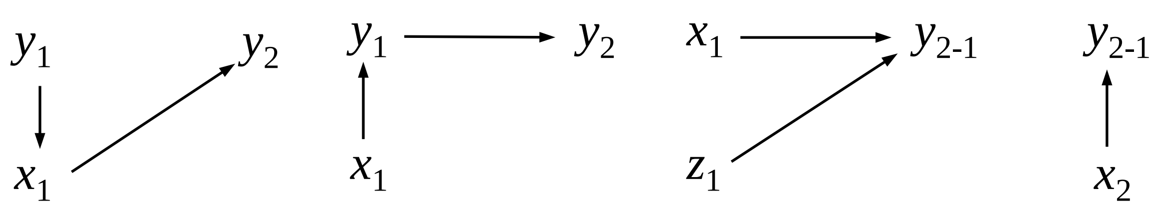
- Oddly enough, greater variability in the difference score will be associated with greater power to detect effect of x_1 on $y_{2\text{-}1}$ (Collins, 1996)

- Reliability of difference scores impacts inferences about individual differences but not the expected values, so noncentrality and reliability of difference scores not the same thing (Thomas & Zumbo, 2012).

Limitations of Both Approaches

- Because both lagged regression and difference score prediction rely on just two scores, impact of measurement error or transient (state) factors can have a large impact on the conclusions

- Either model omits alternative models between x and y. For example,



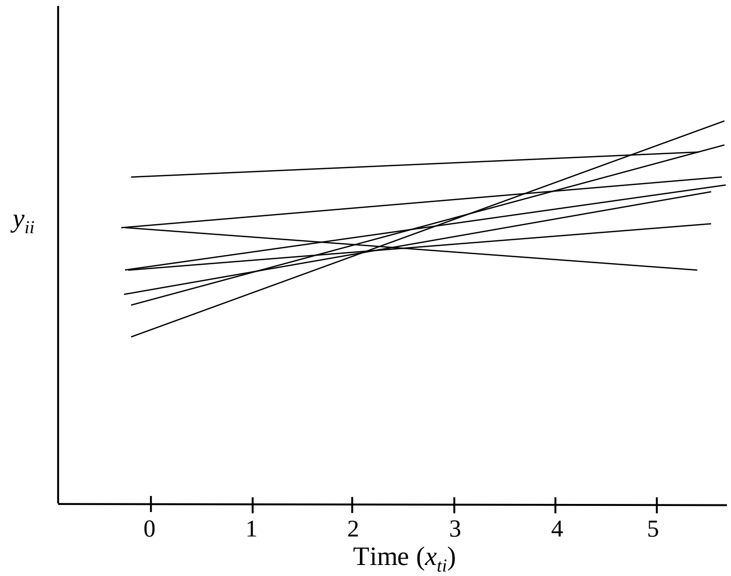
Limitations of Both Approaches

- Lag length must be appropriate
  - Smoking does not cause cancer over one month
  - Paper cuts do not cause pain one year later
- Measurement equivalency (invariance) not addressed
- Correlated errors not taken into account
- Omitted variables, time-invariant or time-varying, may bias estimates of true effects in either type of model
- Form of change is necessarily linear

Connection to Other Longitudinal Analyses

Difference Scores	Lagged Regression
Repeated Measures ANOVA	ANCOVA
Growth Curve Models	Cross-lagged Panel Models
General Estimating Equations	Time Series (autoregressive component)
Fixed Effects Regression Allison, 1990; Liker, Augustyniak, & Duncan, 1985	
Time Series (regression on t, differencing component)	
Latent Difference/Change Models McArdle & Hamagami, 2001	

Growth Curve Models



Portland State UNIVERSITY

Growth Curve Models

- Growth curve model should be less subject to measurement error at each time point and regression toward the mean, resulting in more reliable estimate of absolute change with additional repeated measures

- But they are also related to difference scores, can capture who increases or decrease over time, but do not try to address whether x cause y or y causes x.

Growth Curve Model

y_{ti}	t	
5	1	
3	1	
5	1	
0	1	
2	1	$\bar{y}_1 = 3$
8	2	
4	2	
6	2	
2	2	
9	2	$\bar{y}_2 = 5.8$
5	3	
9	3	
10	3	
7	3	
9	3	$\bar{y}_3 = 8$
$b_{yt} =$	2.5	

$\bar{y}_2 - \bar{y}_1 = 2.8$

$\bar{y}_3 - \bar{y}_2 = 2.2$

$\left[\left(\bar{y}_2 - \bar{y}_1\right) + \left(\bar{y}_3 - \bar{y}_2\right)\right] / 2 = 2.5$

Fixed Effects Regression

- Also, known as the method of first difference (Allison, 1990; Liker, Augustyniak, & Duncan, 1985), is a regression model testing whether the change in x predicts the change in y

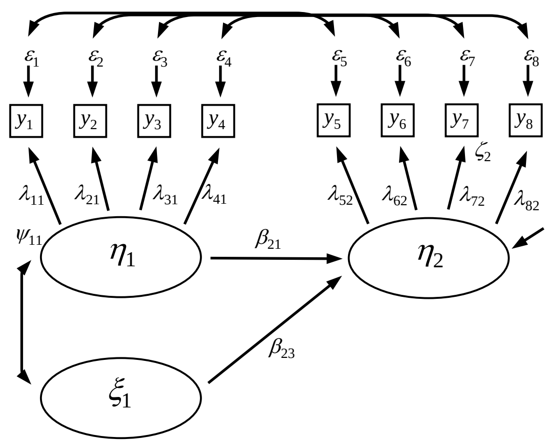
$$y_2 - y_1 = b_1\left(x_2 - x_1\right) + \left(e_2 - e_1\right)$$

- This regression model is the same as two separate regressions with two synchronous paths, assuming equal regression coefficients

$$y_1 = b_{(1)}x_1 + e_1$$

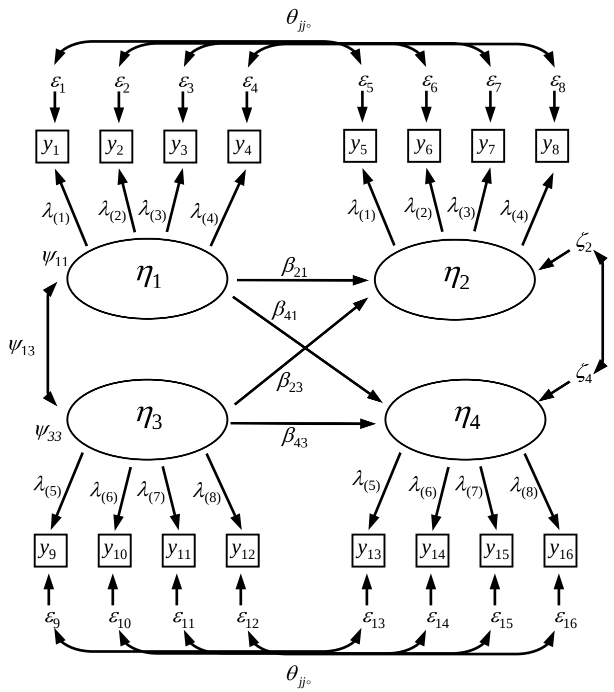
$$y_2 = b_{(1)}x_2 + e_2$$

Structural Equation Modeling Approach to Lagged Regression



- Correction for measurement error: dissattenuation of cross-lagged effect of x_1 on y_2 the autoregression effect, and the correlation of x_1 with y_1 are addressed
- Can investigate measurement equivalence over time
- Can take into account method-specific variance over time with correlated measurement residuals

SEM Cross-lagged Panel Model



Recommended Readings

Campbell, D. T., & Kenny, D. A. (1999). *A primer on regression artifacts*. Guilford Publications.
Perhaps the most thorough analysis of regression toward the mean and the difference between statistical control and difference score prediction.

Finkel, S. E. (1995). *Causal analysis with panel data*. Thousand Oaks, CA: Sage.
Discussion of cross-lagged panel models with SEM with some excellent observations about the contrast of predicting differences scores and lagged models

Lord, F. M. (1967). A paradox in the interpretation of group comparisons. *Psychological Bulletin*, *68*(5), 304.
Classic paper on the difference between difference scores and statistical control.

MacKinnon, D. P. (2008). *Introduction to Statistical Mediation Analysis*. New York: Erlbaum.
pp. 193–199 provides an excellent brief summary and many of the key references to the change-score-versus-lagged-regression debate.

Newsom, J.T. (2015). *Longitudinal Structural Equation Modeling: A Comprehensive Introduction.* New York: Routledge.
Chapter 4 is in-depth discussion of the stability and change concepts presented here with many more references

Plewis, I. (1985). *Analysing change: Measurement and exploration using longitudinal data*. Chichester, UK: Wiley.
Good overview of fundamental issues related to analysis of change, including change scores, lagged regression analysis, discussion of continuous and categorical variables, and basic SEM concepts for longitudinal data.

References

Allison, P. D. (1990). Change scores as dependent variables in regression analysis. *Sociological methodology*, 93-114.

Allison, P. D. (2005). *Fixed effects regression methods for longitudinal data using SAS*. Sas Institute.

Campbell, D. T., & Kenny, D. A. (1999). *A primer on regression artifacts*. Guilford Publications.

Collins, L. M. (1996). Is reliability obsolete? A commentary on" Are simple gain scores obsolete?". *Applied Psychological Measurement*, *20*(3), 289-292.

DuBois, P. H. (1957). *Multivariate correlational analysis.* Harper.

Liker, J. K., Augustyniak, S., & Duncan, G. J. (1985). Panel data and models of change: A comparison of first difference and conventional two-wave models. *Social Science Research*, *14*(1), 80-101.

Lord, F. M. (1956). The measurement of growth. *ETS Research Bulletin Series*, *1956*(1), i-22.

Lord, F. M. (1967). A paradox in the interpretation of group comparisons. *Psychological bulletin*, *68*(5), 304.

Maassen, G. H. (2001). The unreliable change of reliable change indices. *Behaviour Research and Therapy*, *39*(4), 495-498.

McArdle, J. J., & Hamagami, F. (2001). Latent difference score structural models for linear dynamic analyses with incomplete longitudinal data. In Collins & Sayer... Collins, Linda M. (Ed); Sayer, Aline G. (Ed), (2001). *New methods for the analysis of change. Decade of behavior*, 139-175. Washington, DC, US: American Psychological Association.

McNemar, Q. (1958). On growth measurement. *Educational and Psychological Measurement*, *18*(1), 47-55.

Pituch, K. A., Stevens, J., & James, P. (2016). *Applied Multivariate Statistics for the Social Sciences: Analyses with SAS and IBM's SPSS*.

Thomas, D. R., & Zumbo, B. D. (2012). Difference scores from the point of view of reliability and repeated-measures ANOVA: In defense of difference scores for data analysis. *Educational and Psychological Measurement*, *72*(1), 37-43.

Thorndike, R., & Hagen, E. (1977). *Measurement and evaluation in psychology and education*. New York, NY: Wiley.

Wänström, L. (2009). Sample sizes for two-group second-order latent growth curve models. *Multivariate Behavioral Research*, *44*, 588-619.

Portland State University

Thanks for Listening!

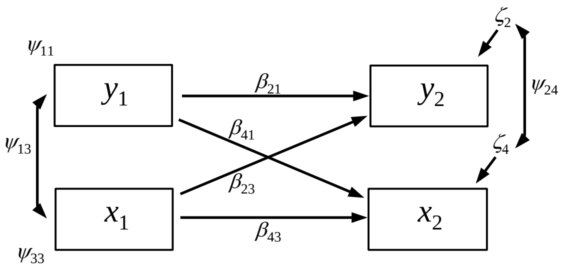
newsomj@pdx.edu

Portland State University

Extra Slides

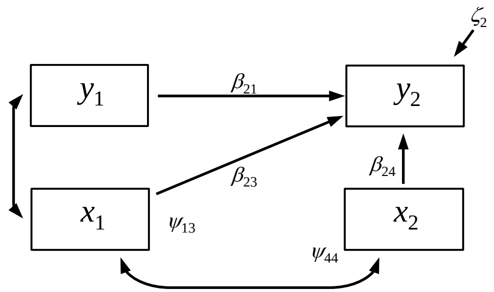
Some SEM models

Cross-lagged Panel Model



Lagged + Synchronous Regression

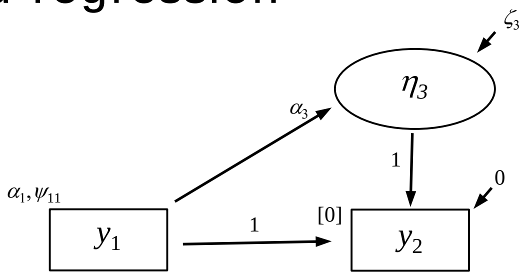
- The lagged regression could include synchronous effects too



- Interpretation: synchronous impact of x_2 controls for x_1, and the lagged effect of x_1 controls for synchronous effect of x_2

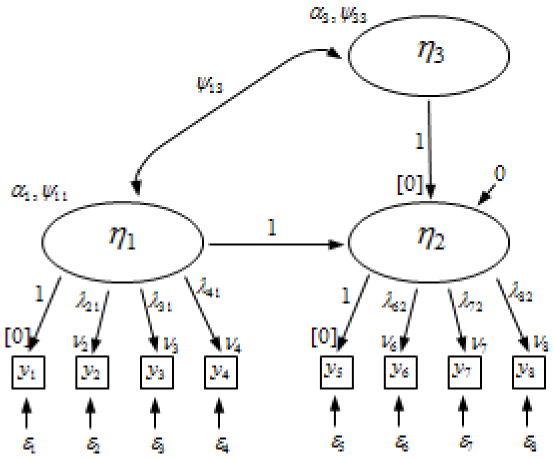
Latent Difference Score Model

- Note that if we include a predictive path instead, we have the redundant model where the difference score is regressed on the Time 1 dv, which is no different from lagged regression



Latent Difference Score Model

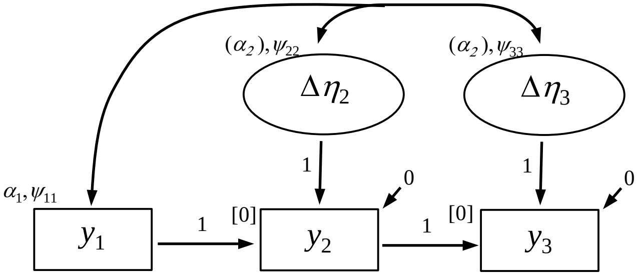
- Inclusion of latent variables to account for measurement error



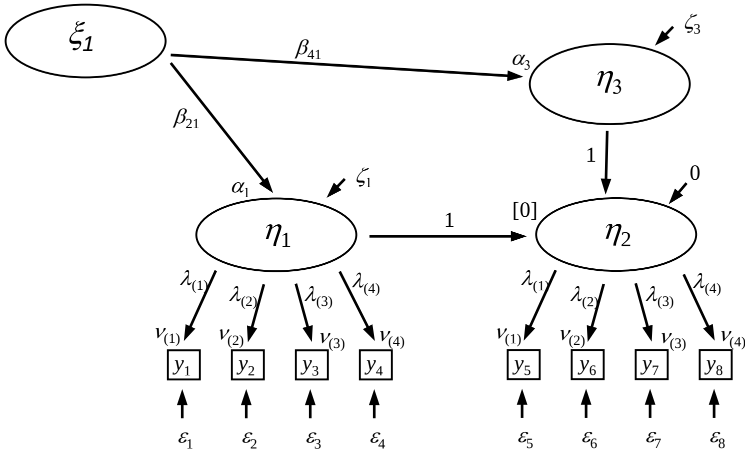
- Remember there is no impact on average mean difference estimate, but the variance (and consequently statistical power) is improved

Latent Difference Score Model (again)

- So, if the mean difference is constrained to be equal across waves, we have the same estimate as the slope in the latent growth curve model (McArdle & Hamagami, 2001)

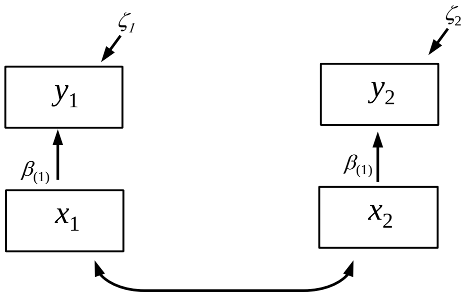


SEM Mixed Between x Repeated Measures ANOVA with Latent Variables



- Correction for measurement error: addresses attenuation due to x, but does not affect average mean difference

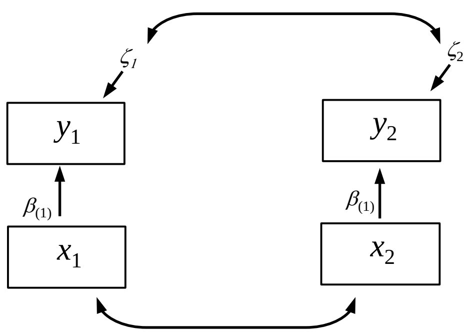
Fixed Effects Model



- Could incorporate latent variables here too

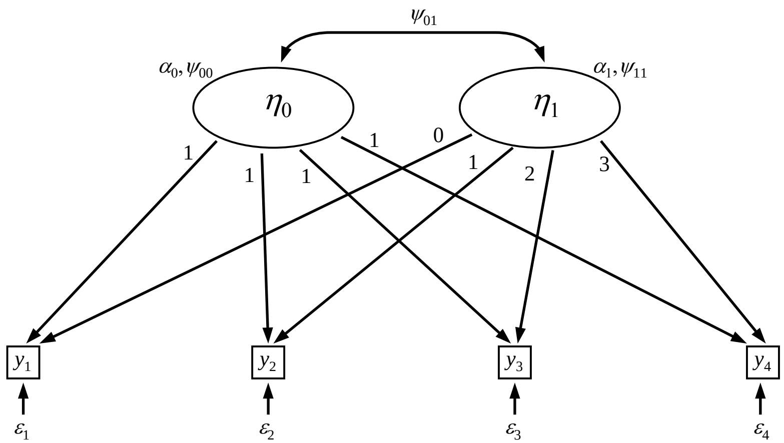
General Estimating Equations (GEE)

- The GEE model for longitudinal data in a simple case with two waves is conceptually similar to the fixed effects regression (Allison, 2005)



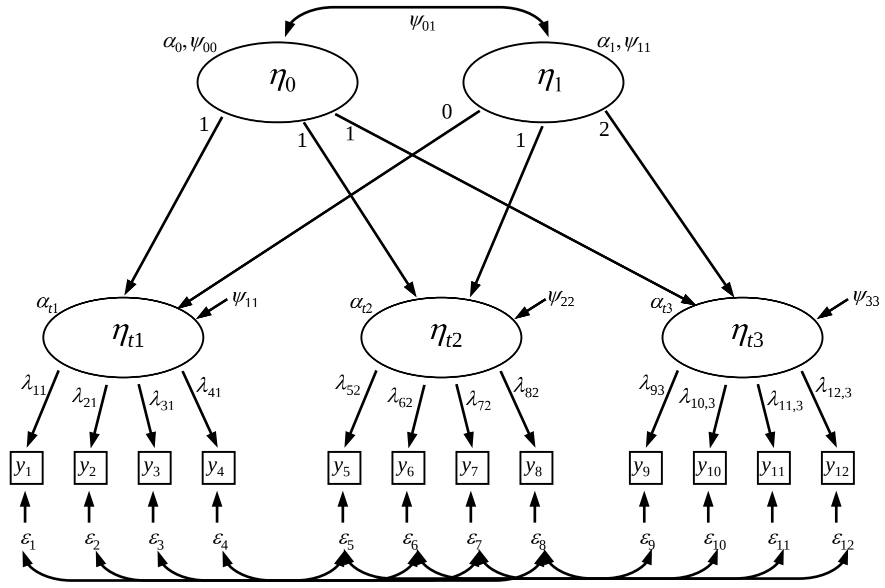
- Estimation methodology (GLS rather than ML) as well as a few other details differ

Latent Growth Curve Model



Latent Growth Curve Model

- Can be used in conjunction with latent variables at each time point (second-order LGC) to account for measurement error



- Remember there is no impact on mean change estimate, but the variance (and consequently statistical power) is improved (Wainstrom, 2009)