Statistical Models for the Inference of Within-person Relations in Longitudinal Design

Satoshi Usami

University of Tokyo

Email: usami_s@p.u-tokyo.ac.jp
HP: http://usami-lab.com/index-e.html/

Overview of presentation

- Inference of (within-person) reciprocal relation (Xt-1⇒Yt and/or Yt-1⇒Xt) between variables is popular in observational studies of behavioral science.

- Various longitudinal models have been developed in different contexts and disciplines, and an issue of model choice is still under discussion.

- Researchers in psychology often use CLPM (or RI-CLPM) by SEM, while other statistical models and estimation procedures are available.

In this presentation:

(1) Introducing various statistical models for reciprocal relation and explaining that (predetermined) RI-CLPM and DPM are safety options.

(2) Clarifying the potential problems of applying other models (LCM-SR, LCS, GCLM).

(3) Briefly explaining an alternative estimation approach (Usami, 2022) to account for time-varying treatments/predictors and confounders.

Outline

- Introduction

- Longitudinal models to examine reciprocal relation

CLPM　RI-CLPM　LCM-SR　LCS　GCLM　DPM

- Issue of time-varying confounders (Usami, 2022)

- Summary and future research agenda

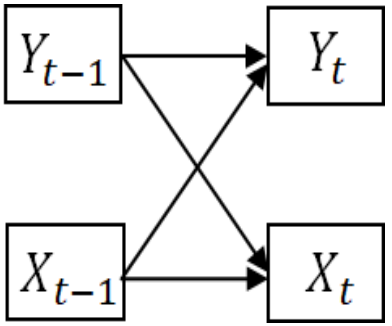
Introduction

- In educational and behavioral research, longitudinal data analysis has been widely used to understand patterns, individual differences, and relations of change in variables.

- One of the primary interests is the inference about reciprocal relation or causality between variables. RQ: How does change in one variable (X_{t-1}) affect change in another variable (Y_t)?

⇒Controlling lagged variables (Y_{t-1}) in dynamic process is a key.

- Also, researchers have shown strong interest in disaggregating within-person relations from between-person differences (or unobserved heterogeneity).



RQ: How changes in a variable influence another for the same person? (≠ group-level relation)

⇒Controlling (time-invariant) latent factors is also a key.

Introduction

- In psychology, structural equation modeling (SEM)-based approach has been popular for uncovering reciprocal relation in observation studies, and a cross-lagged panel model (CLPM) has been a gold standard.

- However, Hamaker et al (2015) criticizes the use of CLPM for the inference of within-person relation, and suggested a random-intercept CLPM (RI-CLPM), which includes common factors called stable trait factors (reaching more than 1600 citations on Google as of October 2022).

- Increase of research that conducts secondary analysis and/or compares estimation results through CLPM, RI-CLPM and other models (e.g., Orth et al., 2021).

Introduction

Smoking intensity and exposure to smoking behavior through movie (Usami et al 2019)

Parameters and model fit indices	CLPM			RI-CLPM			LCM-SR			LCS		
	Est.	*SE*	*p*	Est.	*SE*	*p*	Est.	*SE*	*p*	Est.	*SE*	*p*
β_y	.818	.006	.000	.700	.012	.000	.349	.027	.000	.792	.033	.000
γ_y	.037	.013	.003	−.005	.022	.824	.026	.025	.308	−.131	.094	.163
β_x	.511	.006	.000	.166	.011	.000	.082	.014	.000	.736	.047	.000
γ_x	.009	.004	.022	−.008	.007	.250	.015	.011	.178	−.019	.009	.042
CFI	.875			.959			.973			.973		
TLI	.879			.958			.973			.973		
AIC	73465.636			72224.166			72009.638			72015.863		
BIC	73607.512			72385.389			72164.412			72170.637		
RMSEA [95% CI]	.077 [.074, .079]			.045 [.042, .048]			.036 [.033, .039]			.036 [.033, .039]		
SRMR	.113			.066			.042			.044		
Degrees of freedom	68			65			66			66		
Number of parameters	22			25			24			24		

Usami et al (2019)

Magnitude, sign and statistical significance of cross-lagged parameters (γ) estimates differ among models (Hamaker et al., 2015; Orth et al., 2021).

Introduction

- Various longitudinal models for to inferring within-person relation have been developed in different contexts and disciplines.

- However, their relations have not been well recognized among researchers and an issue regarding model choice is still under discussion.

- This lack of insight makes it difficult for researchers to select an appropriate statistical model when analyzing longitudinal data, although estimation results can be largely influenced according to the choice of model.

- Researchers in psychology often apply only CLPM or RI-CLPM, while other statistical models and estimation procedures are available.

Outline

- Introduction

- Longitudinal models to examine reciprocal relation

CLPM RI-CLPM LCM-SR LCS GCLM DPM

- Issue of time-varying confounders (Usami, 2022)

- Summary and future research agenda

Related works

- Usami,S. (2022). Within-person variability scores based causal inference: A two-step estimation for joint effects of time-varying treatments. *Psychometrika,* in press https://link.springer.com/content/pdf/10.1007/s11336-022-09879-1.pdf (open access)

- Usami,S. (2021). On the differences between general cross-lagged panel model and random-intercept cross-lagged panel model: Interpretation of cross-lagged parameters and model choice. *Structural Equation Modeling, 28*, 331-344. https://www.tandfonline.com/doi/pdf/10.1080/10705511.2020.1821690 (open access)
- Usami,S., Todo,N., & Murayama,K. Modeling reciprocal effects in medical research: Critical discussion on the current practices and potential alternative models. *PLOS ONE. 14*(9): e0209133. (open access) https://journals.plos.org/plosone/article/file?id=10.1371/journal.pone.0209133&type=printable

- Usami,S., Murayama,K., & Hamaker,E.L. (2019). A unified framework of longitudinal models to examine reciprocal relations. *Psychological Methods, 24*, 637-657. https://psycnet.apa.org/fulltext/2019-21491-001.pdf (open access)

Coming soon (English version is under preparation)

- Usami,S. (in press). Statistical models for the inference of within-person relations: A random intercept cross-lagged panel model and its interpretation. *The Japanese Journal of Developmental Psychology* (written in Japanese).

CLPM

$$y_{it} = \alpha_{yt} + \beta_{yt} y_{i(t-1)} + \gamma_{yt} x_{i(t-1)} + d_{yit}$$
$$x_{it} = \alpha_{xt} + \beta_{xt} x_{i(t-1)} + \gamma_{xt} y_{i(t-1)} + d_{xit}$$
$$\begin{pmatrix} d_{yit} \\ d_{xit} \end{pmatrix} \sim N\left(\begin{pmatrix} 0 \\ 0 \end{pmatrix}, \begin{pmatrix} \omega_{yt}^2 & \omega_{xyt} \\ \omega_{xyt} & \omega_{xt}^2 \end{pmatrix} \right)$$

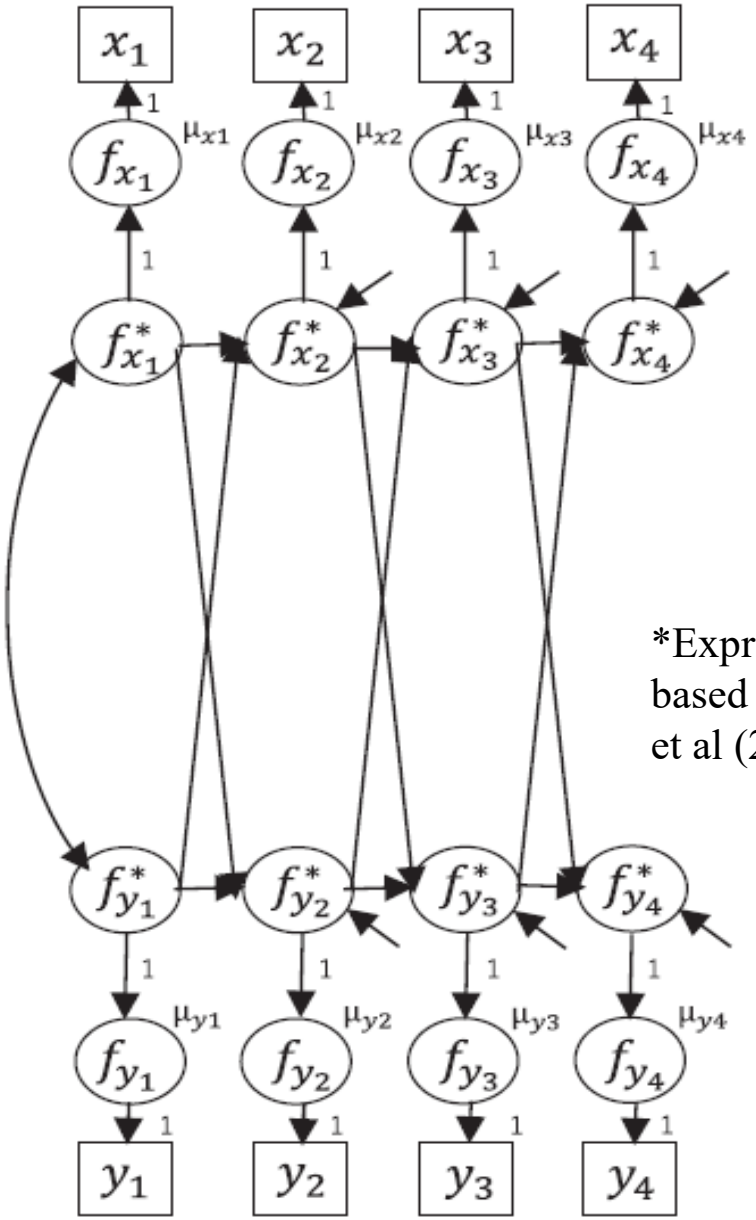
i: individual t: time point x, y: observations d: residuals
α: intercept β: autoregressive coefficient, γ: cross-lagged coefficient

Other formulation (use temporal group means μ instead of intercept α).

$$y_{it} = \mu_{yt} + y_{it}^* \qquad y_{it}^* = \beta_{yt} y_{i(t-1)}^* + \gamma_{yt} x_{i(t-1)}^* + d_{yit}$$
$$x_{it} = \mu_{xt} + x_{it}^* \qquad x_{it}^* = \beta_{xt} x_{i(t-1)}^* + \gamma_{xt} y_{i(t-1)}^* + d_{xit}$$

μ: temporal group means x^*, y^* : temporal deviations

CLPM



*Expression of this path diagram is based on a unified framework of Usami et al (2019)

RI-CLPM (Hamaker et al., 2015)

$$y_{it} = \mu_{yt} + I_{yi} + y^*_{it} \qquad y^*_{it} = \beta_{yt} y^*_{i(t-1)} + \gamma_{yt} x^*_{i(t-1)} + d_{yit}$$
$$x_{it} = \mu_{xt} + I_{xi} + x^*_{it} \qquad x^*_{it} = \beta_{xt} x^*_{i(t-1)} + \gamma_{xt} y^*_{i(t-1)} + d_{xit}$$

μ_{yt} , μ_{xt}: temporal group means at t

I_{yi}, I_{xi}: stable trait factors of i $\quad E(I_{yi}) = E(I_{xi}) = 0$

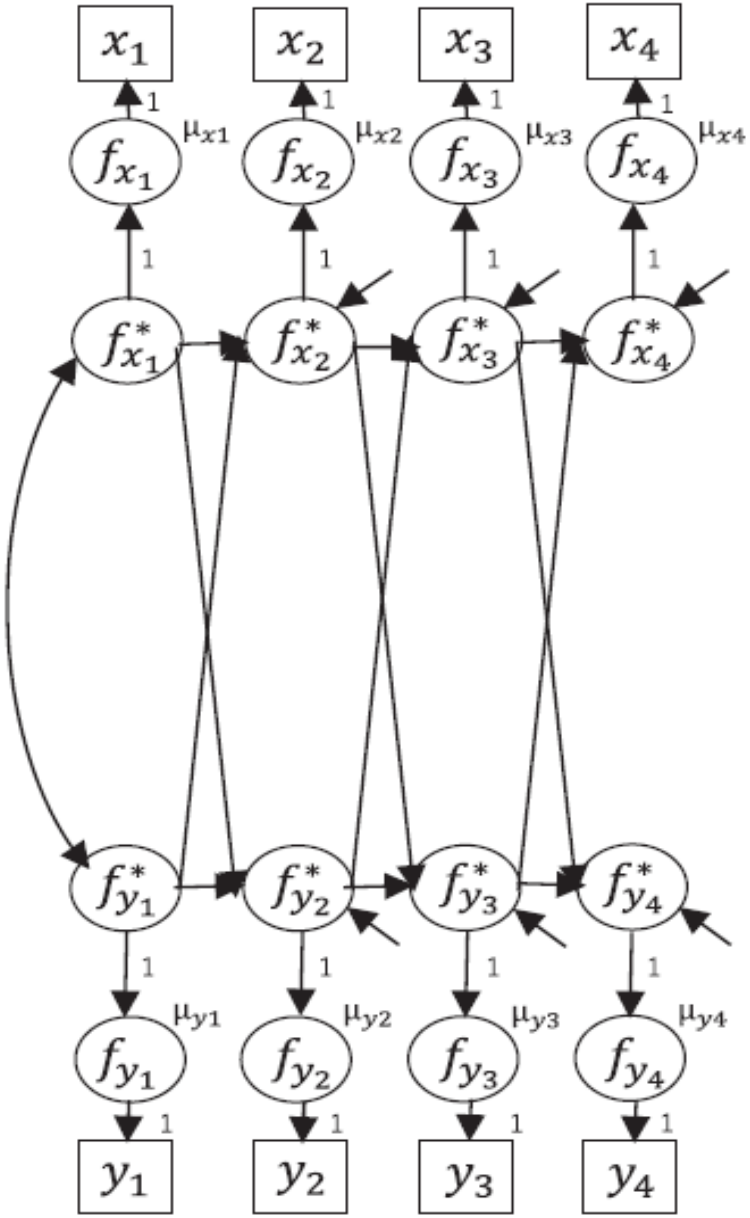
x^*_{it}, y^*_{it}: temporal deviation from expected value of i ($\mu_{yt} + I_{yi}$, $\mu_{xt} + I_{xi}$)

$\Rightarrow x^*_{it}, y^*_{it}$ represent within-person variability.

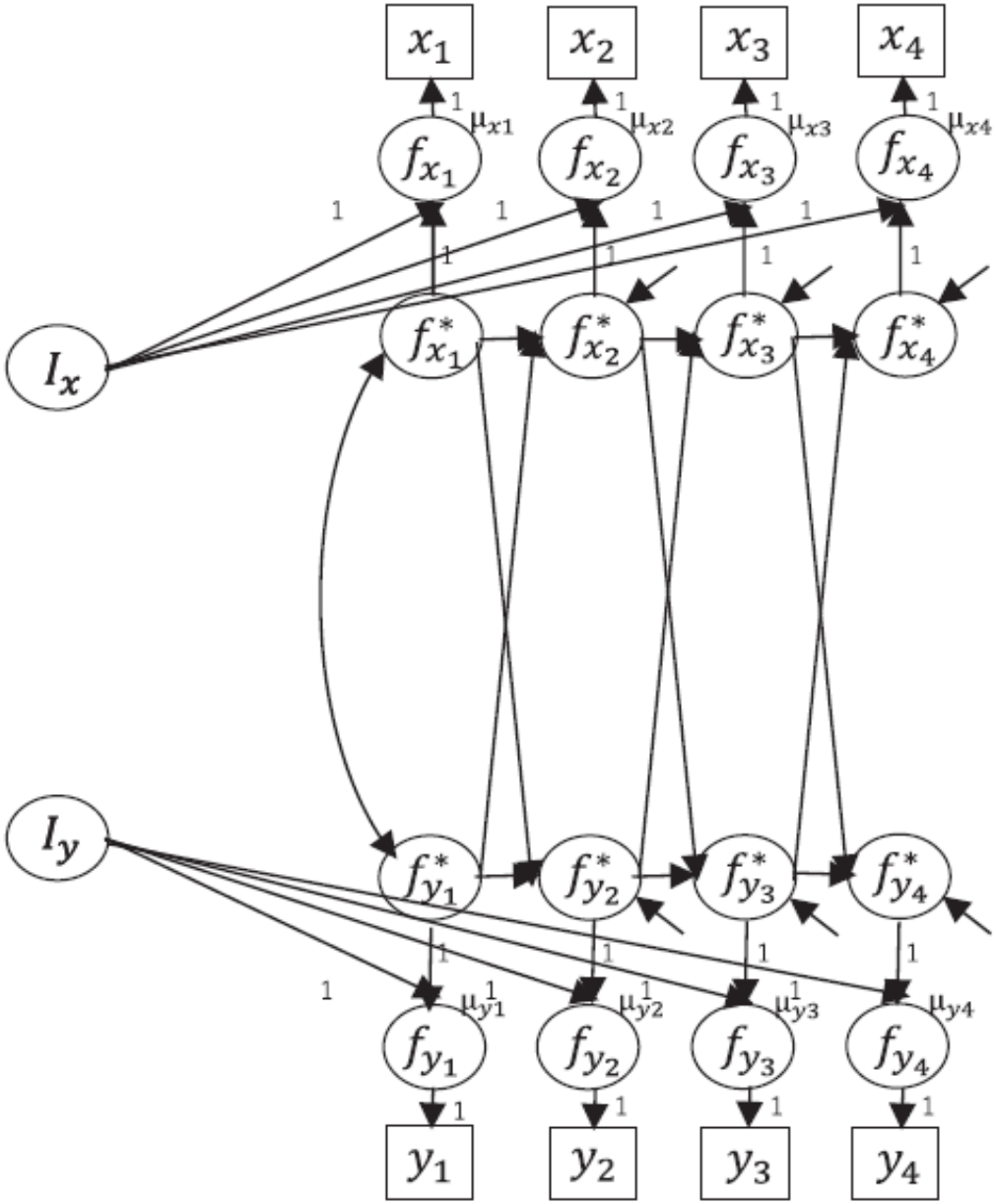
・Stable trait factors represent stable individual differences over time.

・Stable trait factors are assumed to be uncorrelated with within-person variability: $cov(I_{yi}, y^*_{it}) = cov(I_{yi}, x^*_{it}) = cov(I_{xi}, y^*_{it}) = cov(I_{xi}, x^*_{it}) = 0$

CLPM



RI-CLPM



*Expressions of diagrams are based on a unified framework of Usami et al (2019)

RI-CLPM

・Data with $T \geqq 3$ is required for model identification.

・RI-CLPM often shows better model fit than CLPM.

・Typically, estimates of β in the RI-CLPM become smaller than that of CLPM. Estimates of γ would also differ between models.

・RI-CLPM is a possible option, but there are many other models.

LCM-SR (latent curve model with structured residuals; Curran et al, 2013)

$$y_{it} = I_{yi} + (t-1)S_{yi} + y^*_{it} \qquad y^*_{it} = \beta_{yt} y^*_{i(t-1)} + \gamma_{yt} x^*_{i(t-1)} + d_{yit}$$
$$x_{it} = I_{xi} + (t-1)S_{xi} + x^*_{it} \qquad x^*_{it} = \beta_{xt} x^*_{i(t-1)} + \gamma_{xt} y^*_{i(t-1)} + d_{xit}$$

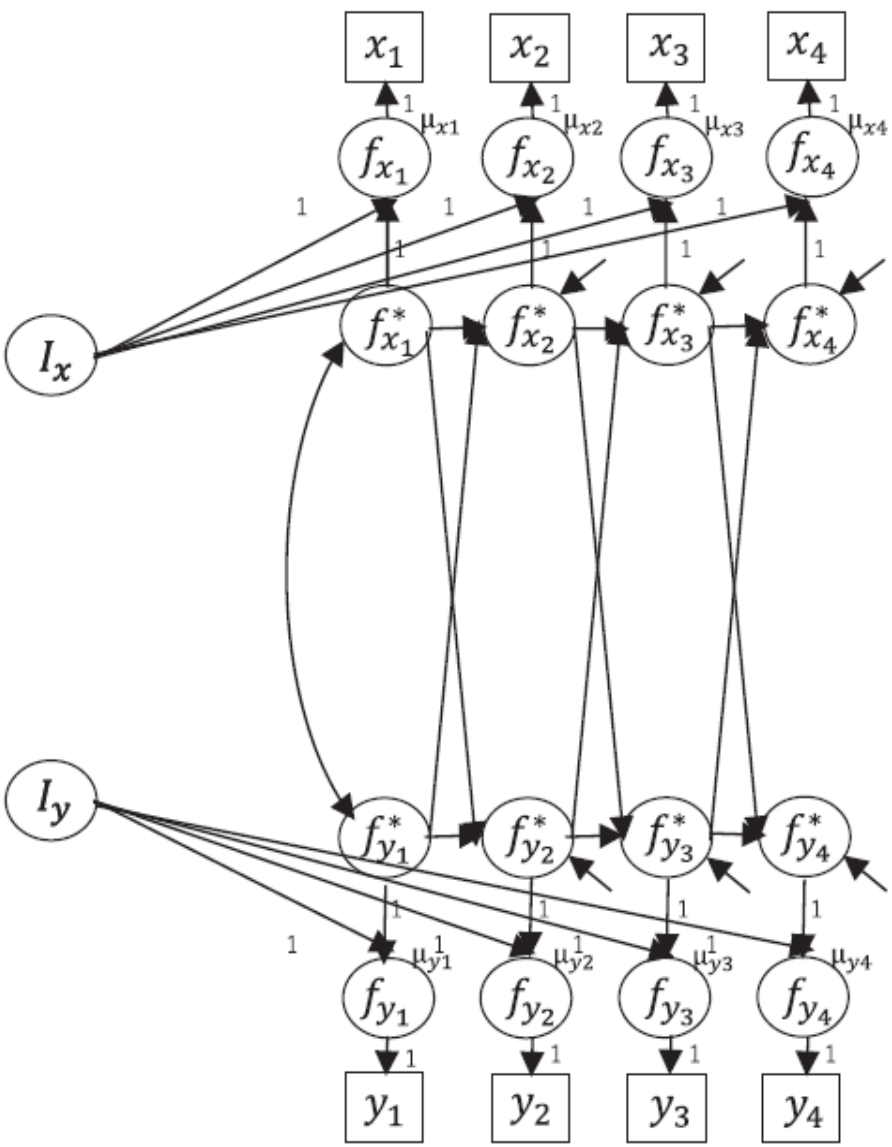
・Growth trajectory (mean structure) is modeled with growth factors (intercept factor *I* and slope factor *S*). ⇒ group mean μ is not included.

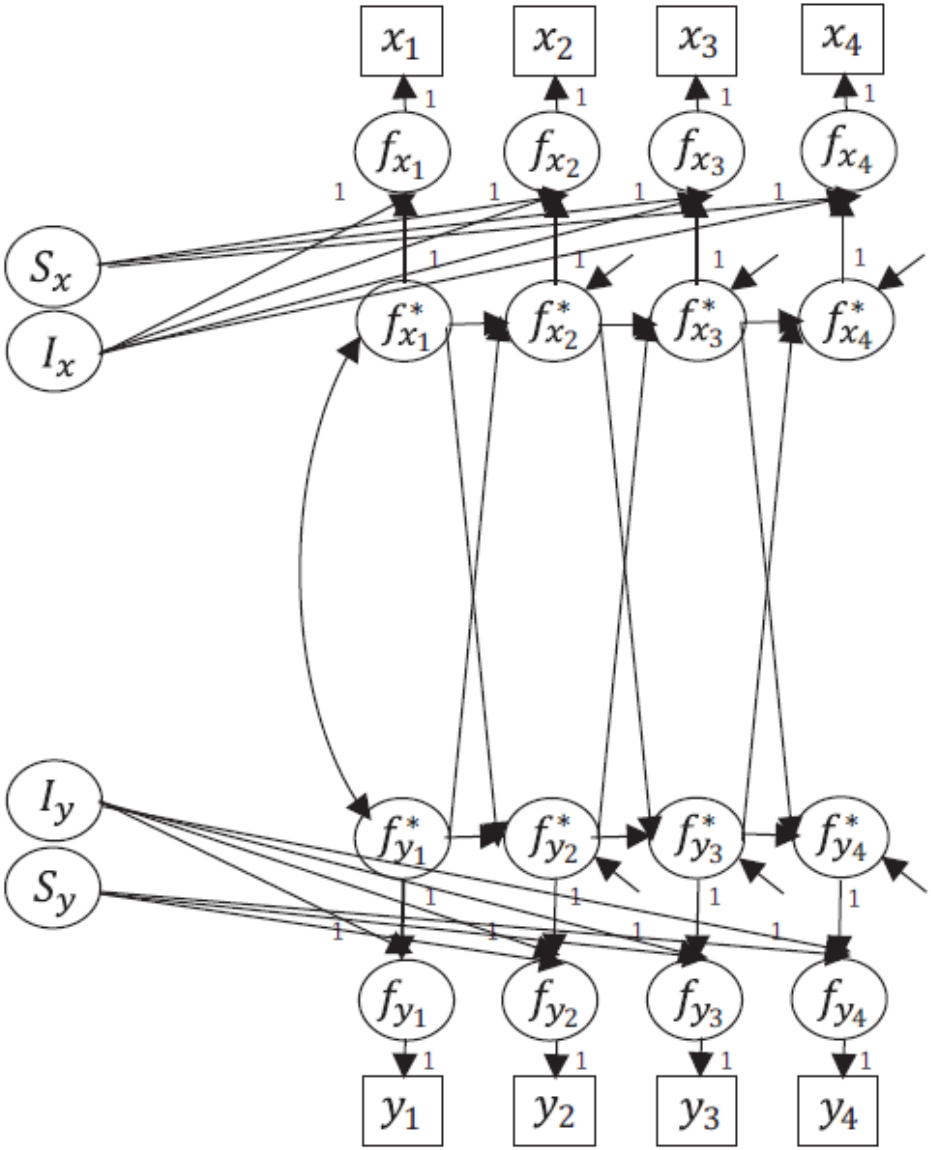
・Growth factors further decompose within-person variability of RI-CLPM into $(t-1)S_{yi}$ and y^*_{it} (or $(t-1)S_{xi}$ and x^*_{it}).

・*S* indicates individual differences of changes as (linear) trend. If this individual differences in growth trajectories are considered a critical component of reciprocal relations, γ would be biased.
⇒The problem of overadjustment.
⇒"Throwing the baby out with the bathwater" (Usami et al., 2019).

RI-CLPM

LCM-SR





*Expressions of diagrams are based on a unified framework of Usami et al (2019)

LCS (latent change score model; McArdle & Hamagami, 2001)

$$y_{it} = f_{yit} + \varepsilon_{yit} \quad f_{yit} = A_{yi} + \beta_y f_{yi(t-1)} + \gamma_y f_{xi(t-1)} + d_{yit}$$
$$x_{it} = f_{xit} + \varepsilon_{xit} \quad f_{xit} = A_{xi} + \beta_x f_{xi(t-1)} + \gamma_x f_{yi(t-1)} + d_{xit}$$

・f: true (or common factor) score　ε: measurement error (or unique factor)

・A: accumulating factor (Usami et al., 2019).

⇒has non-zero mean and is allowed to be correlated with f.

⇒is included in the lagged regression and it has both direct and indirect effects on observations, while stable trait factor has only direct effect.

・Time-invariant β and γ parameters are usually assumed. Estimates are influenced by linear transformation of observations if time-varying parameters are assumed.

LCS (latent change score model)

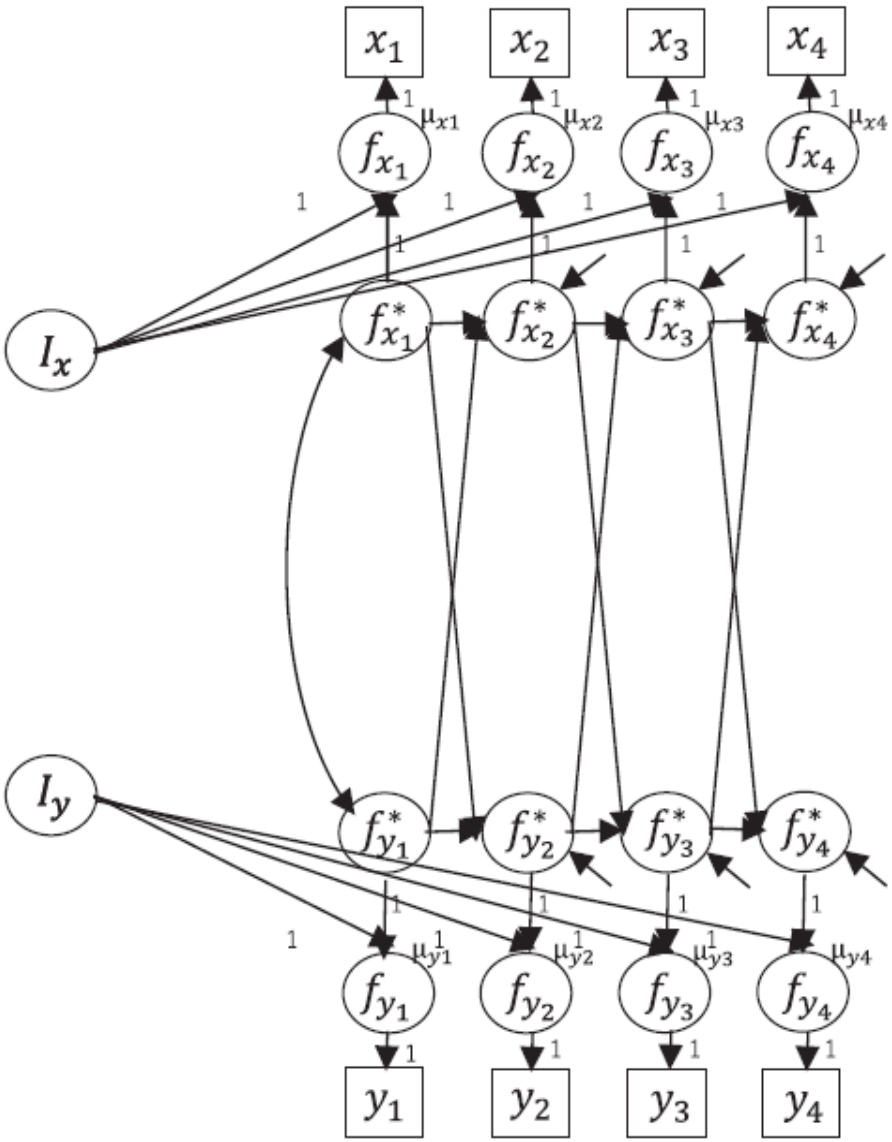
・Two primary analytic purposes (i.e., inferring reciprocal relations between variables, and describing growth trajectories/mean structures) are intertwined.

Common potential problem in LCM-SR and LCS:

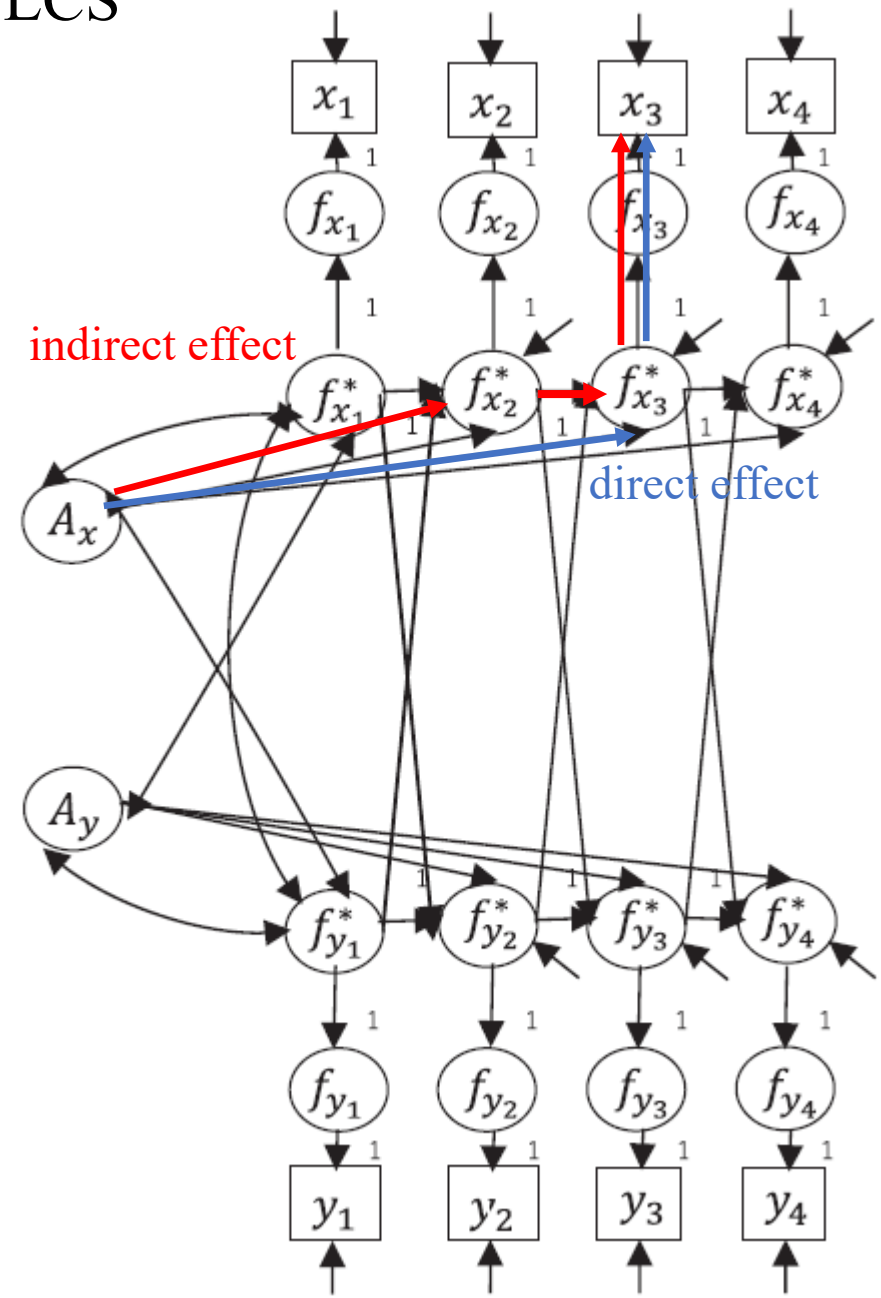
・Misspecification in growth trajectories/mean structure through latent factors (S or A) could lead to biased estimates of γ , while the growth trajectory can take on any shape in (RI-)CLPM.

・ALT (autoregressive trajectory model; e.g., Curran & Bollen, 2001) has similar potential problems as LCS.

RI-CLPM



LCS



*Expressions of diagrams are based on a unified framework of Usami et al (2019)

GCLM (general cross-lagged panel model; Zyphur et al., 2020ab)

$$y_{it} = \alpha_{yt} + \lambda_{yt}B_{yi} + \beta_{yt}y_{i(t-1)} + \gamma_{yt}x_{i(t-1)} + \delta_{yt}d_{yi(t-1)} + \zeta_{yt}d_{xi(t-1)} + d_{yit}$$
$$x_{it} = \alpha_{xt} + \lambda_{xt}B_{xi} + \beta_{xt}x_{i(t-1)} + \gamma_{xt}y_{i(t-1)} + \delta_{xt}d_{xi(t-1)} + \zeta_{xt}d_{yi(t-1)} + d_{xit}$$

- Moving average (MA) terms ($\delta_{yt}d_{yi(t-1)} + \zeta_{yt}d_{xi(t-1)}$) are included.
- Accumulating factors with time-varying loadings λ_t are included.
- If time-varying loadings λ_t are assumed, like LCM-SR, "throwing the baby out with the bathwater" problem might occur.
- Exact meanings of MA terms (or residuals) are obscure in general, while including these can be useful for prediction.
- These aspects make interpretation of cross-lagged parameters much more difficult, and it can also lead to biased estimates of γ because of possible overadjustment (Usami, 2021).

DPM (dynamic panel model)

$$y_{it} = \alpha_{yt} + A_{yi} + \beta_y y_{i(t-1)} + \gamma_y x_{i(t-1)} + d_{yit}$$
$$x_{it} = \alpha_{xt} + A_{xi} + \beta_x x_{i(t-1)} + \gamma_x y_{i(t-1)} + d_{xit}$$

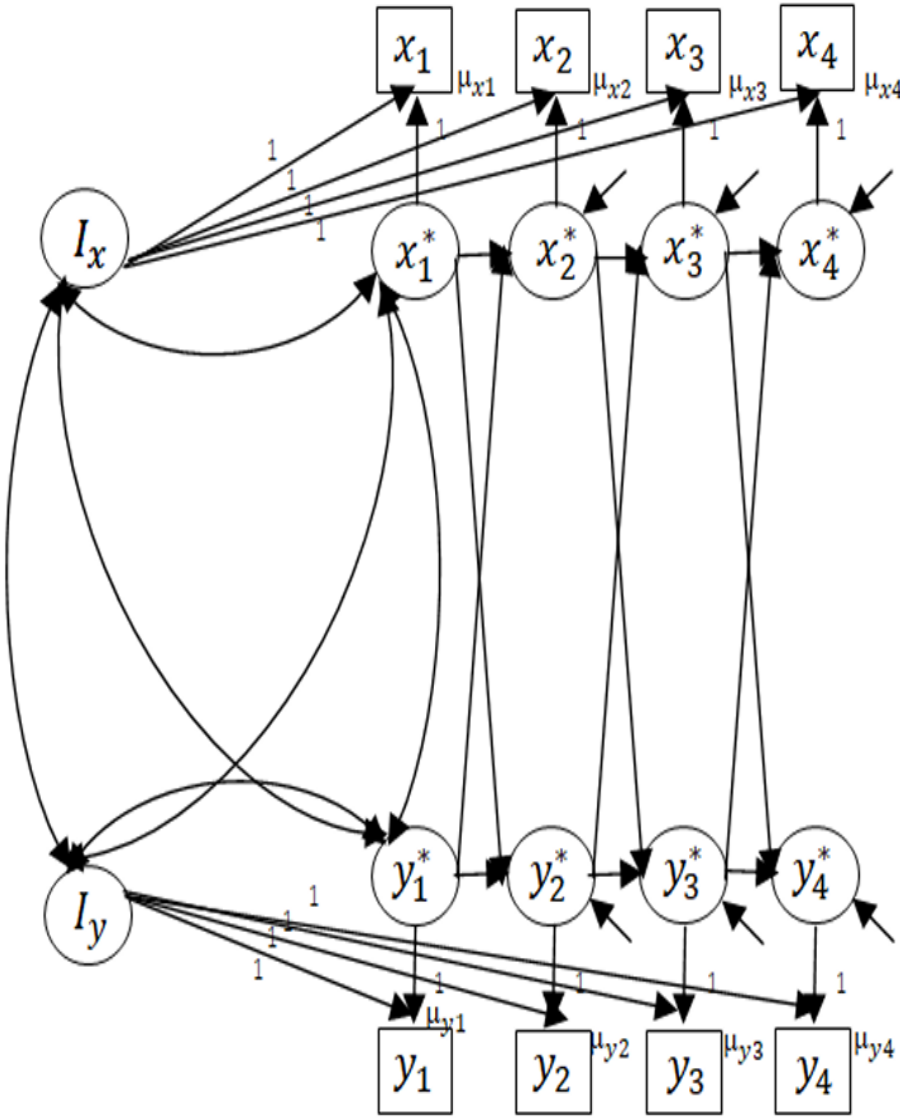
・Accumulating factor (usually called as unit effect) *A* has zero mean.

・In econometrics, usually time-invariant β and γ are assumed and unidirectional relation (X⇒Y) is considered. Other terms (e.g., trend) might be included.

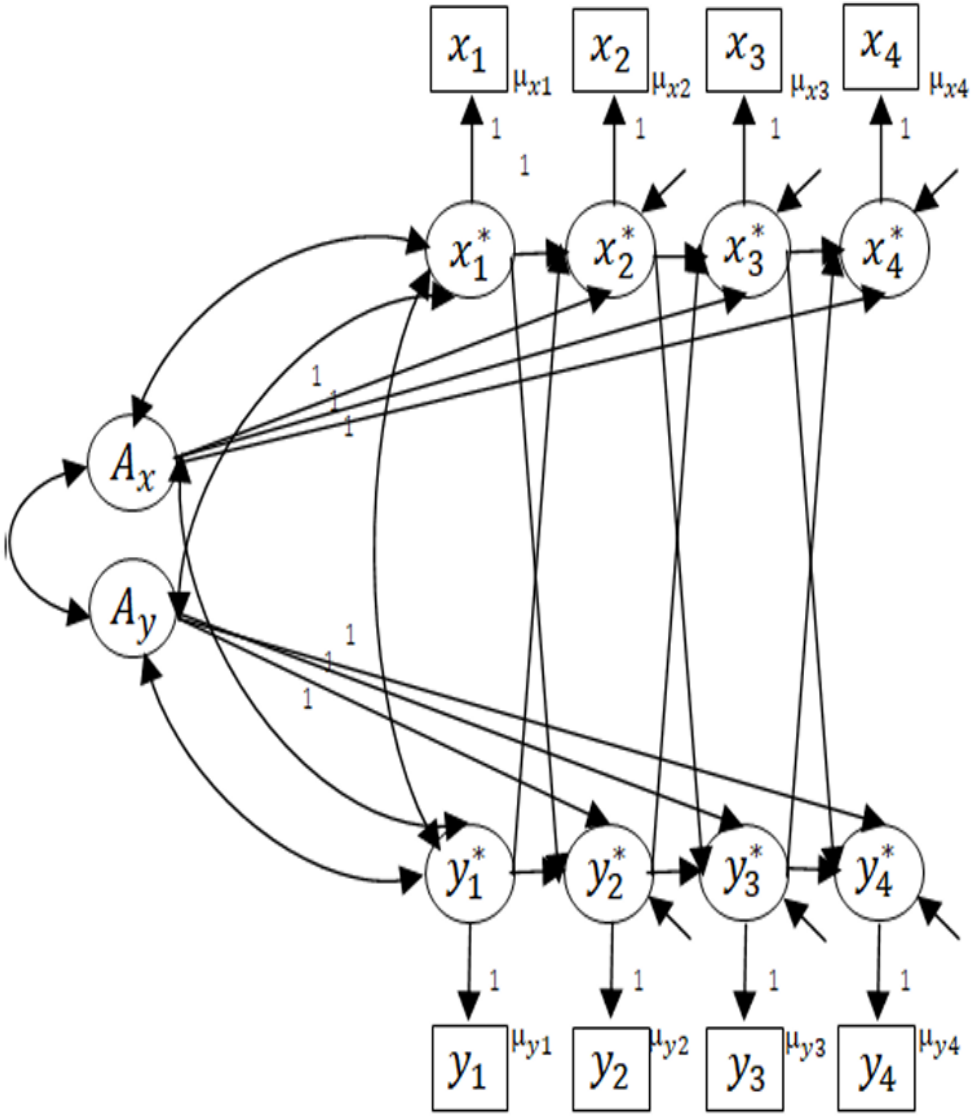
・Interestingly, if stable trait factor and within-person variability are allowed to be correlated in RI-CLPM (called predetermined RI-CLPM) and time-invariant parameters are assumed for β , γ and residual (co)variances, in this situation estimates of γ would be the same between predetermined RI-CLPM and DPM (Andersen, 2021).

・Model choice to infer within-person relation: DPM, (predetermined) RI-CLPM, STARTS (ε is included in the RI-CLPM). But see Usami et al (2019) for more discussion.

predetermined RI-CLPM



DPM



Unified Framework （Usami et al., 2019）

・To clarify the conceptual and mathematical similarities and differences among models:

$$y_{it} = f_{yit} + \varepsilon_{yit} \qquad x_{it} = f_{xit} + \varepsilon_{xit}$$

$$f_{yit} = [\mu_{yt} + \{I_{yi} + (t-1)S_{yi}]f^*_{yit}, \qquad f_{xit} = [\mu_{xt} + \{I_{xi} + (t-1)S_{xi}]f^*_{xit}$$

$$f^*_{yit} = \{A_{yi} + (t-1)B_{yi}\} + \beta_{yt} f^*_{yi(t-1)} + \gamma_{yt} f^*_{xi(t-1)} + d_{yit}$$

$$f^*_{xit} = \{A_{xi} + (t-1)B_{xi}\} + \beta_{xt} f^*_{xi(t-1)} + \gamma_{xt} f^*_{yi(t-1)} + d_{xit}$$

Three key aspects:

・Whether measurement error ε is included.

・Whether temporal group mean μ is included (or growth trajectory is modeled) for mean structure.

・Whether stable trait factor (*I*) or growth factors (*I* and/or *S*), or accumulating factors (*A* and/or *B*) are included.

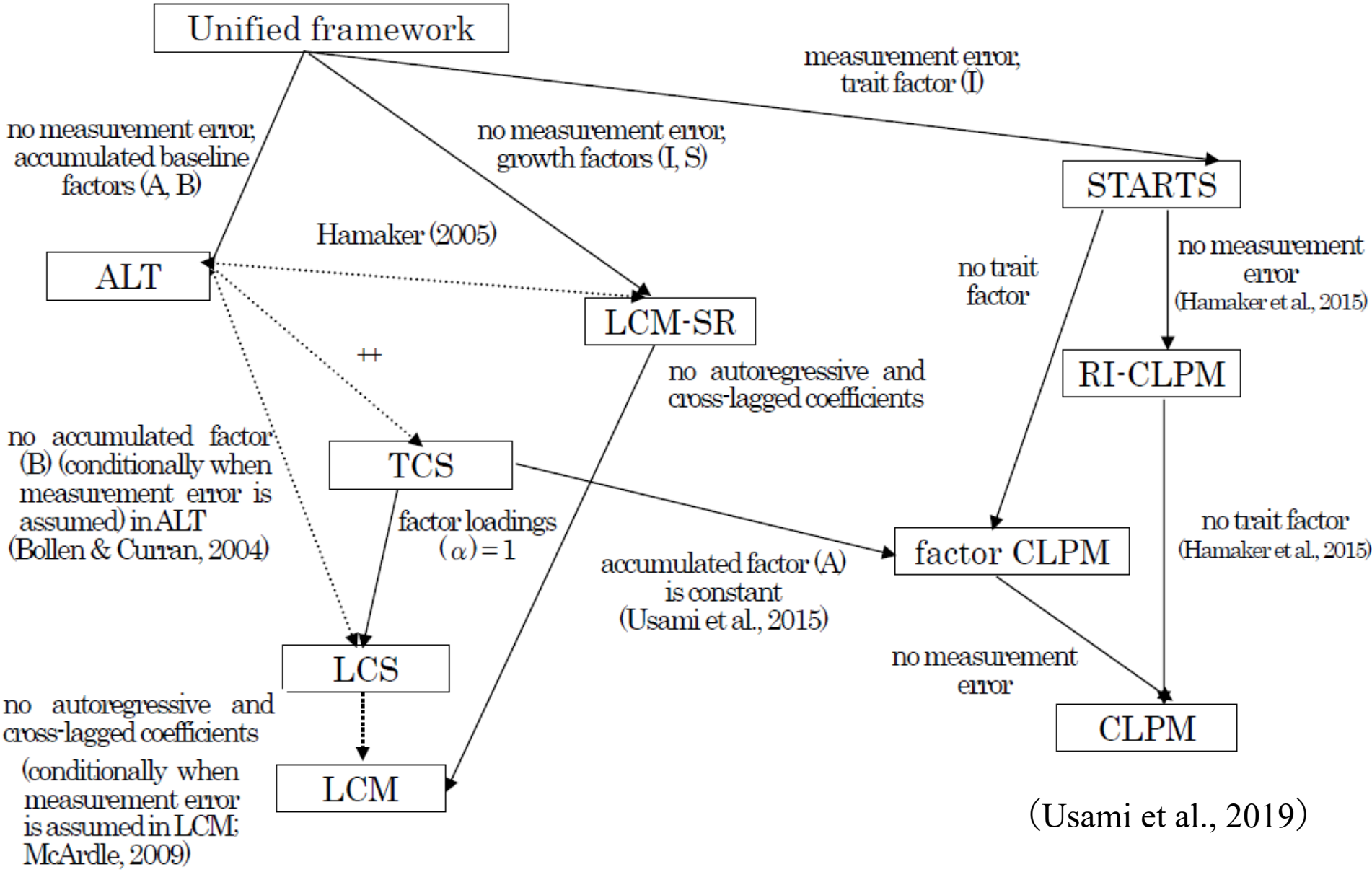
An Overview of Cross-Lagged Models Indicating Which of the Components They Include

	Measurement equations	Decomposition equations			Dynamic equations		
Model	ϵ	μ	I	S	A	B	d
CLPM	—	+	—	—	—	—	+
Factor CLPM	+	+	—	—	—	—	+
RI-CLPM	—	+	+[a]	—	—	—	+
STARTS	+	+	+[a]	—	—	—	+
LCM-SR	—	—	+	+	—	—	+
ALT	—	—	—	—	+	+	+
LCS	+	—	—	—	+	—	+[b]

Note. CLPM = cross-lagged panel model; RI-CLPM = random-intercept CLPM; STARTS = stable trait autoregressive trait and state model; LCM-SR = latent curve model with structured residuals; ALT = autoregressive latent trajectory model; LCS = latent change score model.
[a] These factors have a mean of zero, unless the μ's are fixed to zero. [b] This residual is typically fixed to zero for estimation purposes.

Usami et al (2019)

Conceptual diagrams



(Usami et al., 2019)

Outline

- Introduction

- Longitudinal models to examine reciprocal relation

CLPM　RI-CLPM　LCM-SR　LCS　GCLM　DPM

- Issue of time-varying confounders (Usami, 2022)

- Summary and future research agenda

Time-varying confounders

- Models described so far do not account for time-varying confounders. However, within-person relation is different from causal relation.

- In SEM approach, one needs to correctly specify how time-varying observed confounders are functionally related to treatments/predictors and outcomes at the within-person level.

- The RI-CLPM as SEM (or, covariance structure analysis) has obvious drawback in that it generally demands linear regressions for latent variables (i.e., within-person variability/temporal deviations) .

Epidemiological approach for time-varying treatments and confounders

- Potential outcome approach such as marginal structural models (MSMs; Robins, 1999; Robins, Hernan, & Brumback, 2000) has been popular in epidemiology to account for time-varying treatments.

- Although the number of application has been relatively infrequent, structural nested models (SNMs: Robins, 1989, 1994) with G-estimation are in principle better tailored for dealing with failure of the usual assumptions of no unobserved confounders or sequential ignorability (Robins, 1999; Robins & Hernan, 2009; Vansteelandt & Joffe, 2014).

- MSMs and SNMs use observations in the model and they do not explicitly include latent factors (e.g., stable traits). But estimation procedures used in MSMs and SNMs can be utilized to infer reciprocal relations at the within-person level.

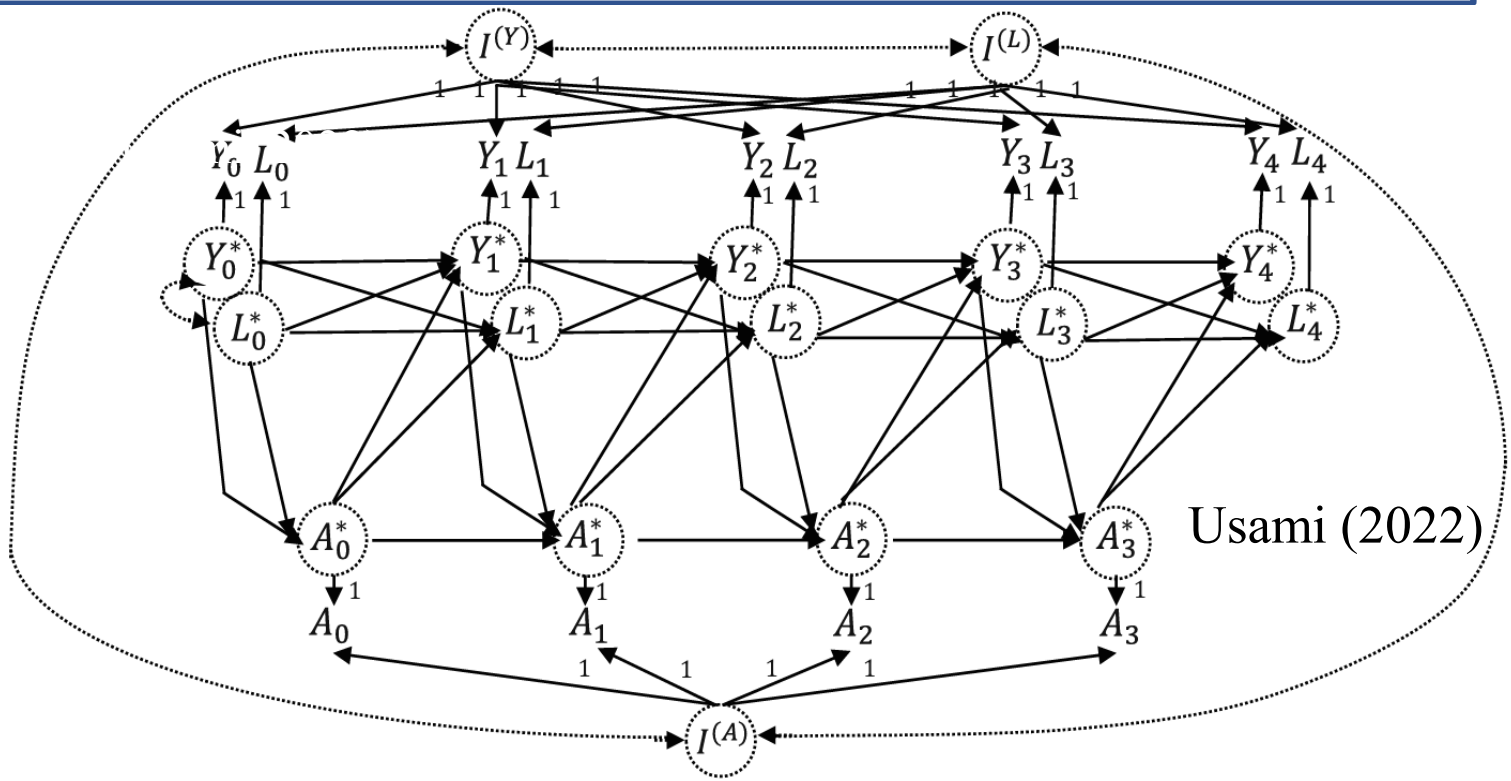
Two step estimation (Usami, 2022)

- Assuming data generating process like RI-CLPM (stable trait factor).

1: *Within-person variability score* of each individual is predicted for each variable (outcome Y, treatment A, confounder L) through SEM.

2: Causal parameters (or cross-lagged parameters) are then estimated by MSMs or SNMs, using calculated within-person variability scores.

Y: outcome
A: treatment(X)
L: observed confounder



Two step estimation (Usami, 2022)

・MSM/SNM is more flexible than the RI-CLPM in modeling how observed confounders are functionally related to outcomes and treatments/predictors.

・Notably, estimator of SNM through estimating equation has doubly robust property: consistent estimator of (joint) treatment effect if either outcome model or treatment model can be correctly specified.

・SNMs can allow one to directly model interactions/moderations effects of treatments/predictors A with observed confounders L.

・ Through simulation it was shown that the proposed method can recover causal parameters well and that causal estimates might be severely biased if one does not properly account for stable traits.

Outline

- Introduction

- Longitudinal models to examine reciprocal relation

CLPM　RI-CLPM　LCM-SR　LCS　GCLM　DPM

- Issue of time-varying confounders (Usami, 2022)

- Summary and future research agenda

Summary

- Various longitudinal models for examining reciprocal relations at the within-person level exist, and an issue regarding model choice is still under discussion.

- Researchers in psychology often use CLPM (or RI-CLPM) by SEM, while other statistical models and estimation procedures are available.

Potential problems of existing models for reciprocal relation:

- Overadjustment by including time-varying latent factors S or B (LCS and GCLM).
- Estimation results of within-person relation (γ) could be biased if growth trajectory/mean structure is misspecified (LCS and LCM-SR).
- Interpretative difficulty of γ because of inclusion of MA terms (GCLM).

Conclusion and future research agenda

・(predetermined) RI-CLPM and DPM seem to be better options for the inference of reciprocal relation at the within-person level, and comparing estimation results from these models could be especially useful.

・From the view of causal inference, SNM with G-estimation as used in Usami (2022) could be a potential estimation option to account for time-varying confounders in inferring (within-person) reciprocal relation.

Future research:

・Comparison between DPM and (predetermined) RI-CLPM in time-varying parameters setting.

・Extension of Usami (2022):

Additional simulation research/ Including measurement errors/ Assuming data generating process as DPM/ Extension to other research questions (e.g., mediation effects).

Reference

- Andersen, H.K. (2021). Equivalent approaches to dealing with unobserved heterogeneity in cross-lagged panel models? Investigating the benefits and drawbacks of the latent curve model with structured residuals and the random intercept cross-lagged panel model. *Psychological Methods*. https://doi.org/10.1037/met0000285
- Curran, P.J., & Bollen, K.A. (2001). The best of both worlds: Combining autoregressive and latent curve models. In L.M. Collins & A. G. Sayer (Eds.), *New methods for the analysis of change* (pp. 105–136). Washington, DC: American Psychological Association.
- Curran, P.J., Howard, A.L., Bainter, S.A., Lane, S.T., & McGinley, J.S. (2013). The separation of between-person and within-person components of individual change over time: A latent curve model with structured residuals. *Journal of Consulting and Clinical Psychology, 82*, 879–894.
- Hamaker, E.L., Kuiper, R.M., Grasman, R.P.P.P. (2015). A critique of the cross-lagged panel model. *Psychological. Methods, 20*, 102–116.
- McArdle, J.J., & Hamagami, F. (2001). Latent difference score structural models for linear dynamic analyses with incomplete longitudinal data. In L. Collins & A. Sayer (Eds.), *New methods for the analysis of change* (pp. 137–175). Washington, DC: American Psychological Association.
- Orth, U.D., Clark, A.M., Donnellan, B., & Robins, R. W. (2021). Testing prospective effects in longitudinal research: Comparing seven competing cross-lagged models. *Journal of Personality and Social Psychology. 120,* 1013–1034.

Reference

- Robins, J. M. (1989). The analysis of randomized and non-randomized AIDS treatment trials using a new approach to causal inference in longitudinal studies. In L. Sechrest, H. Freeman, & A. Mulley (Eds.), *Health service research methodology: A focus on AIDS* (pp. 113–159). U.S. Public Health Service, National Center for Health Services Research.
- Robins, J. M. (1994). Correcting for non-compliance in randomized trials using structural nested mean models. *Communications in Statistics-Theory and Methods, 23*, 2379–2412
- Robins, J.M. (1999). Marginal structural models versus structural nested models as tools for causal inference. *Epidemiology, 116,* 95-134.
- Robins, J.M., Hernán, M.A., & Brumback, B. (2000). Marginal structural models and causal inference in epidemiology. *Epidemiology, 11,* 550-560.
- Robins, J.M., & Hernán, M.A. (2009). Estimation of the causal effects of time-varying exposures. In G. Fitzmaurice et al. (Eds.), *Handbooks of modern statistical methods: Longitudinal data analysis* (pp. 553-599). Boca Raton: CRC Press.
- Usami,S. (2021). On the differences between general cross-lagged panel model and random-intercept cross-lagged panel model: Interpretation of cross-lagged parameters and model choice. *Structural Equation Modeling, 28*, 331-344. https://www.tandfonline.com/doi/pdf/10.1080/10705511.2020.1821690
- Usami,S. (2022). Within-person variability scores based causal inference: A two-step estimation for joint effects of time-varying treatments. *Psychometrika,* in press. https://link.springer.com/content/pdf/10.1007/s11336-022-09879-1.pdf
- Usami,S., Murayama,K., & Hamaker,E.L. (2019). A unified framework of longitudinal models to examine reciprocal relations. *Psychological Methods, 24*, 637-657. https://psycnet.apa.org/fulltext/2019-21491-001.pdf

Reference

- Vansteelandt, S., & Joffe, M. (2014). Structural nested models and g-estimation: The partially realized promise. *Statistical Science, 29*, 707–731.
- Zyphur, M.J., Allison, P.D., Tay, L., Voelkle, M.C., Preacher, K.J., Zhang, Z., Hamaker, E.L., Shamsollahi, A., Pierides, D.C., Koval, P., & Diener, E. (2020a). From data to causes I: Building a general cross-lagged panel model (GCLM). *Organizational Research Methods, 23*, 651–687.
- Zyphur, M.J., Voelkle, M.C., Tay, L., Allison, P.D., Preacher, K.J., Zhang, Z., Hamaker, E.L., Shamsollahi, A., Pierides, D.C., Koval, P., & Diener, E. (2020b). From data to causes II: Comparing approaches to panel data analysis. *Organizational Research Methods, 23*, 688–716.