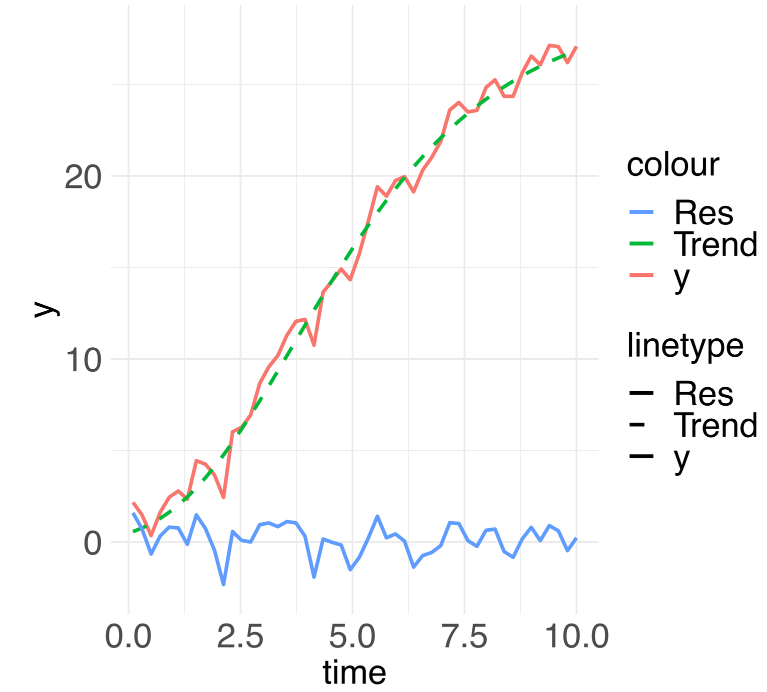
Detrending Multi-Subject, Short Time Series Data for Dynamic Structural Equation Model

Xiaoyue Xiong, Sy-Miin Chow, and Yanling Li

Department of Human Development and Family Studies, The Pennsylvania State University

BACKGROUND

Trend in a Time Series



$$y_{t,i} = \text{Trend} + \text{Residual} = \mu_{t,i} + \eta_{t,i}$$ (Craigmile, 2009)

- **Trend** $\mu_{t,i}$ long-term variation, usually a smoothly varying function of time.
- **Residual** $\eta_{t,i}$ nuanced patterns of change and random, independent and identically (i.i.d.) noise.

Problem of Trended Time Series

- Cause spurious long-range correlations in the error structures.
- Make a time series non-stationary. This violates a fundamental assumption of standard time series models such as the Vector Auto-Regression (VAR) model, and other related multi-subject extensions, such as Dynamic Structural Equation Models (DSEM).

QUESTIONS

Q1 In what ways are estimation results involving the DSEMs affected by the presence of trends?

Q2 How do different approaches of accounting for trends influence the estimation of the DSEMs?

METHOD

Q1

- **Step1:** Given nT and nP, we generate 100 Monte Carlo samples each of **no-trend** and **trended** data, corresponding to the two-level DSEM and two-level Gompertz-AR(1) models, respectively. The trend in the Gompertz-AR(1) model follows a Gompertz curve (Browne, 1993), and the residuals have an autoregressive pattern.

Table 1: Summary of Models in 100 Monte Carlo replications.

DSEM	Trend	Residual
Level1 (measurement)	$\mu_{t,i} = 0$	$\eta_{t,i} = \phi_i \eta_{t-1,i} + e_{t,i}$ $e_{t,i} \sim N(0, 1)$
Level2 (person)		$\phi_i \sim N(0.3, 0.01)$
Gompertz-AR(1) Model	**Trend**	**Residual**
Level1 (measurement)	$\mu_{t,i} = \theta_{1,i} e^{\left[-\theta_{2,i}\, e^{exp(-t\,\theta_{3,i})}\right]}$	$\eta_{t,i} = \phi_i \eta_{t-1,i} + e_{t,i}$ $e_{t,i} \sim N(0, 1)$
Level2 (person)	$\theta_{1,i} \sim N(35, 81)$ $\theta_{2,i} \sim N(4, 0.25)$ $\theta_{3,i} \sim N(0.8, 0.01)$	$\phi_i \sim N(0.3, 0.01)$

ϕ_i is person-specific autoregression coefficient.
$\theta_{1,i}$ is capability/asymptotic maximum of the person-specific Gompertz curve,
$\theta_{2,i}$ controls the displacement of the person-specific Gompertz curve along x-axis.
$\theta_{3,i}$ is growth rate of the person-specific Gompertz curve.

- **Step2:** Compare DSEM performances on no-trend samples and trended samples for different nT and nP. nT = 5, 15 or 50, and nP =150 or 500.

Q2

- **Step3:** Five approaches in total are applied to analyze the trended Monte Carlo samples generated in the step 2.

Table 2: Summary of 5 Approaches.

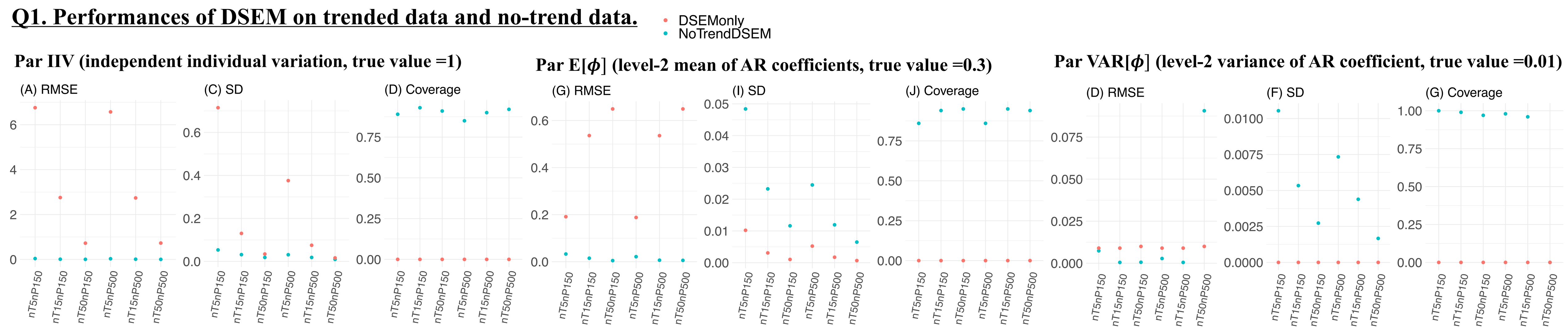
Approaches		1) Single-Stage		2) DSEM-Only	3) Two-Stage-Linear	4) Two-Stage-NLME	5) Two-Stage-NLSLIST
Procedure		Fit Data with Gompertz-AR(1) model.	Stage 1: Remove person-specific trends from data.	-	Fit a **linear model** for each person.	Fit a **Non-Linear Mixed-Effect (NLME) model** for people.	Fit one non-linear model for each person. Thus, a list of non-linear person-specific models are obtained. (**NLSLIST**)
			Stage 2: Fit residuals.	Fit data with DSEM.	Fit detrended data with DSEM.	Fit detrended data with DSEM.	Fit detrended data with DSEM.
Modes for Estimation	$\mu_{t,i} =$	$\theta_{1,i} e^{\left[-\theta_{2,i}\, e^{exp(-t\,\theta_{3,i})}\right]}$		0	$B_{0,i} + B_{1,i} \times t$	$\theta_{1,i} e^{\left[-\theta_{2,i}\, e^{exp(-t\,\theta_{3,i})}\right]}$	$\theta_{1,i} e^{\left[-\theta_{2,i}\, e^{exp(-t\,\theta_{3,i})}\right]}$
	$\eta_{t,i} =$	$\phi_i \eta_{t-1,i} + e_{t,i}$		$\phi_i \eta_{t-1,i} + e_{t,i}$	$\phi_i \eta_{t-1,i} + e_{t,i}$	$\phi_i \eta_{t-1,i} + e_{t,i}$	$\phi_i \eta_{t-1,i} + e_{t,i}$

Main R pacakges: rJAGS, MplusAutomation, nlme.

- **Step 4:** Assess performances of five approaches according to multiple indexes (Chow & Zhang, 2013).

RESULTS

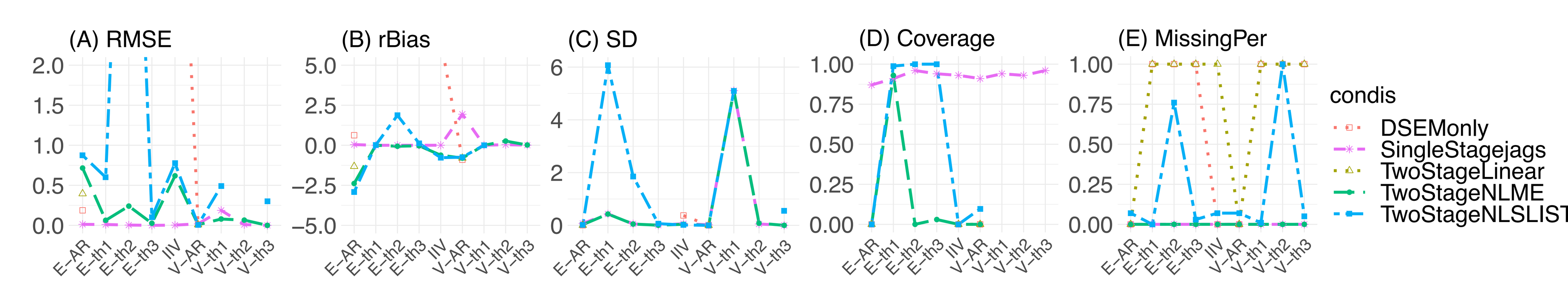
Q1. Performances of DSEM on trended data and no-trend data.



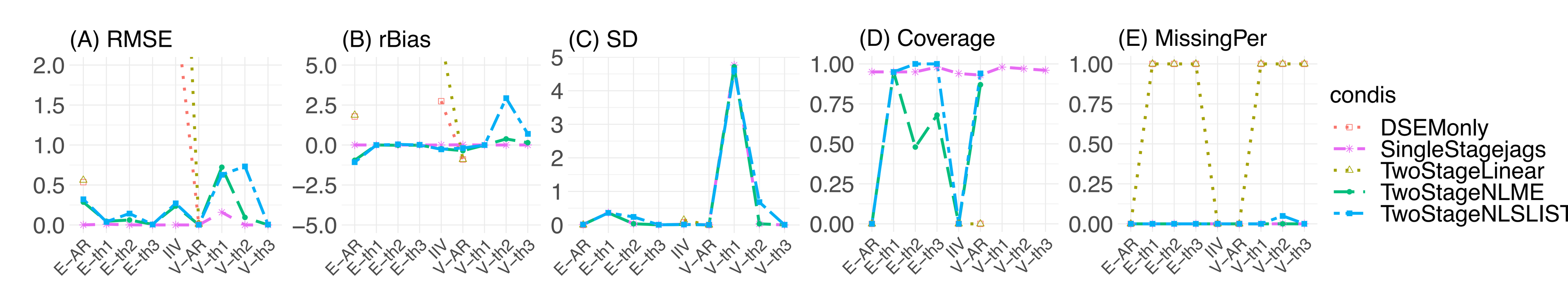
- When there is no trend in data (condition = NotrendDSEM), the DSEM model have acceptable performances even when nT=5.
- When there is trend in data (condition = DSEMonly), the DSEM model have terrible performances, only except for when estimating VAR[ϕ] with nT=50 and nP=500.
- When nT increases, the performances of DSEM across all situations get better, only except for when estimating VAR[ϕ] with nT=50 and nP=500.
- When nP increases, the SD of DSEM estimates tend to be smaller across all situations, which suggest a more valid performance.

Q2 (a). Performances of different approaches of accounting for trends.

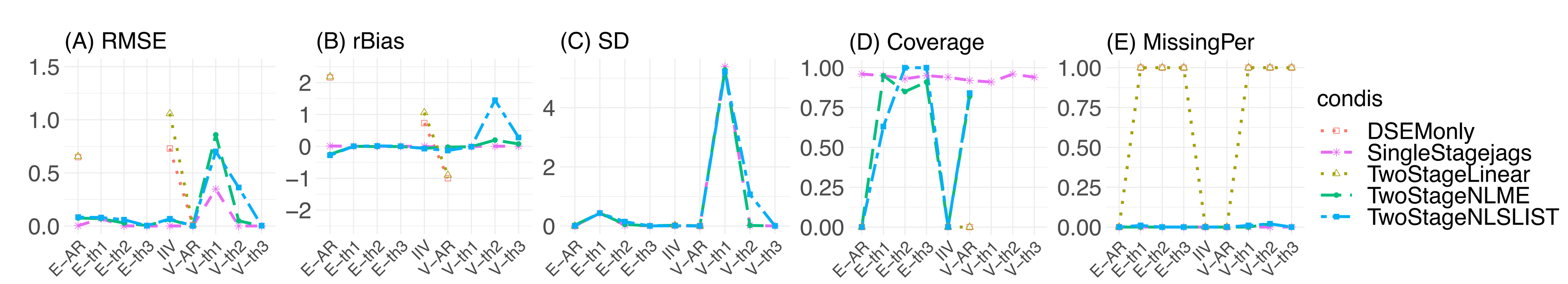
- **nT=5 and nP=500, all approaches.**
- DSEM & LINEAR are terrible.
- Accuracy (RMSE, rBias, and Coverage)
  Single-Stage >> NLME ≈ NLSLIST
- Stability (MissingPer and SD)
  Single-Stage ≈ NLME >> NLSLIST
- Overall
  Single-Stage >> NLME > NLSLIST



- **nT=15 and nP=500, all approaches.**
- DSEM & LINEAR are still terrible.
- Accuracy (RMSE, rBias, and Coverage)
  Single-Stage >> NLSLIST ≈ NLME
  NLME is better on V-(parameters)
  NLSLIST is better on Coverage
- Stability: three are similar.

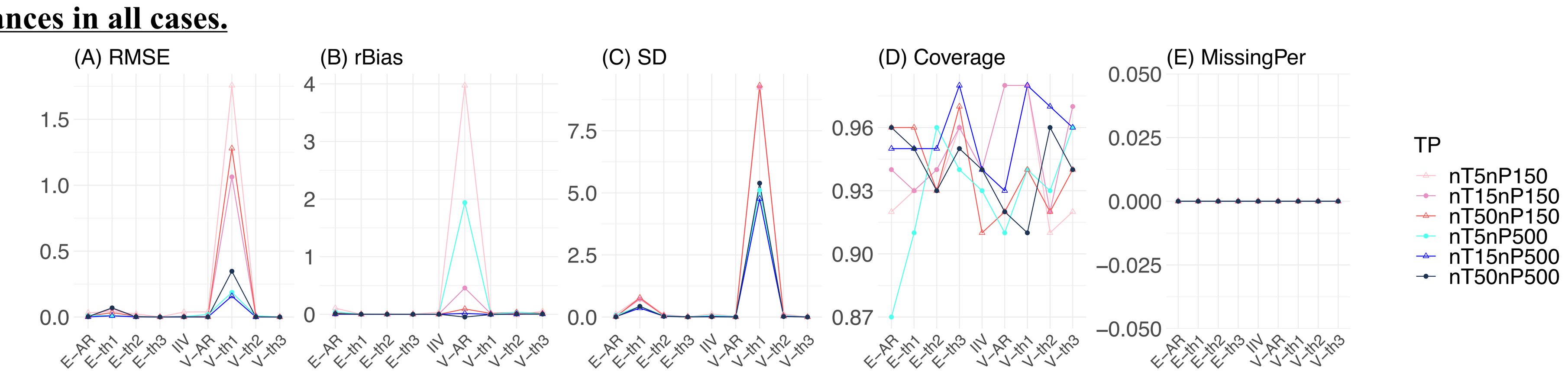


- **nT=50 and nP=500, all approaches.**
- DSEM & LINEAR are still terrible.
- Accuracy (RMSE, rBias, and Coverage)
  Single-Stage >> NLME ≈ NLSLIST
- Stability: three are similar.



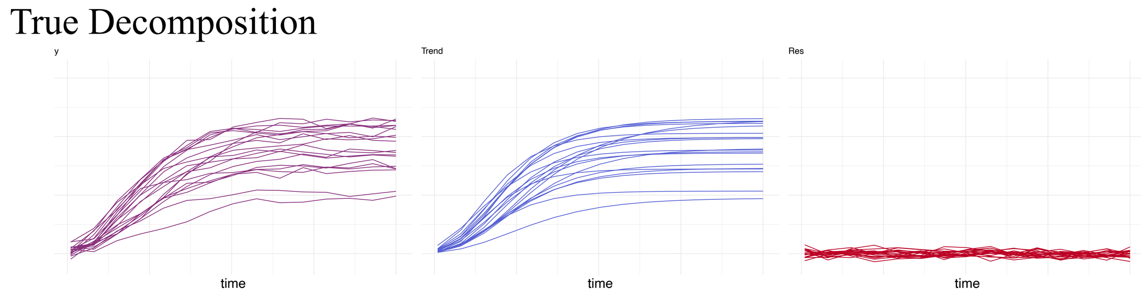
Q2 (b). Singe-Stage Approach Performances in all cases.

- Across all situations, Single-Stage Approach is accurate and stable on both trend parameters and residual parameters.
- The approach has biggest trouble when estimating E-AR (level-2 variance of AR coefficients, true value = 0.01).

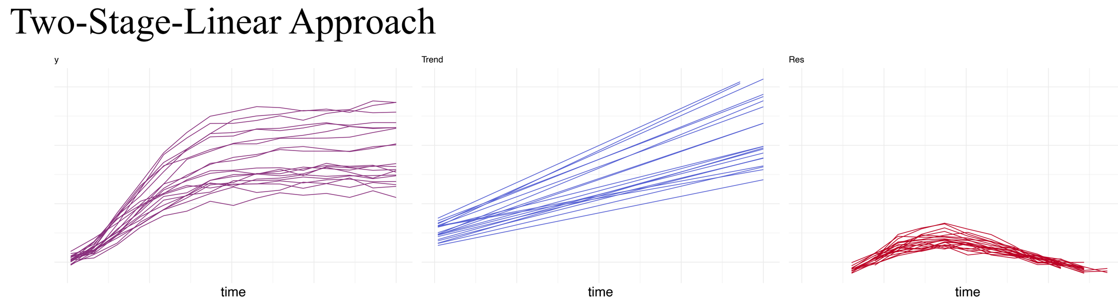


Que2 (c). Performances of detrending for two-stage methods.

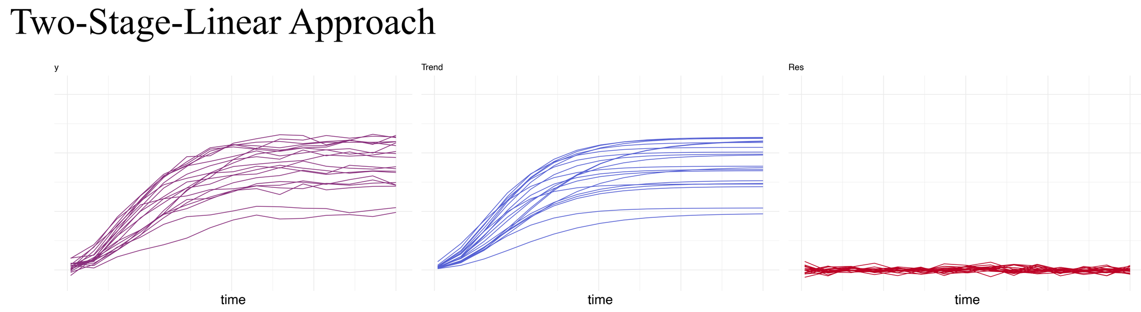
True Decomposition



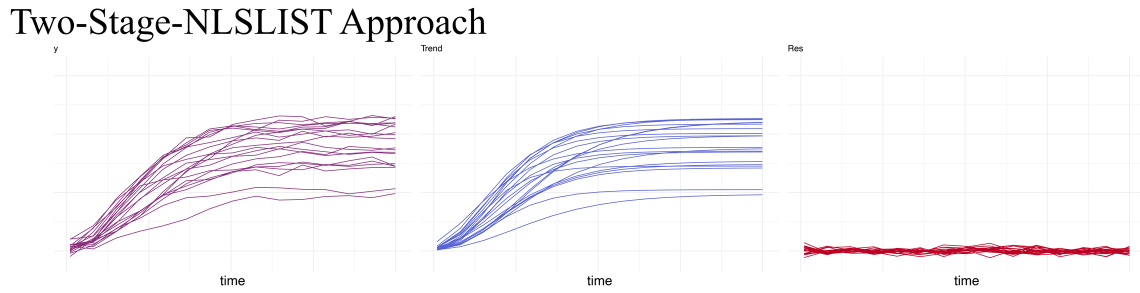
Two-Stage-Linear Approach



Two-Stage-Linear Approach



Two-Stage-NLSLIST Approach



Trends and residuals in Two-Stage Approaches (nT=15, nP=500). The detrending stage is based on whole 500 people. But only 20 people are sampled for clean plots.

Two-Stage-NLSLIST and Two-Stage-NLME can remove trend from data, although lose some information in residuals.

CONCLUSION

- Upon evaluating the performance of DSEM on 100 trended and 100 non-trended datasets, we discovered that the presence of trend in time series data could invalidate the estimates. Thus, **not accounting for trends is not recommended.**
- Subsequently, after comparing five different strategies to fit 100 multi-subject, short time series data replications, we concluded that **the single-stage approach exhibited the most efficient and stable performances, even when nT=5.** Two-stage approaches estimated trend parameters well if the curve type was correctly specified, but they failed to recover information of residuals.

FUTURE WORK

- The degree of autoregressive effects, the inter-person variation of autoregressive effects, the number of observed variables and their correlations are fixed in the current study and may be extended in the future.
- Current study only includes simulation experiments. The performances of five methods on substantive data need to be assessed.
- It will be worthwhile to explore appropriate measures to assess how "clean" the data are detrended.

ACKNOWLEDGEMENT

Funding for this study was provided by the NIH Intensive Longitudinal Health Behavior Cooperative Agreement Program U24AA027684, National Science Foundation grants BCS-1052736 and IGE-1806874, and the Pennsylvania State University Quantitative Social Sciences Initiative and UL TR000127 from the National Center for Advancing Translational Sciences.

REFERENCES

Browne, M. W. (1993). Structured latent curve models. In C. M. Cuadras & C. R. Rao (Eds.), *Multivariate Analysis: Future Directions 2* (pp. 171–197). North-Holland. https://doi.org/10.1016/B978-0-444-81531-6.50016-7

Chow, S.-M., & Zhang, G. (2013). Nonlinear Regime-Switching State-Space (RSSS) Models. *Psychometrika*, *78*(4), 740–768. https://doi.org/10.1007/s11336-013-9330-8

Craigmile, P. F. (2009). Detrending Response Time Series. In *Statistical methods for modeling human dynamics: An interdisciplinary dialogue.* (pp. 213–240).

The poster is presented in the 2023 Modern Modeling Methods conference at UCONN.
For any comments and suggestions, please feel free to contact xiaoyue_xiong@psu.edu