A simple method to estimate and test the correlation between momentary assessments in the presence of subject-level heterogeneity

Society for Ambulatory Assessment
2025 Annual Conference
Leuven, Belgium
May 26-28, 2025

Wolfgang Viechtbauer
Maastricht University
2025-05-27

Introduction

- **standard ESM/EMA analysis:** examine how two variables measured repeatedly relate to each other (within subjects)
- **examples:** stress ↔ affect, PA ↔ NA, affect ↔ symptoms, passive sensor data ↔ some self-report variable
- **random intercepts and slopes model:**

$$y_{ij} = \beta_{0i} + \beta_{1i} x_{ij} + \varepsilon_{ij}$$

where $\begin{bmatrix}\beta_{0i} \\ \beta_{1i}\end{bmatrix} \sim \mathrm{N}\left(\begin{bmatrix}\beta_0 \\ \beta_1\end{bmatrix}, \begin{bmatrix}\tau_0^2 & \tau_{01} \\ \tau_{01} & \tau_1^2\end{bmatrix}\right)$ and $\varepsilon_{ij} \sim \mathrm{N}(0, \sigma^2)$

- some extensions:
  - add an AR(1) structure on ε_{ij}
  - use within-person mean centering $(x_{ij} - \bar{x}_i)$

Example

- say we have data from **150 subjects** that were prompted to report on their mood 10 times per day for 6 days
- will use the data from **6921 beeps** that were responded to
- examine the association between **positive and negative affect** (average of several mood items measured on a 1-7 scale)

subjno	beep	pa	na
1	1	4.3	1.1
1	2	3.8	2.4
1	3	5.4	1.4
1	4	4.5	2.1
1	5	4.2	1.7
...	...	...	...

Example

- **problem:** which variable should be x and which should be y?
- **common approach:** let's try both!
- predict positive affect from negative affect:

	Value	Std.Error	t-value	p-value
(Intercept)	6.55	0.10	67.89	0.00
na	-0.88	0.05	-19.36	0.00

- predict negative affect from positive affect:

	Value	Std.Error	t-value	p-value
(Intercept)	3.18	0.07	44.40	0.00
pa	-0.27	0.01	-19.82	0.00

- AR(1) + within-person mean centering yields very similar results

Multivariate Data

- the coefficients differ considerably!
- sometimes the p-values can even be contradictory
- but both variables were simply observed
- so maybe we should treat them as such
- restructure the data into 'extra long' format

subjno	beep	outcome	value
1	1	pa	4.3
1	1	na	1.1
1	2	pa	3.8
1	2	na	2.4
1	3	pa	5.4
1	3	na	1.4
...	...	...	...

Multivariate Model

- multivariate model:

$$\begin{bmatrix} y_{ij1} \\ y_{ij2} \end{bmatrix} = \begin{bmatrix} \mu_{i1} \\ \mu_{i2} \end{bmatrix} + \begin{bmatrix} \varepsilon_{ij1} \\ \varepsilon_{ij2} \end{bmatrix}$$

where

$$\begin{bmatrix} \mu_{i1} \\ \mu_{i2} \end{bmatrix} \sim \mathrm{N}\left(\begin{bmatrix} \mu_1 \\ \mu_2 \end{bmatrix}, \begin{bmatrix} \tau_1^2 & \rho_b \tau_1 \tau_2 \\ \rho_b \tau_1 \tau_2 & \tau_2^2 \end{bmatrix}\right)$$

$$\begin{bmatrix} \varepsilon_{ij1} \\ \varepsilon_{ij2} \end{bmatrix} \sim \mathrm{N}\left(\begin{bmatrix} 0 \\ 0 \end{bmatrix}, \begin{bmatrix} \sigma_1^2 & \rho_w \sigma_1 \sigma_2 \\ \rho_w \sigma_1 \sigma_2 & \sigma_2^2 \end{bmatrix}\right)$$

- ρ_b denotes the **between-subject correlation**
- ρ_w denotes the **within-subject correlation** (i.e., the correlation between the two outcomes at the same beep)
- this is what we are interested in

Example

```
## Linear mixed-effects model fit by REML
##
## Random effects:
##          StdDev    Corr
## y_na     0.2929783 y_na
## y_pa     0.7190153 -0.422
##
## Correlation Structure: General
##   1
## 2 -0.485
##
## Variance function:
##  Structure: Different standard deviations per stratum
##        pa        na
## 0.8929284 0.5001988
##
## Approximate 95% confidence intervals
##  Correlation structure:
##               lower       est.      upper
## cor(1,2) -0.5031847 -0.4851379 -0.4666681
```


Multivariate Model

- **problem:** the model assumes that the within-subject correlation ρ_w is constant across subjects
- not easy/possible to add a random effect for ρ_w to account for subject-level heterogeneity in mixed-effects model software
- can do this in a Bayesian framework (e.g., with JAGS or Stan)[1], in SAS with `proc nlmixed`, and in Mplus with some trickery[2]

[1] but model fitting can take hours

[2] have to assume that either all within-subject correlations are positive or negative

Easier Solution

- we know from standard regression models:
  - let $z()$ denote the z-transformation ('standardization')
  - then the following two models yield identical results:
    - $z(y_i) = \beta_0 + \beta_1 z(x_i) + \varepsilon_i$
    - $z(x_i) = \beta_0 + \beta_1 z(y_i) + \varepsilon_i$
  - $\beta_0 = 0$ by definition
  - $\beta_1 = r(x, y)$ (i.e., the correlation between x and y)
- apply the principle to the random intercepts/slopes model:

$$z_i(y_{ij}) = \beta_{1i} z_i(x_{ij}) + \varepsilon_{ij}$$

  where $\beta_{1i} \sim \mathrm{N}(\beta_1, \tau^2)$ and $\varepsilon_{ij} \sim \mathrm{N}(0, \sigma^2)$
- β_1 denotes the **average correlation** and τ^2 the **variance**
- note: $z_i()$ denotes within-person standardization

Example

```
## Linear mixed-effects model fit by REML
##
## Random effects:
##              z_na  Residual
## StdDev: 0.2780999 0.8202982
##
## Fixed effects:  z_pa ~ 0 + z_na
##          Value Std.Error   DF   t-value p-value
## z_na -0.4884274 0.02487607 6771 -19.63443       0
##
## Approximate 95% confidence intervals
##
##  Fixed effects:
##           lower       est.      upper
## z_na -0.5371923 -0.4884274 -0.4396625
```

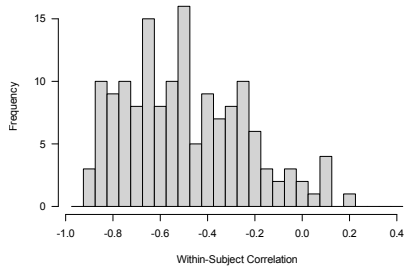

Example

```
## Linear mixed-effects model fit by REML
##
## Random effects:
##              z_na  Residual
## StdDev: 0.2780999 0.8202982
##
## Fixed effects:  z_pa ~ 0 + z_na
##          Value Std.Error   DF   t-value p-value
## z_na -0.4884274 0.02487607 6771 -19.63443       0
##
## Approximate 95% confidence intervals
##
##  Fixed effects:
##           lower       est.      upper
## z_na -0.5371923 -0.4884274 -0.4396625
```

- note: data were simulated from a bivariate model with $\rho_i \sim \mathrm{N}(-0.5, 0.3^2)$

Results

- multivariate model: $\rho_w = -0.49$ (95% CI: -0.50 to -0.47)
- simple method: $\rho_w = -0.49$ (95% CI: -0.54 to -0.44)
- Bayesian model: $\rho_w = -0.49$ (95% CI: -0.54 to -0.45)
- distribution of the BLUPs for β_{1i}:



Further Applications / Extensions

- moderators of the association
- network analysis
- dyadic data analysis

Moderator Analysis

- 324 subjects in three groups (healthy controls, depressed, psychotic) did 6 days of ESM with 10 beeps per day
- want to exmine the association between:
  - **positive affect:** cheerful, relaxed, content/satisfied
  - **negative affect:** irritated, anxious, down, guilty, insecure, lonely
- all items measured on a 1-7 scale

```
## Random effects:
##              z_na  Residual
## StdDev: 0.1801105 0.7990651
##
## Fixed effects:  z_pa ~ 0 + z_na:status
##                                Value  Std.Error    DF   t-value p-value
## z_na:statuscontrol    -0.4996309 0.02074291 14169 -24.08683       0
## z_na:statusdepressed -0.6985262 0.02070565 14169 -33.73602       0
## z_na:statuspsychotic -0.4308949 0.02153528 14169 -20.00879       0
```


Notes / Discussion

- excludes subjects with no variation in responses
- correlations are bounded $(-1, 1)$, but the model assumes $\rho_i \sim \mathrm{N}(\rho, \tau^2)$, which is questionable when $|\rho|$ is close to 1
- overall a simple and easy method for estimating and testing the correlation between two momentary assessments

Thank You for Your Attention!

Questions, Comments, Suggestions?

wolfgang.viechtbauer@maastrichtuniversity.nl

https://www.wvbauer.com

@wviechtb @wviechtb