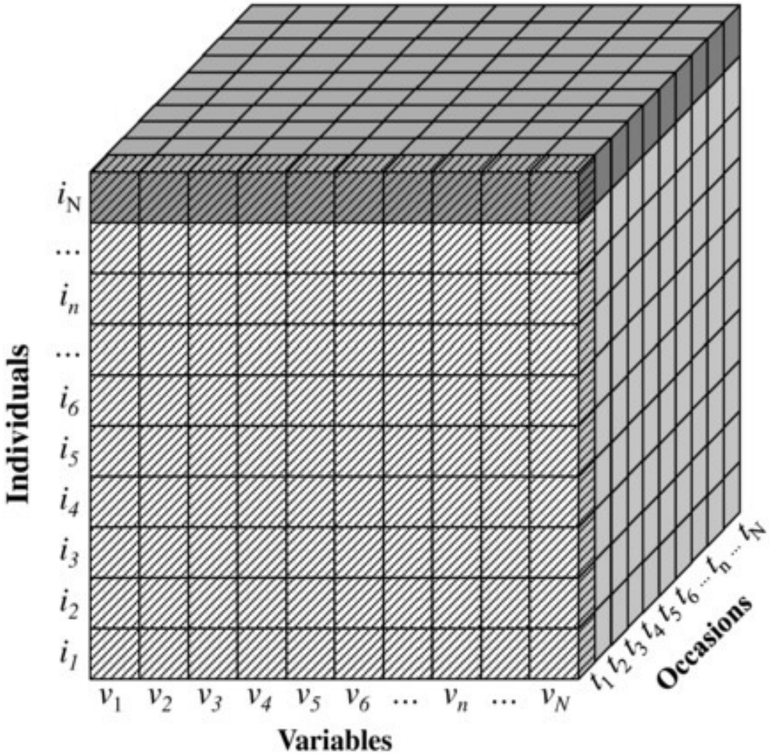
Network modeling of time-series data

Ria Hoekstra

Network Analysis 2023

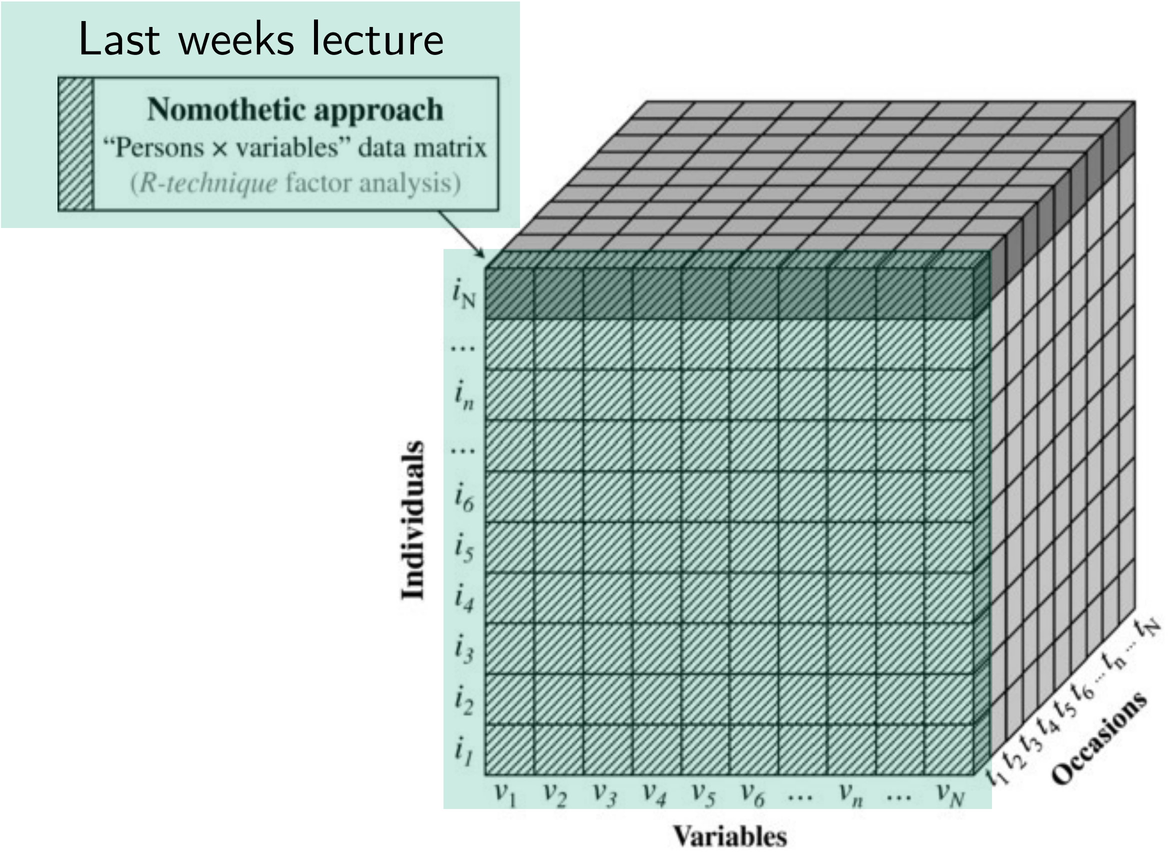
University of Amsterdam

Cross-sectional versus time-series data



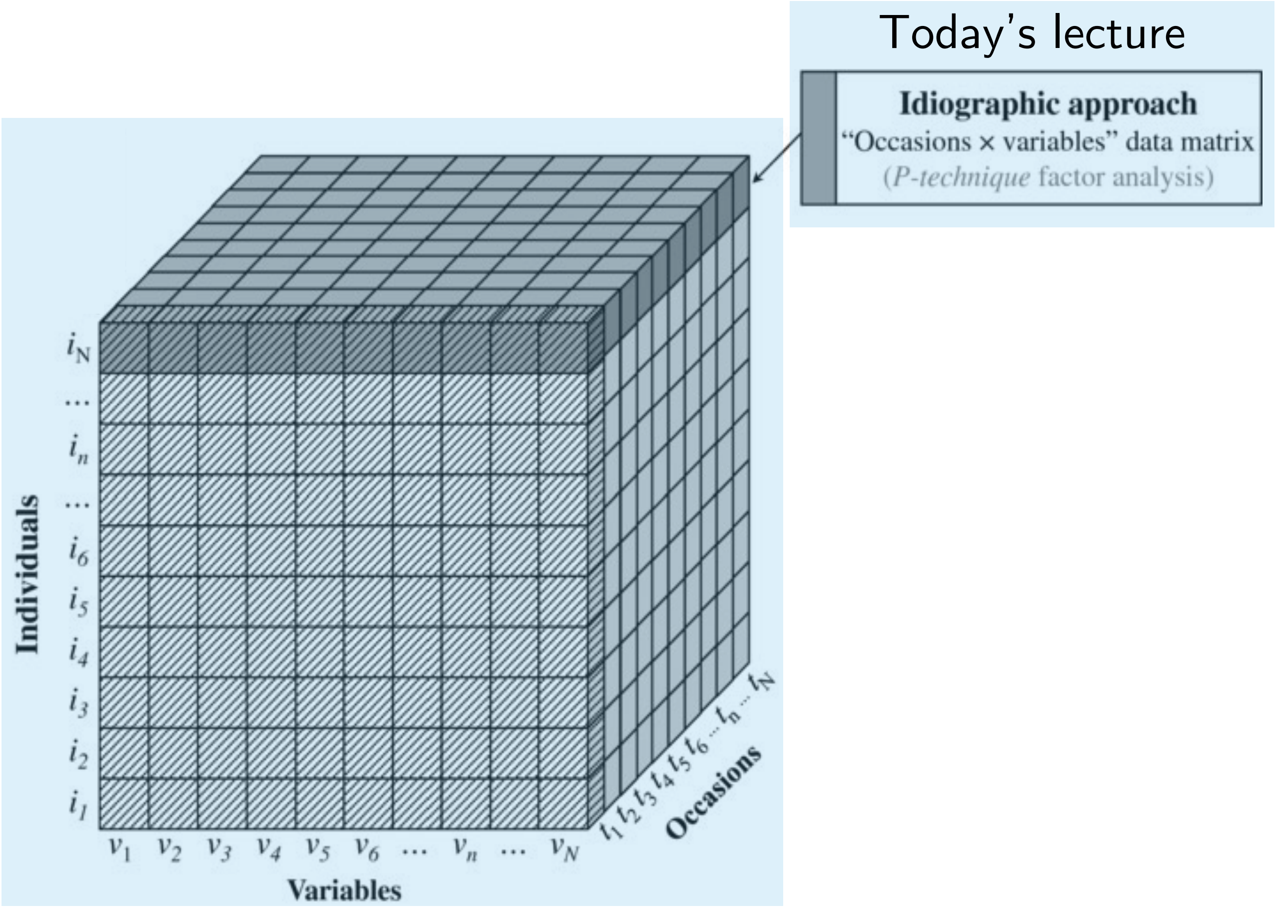
Cattell's databox in Barbot, B., & Perchec, C. (2015). New Directions for the Study of Within-Individual Variability in Development: The Power of "N= 1". *New Directions for Child and Adolescent Development*, *2015*(147), 57-67.

Cross-sectional versus time-series data



Cattell's databox in Barbot, B., & Perchec, C. (2015). New Directions for the Study of Within-Individual Variability in Development: The Power of "N= 1". *New Directions for Child and Adolescent Development*, *2015*(147), 57-67.

Cross-sectional versus time-series data



Cattell's databox in Barbot, B., & Perchec, C. (2015). New Directions for the Study of Within-Individual Variability in Development: The Power of "N= 1". *New Directions for Child and Adolescent Development*, *2015*(147), 57-67.

What is time-series data?

- One or multiple individuals are measured multiple times

What is time-series data?

- Different forms of time-series data:
  - Daily Diary (DD)
  - Experience Sampling Method (ESM)
  - Ecological Momentary Assessment (EMA)

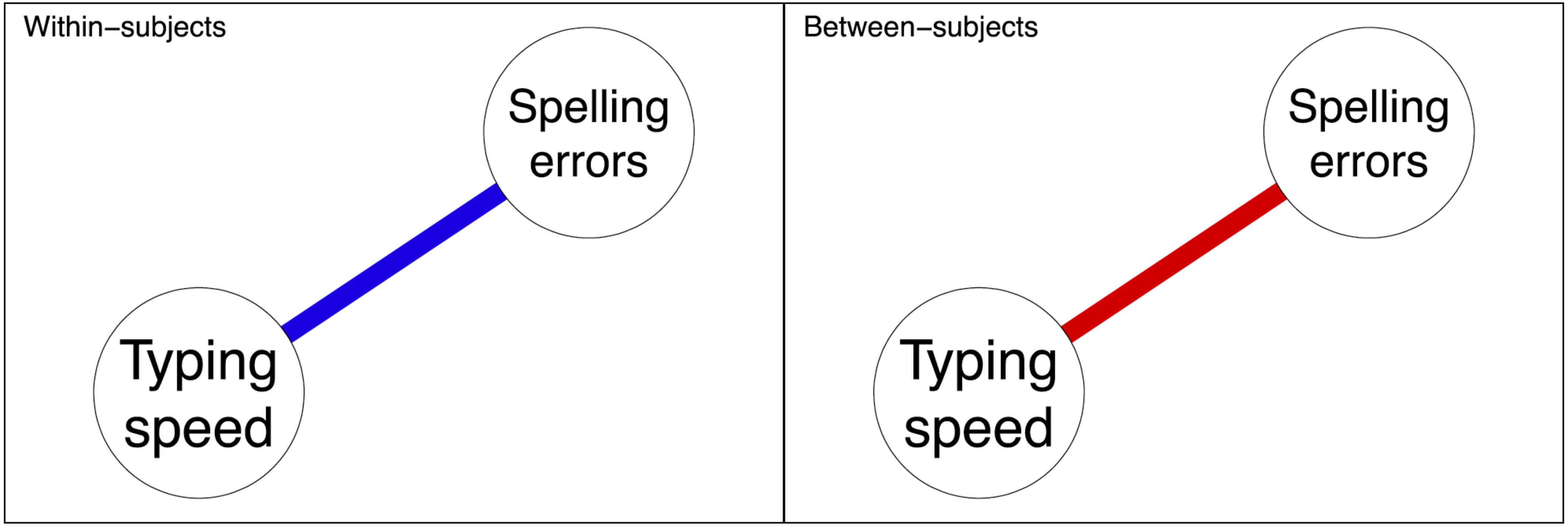
- Cases are not independent
  - There is a natural temporal ordering in the data
    - Knowing someone's mood at one time point helps predict their mood at a next time point
    - Observations closer together in time will be more closely related than observations further apart in time

Why time-series modeling?

- Gives insight into the **dynamic** relations between variables
  - Granger causality
  - Can be part of eHealth

- Difficult to translate results from cross-sectional analysis to individual
  - Results are only identical when a system is **ergodic**:
    - Stationarity
    - Homogenous
    - i.e., there are no between-person differences
  - Simpsons paradox

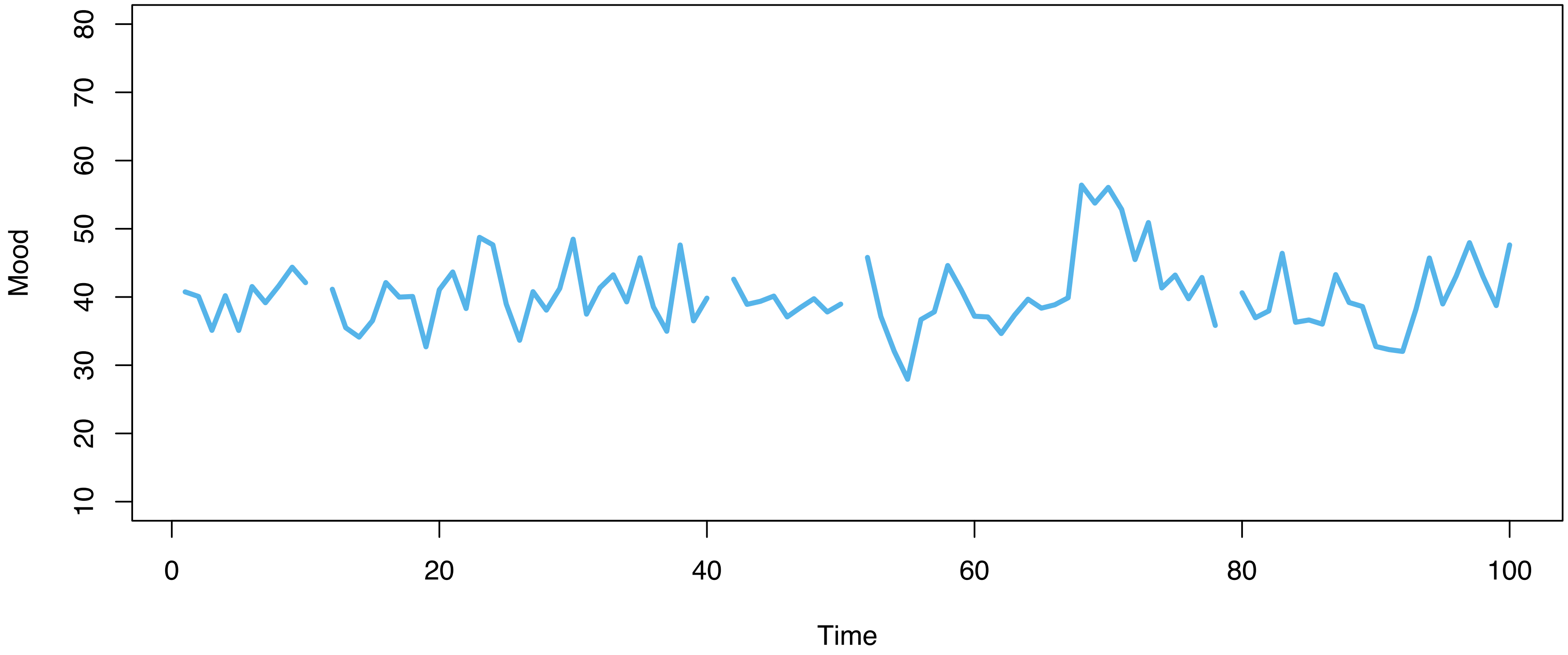
Simpsons Paradox



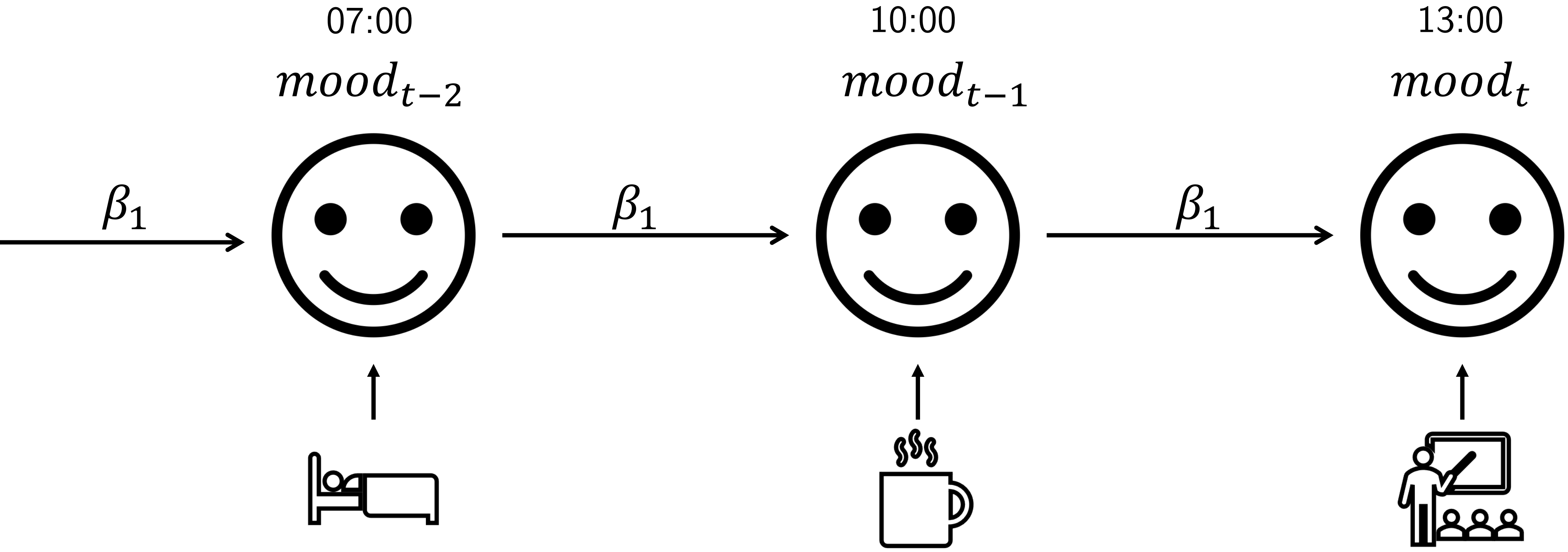
Example based on Hamaker, E. L. (2012). Why Researchers Should Think ‘Within-Person’: A Paradigmatic Rationale. Handbook of Research Methods for Studying Daily Life. The Guilford Press New York, NY, 43–61.

Modeling time-series data: $N = 1$ Networks

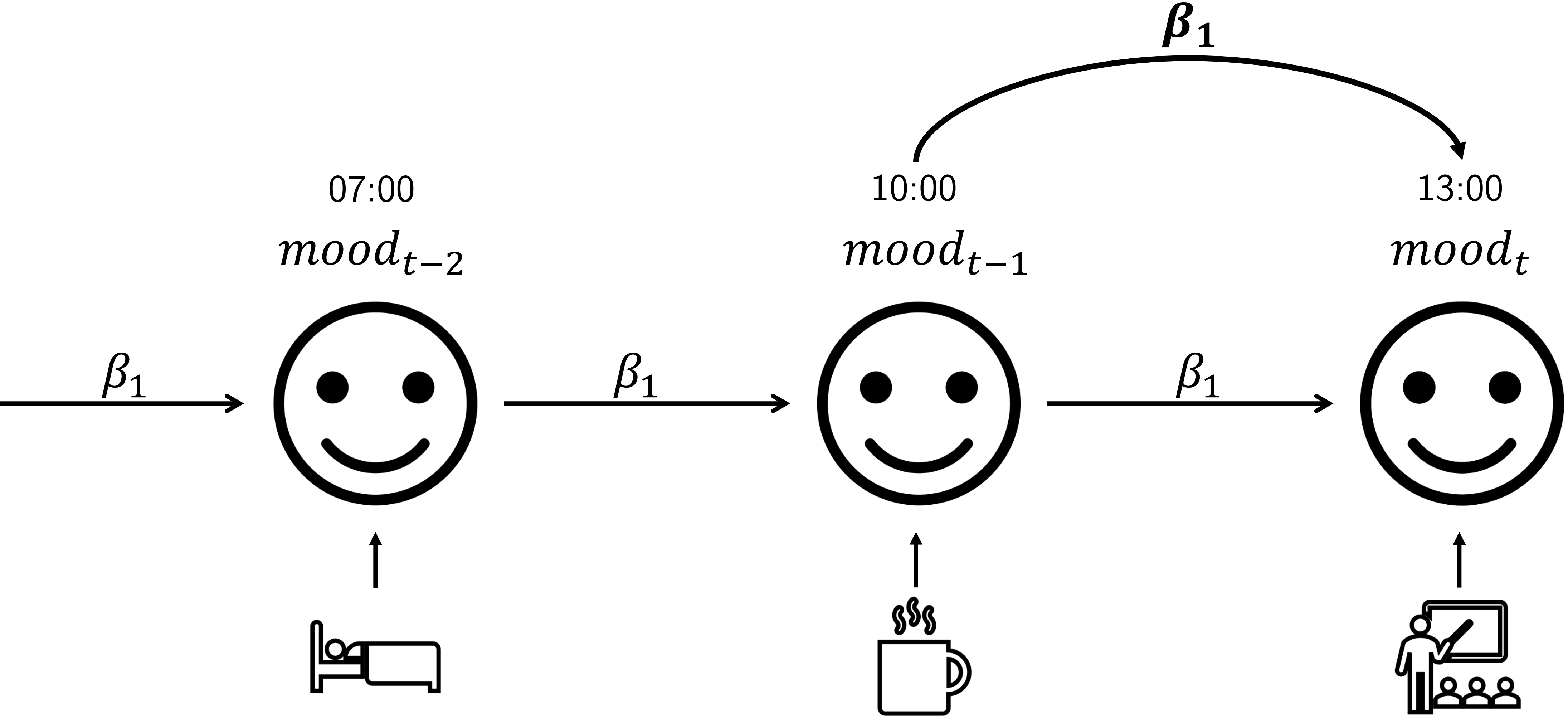
How to model **univariate** time-series?



Auto Regressive (AR) Model

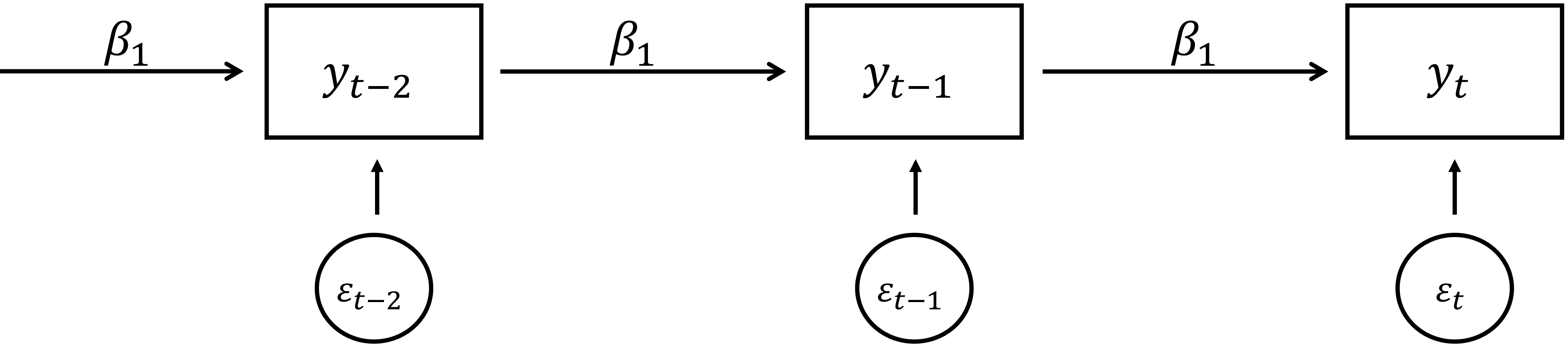


AR (1): Autoregressive **Lag 1** Model

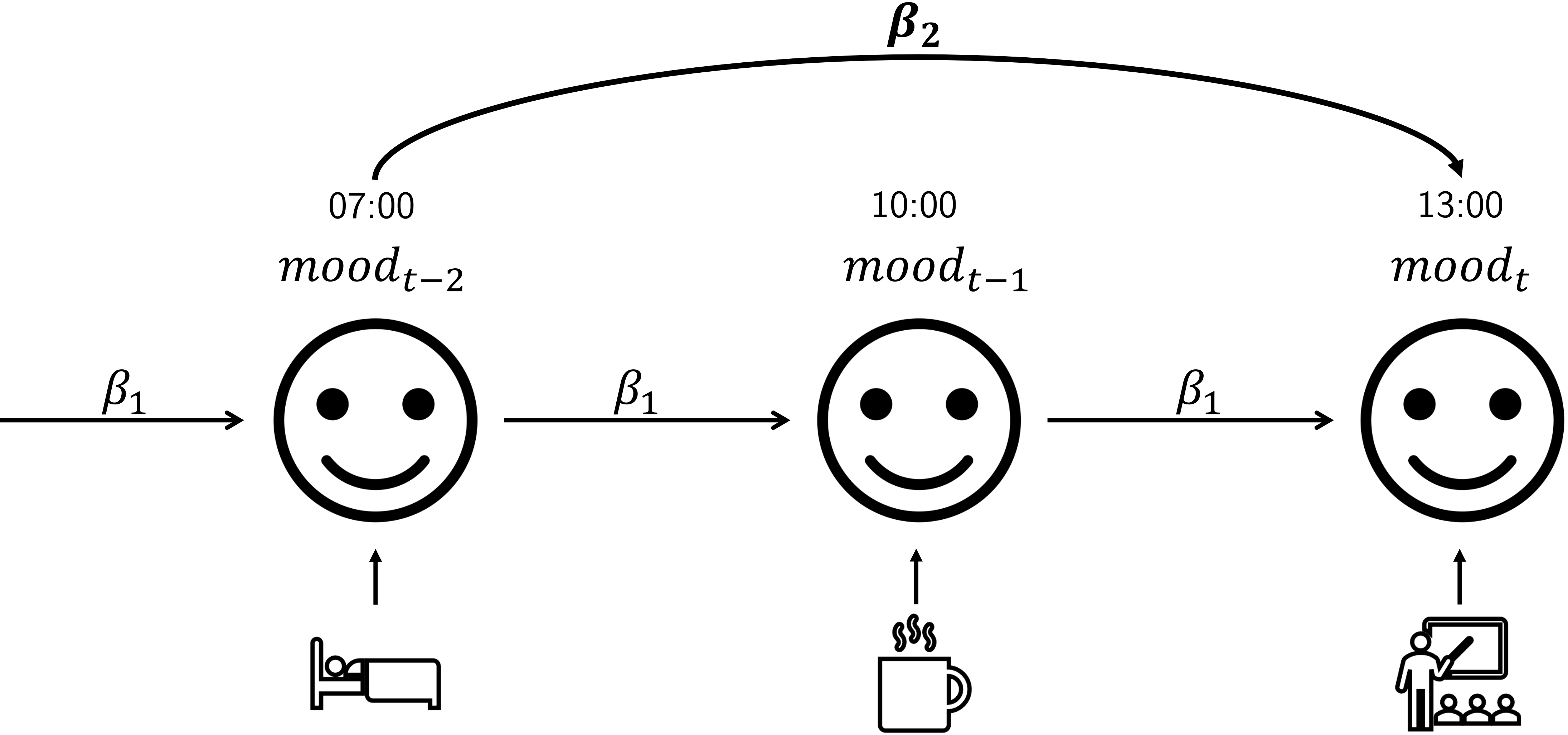


AR (1) Model:

$$y_t = \beta_1 y_{t-1} + \varepsilon_t$$

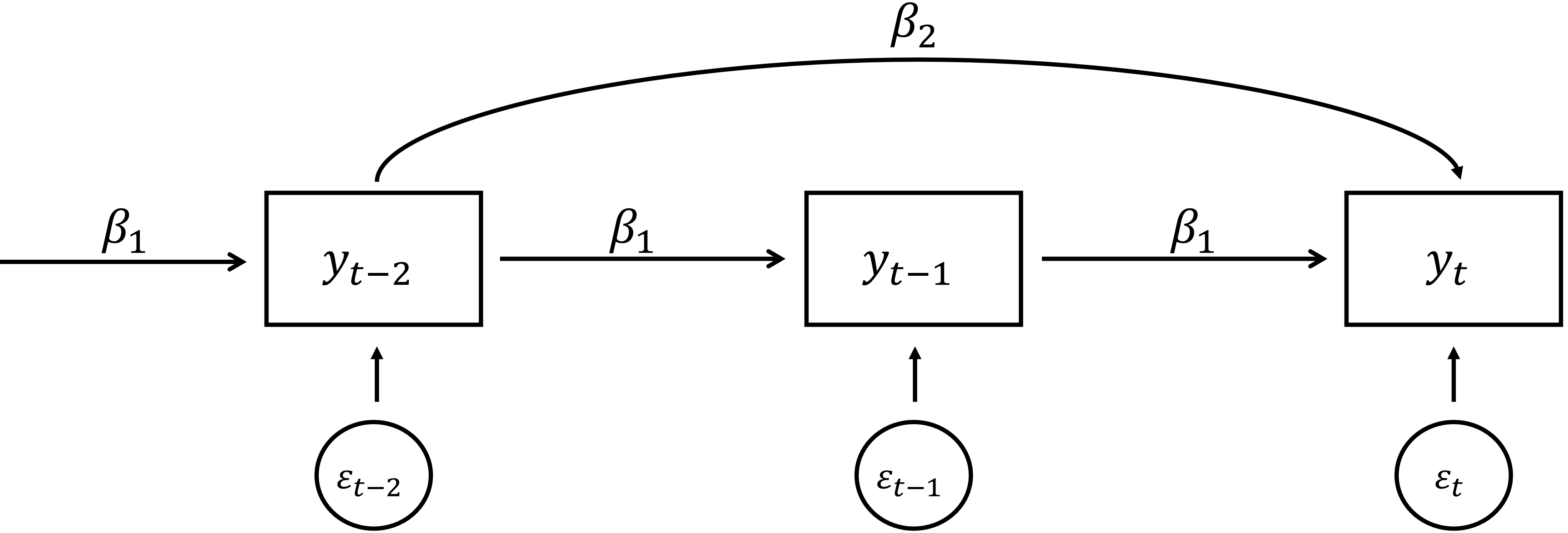


AR (2): Autoregressive **Lag 2** Model

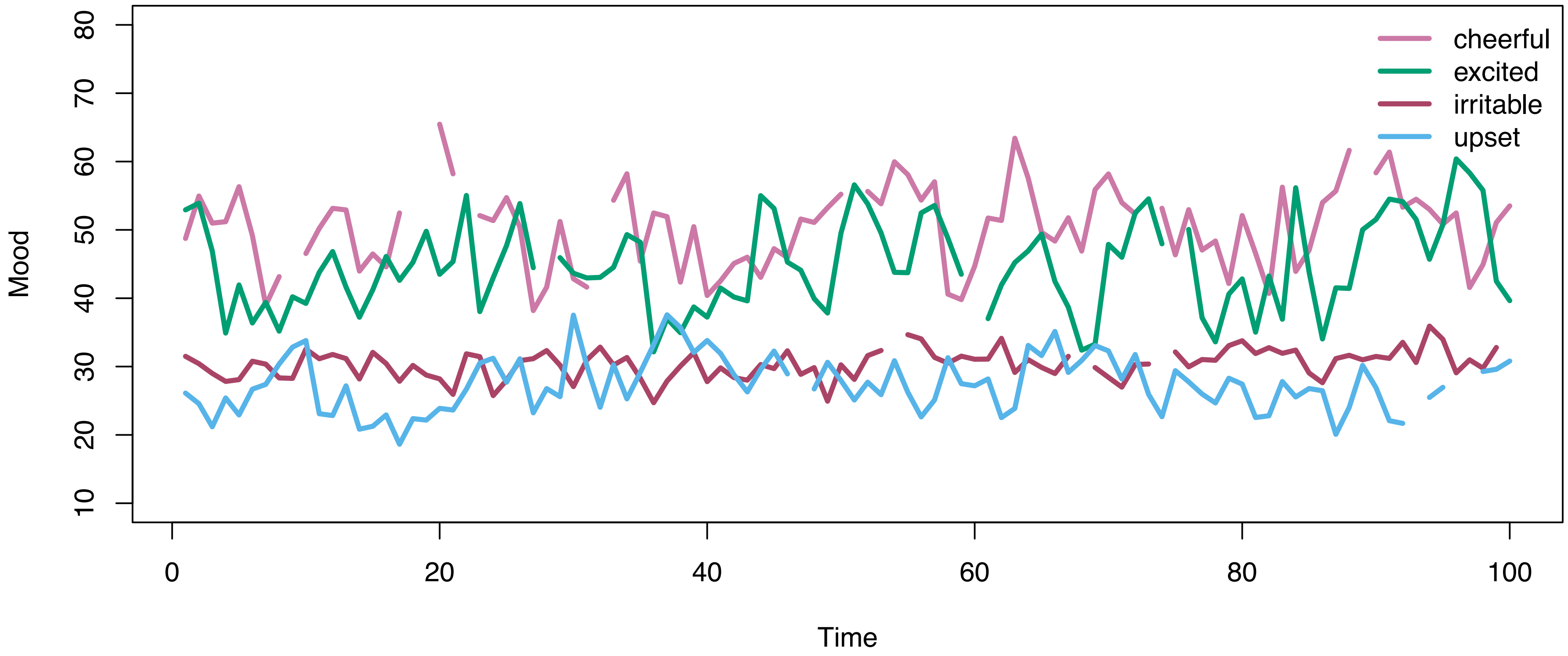


AR (2) model:

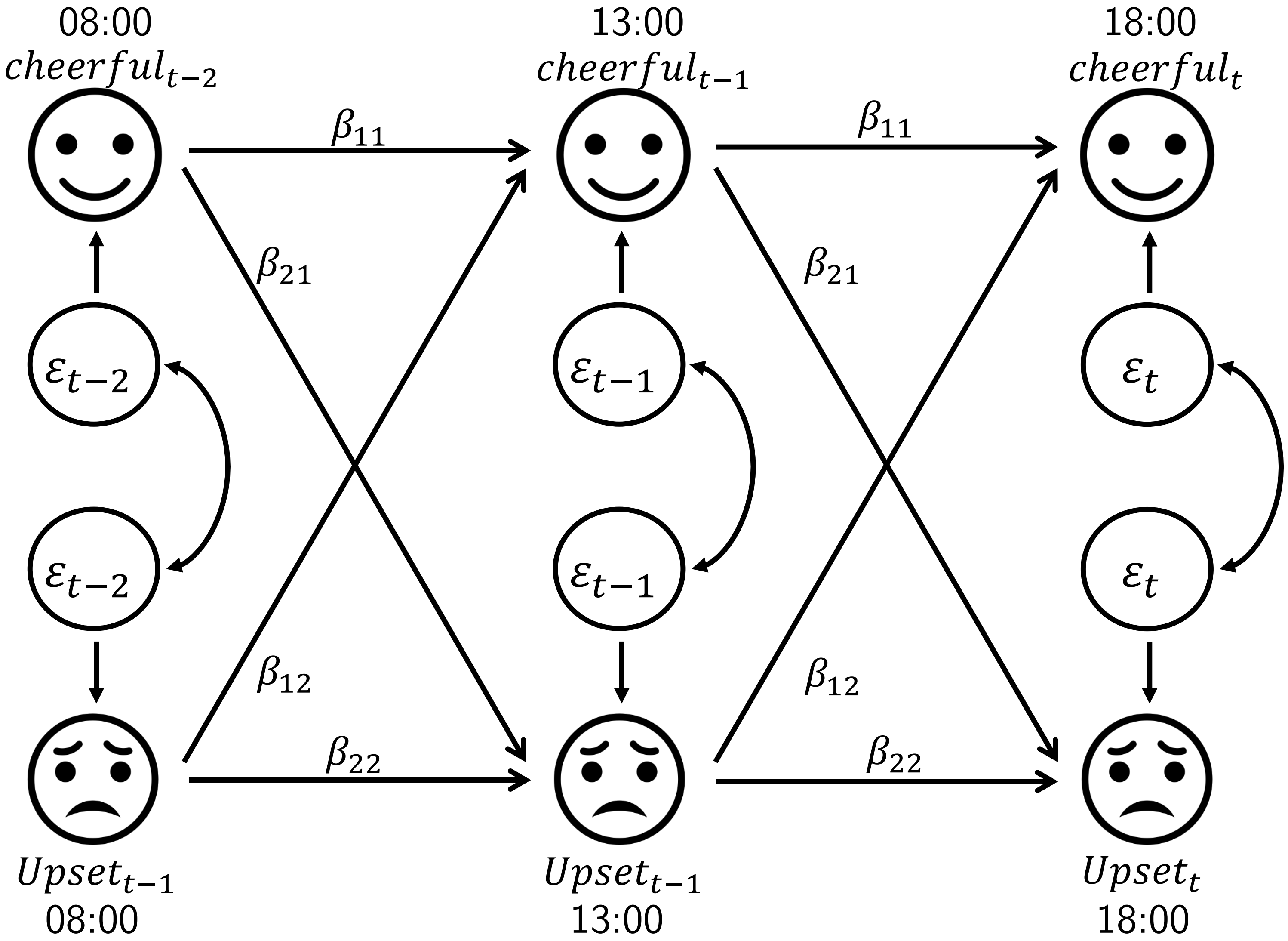
$$y_t = \beta_1 y_{t-1} + \beta_2 y_{t-2} + \varepsilon_t$$



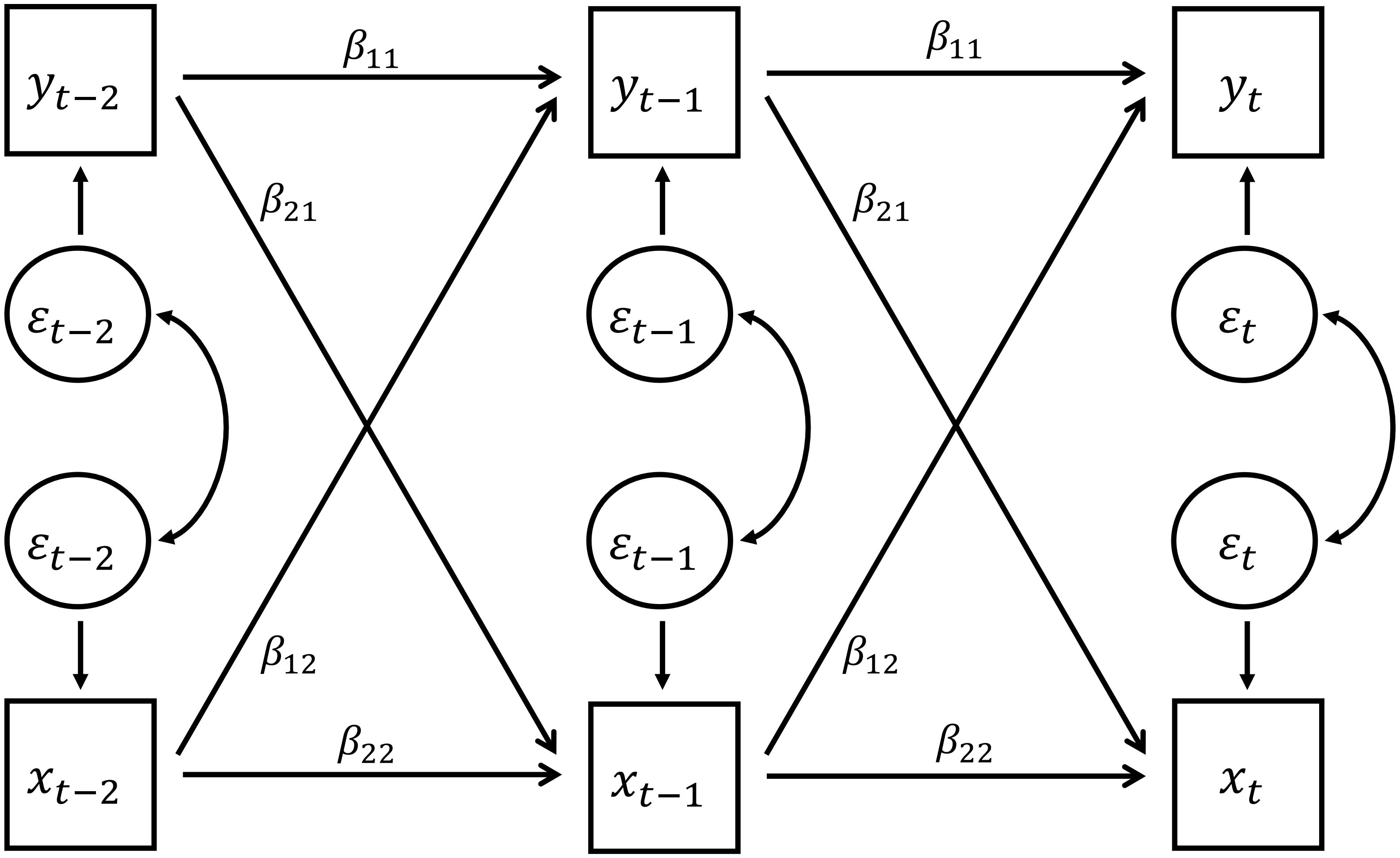
How to model **multivariate** time-series?



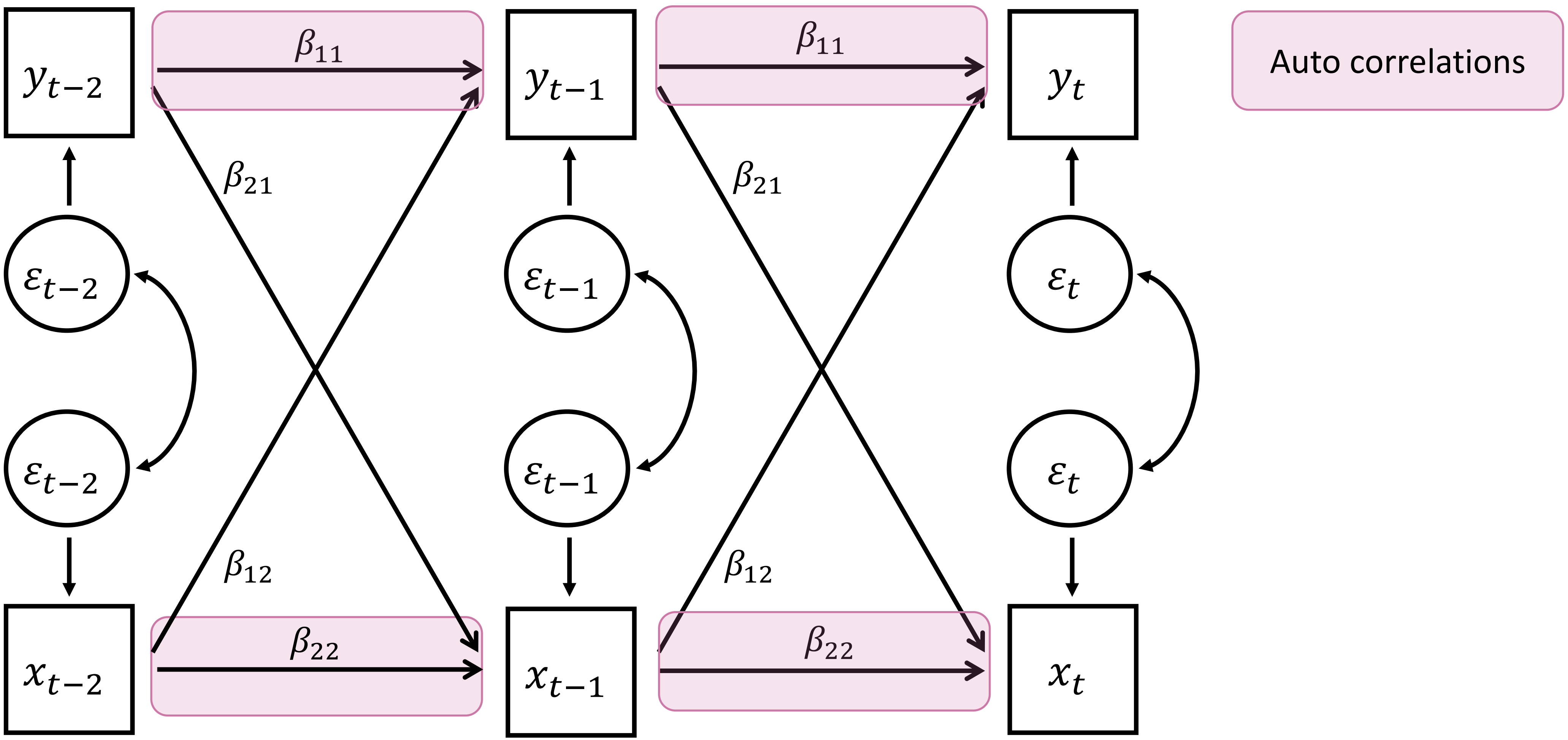
Vector Auto Regressive (VAR) model



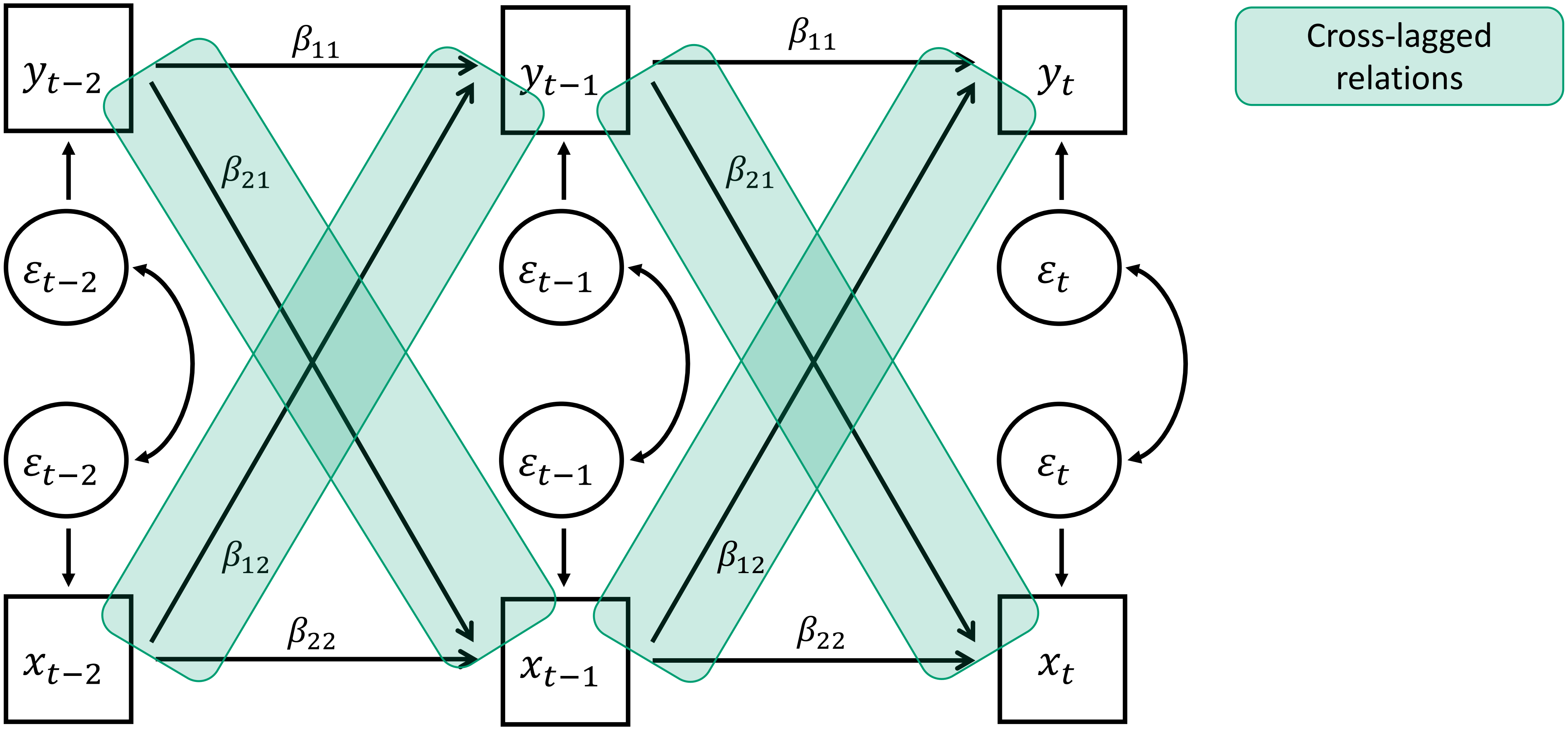
Vector Auto Regressive (VAR) model



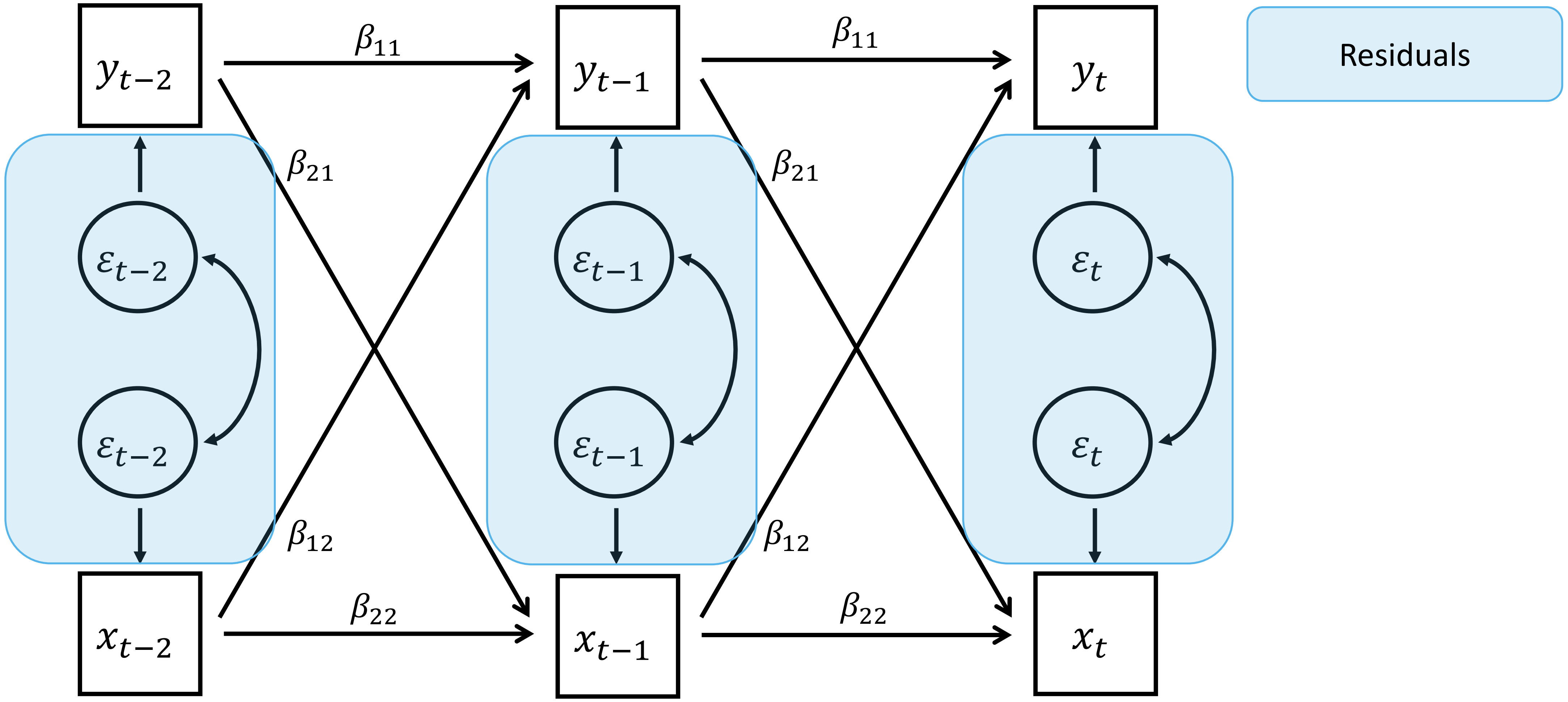
Vector Auto Regressive (VAR) model



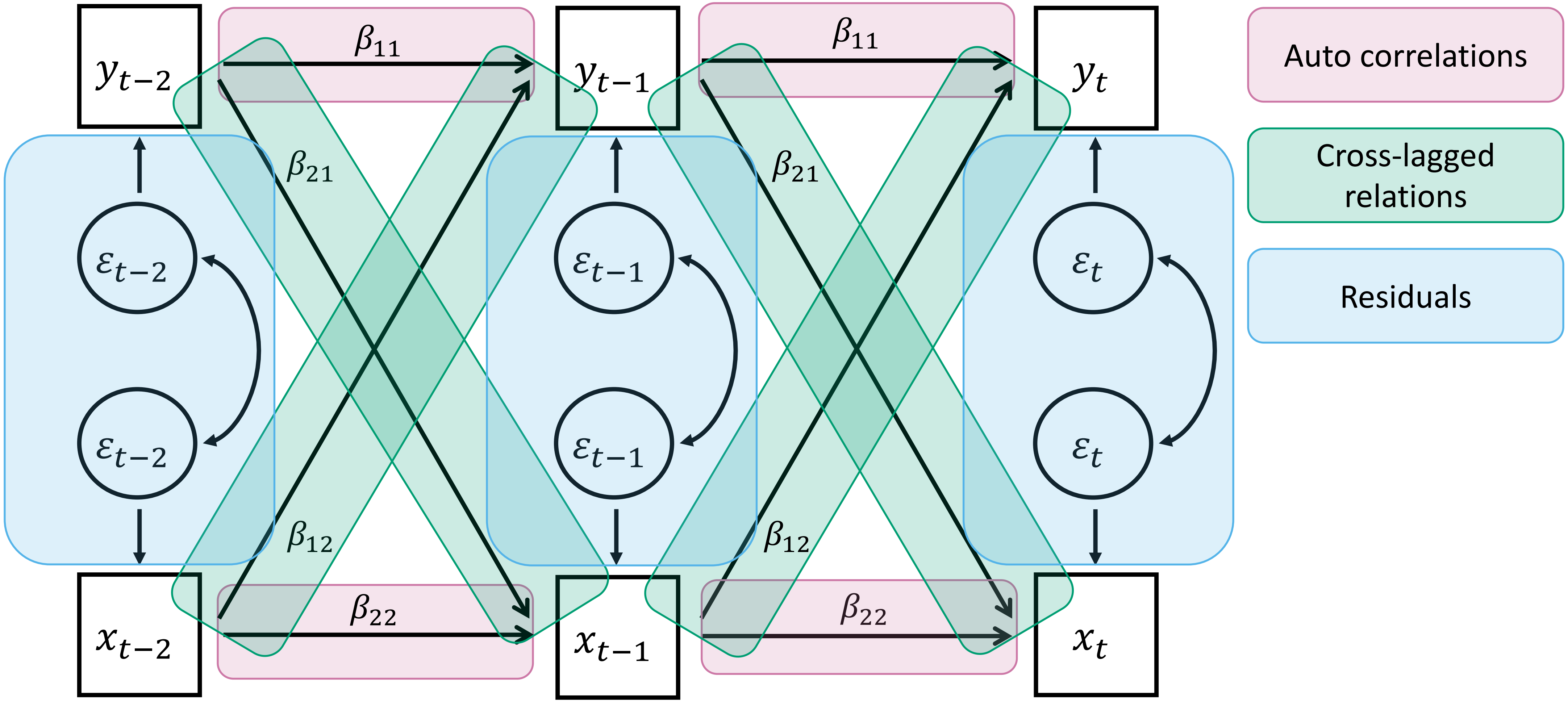
Vector Auto Regressive (VAR) model



Vector Auto Regressive (VAR) model



Vector Auto Regressive (VAR) model



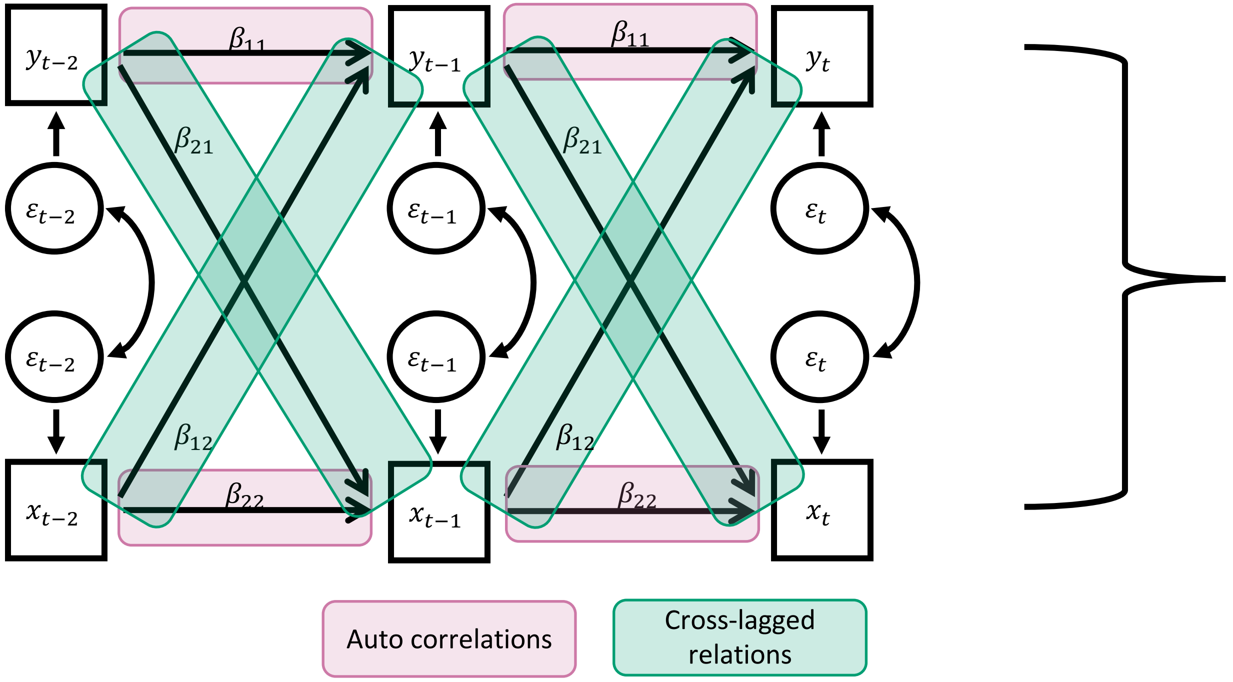
From VAR to **Graphical** VAR (GVAR)

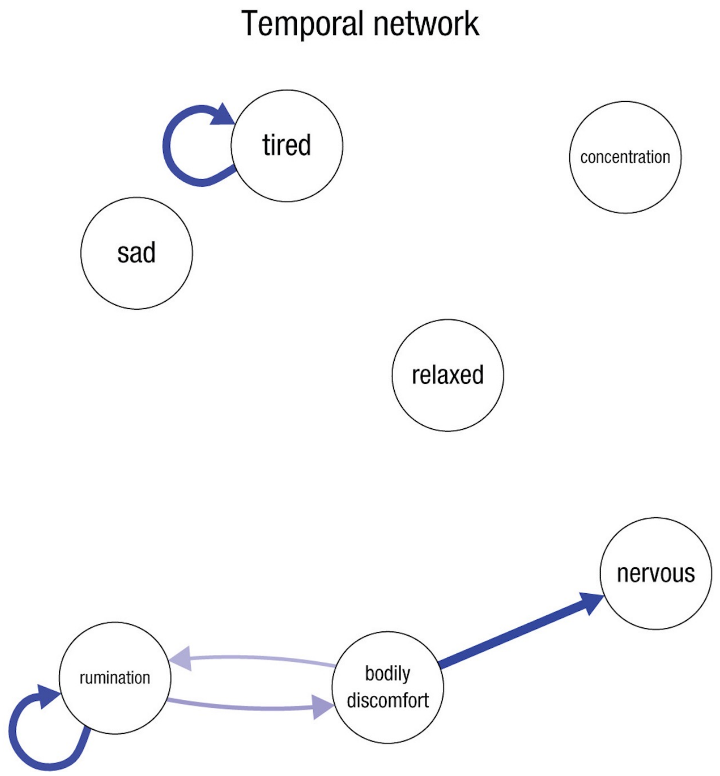
- **Temporal network:** shows that one variable predicts another variable in the next measurement occasion
  - Granger causality

	relaxed	sad	nervous	concentration	tired	rumination	bodily.discomfort	time
31	5	6	4	5	6	4	4	2014-05-01 10:15:00
32	3	5	4	5	6	4	5	2014-05-01 13:15:00
33	NA	NA	NA	NA	NA	NA	NA	2014-05-01 16:15:00
34	3	5	4	5	6	4	4	2014-05-01 19:15:00
35	4	6	4	4	7	5	4	2014-05-01 22:15:00
36	4	3	3	5	5	2	3	2014-05-02 10:15:00
37	4	3	4	5	5	3	2	2014-05-02 13:15:00

From VAR to **Graphical** VAR (GVAR)

- **Temporal network:** shows that one variable predicts another variable in the next measurement occasion





Epskamp, S., van Borkulo, C. D., van der Veen, D. C., Servaas, M. N., Isvoranu, A. M., Riese, H., & Cramer, A. O. (2018). Personalized network modeling in psychopathology: The importance of contemporaneous and temporal connections. Clinical Psychological Science, 6(3), 416-427.

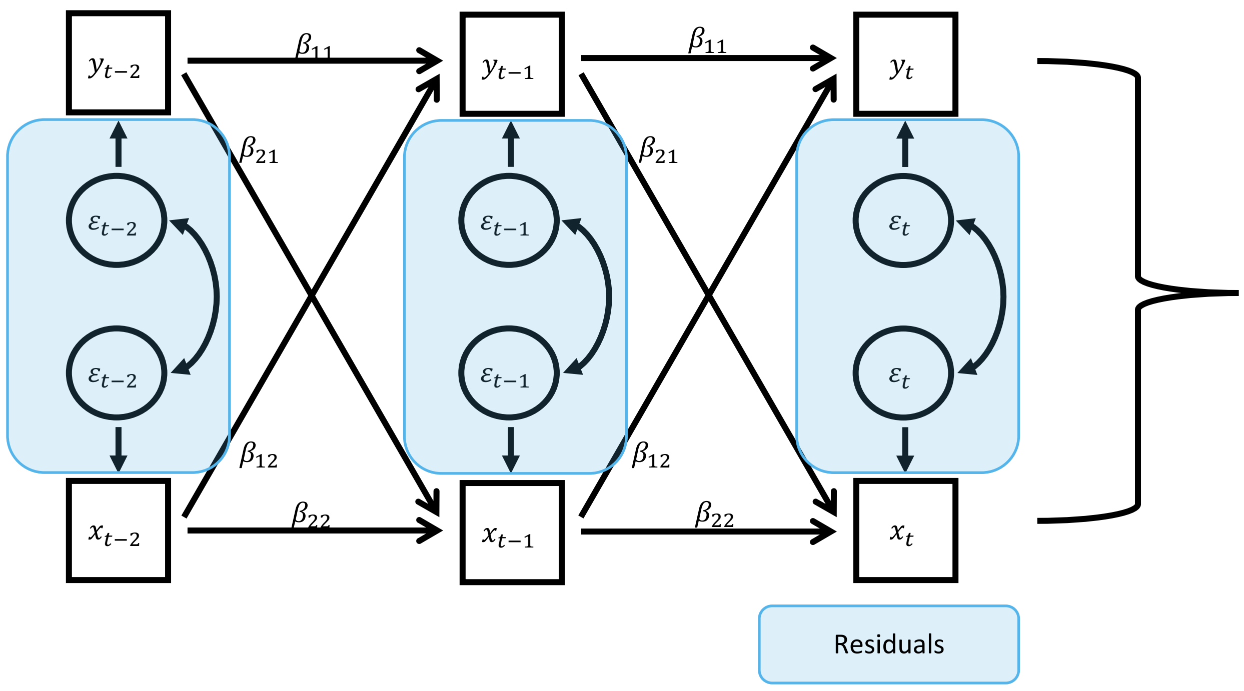
From VAR to **Graphical** VAR (GVAR)

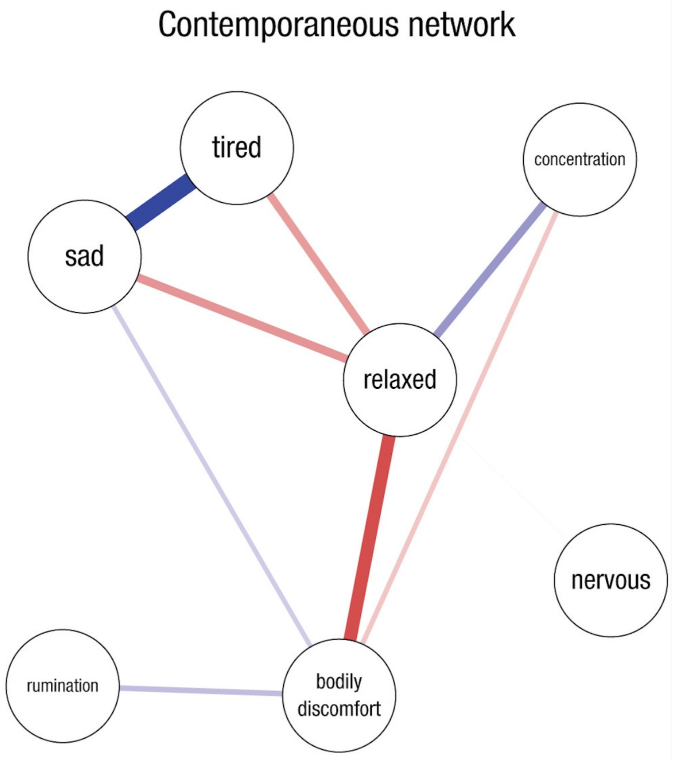
- **Contemporaneous network:** shows if two variables predict one another after taking temporal information into account
  - Contains *effects faster* than the time-window of measurement

	relaxed	sad	nervous	concentration	tired	rumination	bodily.discomfort	time
31	5	6	4	5	6	4	4	2014-05-01 10:15:00
32	3	5	4	5	6	4	5	2014-05-01 13:15:00
33	NA	NA	NA	NA	NA	NA	NA	2014-05-01 16:15:00
34	3	5	4	5	6	4	4	2014-05-01 19:15:00
35	4	6	4	4	7	5	4	2014-05-01 22:15:00
36	4	3	3	5	5	2	3	2014-05-02 10:15:00
37	4	3	4	5	5	3	2	2014-05-02 13:15:00

From VAR to **Graphical** VAR (GVAR)

- **Contemporaneous network:** shows if two variables predict one another after taking temporal information into account

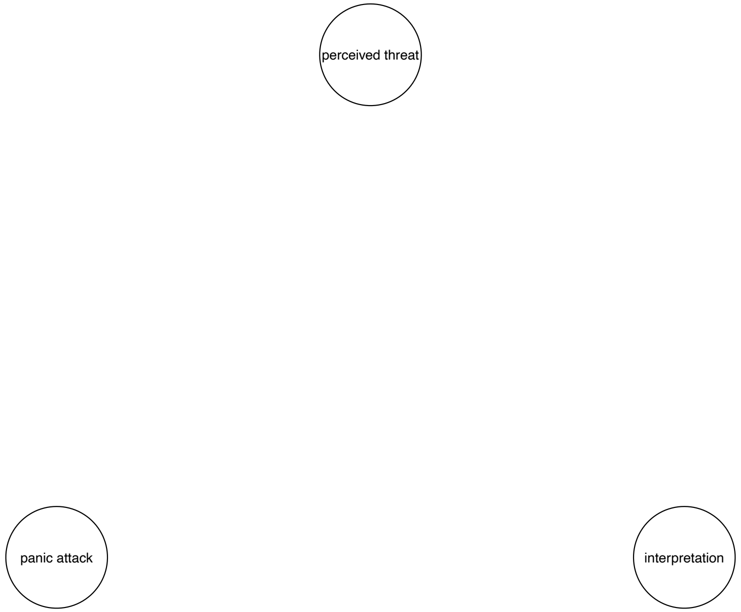




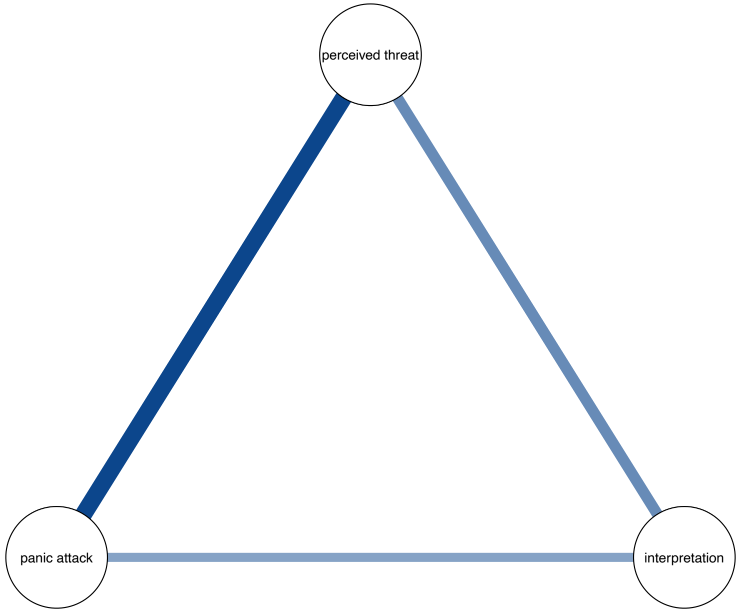
Epskamp, S., van Borkulo, C. D., van der Veen, D. C., Servaas, M. N., Isvoranu, A. M., Riese, H., & Cramer, A. O. (2018). Personalized network modeling in psychopathology: The importance of contemporaneous and temporal connections. Clinical Psychological Science, 6(3), 416-427.

From VAR to **Graphical** VAR (GVAR)

- **Contemporaneous network:** shows if two variables predict one another after taking temporal information into account
  - Contains *effects faster* than the time-window of measurement
  - Example: Dynamics of panic disorder



Temporal network



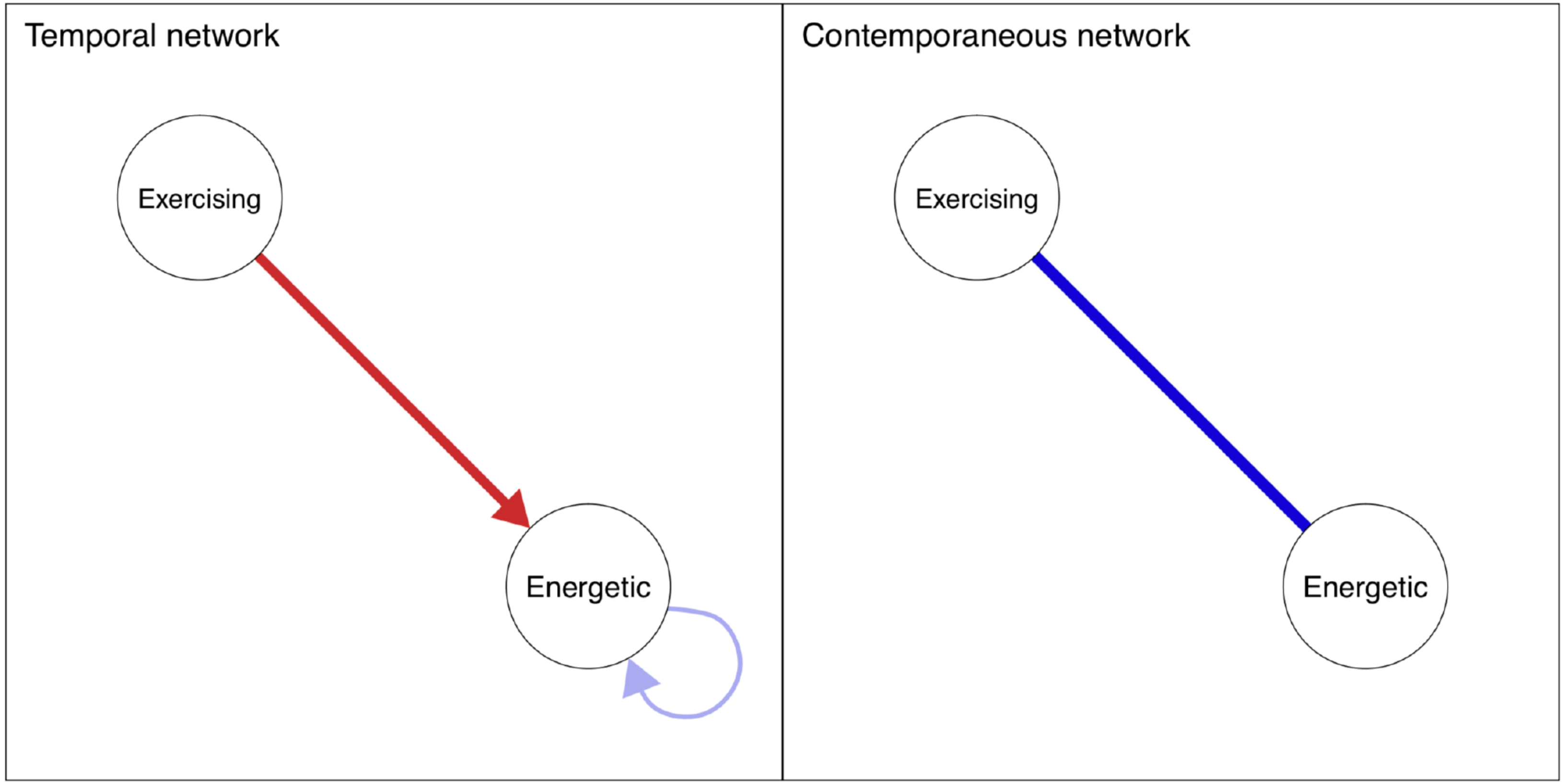
Contemporaneous network

From VAR to **Graphical** VAR (GVAR)

- **Contemporaneous network:** shows if two variables predict one another after taking temporal information into account
  - Contains *effects faster* than the time-window of measurement
  - Example: Dynamics of panic disorder

- Note: not equal to **lag (0)**
  - Takes the temporal information into account
  - GGM of residuals

Temporal versus contemporaneous effects



"Exercising now will lead me to be less energetic in the next time frame. I am energetic now, I will probably still be energetic in the next time frame."

"While I am exercising, I feel energetic."

Graphical VAR model:

$$\boldsymbol{y}_t = \boldsymbol{B}\boldsymbol{y}_{t-1} + \boldsymbol{\varepsilon}_t$$

$$\boldsymbol{\varepsilon}_t \sim N(0, \boldsymbol{\Theta})$$

- $\boldsymbol{y}$ represents a p-variate vector of scores on p measures for an individual at time t
- $\boldsymbol{B}$ is the p × p matrix containing the lagged regression coefficients
    - encodes the **temporal network**
- $\boldsymbol{\varepsilon}$ is a p-variate vector of residuals with a mean of zero and a time-invariant p × p dimensional covariance
- $\boldsymbol{K} = \boldsymbol{\Theta}^{-1}$
    - encodes the **contemporaneous network**

Graphical VAR estimation

- Estimation without regularization/model selection is possible by using multiple regression
- Estimation with regularization: finding sparse solutions for $\boldsymbol{B}$ and $\boldsymbol{\Theta}^{-1}$
  - LASSO estimation with EBIC model selection implemented in the `graphicalVAR` package
    - Missings can be handled using the Kahlman filter in the `impute.ts` package
  - Stepwise non-regularized estimation implemented in the `psychonetrics` package
    - Includes full information maximum likelihood (FIML) for missing data handling

Graphical VAR estimation in R using `graphicalVAR` package

`graphicalVAR()`

- `gamma`: EBIC hyper-parameter
- `vars`: vector of variables to include in the analysis
- `beepvar`: string indicating assessment beep per day
- `dayvar`: string indicating assessment day

Graphical VAR estimation in R using `graphicalVAR` package

```
res_gvar <- graphicalVAR(data, gamma = 0, vars =
vars, dayvar = dayvar, beepvar = beepvar)
```

- Estimated **temporal** network:

  ```
  res_gvar$PDC
  ```

- Estimated **contemporaneous** network:

  ```
  res_gvar$PCC
  ```

Graphical VAR estimation in R using `psychonetrics` package

- Specify model:

```
mod <- gVAR(data, vars = vars, dayvar = dayvar, beepvar = beepvar,
estimator = "FIML")
```

- Unregularized model estimation:

```
mod <- mod %>% runmodel()
```

- "Regularization" i.e., pruning the model:

```
mod <- mod %>% prune(alpha = 0.05, recursive = FALSE, runmodel = FALSE)
```

- Evaluate model fit:

```
mod %>% fit()
```

- Estimated **contemporaneous** network:

```
mod %>% getmatrix("omega_zeta")
```

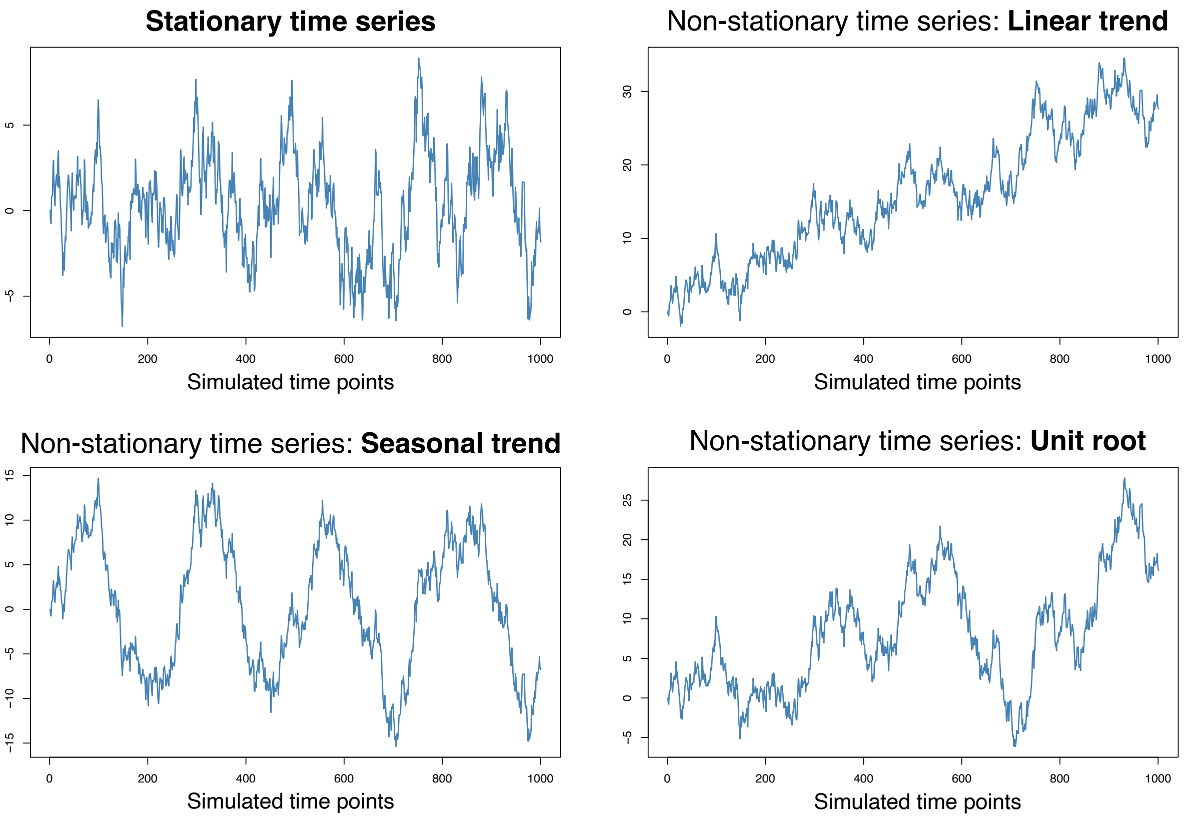
- Estimated **temporal** network:

```
mod %>% getmatrix("PDC")
```

Note: different pruning options here such as FDR-thresholding, greedy set up search and within each of these pruning options you can make different choices again such as correcting for multiple testing

Assumptions of VAR (1): **stationarity**

- We speak of stationary time series when statistical properties such as mean, variance, autocorrelation, etc. are all *constant* over time



Assumptions of VAR (1): **stationarity**

- We speak of stationary time series when statistical properties such as mean, variance, autocorrelation, etc. are all *constant* over time
  - Plausible when data is obtained within a short time span but is often not what people are interested in
  - Can be assessed by regressing each time-series on itself as predictor
    - `kpp.test(data)` in R

- Solution:
  - Detrending is possible: for example you can remove linear trends from the data
    - However often it is these trends that people are interested in
  - Non stationary network model estimation such as time-varying network models (Haslbeck, Bringmann, & Warldorp, 2021)

Assumptions of VAR (1): **stationarity**

- Check for linear trend in R
  - Regress variable var1 on time:

```
lm_var1 <- lm(var1 ~ time, data = data)
Summary(lm_var1)
```

  - If significant, detrend data and take residuals as data:

```
Data$var1[!is.na(data$var1)] <- residuals(lm_var1)
```

Note this is not a magic trick, there can still be non-linear trends in your data, or trends in your means, variances etc.

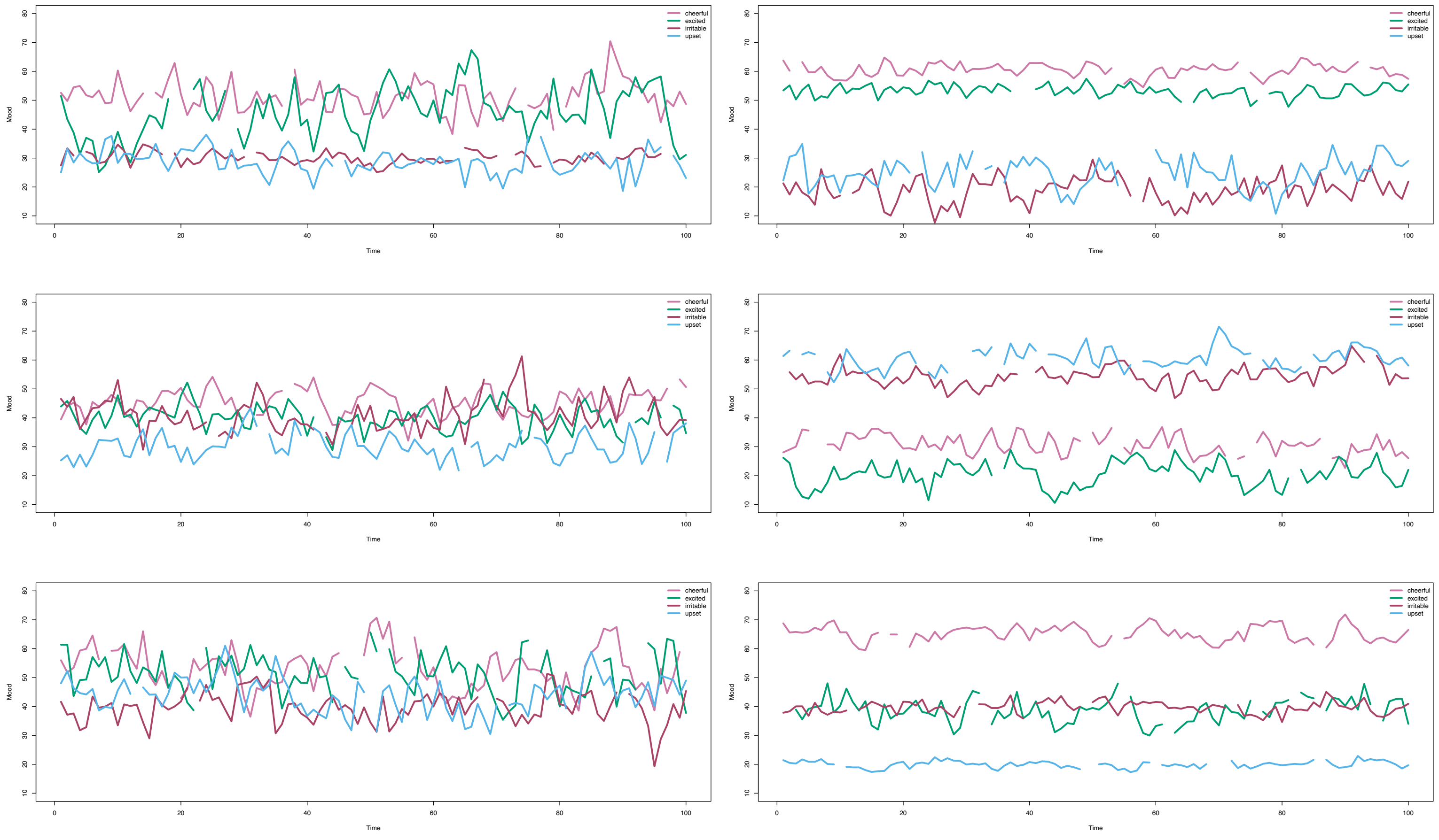
Assumptions of VAR (2): **Equidistant measures**

- Equal distance between measures
  - When measuring over multiple days this assumption is by definition violated due to the nights

- Possible solutions:
  - Remove nights, or model nights as missing observations (can be done in `graphicalVAR` package)
  - Continuous time-series modeling (Ryan & Hamaker, 2022)

Modeling time-series data: $N > 1$ Networks

How to model **multivariate** time-series for **multiple** people?



Extend VAR for multiple people

- With time-series data of multiple persons, a GVAR model could be estimated per person
  - $N = 1$ analyses are typically low-powered → might be lacking the power to pick up effects in the data
- Multilevel VAR modeling allows us to use all information in a single analysis
  - Instead of many under powered analyses → 1 powerful analysis

Why multilevel VAR modeling?

- Several reasons:
  - People are similar, people are different
  - Borrowing information from group-level data
  - Separate *between* from *within* variance
  - Requires fewer observations (compared to averaging the parameters for each individual network)

- Shrinkage
  - Estimates of individuals are somewhat “pulled” together

- Less parameters need to be estimated
  - Only fixed effects and variance–covariance of the random effects need to be estimated

How to extend the VAR model over multiple people?

- We add an extra level to the measurements

- Level 1:
  - Variation *within* an individual over time
  - **Random effects**

- Level 2:
  - Variation *between* individuals over time
  - **Fixed effects**

Multilevel VAR model

- **Level 1:** Each individual is assumed to have their own temporal and contemporaneous VAR model:

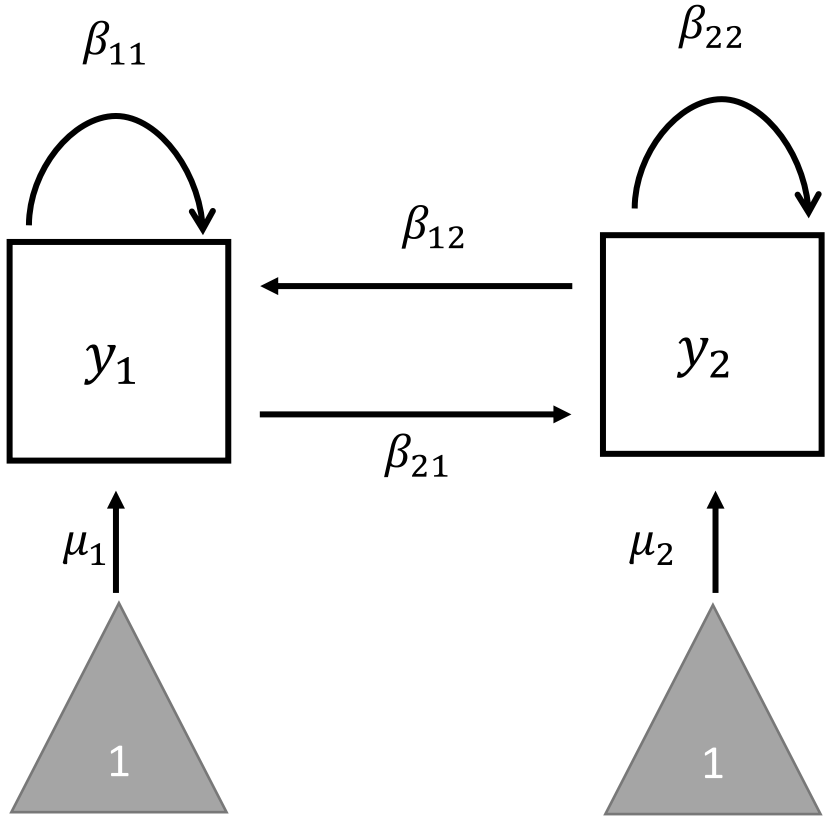
$$\boldsymbol{Y}_t^{(p)} | \boldsymbol{y}_t^{(p)} = N(\boldsymbol{\mu}^{(p)} + \boldsymbol{B}^{(p)}\left(\boldsymbol{y}_{t-1}^{(p)} - \boldsymbol{\mu}^{(p)}\right), \boldsymbol{\Theta}^{(p)})$$

- **Level 2:** Between-subject effects model:

$$\begin{bmatrix} \boldsymbol{\mu}^{(p)} \\ Vec(\boldsymbol{B}^{(p)}) \end{bmatrix} \sim N(\boldsymbol{f}, \boldsymbol{\Omega})$$

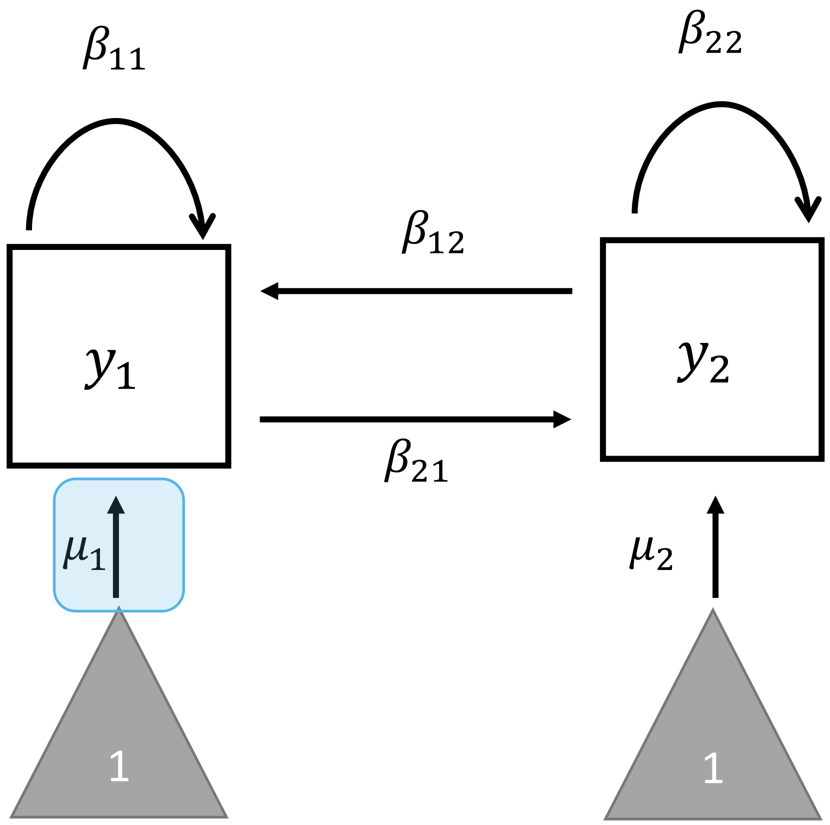
- $\boldsymbol{f}$ encodes fixed effects and Ω the distribution of random effects

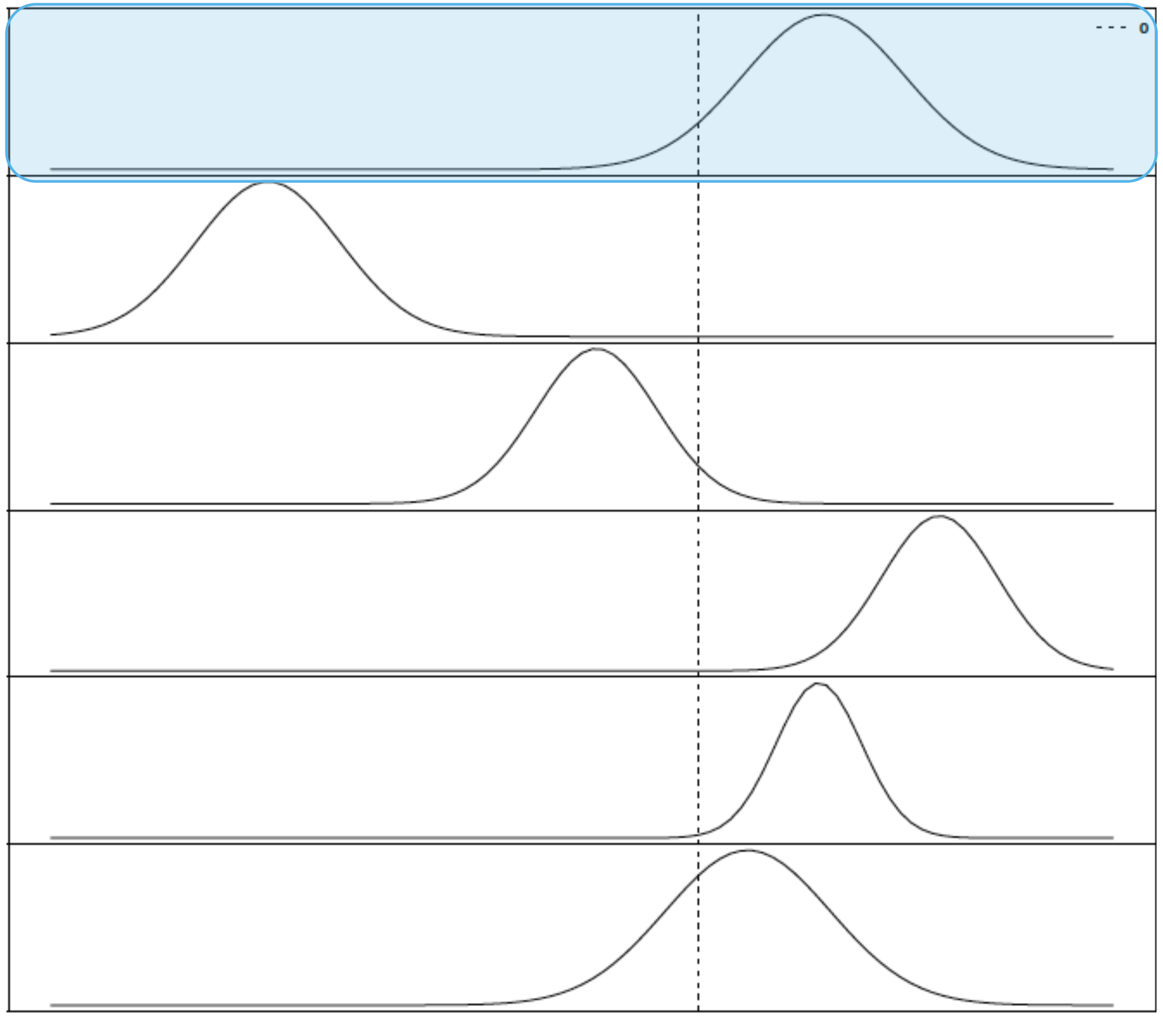
Multilevel VAR model: Basic idea



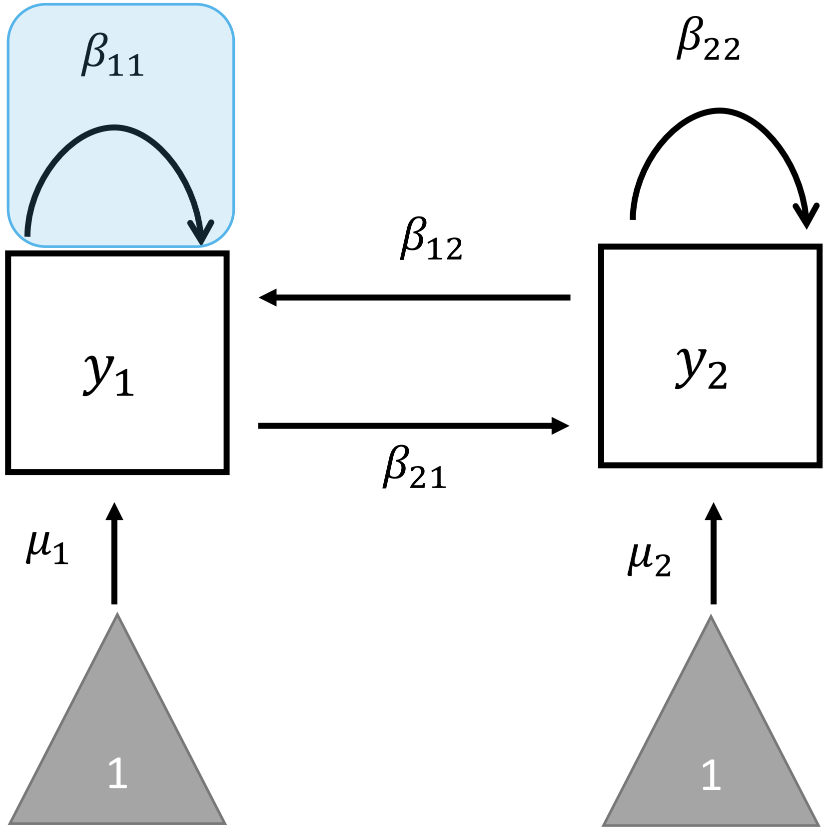
- Each individual has their own **mean** and **(cross-)lagged** correlations
- Individual parameters are sampled from a **shared probability distribution**

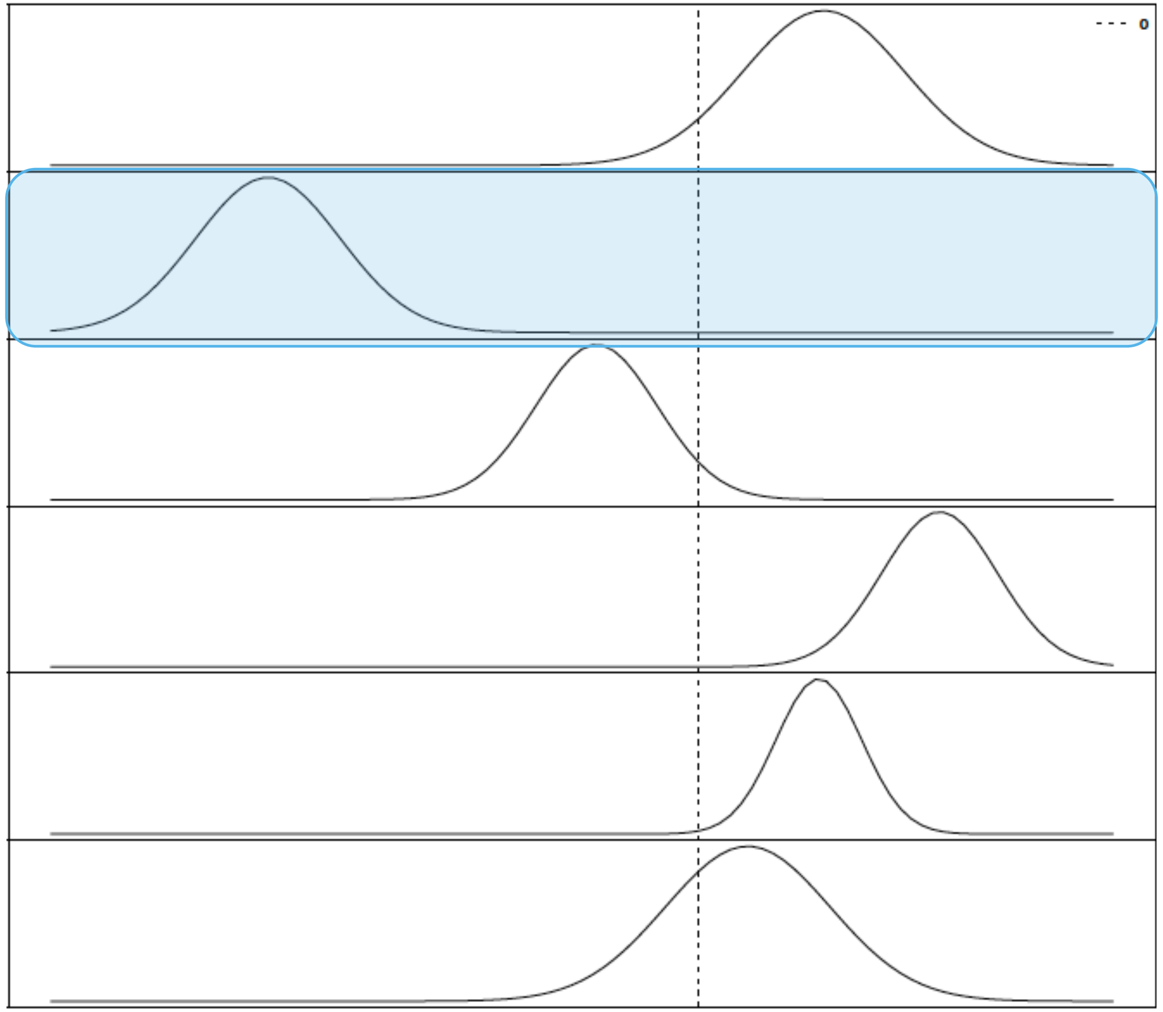
Each parameter has a distribution



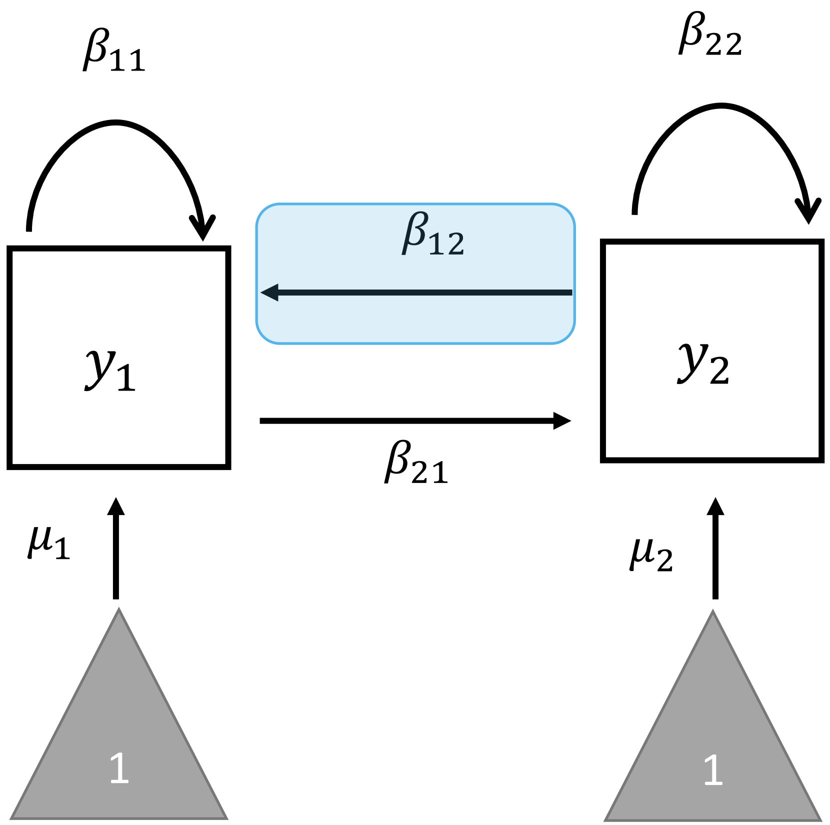


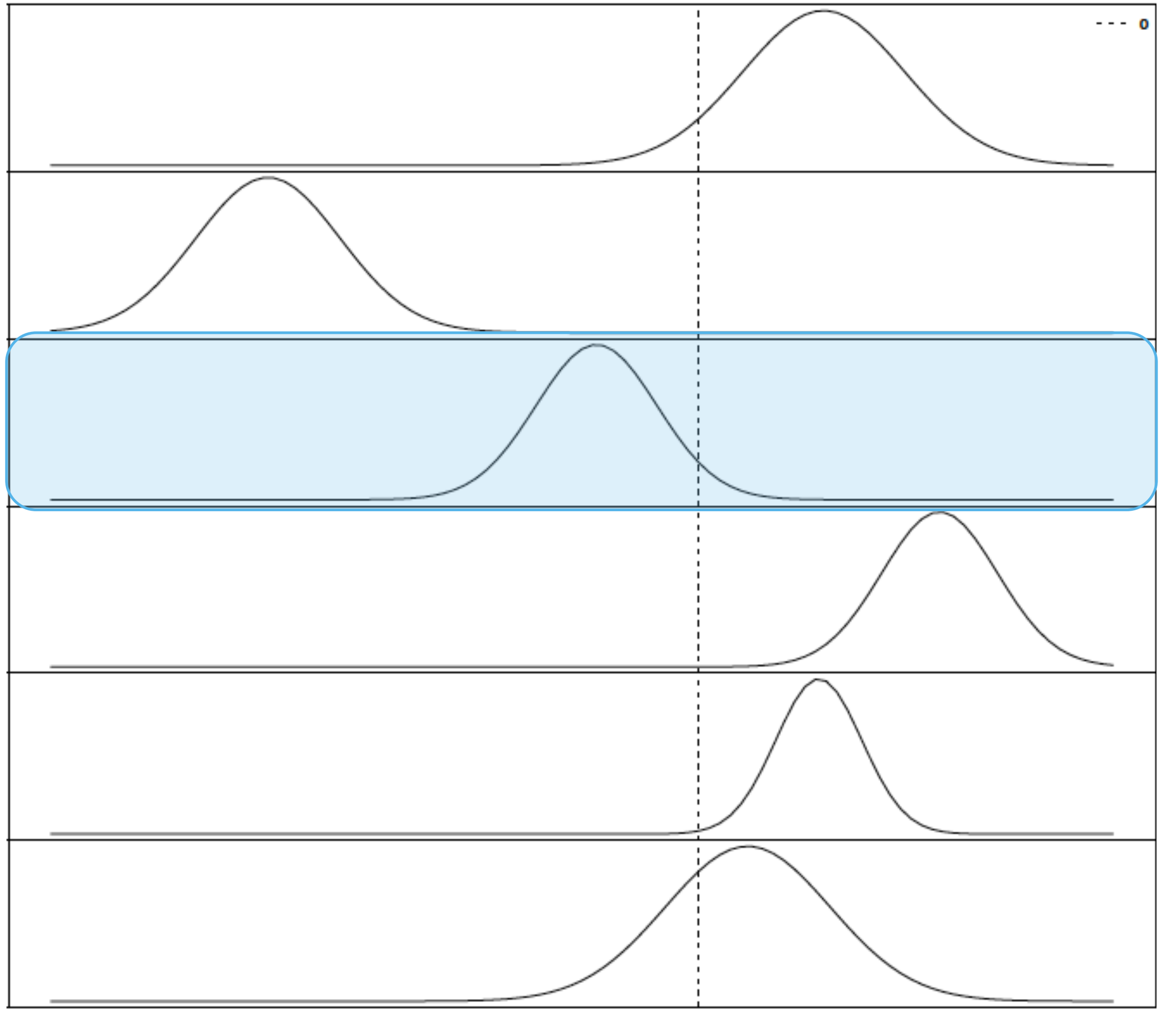
Each parameter has a distribution



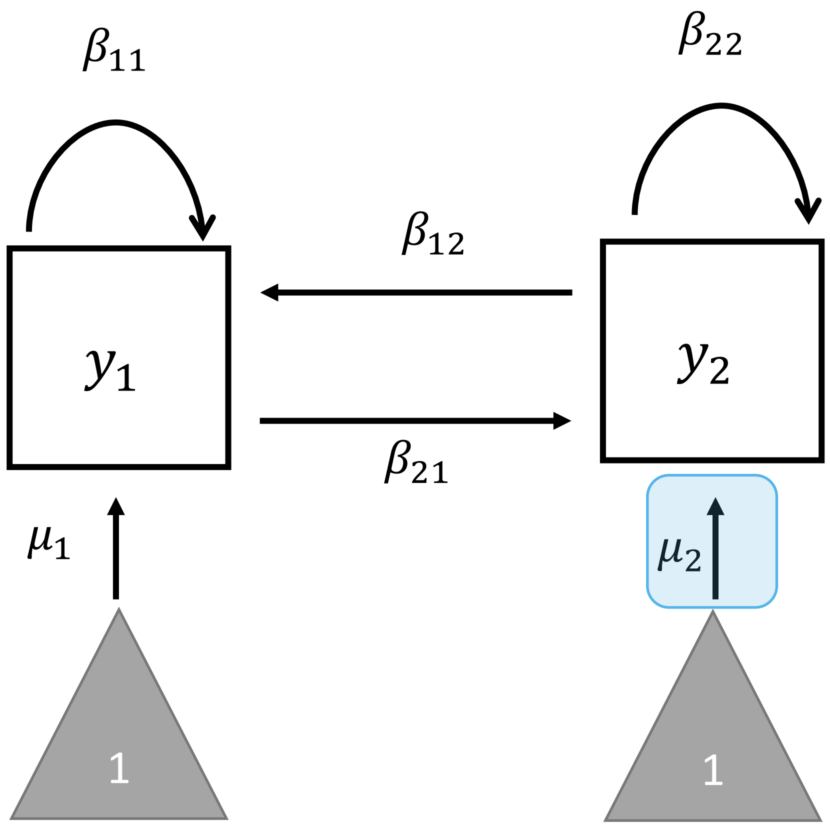


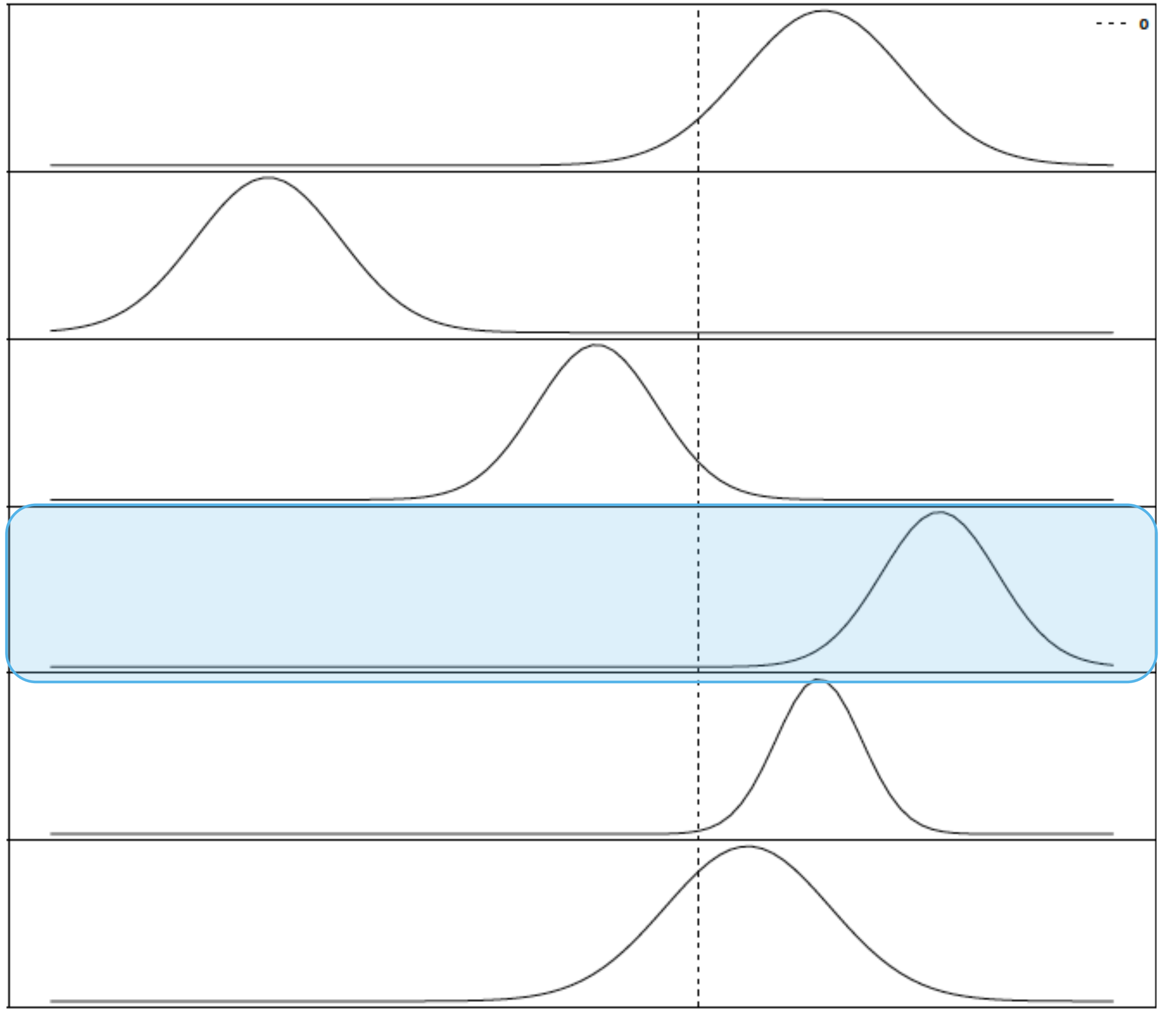
Each parameter has a distribution



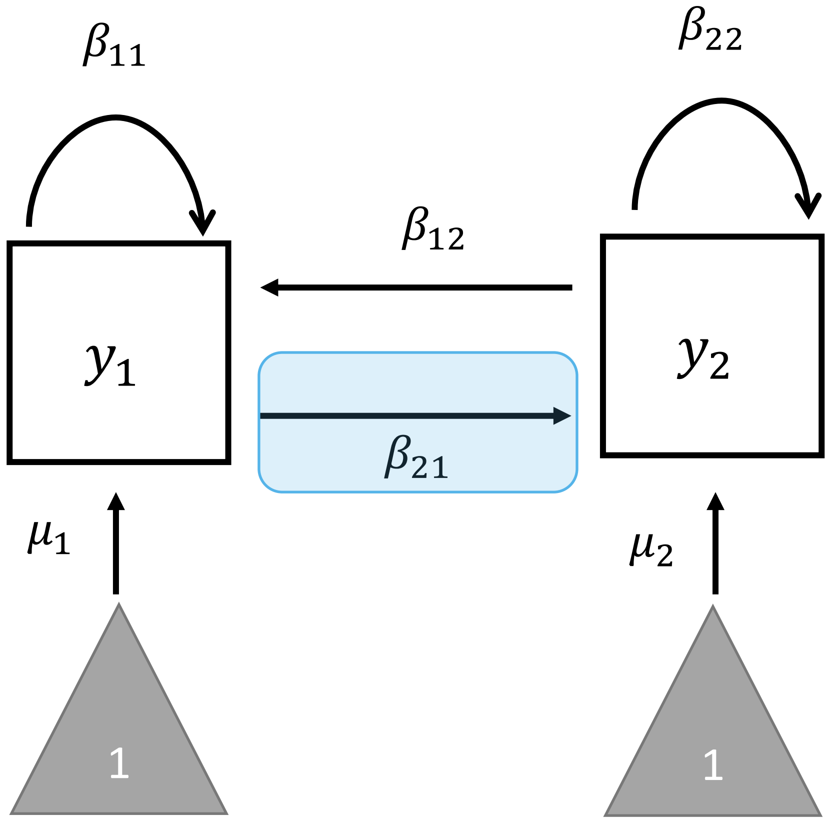


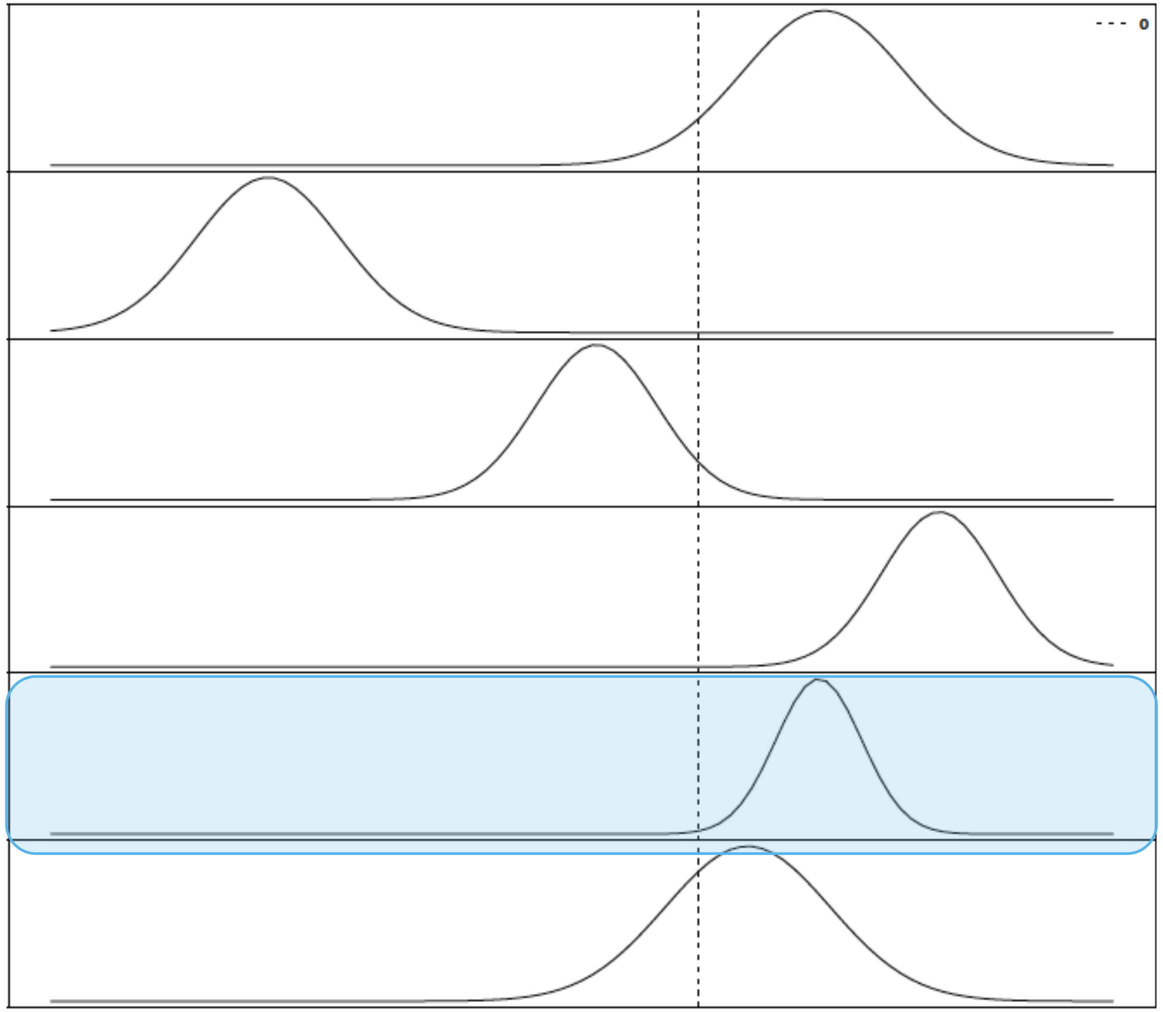
Each parameter has a distribution



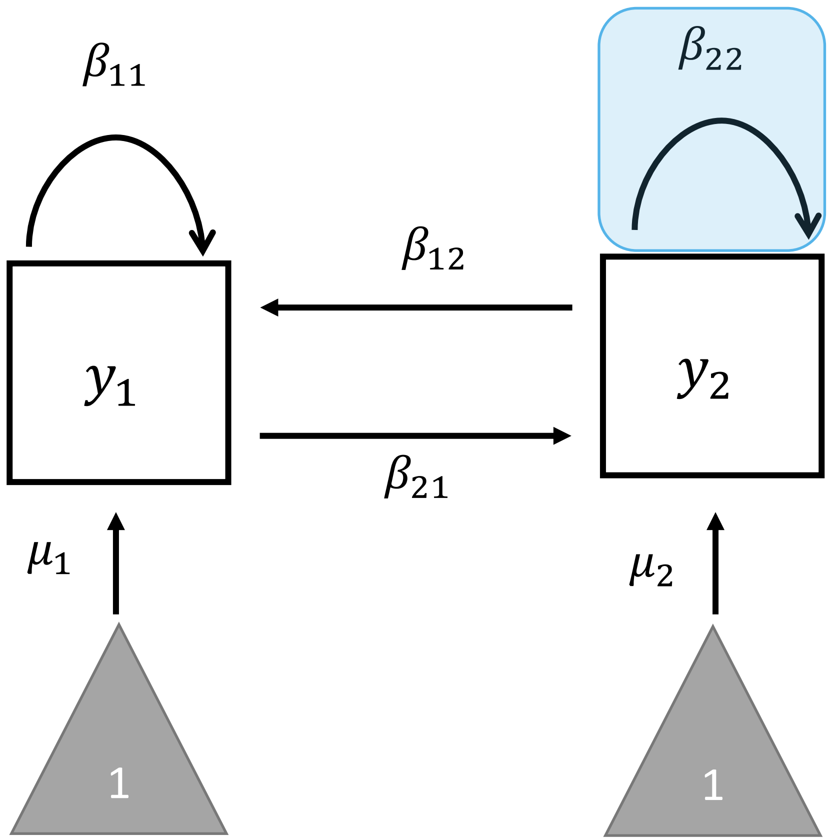


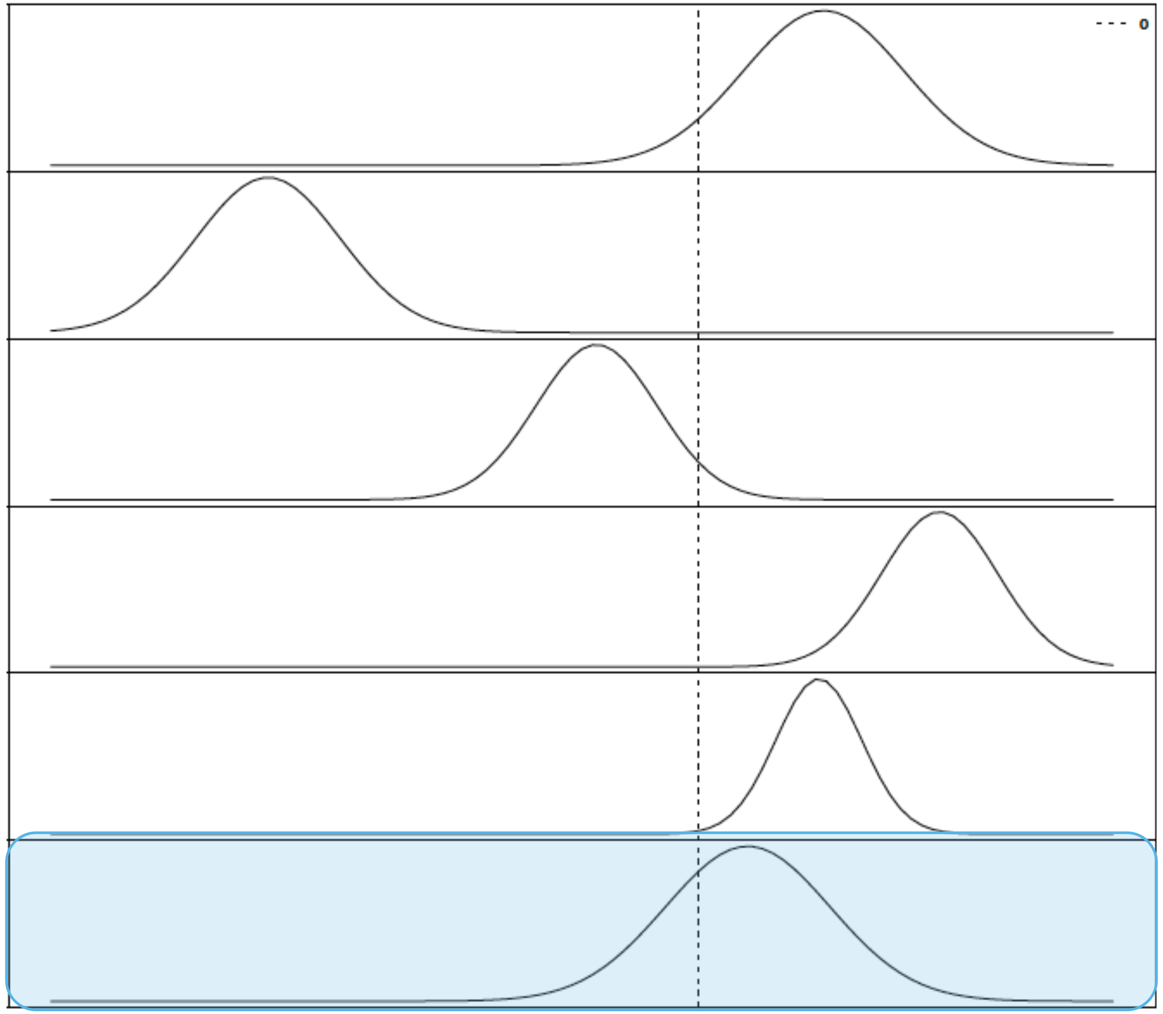
Each parameter has a distribution



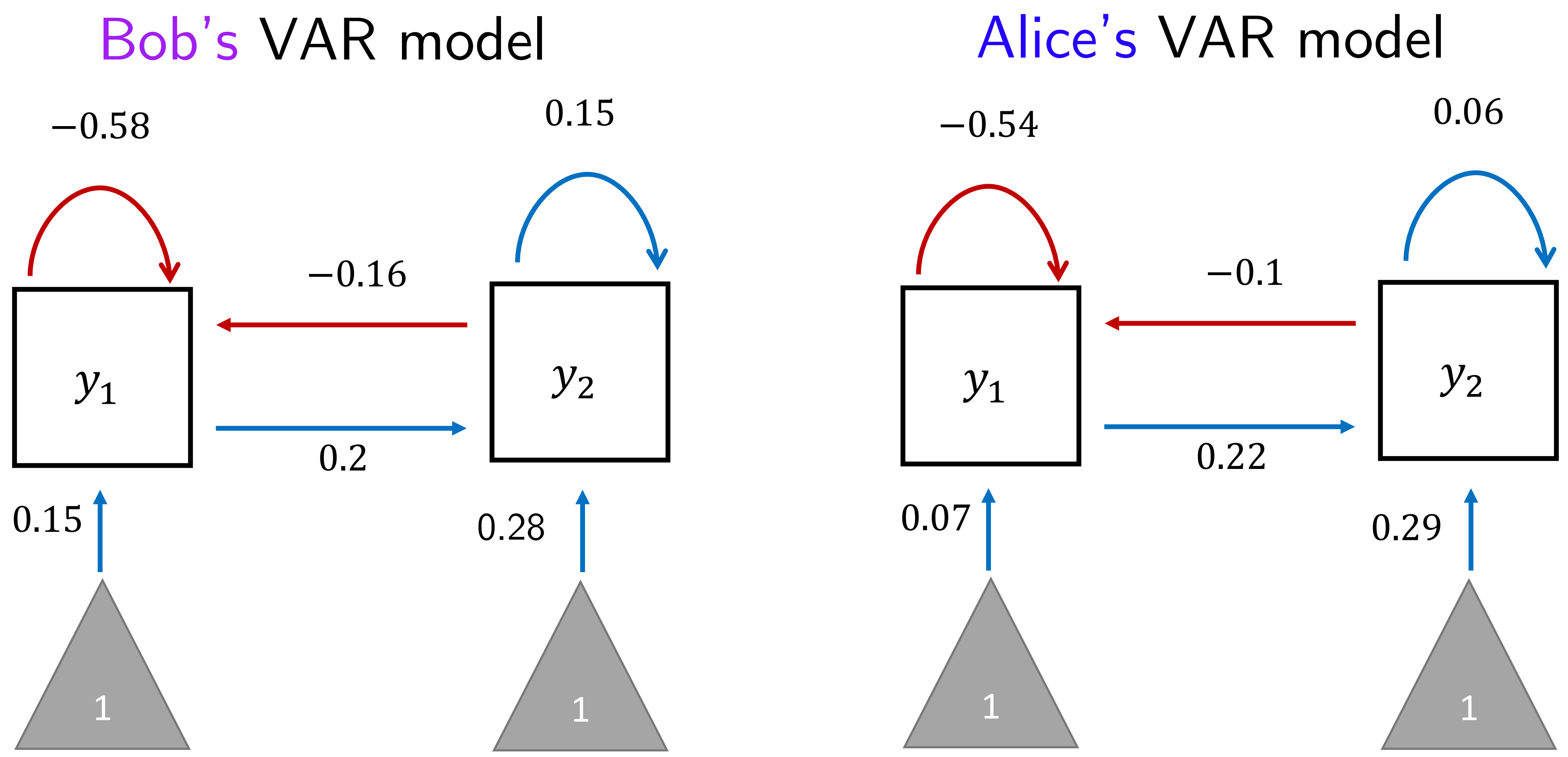


Each parameter has a distribution

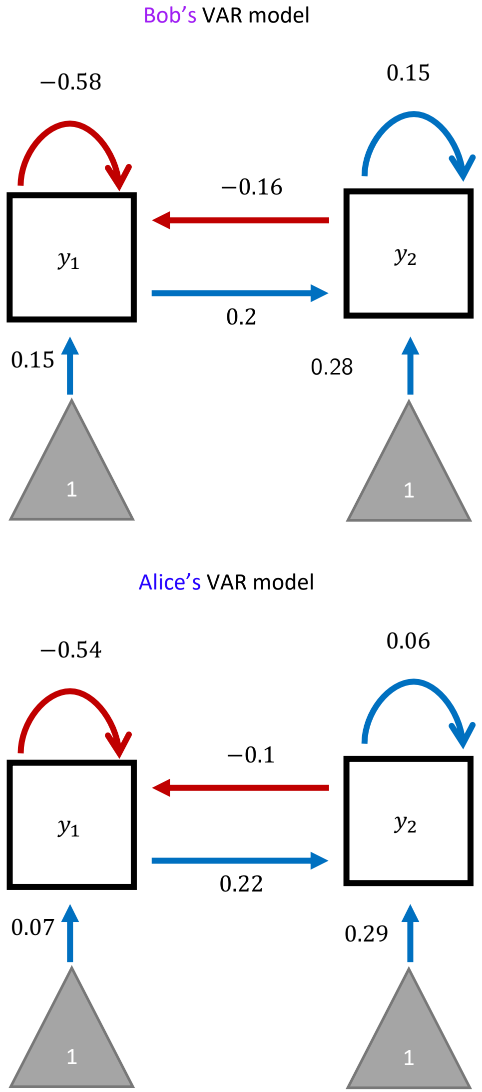


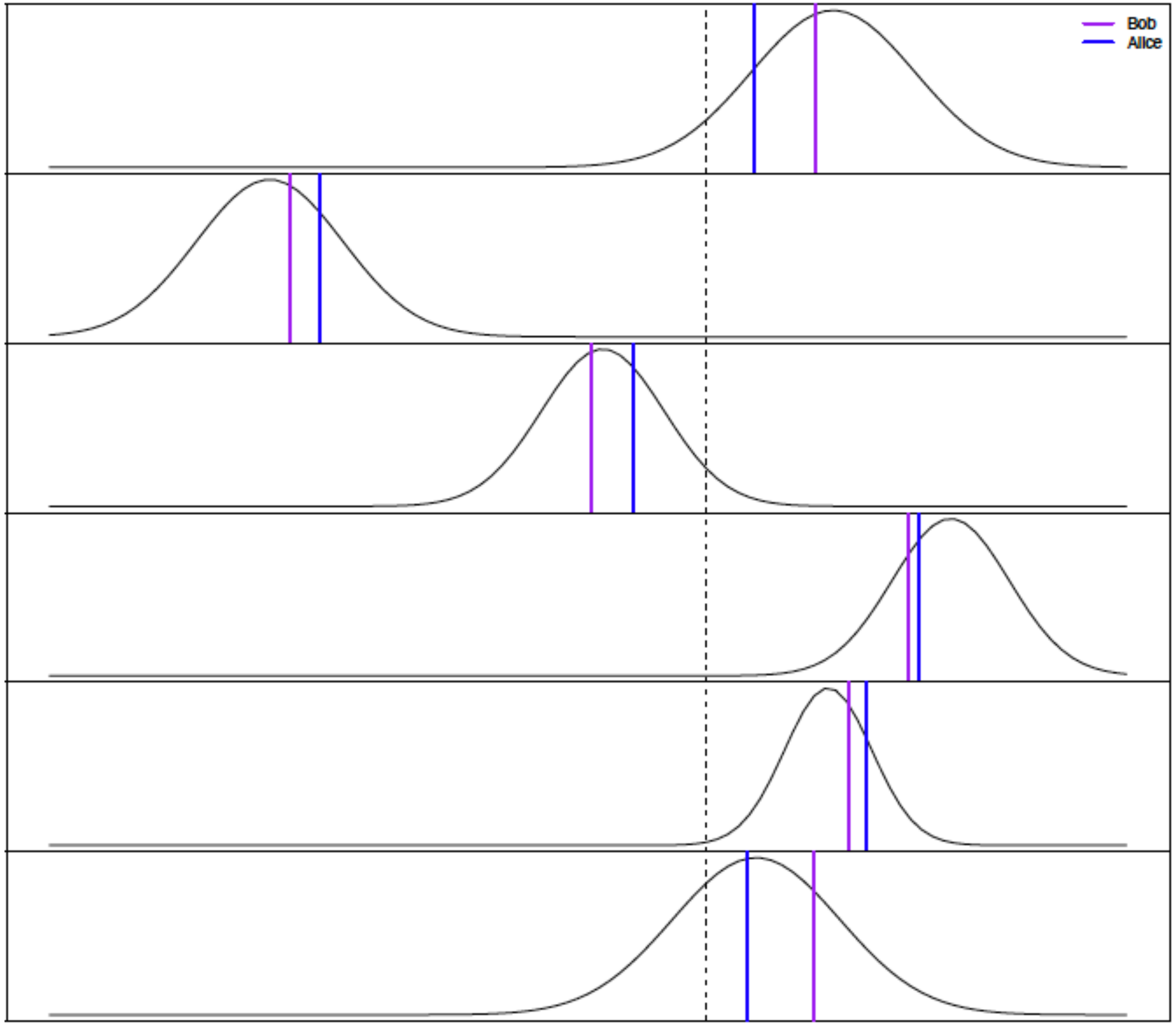


Individual VAR models

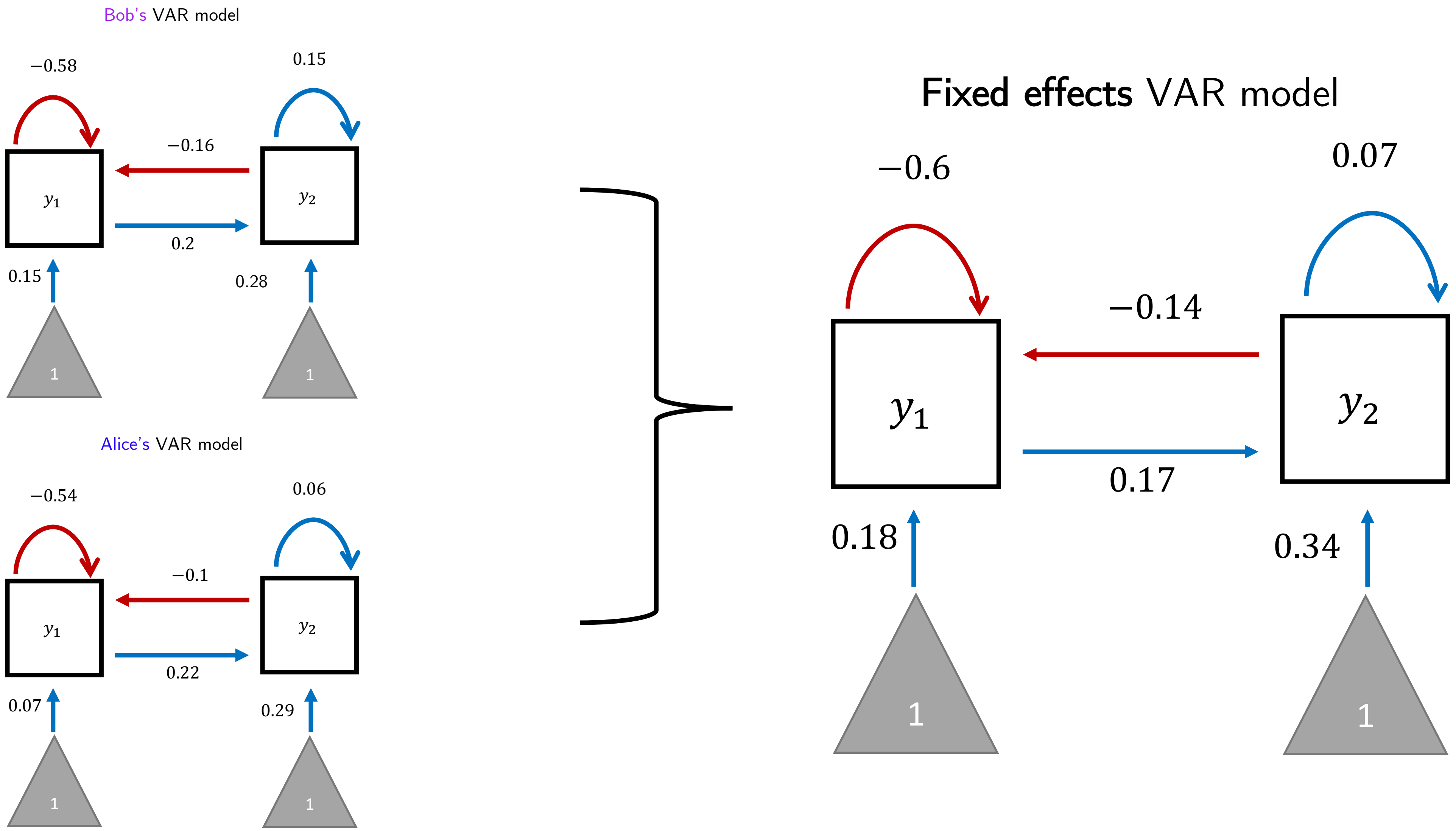


mlVAR: Random effects

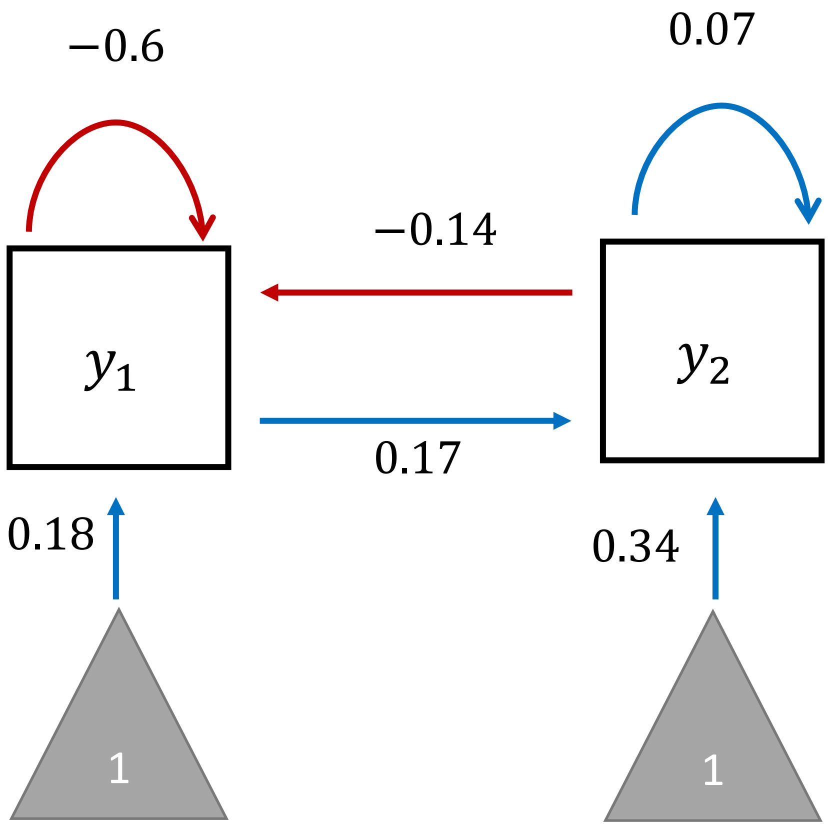


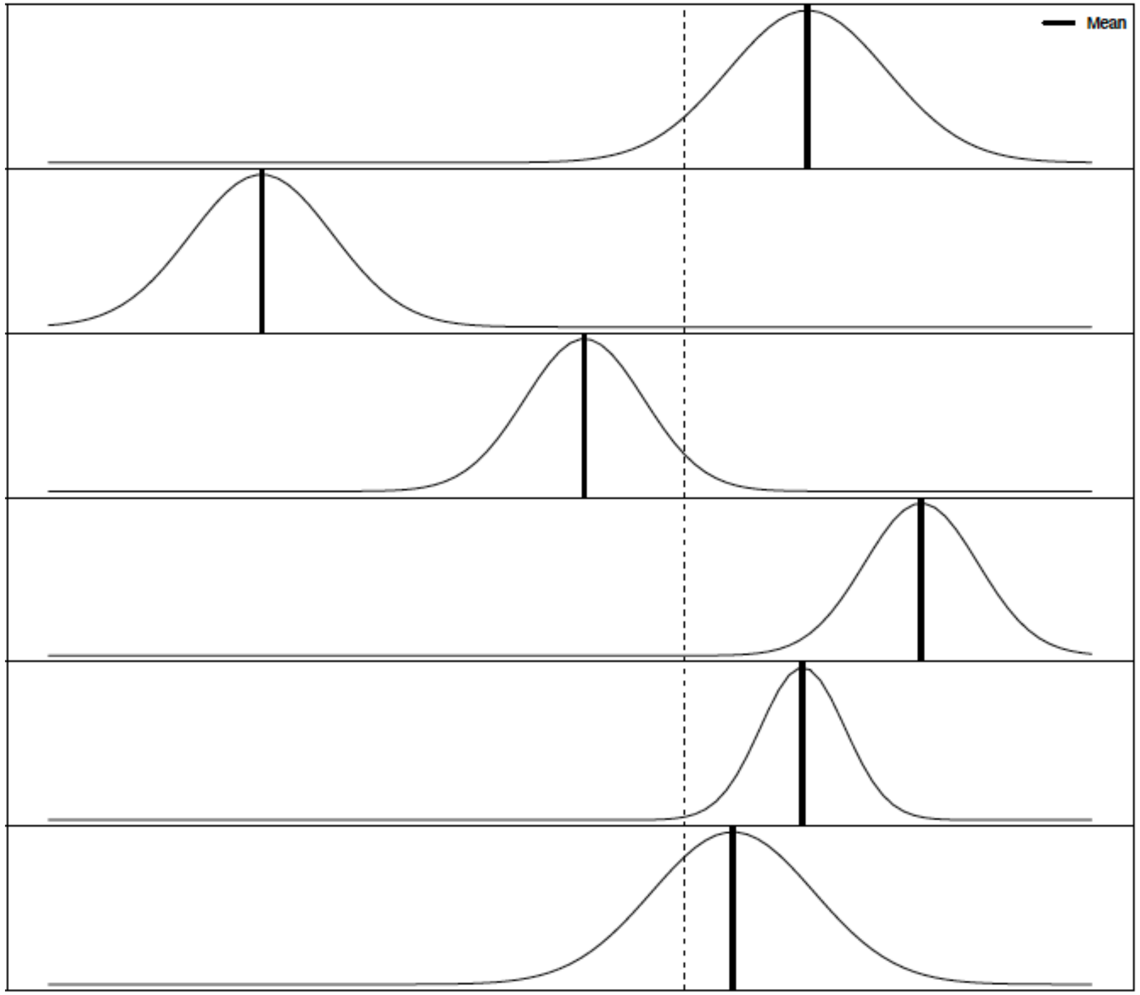


mlVAR: Random effects vs fixed effects

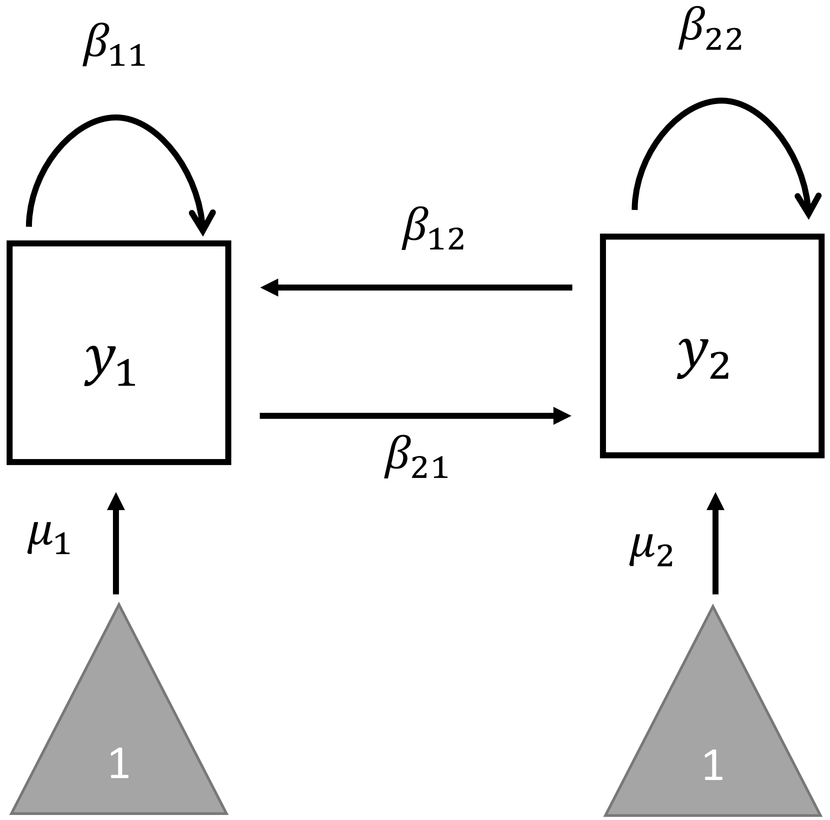


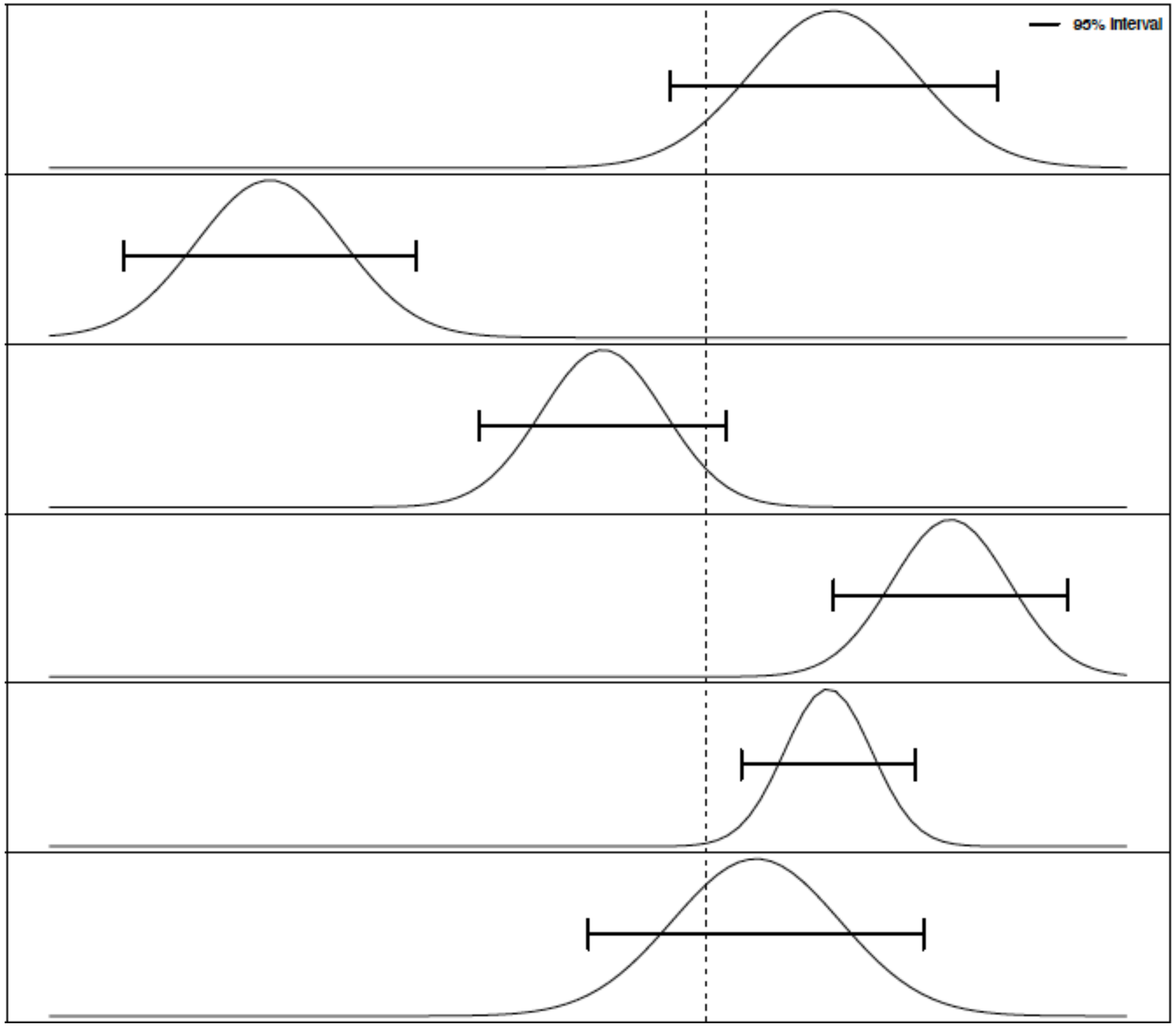
mlVAR: Fixed effects





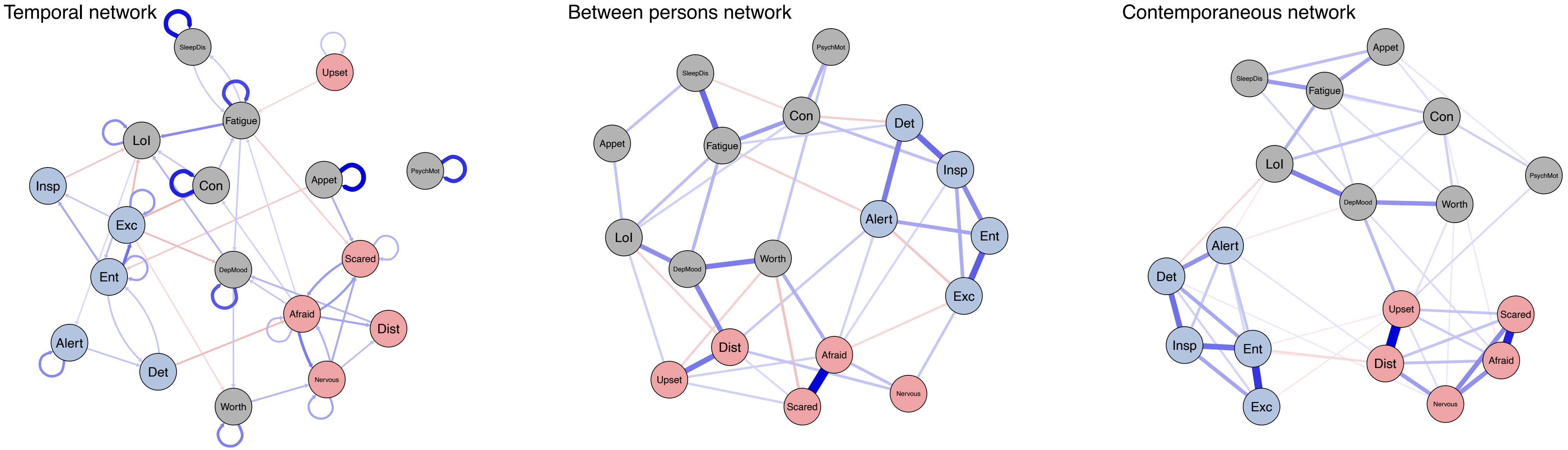
mlVAR: Individual differences





Two step multilevel VAR estimation

- **Step 1:** Node-wise multilevel regressions of variables on within-person centered lagged predictors (temporal effects) and person-wise means (between-subject effects)
- **Step 2:** Node-wise multilevel regressions using residuals from step 1 (contemporaneous effects)
- Implemented in `mlVAR` package



Lunansky, G.,* Hoekstra, R. H. A.,* & Blanken, T. F. (2022). Disentangling the role of affect in the evolution of depressive complaints using complex dynamical networks.

Two step multilevel VAR estimation

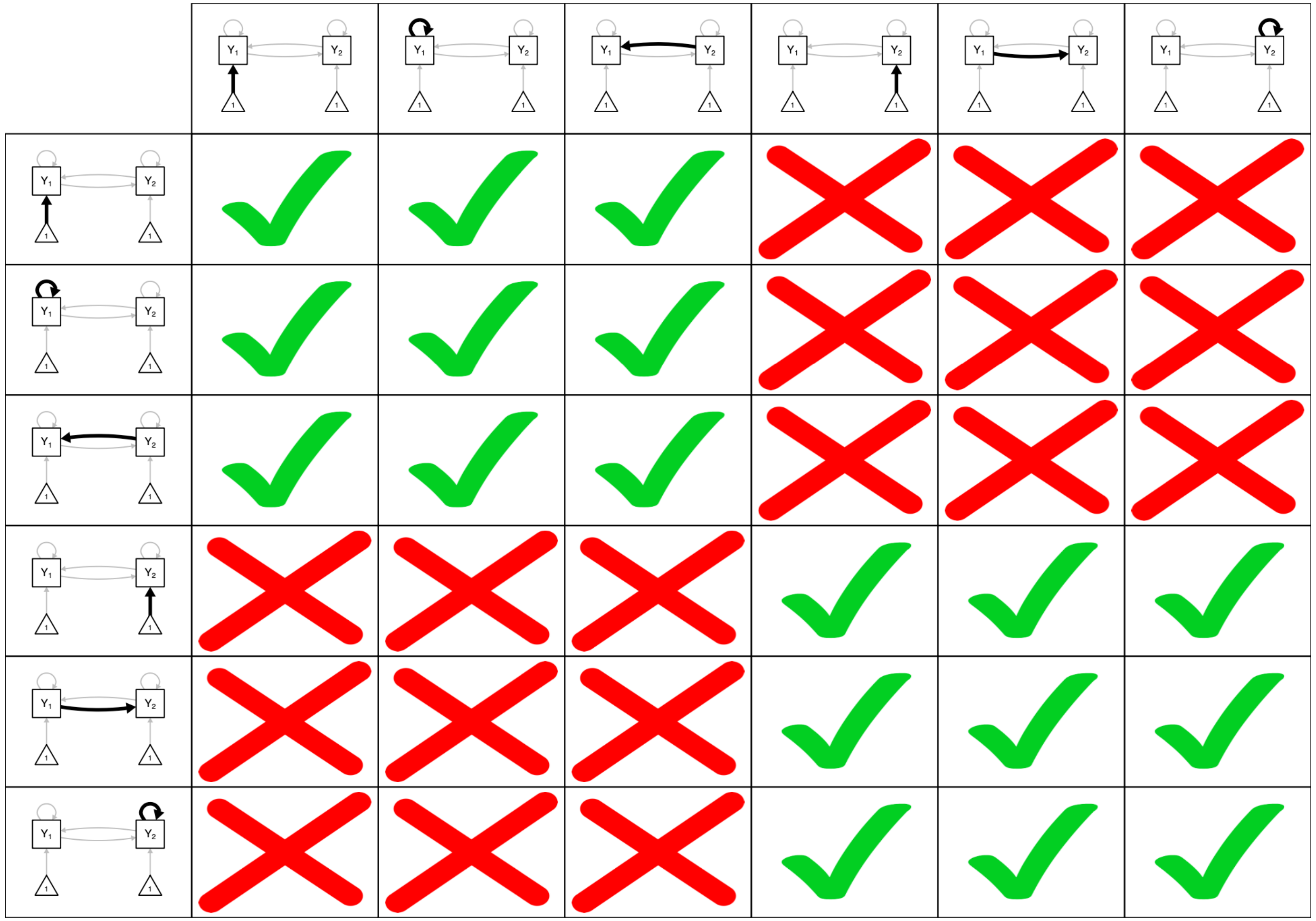
- Full multivariate estimation of the ml VAR model is not possible in open source software

- Solution:
  - Estimating sequentially univariate models: estimate all incoming edges, per node (bringmann et al., 2013)

- Two options for **univariate** multilevel regression models:
  - **Correlated** estimation
  - **Orthogonal** estimation

Correlated versus orthogonal estimation

- **Correlated** estimation in mlVAR

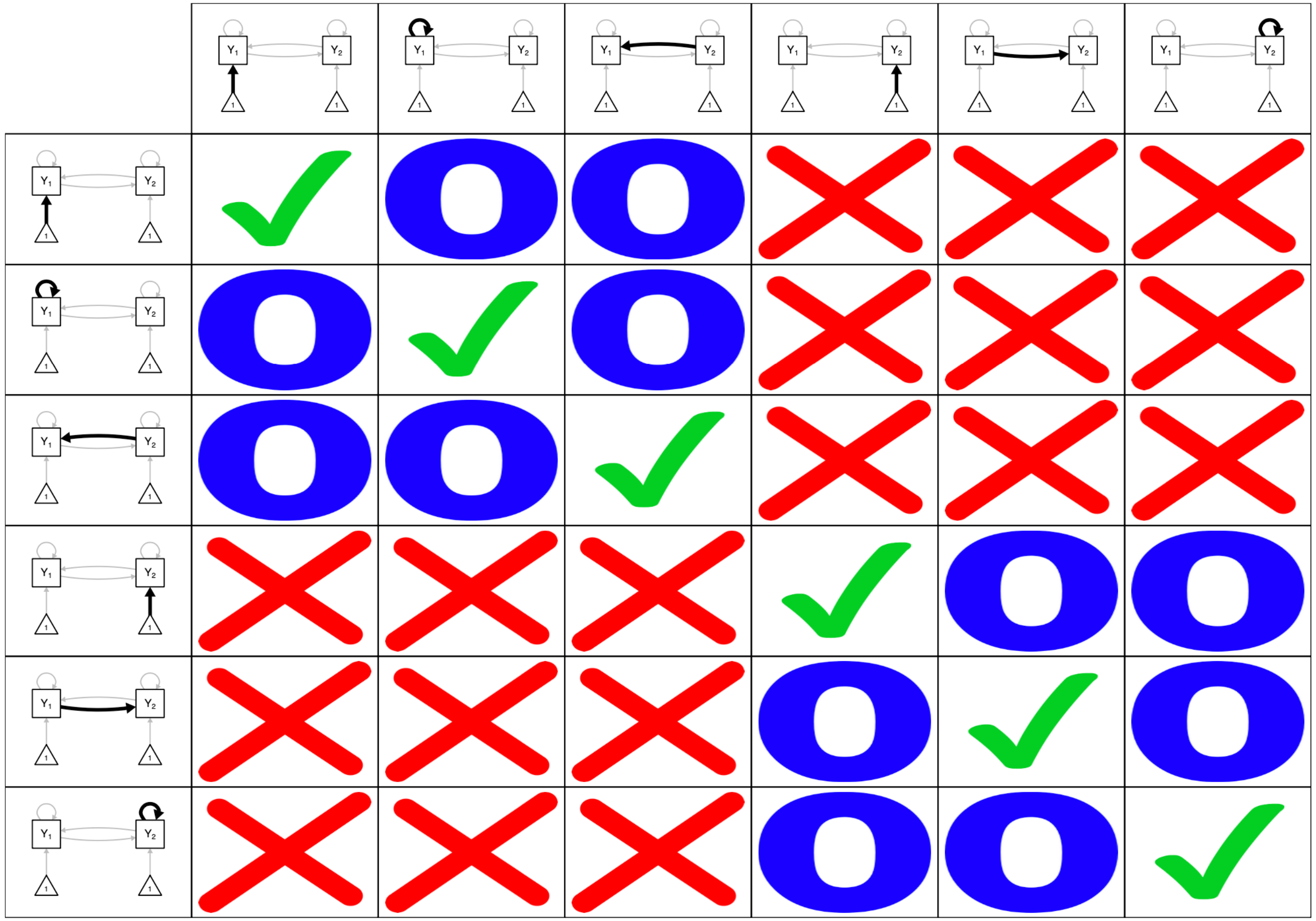


Correlated versus orthogonal estimation

- **Correlated** estimation in mlVAR
  - Some (not all) parameter covariances are being estimated
  - This needs to be integrated out a high-dimensional distribution over parameters
  - Only feasible for up to 6 nodes

Correlated versus orthogonal estimation

- **Orthogonal** estimation in mlVAR



Correlated versus orthogonal estimation

- **Orthogonal** estimation in mlVAR
  - Parameter covariances can be fixed to zero
  - Fast and works for many variables
  - Doesn't return any parameter co-dependencies

Multilevel VAR estimation in R

`mlVAR()`

- `estimator`: "LMER" for sequential univariate multi-level estimation, "mplus" for multivariate Bayesian estimation (requires Mplus), and "lm" for fixed effects estimation
- `contemporaneous/temporal`: estimating correlated or orthogonal contemporaneous/temporal networks?

Multilevel VAR estimation in R

```
res_ml <- mlVAR(data, vars = vars, idvar = idvar,
contemporaneous = ”correlated”, temporal =
“correlated”)
```

- Fixed effects temporal network:

  ```
  plot(res_ml , “temporal”, title = “Temporal Network”)
  ```
- Fixed effects contemporaneous network:
  - `plot(res_ml , “Contemporaneous”, title = “Contemporaneous network”)`
- Between-subjects network:

  ```
  plot(res_ml , “between”, title = “Between-subjects
  network”)
  ```

Multilevel VAR estimation in R

```
res_ml <- mlVAR(data, vars = vars, idvar = idvar,
contemporaneous = ”correlated”, temporal = “correlated”)
```

- Random effects temporal network:

  ```
  plot(res_ml , type= “temporal”, subject = 1, title = “Temporal
  Network”)
  ```

- Random effects contemporaneous network:

  ```
  plot(res_ml , type = Contemporaneous”, subject = 1, title =
  “Contemporaneous network”)
  ```

- Individual differences network:

  ```
  plot(res_ml , type = “temporal”, SD = TRUE, title = “Individual
  differences temporal network”)
  plot(res_ml , type = “contemporaneous”, SD = TRUE, title =
  “Individual differences contemporaneous network”)
  ```

Disadvantages of multilevel VAR

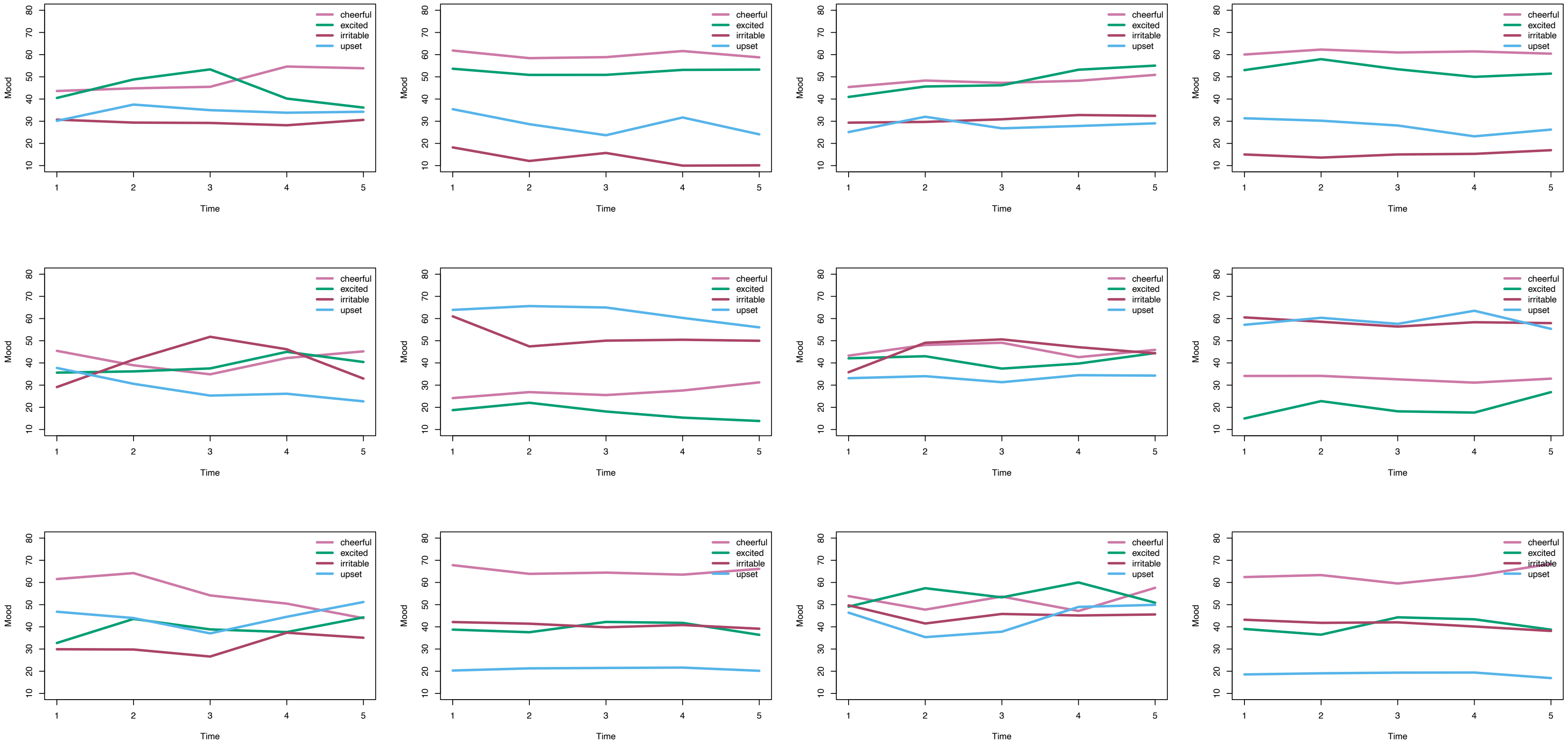
- Some parameters in the model might be correlated.
  - For example: People who have a high mean of worry could also have a high auto-correlation of worry

- This can be modeled in a joint multivariate distribution.
  - However, multi-level VAR uses univariate multilevel regression models $\rightarrow$ forces some random effects to be uncorrelated.

+ Computationally less expensive

- Not “fully” multivariate

Extending VAR over multiple people with fewer time-points



Panel VAR model

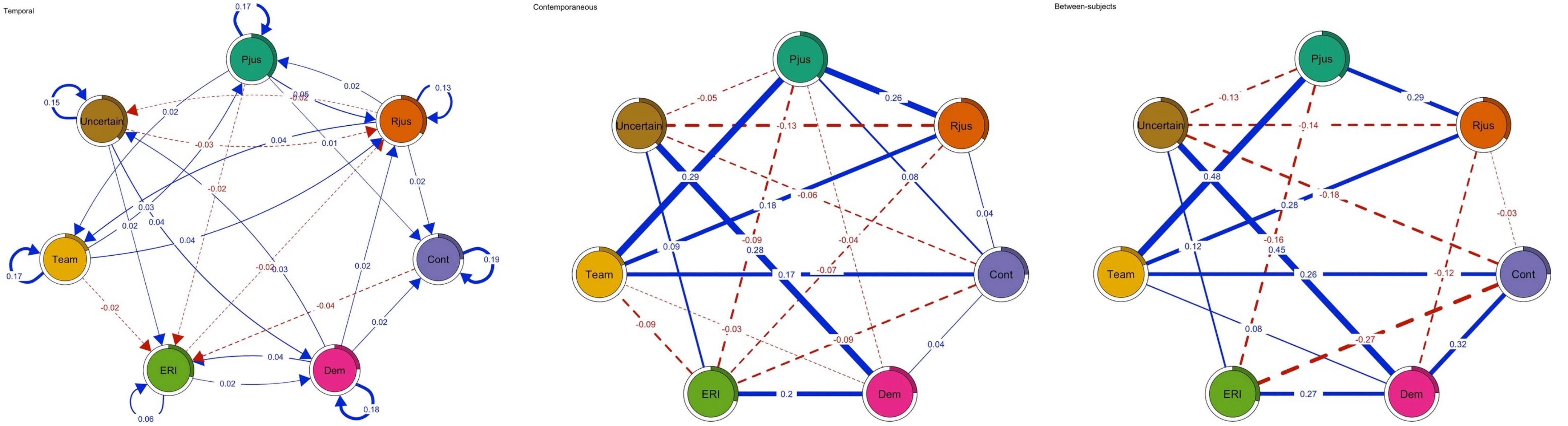
- An often occurring datatype between cross-sectional and time-series data is panel data
  - Many people measured on a few repeated occasions ($t = 3$ or $t = 5$)

- Estimating a VAR model with panel data cannot be done in mlVAR estimation
  - Recently the ml VAR framework of contemporaneous, temporal and between-subjects network is extended to model panel VAR models

- **Advantage**: if interested in the fixed effect structures, fewer data-points are required

Panel VAR model estimation

- Fixed effect within-subject temporal network modeled with $\boldsymbol{B}$
- Fixed-effect within-subject contemporaneous network modeled with $\boldsymbol{\Theta}^{-1}$
- Between-subject network modeled with $\boldsymbol{\Omega}$



Elovainio, M., Hakulinen, C., Komulainen, K., Kivimäki, M., Virtanen, M., Ervasti, J., & Oksanen, T. (2022). Psychosocial work environment as a dynamic network: a multi-wave cohort study. *Scientific reports*, *12*(1), 1-11.

Panel VAR model estimation

- Fixed effect within-subject temporal network modeled with $\boldsymbol{B}$
- Fixed-effect within-subject contemporaneous network modeled with $\boldsymbol{\Theta}^{-1}$
- Between-subject network modeled with Ω

- Temporal, contemporaneous and between-subject network are estimated using FIML estimation
  - Implemented in `psychometrics` package

Panel VAR model estimation in R

- Form model:
  ```
  mod <- panelgvar(data, vars = vars)
  ```

- Estimate model:
  ```
  mod <- mod %>% runmodel()
  ```
- Estimate pruned model:
  ```
  mod %>% prune(alpha = 0.05, recursive = FALSE)
  ```
- Evaluate model fit:
  ```
  Mod %>% fit()
  ```

- Estimated fixed effects **contemporaneous** network:
  ```
  mod %>% getmatrix(”omega_zeta_within")
  ```
- Estimated fixed effects **temporal** network:
  ```
  mod %>% getmatrix(”PDC")
  ```
- Estimated **between-subjects** network:
  ```
  mod %>% getmatrix(”omega_zeta_between”)
  ```

Note: different pruning options here such as FDR-thresholding, greedy set up search and within each of these pruning options you can make different choices again such as correcting for multiple testing

Challenges in time-series modeling

Challenge: **Measurement**

- Response scales:
  - What scales should we use and how do they impact measurement?
    - VAS-scales vs. Likert-type ordinal scales
  - Recent study showing that different scale types seem to affect distribution (Haslbeck et al., 2022)

- Response shifts:
  - Will measurement of a construct stay the same across the assessment period? If not, why not? How to integrate it into assessments and analyses?
    - If people change their internal standards or redefine constructs, then variability in scores may reflect response shifts in measurement, not changes in the actual construct

Challenge: **Measurement**

- Does continuous administration of intensive longitudinal questionnaires evoke any kind of answer tendencies?
  - If measurement leads to changes in these construct, we don't measure ecologically valid behavior, we measure a "new reality"(Ram et al., 2017)

- Inter-individual differences in scale use:
  - How do people actually decide on their score? (e.g., do they compare with previous scores or personal anchors?)
    - Differences in response strategies may bias measurement in construct-irrelevant ways.
  - How do we take into account inter-individual differences in scale use/interpretation?
    - Recent study: a 50 on a scale of 0-100 meant fundamentally different things to individuals (based on qualitative interviews)

Challenge: **No covariance without variance**

- **Variance** in a variable is necessary in order to show **covariance** with another variable
  - Note: no variance means no connections in the VAR network

- Not all variables we are interested in show variance (within the measurement period)
  - e.g., childhood trauma, economic status, income

- As a consequence: some of the variables that we might be interested in, we cannot model in time-series models

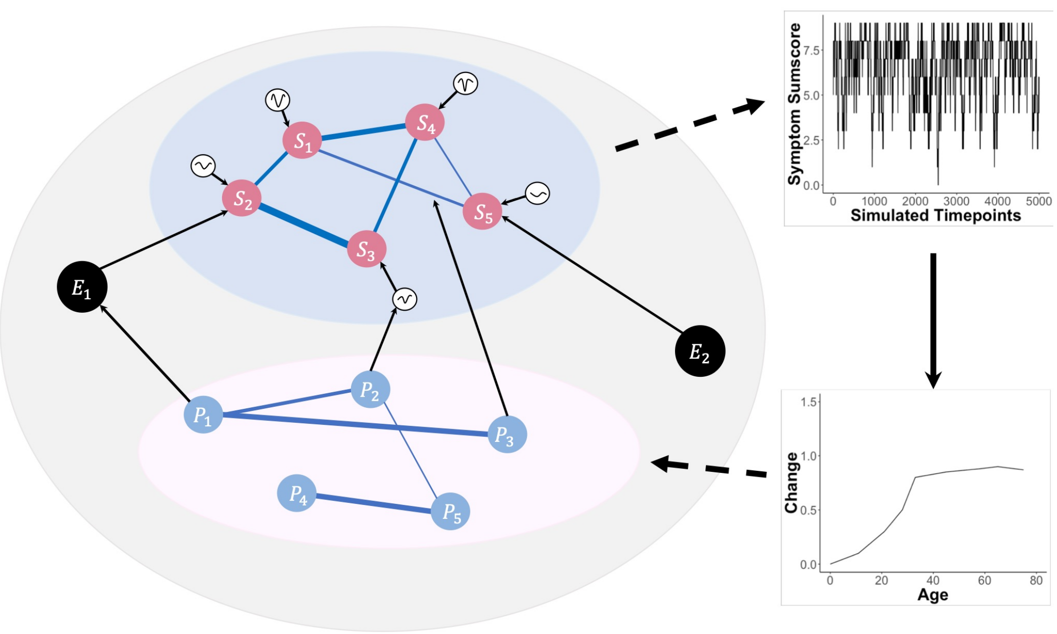
Challenge: **Time-scale**

- Did you get the time scale right?
  - In many cases, we lack knowledge regarding the exact time-scale on which processes of interest operate
  - For practical considerations, ESM questionnaires are often administered every 2-4 hours, or once a day

- Choosing the lag
  - Most commonly, we use lag1 for modeling in idiographic networks
  - This raises the question: Are we capturing the dynamics, or are they faster/slower than our assessment?

Challenge: **Time-scale**

- One time scale for all variables?
  - Even if we capture some of the relevant processes with our measurements, it seems unlikely to assume that all processes within one network operate on the same time scale:
    - E.g., clin … combination with slow processe



Lunansky, G., van Borkulo, C., & Borsboom, D. (2020). Personality, resilience, and psychopathology: A model for the interaction between slow and fast network processes in the context of mental health. *European Journal of Personality*, *34*(6), 969-987.

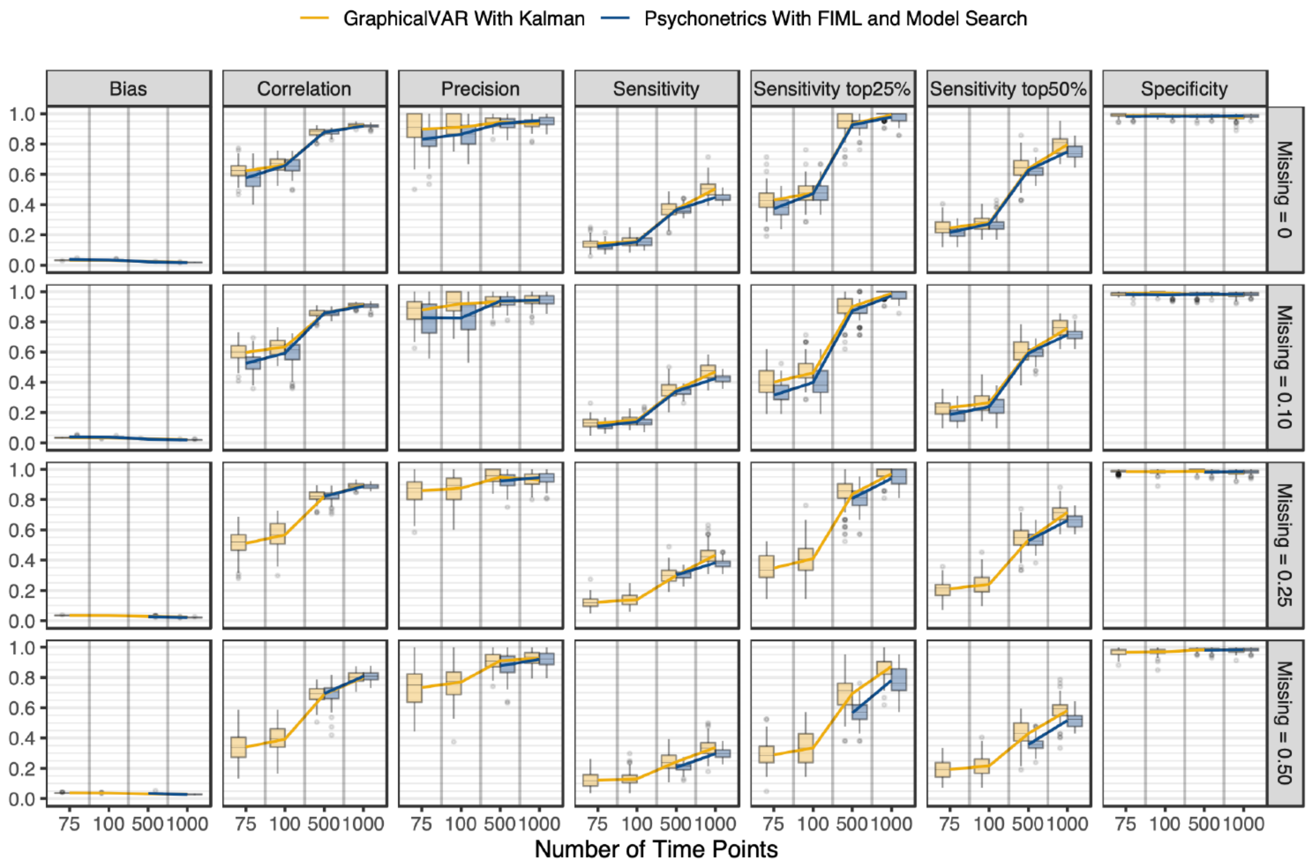
Challenge: **Systematic missingness**

- Often time-series data contains a lot of missingness
  - Is the data missing completely at random?
  - Missing at random?
  - Missing not at random?

- How to handle missing data?
  - Imputation?
    - Which imputation technique to use?
  - FIML estimation
- What percentage of missing data is acceptable?

Challenge: **Systematic missingness**

- What percentage of missing data is acceptable?



Mansueto, A. C., Wiers, R. W., van Weert, J., Schouten, B. C., & Epskamp, S. (2022). Investigating the feasibility of idiographic network models. *Psychological methods*.

Challenge: **Trends in the data**

- How to account for trends in the data?
  - We saw an example of how to detrend for linear trends, but what about all other possible trends in the data?
  - What effect has detrending on the **interpretation**?

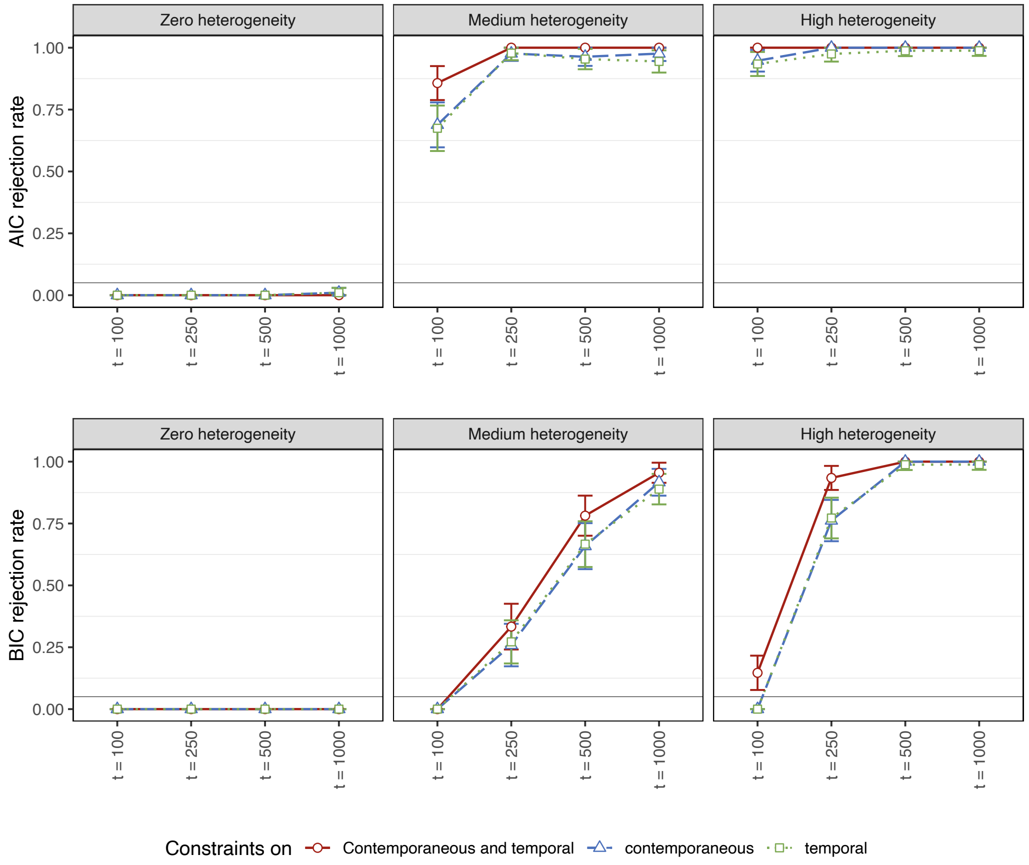
- What if our research question implies non-stationarity?
  - Often we are interested in the trend in the data (e.g., change after treatment)

- **Possible alternative**:
  - Time-varying VAR (Haslbeck, Bringmann, & Waldorp, 2021)

Challenge: **Heterogeneity**

- How to test heterogeneity *between* individuals?
- How to test heterogeneity *within* individuals?

- **Individual Network Invariance Test**: place equality constraints on network models and compare model fit (Hoekstra et al. in preparation)



Hoekstra, R. H. A., Epskamp., S, Niergenberg, A. A., Borsboom, D. & McNally, R. (in preparation). Testing similarity in longitudinal networks: The Individual Network Invariance Test (INIT)

Thank you for your attention!