Tensor regression for dynamic network data

Peter Hoff

Duke University
Supported by the NSF (DMS-1505136)

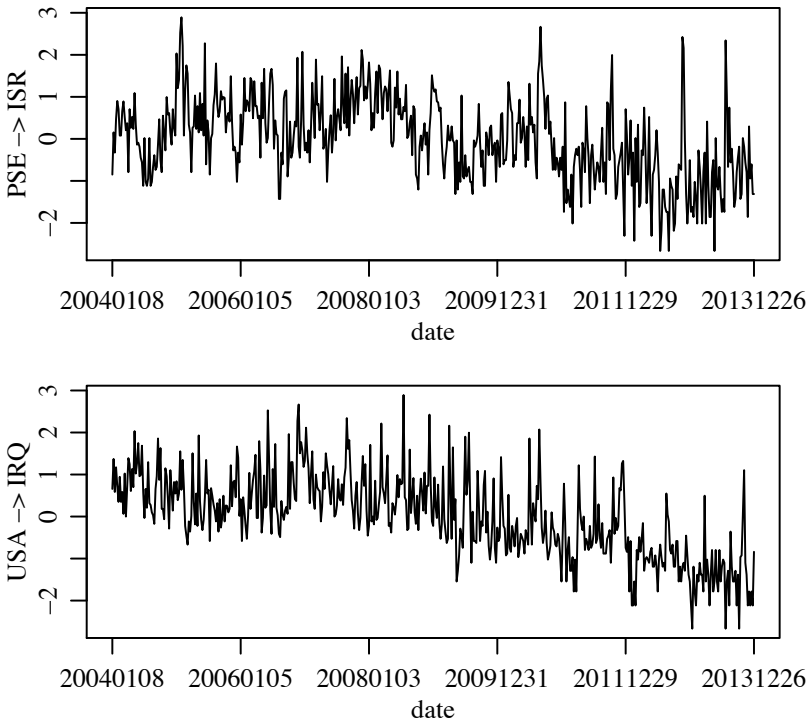
Outline

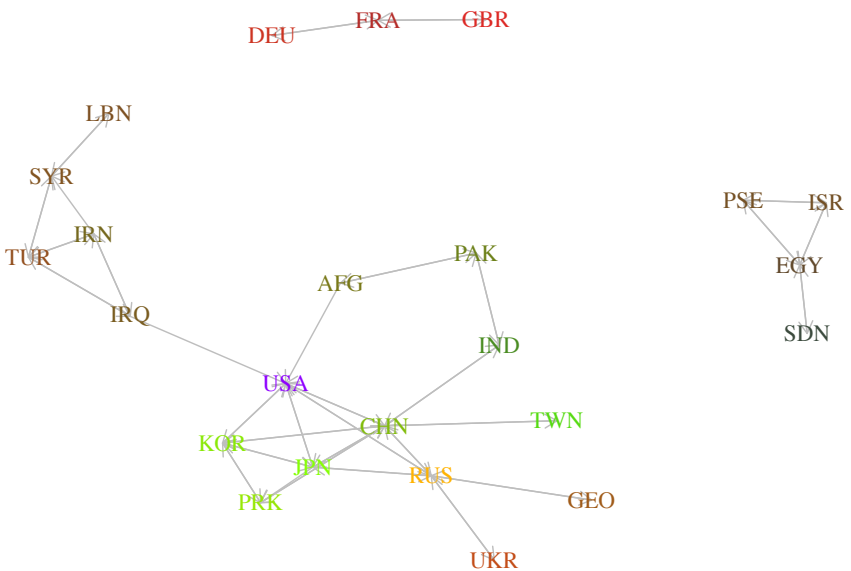
Bilinear regression for matrices

Multilinear regression for tensors

Sparse relational data

ICEWS data





Data: `www.lockheedmartin.com/us/products/W-ICEWS/iData.html`

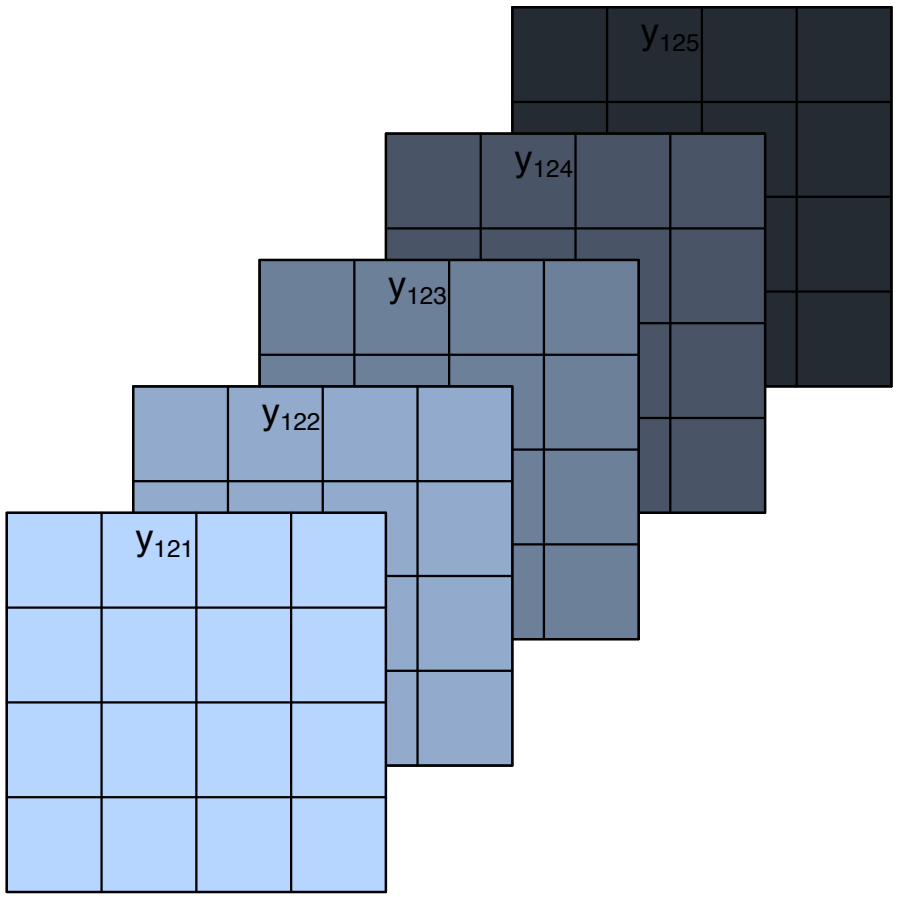
(Thanks to Mike Ward, Duke University, for data access and polysci consulting.)

Dynamic network data

A time series of sociomatrices:

$$\mathbf{Y} = \{\mathbf{Y}_t : t = 1, \ldots, T\}$$

$y_{i,j,t}$ = time-t relation from i to j.

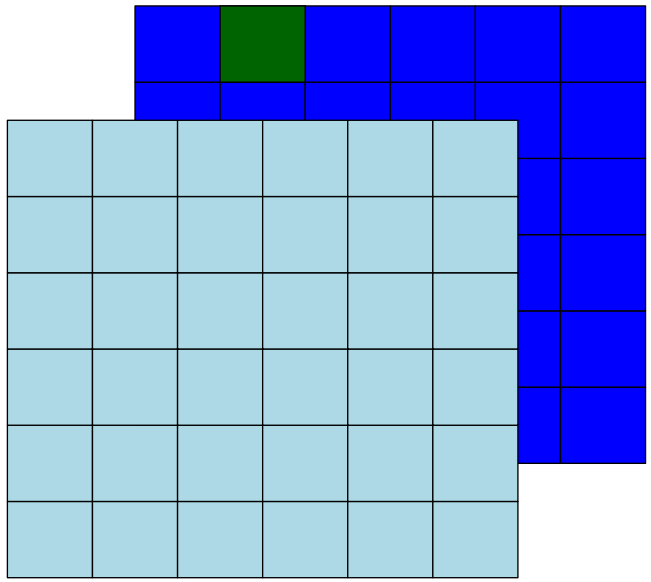


Temporal dependence

How does $y_{i,j,t}$ depend on $\mathbf{Y}_{t-1}$?

Maybe

- $y_{i,j,t-1}$?
- $y_{j,i,t-1}$?
- $y_{i,k,t-1}, y_{k,j,t-1}$?
- $y_{k,i,t-1}, y_{j,k,t-1}$?
- $y_{k,l,t-1}$?

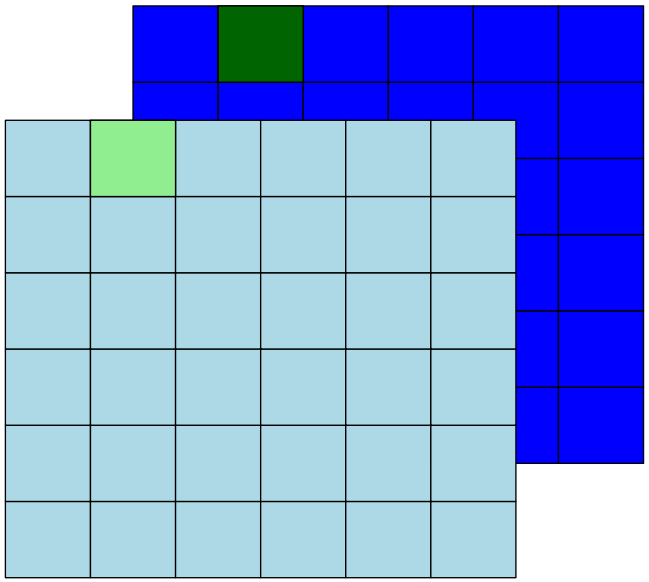


Temporal dependence

How does $y_{i,j,t}$ depend on $\mathbf{Y}_{t-1}$?

Maybe

- $y_{i,j,t-1}$?
- $y_{j,i,t-1}$?
- $y_{i,k,t-1}, y_{k,j,t-1}$?
- $y_{k,i,t-1}, y_{j,k,t-1}$?
- $y_{k,l,t-1}$?

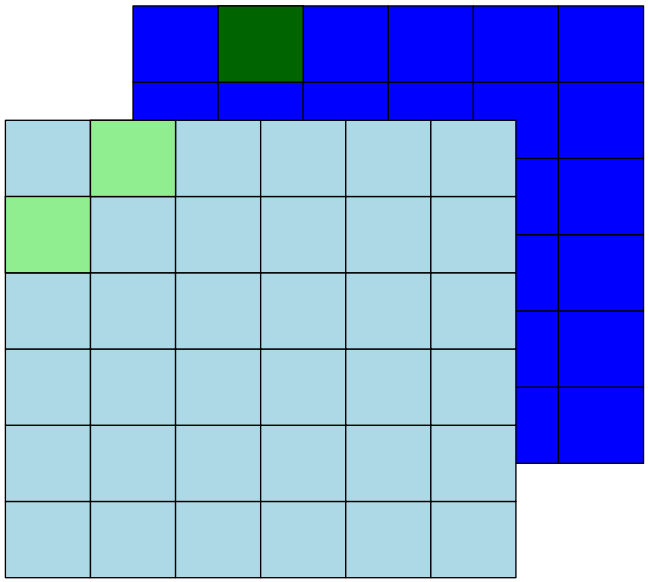


Temporal dependence

How does $y_{i,j,t}$ depend on $\mathbf{Y}_{t-1}$?

Maybe

- $y_{i,j,t-1}$?
- $y_{j,i,t-1}$?
- $y_{i,k,t-1}, y_{k,j,t-1}$?
- $y_{k,i,t-1}, y_{j,k,t-1}$?
- $y_{k,l,t-1}$?

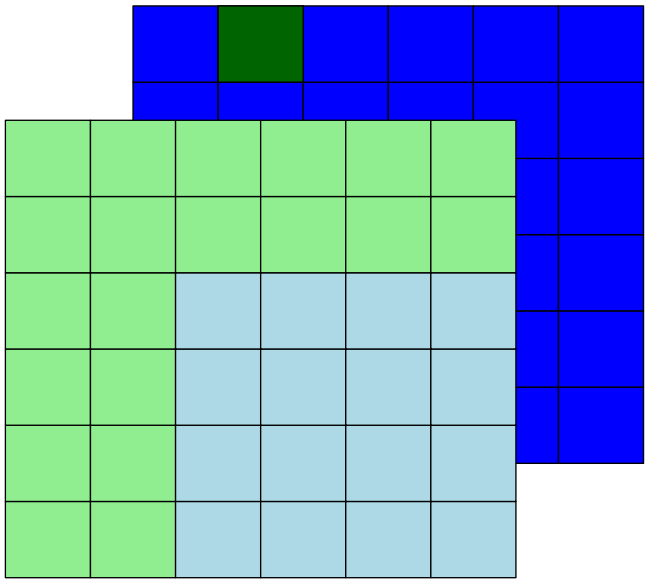


Temporal dependence

How does $y_{i,j,t}$ depend on $\mathbf{Y}_{t-1}$?

Maybe

- $y_{i,j,t-1}$?
- $y_{j,i,t-1}$?
- $y_{i,k,t-1}, y_{k,j,t-1}$?
- $y_{k,i,t-1}, y_{j,k,t-1}$?
- $y_{k,l,t-1}$?

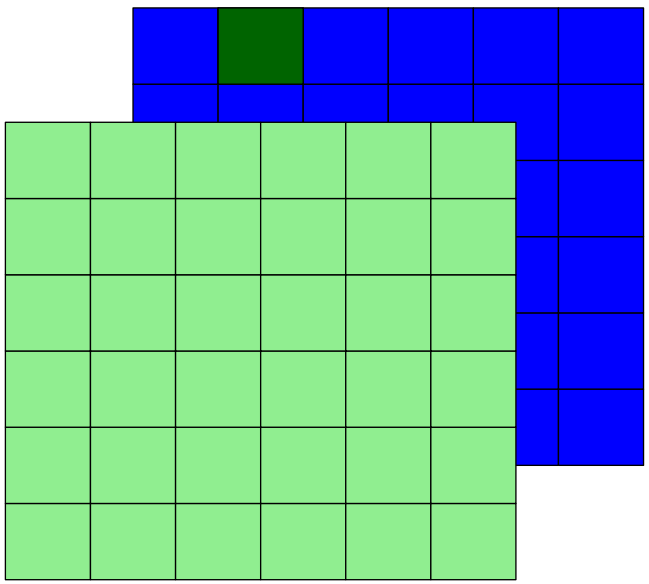


Temporal dependence

How does $y_{i,j,t}$ depend on $\mathbf{Y}_{t-1}$?

Maybe

- $y_{i,j,t-1}$?
- $y_{j,i,t-1}$?
- $y_{i,k,t-1}, y_{k,j,t-1}$?
- $y_{k,i,t-1}, y_{j,k,t-1}$?
- $y_{k,l,t-1}$?

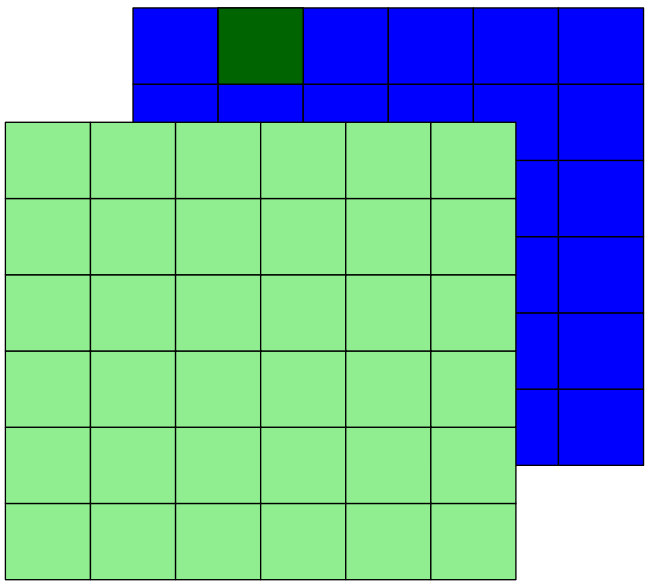


Temporal dependence

How does $y_{i,j,t}$ depend on $\mathbf{Y}_{t-1}$?

Maybe

- $y_{i,j,t-1}$?
- $y_{j,i,t-1}$?
- $y_{i,k,t-1}, y_{k,j,t-1}$?
- $y_{k,i,t-1}, y_{j,k,t-1}$?
- $y_{k,l,t-1}$?

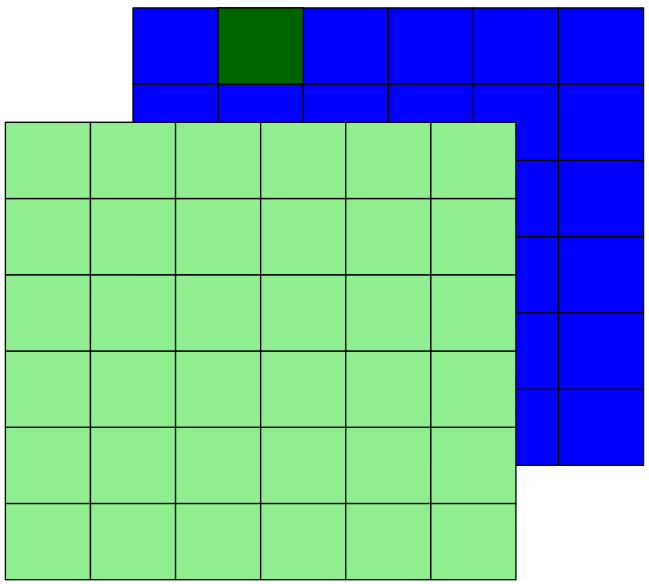


Temporal dependence

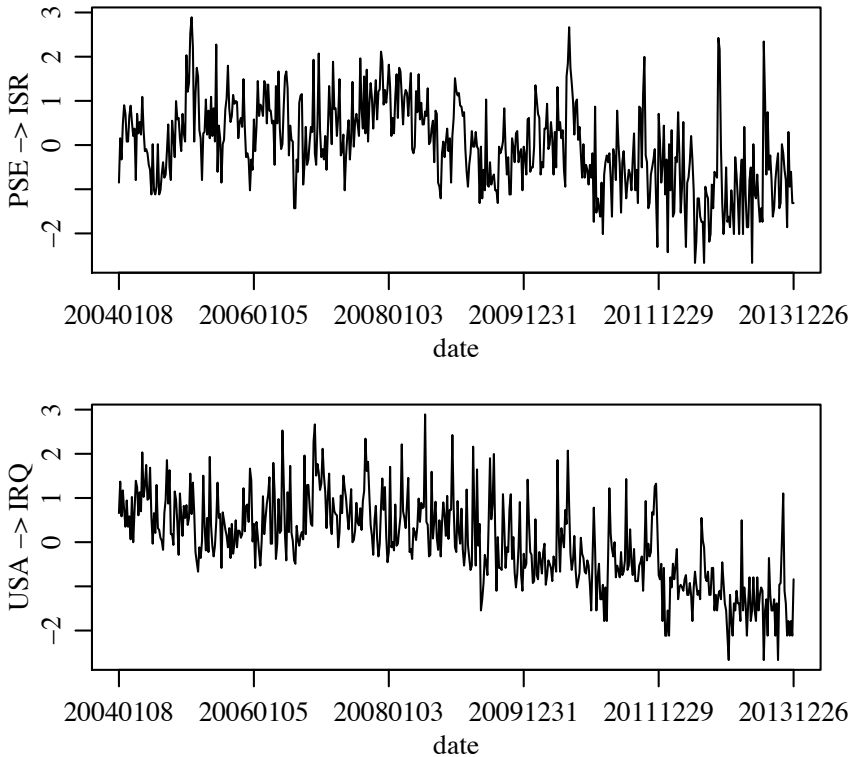
How does $y_{i,j,t}$ depend on $\mathbf{Y}_{t-1}$?

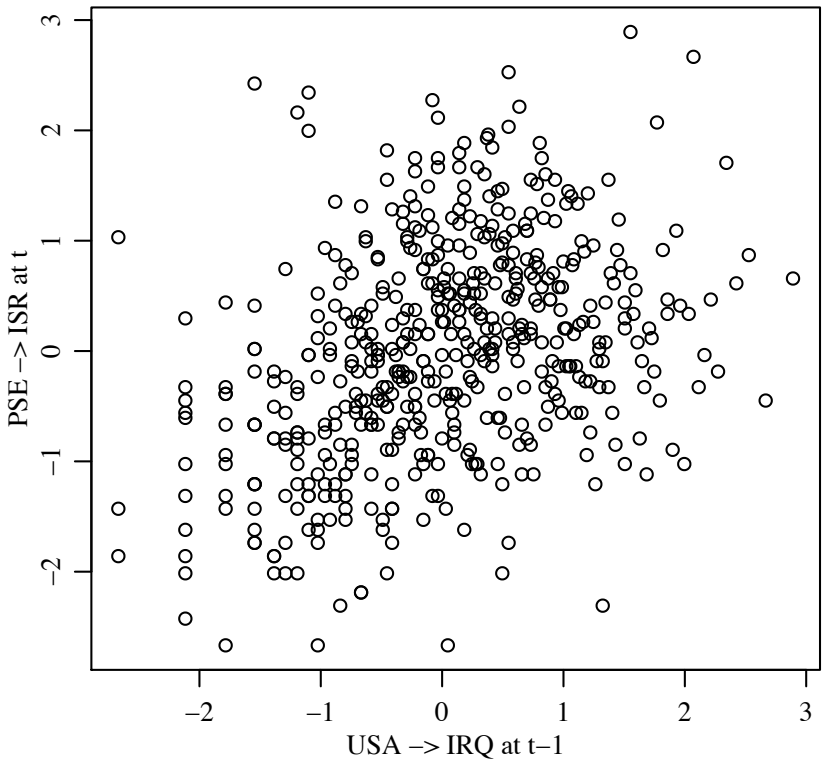
Maybe

- $y_{i,j,t-1}$?
- $y_{j,i,t-1}$?
- $y_{i,k,t-1}, y_{k,j,t-1}$?
- $y_{k,i,t-1}, y_{j,k,t-1}$?
- $y_{k,l,t-1}$?



Across-dyad dependence





$y_{USA,IRQ,t-1}$ is predictive of $y_{PSE,ISR,t}$.

First order VAR model

Let

- $\mathbf{y}_t = \text{vec}(\mathbf{Y}_t)$
- $\mathbf{x}_t = \text{vec}(\mathbf{X}_t)$

VAR: A first-order VAR model posits that

$$\mathbf{y}_t = \mathbf{\Theta}\mathbf{x}_t + \mathbf{e}_t \ , \quad \mathrm{E}[\mathbf{e}_t] = \mathbf{0} \ , \quad \mathrm{E}[\mathbf{e}_t \mathbf{e}_s^T] = \begin{cases} \Sigma & \text{if } t = s \\ \mathbf{0} & \text{if } t \neq s, \end{cases}$$

where $\mathbf{\Theta}$ and Σ are parameters to be estimated.

Is it possible to estimate an effect of each entry of $\mathbf{x}_t$ on each entry of $\mathbf{y}_t$?

First order VAR model

Let

- $\mathbf{y}_t = \text{vec}(\mathbf{Y}_t)$
- $\mathbf{x}_t = \text{vec}(\mathbf{X}_t)$

VAR: A first-order VAR model posits that

$$\mathbf{y}_t = \mathbf{\Theta}\mathbf{x}_t + \mathbf{e}_t \ , \quad \mathrm{E}[\mathbf{e}_t] = \mathbf{0} \ , \quad \mathrm{E}[\mathbf{e}_t \mathbf{e}_s^T] = \begin{cases} \Sigma & \text{if } t = s \\ \mathbf{0} & \text{if } t \neq s, \end{cases}$$

where $\mathbf{\Theta}$ and Σ are parameters to be estimated.

Is it possible to estimate an effect of each entry of $\mathbf{x}_t$ on each entry of $\mathbf{y}_t$?

First order VAR model

Let

- $\mathbf{y}_t = \text{vec}(\mathbf{Y}_t)$
- $\mathbf{x}_t = \text{vec}(\mathbf{X}_t)$

VAR: A first-order VAR model posits that

$$\mathbf{y}_t = \mathbf{\Theta}\mathbf{x}_t + \mathbf{e}_t \ , \quad \mathrm{E}[\mathbf{e}_t] = \mathbf{0} \ , \quad \mathrm{E}[\mathbf{e}_t \mathbf{e}_s^T] = \begin{cases} \Sigma & \text{if } t = s \\ \mathbf{0} & \text{if } t \neq s, \end{cases}$$

where $\mathbf{\Theta}$ and Σ are parameters to be estimated.

Is it possible to estimate an effect of each entry of $\mathbf{x}_t$ on each entry of $\mathbf{y}_t$?

Débauche d'indices

$m = 25$, so $\mathbf{\Theta}$ has $m^2 \times m^2 = m^4 = 390,625$ entries (Σ has half as many)

OLS estimation: The OLS estimator is

$$\begin{aligned}\hat{\Theta} &= \left(\sum_{t=1}^{n} \boldsymbol{y}_t \boldsymbol{x}_t^T\right)\left(\sum_{t=1}^{n} \boldsymbol{x}_t \boldsymbol{x}_t^T\right)^{-1} \\ &= \mathbf{S}_{yx}\mathbf{S}_{xx}^{-1}\end{aligned}$$

We need this $m^2 \times m^2$ matrix $\mathbf{S}_{xx}$ to be non-singular.

Problem: 10 years of weekly data gives $n = 543$, whereas $m^2 = 625$.

One solution: simplify the model.

Débauche d'indices

$m = 25$, so $\mathbf{\Theta}$ has $m^2 \times m^2 = m^4 = 390,625$ entries (Σ has half as many)

OLS estimation: The OLS estimator is

$$\begin{aligned}\hat{\mathbf{\Theta}} &= \left(\sum_{t=1}^{n} \boldsymbol{y}_t \boldsymbol{x}_t^T\right)\left(\sum_{t=1}^{n} \boldsymbol{x}_t \boldsymbol{x}_t^T\right)^{-1} \\ &= \mathbf{S}_{yx}\mathbf{S}_{xx}^{-1}\end{aligned}$$

We need this $m^2 \times m^2$ matrix $\mathbf{S}_{xx}$ to be non-singular.

Problem: 10 years of weekly data gives $n = 543$, whereas $m^2 = 625$.

One solution: simplify the model.

Débauche d'indices

$m = 25$, so $\mathbf{\Theta}$ has $m^2 \times m^2 = m^4 = 390,625$ entries (Σ has half as many)

OLS estimation: The OLS estimator is

$$\begin{aligned}\hat{\mathbf{\Theta}} &= \left(\sum_{t=1}^{n} \boldsymbol{y}_t \boldsymbol{x}_t^T\right)\left(\sum_{t=1}^{n} \boldsymbol{x}_t \boldsymbol{x}_t^T\right)^{-1} \\ &= \mathbf{S}_{yx}\mathbf{S}_{xx}^{-1}\end{aligned}$$

We need this $m^2 \times m^2$ matrix $\mathbf{S}_{xx}$ to be non-singular.

Problem: 10 years of weekly data gives $n = 543$, whereas $m^2 = 625$.

One solution: simplify the model.

Débauche d'indices

$m = 25$, so $\mathbf{\Theta}$ has $m^2 \times m^2 = m^4 = 390,625$ entries (Σ has half as many)

OLS estimation: The OLS estimator is

$$\begin{aligned}\hat{\mathbf{\Theta}} &= \left(\sum_{t=1}^{n} \boldsymbol{y}_t \boldsymbol{x}_t^T\right)\left(\sum_{t=1}^{n} \boldsymbol{x}_t \boldsymbol{x}_t^T\right)^{-1} \\ &= \mathbf{S}_{yx}\mathbf{S}_{xx}^{-1}\end{aligned}$$

We need this $m^2 \times m^2$ matrix $\mathbf{S}_{xx}$ to be non-singular.

Problem: 10 years of weekly data gives $n = 543$, whereas $m^2 = 625$.

One solution: simplify the model.

Débauche d'indices

$m = 25$, so $\mathbf{\Theta}$ has $m^2 \times m^2 = m^4 = 390,625$ entries (Σ has half as many)

OLS estimation: The OLS estimator is

$$\begin{aligned}\hat{\mathbf{\Theta}} &= \left(\sum_{t=1}^{n} \boldsymbol{y}_t \boldsymbol{x}_t^T\right)\left(\sum_{t=1}^{n} \boldsymbol{x}_t \boldsymbol{x}_t^T\right)^{-1} \\ &= \mathbf{S}_{yx}\mathbf{S}_{xx}^{-1}\end{aligned}$$

We need this $m^2 \times m^2$ matrix $\mathbf{S}_{xx}$ to be non-singular.

Problem: 10 years of weekly data gives $n = 543$, whereas $m^2 = 625$.

One solution: simplify the model.

Débauche d'indices

$m = 25$, so $\mathbf{\Theta}$ has $m^2 \times m^2 = m^4 = 390,625$ entries (Σ has half as many)

OLS estimation: The OLS estimator is

$$\begin{aligned}\hat{\mathbf{\Theta}} &= \left(\sum_{t=1}^{n} \boldsymbol{y}_t \boldsymbol{x}_t^T\right)\left(\sum_{t=1}^{n} \boldsymbol{x}_t \boldsymbol{x}_t^T\right)^{-1} \\ &= \mathbf{S}_{yx}\mathbf{S}_{xx}^{-1}\end{aligned}$$

We need this $m^2 \times m^2$ matrix $\mathbf{S}_{xx}$ to be non-singular.

Problem: 10 years of weekly data gives $n = 543$, whereas $m^2 = 625$.

One solution: simplify the model.

Bilinear regression

Q: Do actions of i' predict those of i?

Q: Do actions towards j' predict those towards j?

A multiplicative effects model:

$$y_{i,j,t} = \sum_{i'} \sum_{j'} a_{i,i'} b_{j,j'} x_{i',j',t} + \epsilon_{i,j,t}$$

$$\mathbf{Y}_t = \mathbf{A}\mathbf{X}_t\mathbf{B}^T + \mathbf{E}_t$$

$$\mathbf{y}_t = (\mathbf{B} \otimes \mathbf{A})\,\mathbf{x}_t + \mathbf{e}_t$$

(Not to be confused with the "growth curve" model (Potthoff and Roy, 1964; Gabriel, 1998) or the models of Basu et al. (2012); Shi et al. (2014).)

Bilinear regression

Q: Do actions of i' predict those of i?

Q: Do actions towards j' predict those towards j?

A multiplicative effects model:

$$y_{i,j,t} = \sum_{i'} \sum_{j'} a_{i,i'} b_{j,j'} x_{i',j',t} + \epsilon_{i,j,t}$$

$$\mathbf{Y}_t = \mathbf{A}\mathbf{X}_t\mathbf{B}^T + \mathbf{E}_t$$

$$\mathbf{y}_t = (\mathbf{B} \otimes \mathbf{A})\,\mathbf{x}_t + \mathbf{e}_t$$

(Not to be confused with the "growth curve" model (Potthoff and Roy, 1964; Gabriel, 1998) or the models of Basu et al. (2012); Shi et al. (2014).)

Bilinear regression

Q: Do actions of i' predict those of i?

Q: Do actions towards j' predict those towards j?

A multiplicative effects model:

$$y_{i,j,t} = \sum_{i'} \sum_{j'} a_{i,i'} b_{j,j'} x_{i',j',t} + \epsilon_{i,j,t}$$

$$\mathbf{Y}_t = \mathbf{A}\mathbf{X}_t\mathbf{B}^T + \mathbf{E}_t$$

$$\mathbf{y}_t = (\mathbf{B} \otimes \mathbf{A})\,\mathbf{x}_t + \mathbf{e}_t$$

(Not to be confused with the "growth curve" model (Potthoff and Roy, 1964; Gabriel, 1998) or the models of Basu et al. (2012); Shi et al. (2014).)

Bilinear regression

Q: Do actions of i' predict those of i?

Q: Do actions towards j' predict those towards j?

A multiplicative effects model:

$$y_{i,j,t} = \sum_{i'} \sum_{j'} a_{i,i'} b_{j,j'} x_{i',j',t} + \epsilon_{i,j,t}$$

$$\mathbf{Y}_t = \mathbf{A}\mathbf{X}_t\mathbf{B}^T + \mathbf{E}_t$$

$$\mathbf{y}_t = (\mathbf{B} \otimes \mathbf{A})\,\mathbf{x}_t + \mathbf{e}_t$$

(Not to be confused with the "growth curve" model (Potthoff and Roy, 1964; Gabriel, 1998) or the models of Basu et al. (2012); Shi et al. (2014).)

OLS estimation

$$(\hat{\mathbf{A}}, \hat{\mathbf{B}}) = \arg\min_{\mathbf{A},\mathbf{B}} \sum_i ||\mathbf{Y}_i - \mathbf{A}\mathbf{X}_i\mathbf{B}^T||^2/n$$

For $\mathbf{B} \neq \mathbf{0}$, the minimizer of the residual mean squared error in $\mathbf{A}$ is given by

$$\tilde{\mathbf{A}}(\mathbf{B}) = \left(\sum \mathbf{Y}_i\mathbf{B}\mathbf{X}_i^T\right)\left(\sum \mathbf{X}_i\mathbf{B}^T\mathbf{B}\mathbf{X}_i^T\right)^{-1}.$$

Similarly, for $\mathbf{A} \neq \mathbf{0}$,

$$\tilde{\mathbf{B}}(\mathbf{A}) = \left(\sum \mathbf{Y}_i^T\mathbf{A}\mathbf{X}_i\right)\left(\sum \mathbf{X}_i^T\mathbf{A}^T\mathbf{A}\mathbf{X}_i\right)^{-1}.$$

ALS/block coordinate descent:

Given $\mathbf{B}^{(0)}$, iterate until convergence:

1. $\mathbf{A}^{(s+1)} = \tilde{\mathbf{A}}(\mathbf{B}^{(s)})$
2. $\mathbf{B}^{(s+1)} = \tilde{\mathbf{B}}(\mathbf{A}^{(s+1)})$

OLS estimation

$$(\hat{\mathbf{A}}, \hat{\mathbf{B}}) = \arg\min_{\mathbf{A},\mathbf{B}} \sum_i ||\mathbf{Y}_i - \mathbf{A}\mathbf{X}_i\mathbf{B}^T||^2/n$$

For $\mathbf{B} \neq \mathbf{0}$, the minimizer of the residual mean squared error in $\mathbf{A}$ is given by

$$\tilde{\mathbf{A}}(\mathbf{B}) = \left(\sum \mathbf{Y}_i\mathbf{B}\mathbf{X}_i^T\right)\left(\sum \mathbf{X}_i\mathbf{B}^T\mathbf{B}\mathbf{X}_i^T\right)^{-1}.$$

Similarly, for $\mathbf{A} \neq \mathbf{0}$,

$$\tilde{\mathbf{B}}(\mathbf{A}) = \left(\sum \mathbf{Y}_i^T\mathbf{A}\mathbf{X}_i\right)\left(\sum \mathbf{X}_i^T\mathbf{A}^T\mathbf{A}\mathbf{X}_i\right)^{-1}.$$

ALS/block coordinate descent:

Given $\mathbf{B}^{(0)}$, iterate until convergence:

1. $\mathbf{A}^{(s+1)} = \tilde{\mathbf{A}}(\mathbf{B}^{(s)})$
2. $\mathbf{B}^{(s+1)} = \tilde{\mathbf{B}}(\mathbf{A}^{(s+1)})$

OLS estimation

$$(\hat{\mathbf{A}}, \hat{\mathbf{B}}) = \arg\min_{\mathbf{A},\mathbf{B}} \sum_i ||\mathbf{Y}_i - \mathbf{A}\mathbf{X}_i\mathbf{B}^T||^2/n$$

For $\mathbf{B} \neq \mathbf{0}$, the minimizer of the residual mean squared error in $\mathbf{A}$ is given by

$$\tilde{\mathbf{A}}(\mathbf{B}) = \left(\sum \mathbf{Y}_i\mathbf{B}\mathbf{X}_i^T\right)\left(\sum \mathbf{X}_i\mathbf{B}^T\mathbf{B}\mathbf{X}_i^T\right)^{-1}.$$

Similarly, for $\mathbf{A} \neq \mathbf{0}$,

$$\tilde{\mathbf{B}}(\mathbf{A}) = \left(\sum \mathbf{Y}_i^T\mathbf{A}\mathbf{X}_i\right)\left(\sum \mathbf{X}_i^T\mathbf{A}^T\mathbf{A}\mathbf{X}_i\right)^{-1}.$$

ALS/block coordinate descent:

Given $\mathbf{B}^{(0)}$, iterate until convergence:

1. $\mathbf{A}^{(s+1)} = \tilde{\mathbf{A}}(\mathbf{B}^{(s)})$
2. $\mathbf{B}^{(s+1)} = \tilde{\mathbf{B}}(\mathbf{A}^{(s+1)})$

OLS estimation

$$(\hat{\mathbf{A}}, \hat{\mathbf{B}}) = \arg\min_{\mathbf{A},\mathbf{B}} \sum_i ||\mathbf{Y}_i - \mathbf{A}\mathbf{X}_i\mathbf{B}^T||^2/n$$

For $\mathbf{B} \neq \mathbf{0}$, the minimizer of the residual mean squared error in $\mathbf{A}$ is given by

$$\tilde{\mathbf{A}}(\mathbf{B}) = \left(\sum \mathbf{Y}_i\mathbf{B}\mathbf{X}_i^T\right)\left(\sum \mathbf{X}_i\mathbf{B}^T\mathbf{B}\mathbf{X}_i^T\right)^{-1}.$$

Similarly, for $\mathbf{A} \neq \mathbf{0}$,

$$\tilde{\mathbf{B}}(\mathbf{A}) = \left(\sum \mathbf{Y}_i^T\mathbf{A}\mathbf{X}_i\right)\left(\sum \mathbf{X}_i^T\mathbf{A}^T\mathbf{A}\mathbf{X}_i\right)^{-1}.$$

ALS/block coordinate descent:

Given $\mathbf{B}^{(0)}$, iterate until convergence:

1. $\mathbf{A}^{(s+1)} = \tilde{\mathbf{A}}(\mathbf{B}^{(s)})$
2. $\mathbf{B}^{(s+1)} = \tilde{\mathbf{B}}(\mathbf{A}^{(s+1)})$

OLS estimation

$$(\hat{\mathbf{A}}, \hat{\mathbf{B}}) = \arg\min_{\mathbf{A},\mathbf{B}} \sum_i ||\mathbf{Y}_i - \mathbf{A}\mathbf{X}_i\mathbf{B}^T||^2/n$$

For $\mathbf{B} \neq \mathbf{0}$, the minimizer of the residual mean squared error in $\mathbf{A}$ is given by

$$\tilde{\mathbf{A}}(\mathbf{B}) = \left(\sum \mathbf{Y}_i\mathbf{B}\mathbf{X}_i^T\right)\left(\sum \mathbf{X}_i\mathbf{B}^T\mathbf{B}\mathbf{X}_i^T\right)^{-1}.$$

Similarly, for $\mathbf{A} \neq \mathbf{0}$,

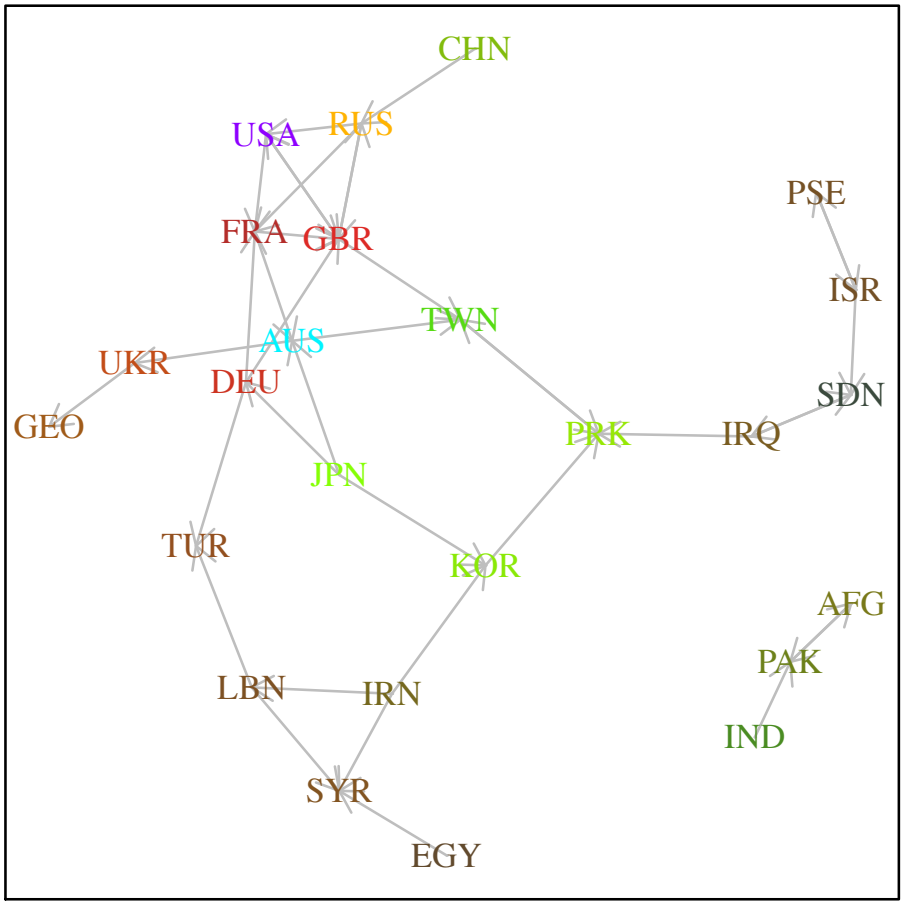
$$\tilde{\mathbf{B}}(\mathbf{A}) = \left(\sum \mathbf{Y}_i^T\mathbf{A}\mathbf{X}_i\right)\left(\sum \mathbf{X}_i^T\mathbf{A}^T\mathbf{A}\mathbf{X}_i\right)^{-1}.$$

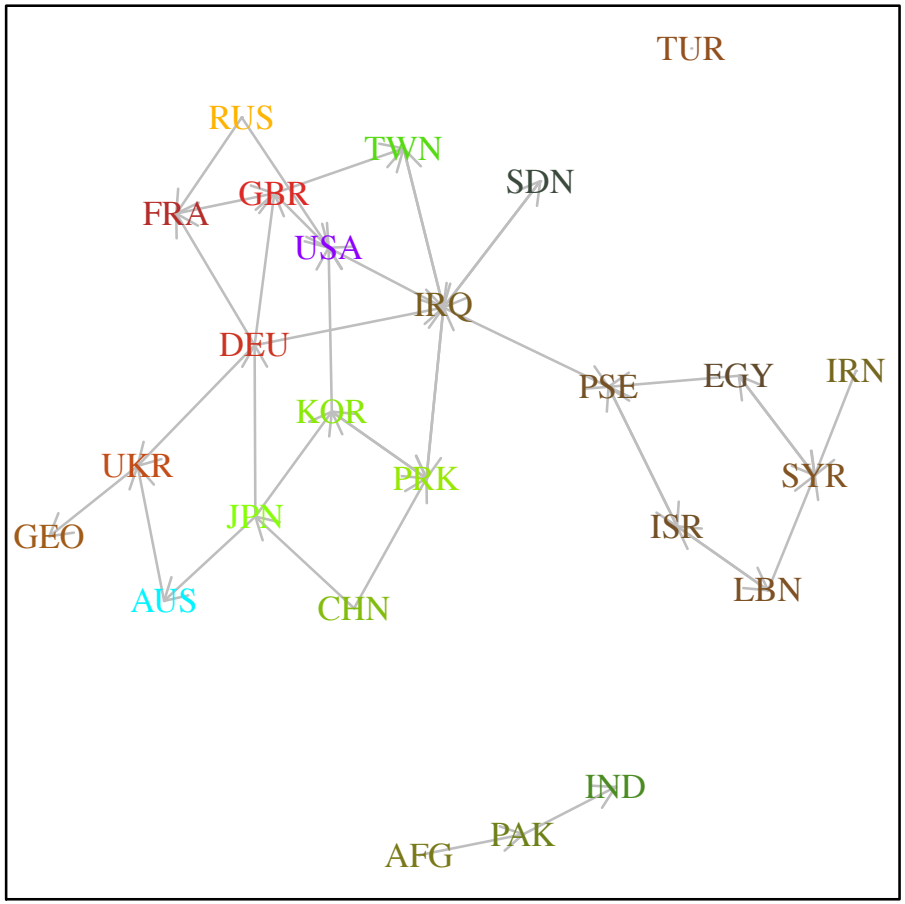
ALS/block coordinate descent:

Given $\mathbf{B}^{(0)}$, iterate until convergence:

1. $\mathbf{A}^{(s+1)} = \tilde{\mathbf{A}}(\mathbf{B}^{(s)})$
2. $\mathbf{B}^{(s+1)} = \tilde{\mathbf{B}}(\mathbf{A}^{(s+1)})$

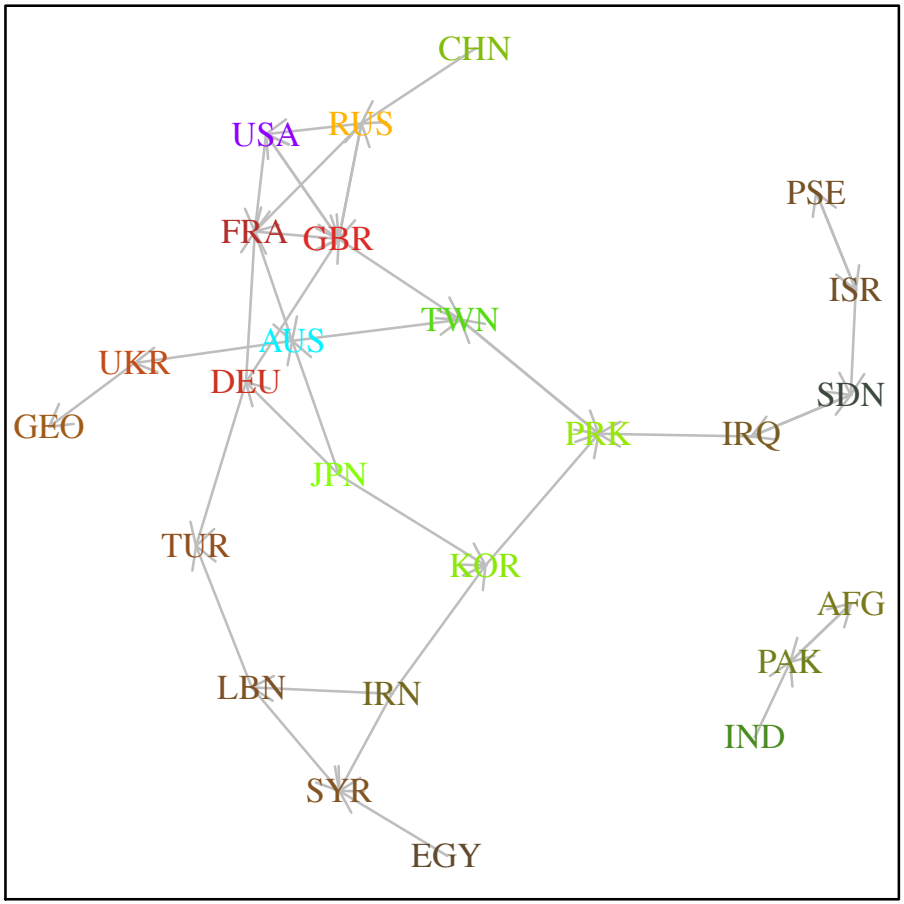
Least squares parameter estimates

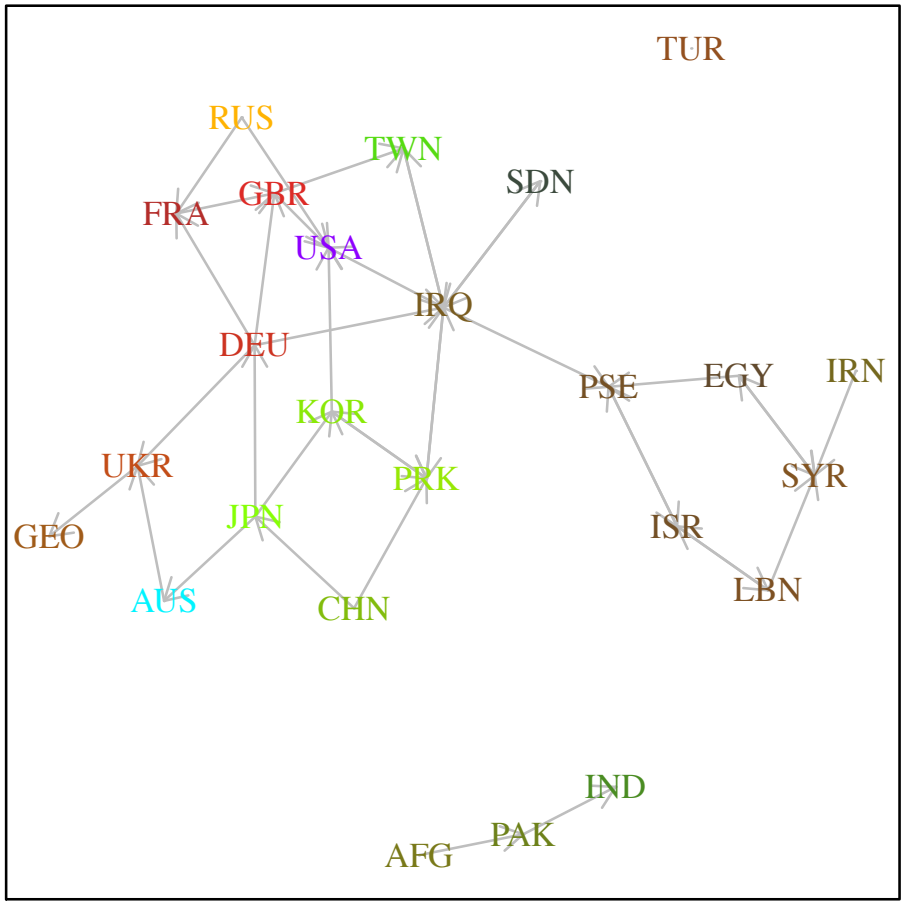




e.g., $y_{FRA,SYR,t-1}$ is predictive of $y_{GBR,IRN,t}$

Least squares parameter estimates





e.g., $y_{FRA,SYR,t-1}$ is predictive of $y_{GBR,IRN,t}$

What are the estimators estimating?

Using the vectorized representation,

$$(\hat{\mathbf{A}}, \hat{\mathbf{B}}) = \arg\min_{\mathbf{A},\mathbf{B}} \sum_i ||\mathbf{y}_i - (\mathbf{B} \otimes \mathbf{A})\mathbf{x}_i||^2 / n$$

$$= \arg\min_{\mathbf{A},\mathbf{B}} \operatorname{tr}\left((\mathbf{B}^T\mathbf{B} \otimes \mathbf{A}^T\mathbf{A}) \sum \mathbf{x}_i\mathbf{x}_i^T / n\right) - 2\operatorname{tr}\left((\mathbf{B} \otimes \mathbf{A}) \sum \mathbf{x}_i\mathbf{y}_i^T / n\right)$$

$$= \arg\min_{\mathbf{A},\mathbf{B}} \operatorname{tr}((\mathbf{B}^T\mathbf{B} \otimes \mathbf{A}^T\mathbf{A})\mathbf{S}_{xx}) - 2\operatorname{tr}((\mathbf{B} \otimes \mathbf{A})\mathbf{S}_{xy})$$

$$\approx \arg\min_{\mathbf{A},\mathbf{B}} \operatorname{tr}((\mathbf{B}^T\mathbf{B} \otimes \mathbf{A}^T\mathbf{A})\Sigma_{xx}) - 2\operatorname{tr}((\mathbf{B} \otimes \mathbf{A})\Sigma_{xy}) = (\tilde{\mathbf{A}}, \tilde{\mathbf{B}})$$

$(\tilde{\mathbf{A}}, \tilde{\mathbf{B}})$ are the *pseudotrue values* of the parameters.

$(\hat{\mathbf{A}}, \hat{\mathbf{B}})$ are estimating $(\tilde{\mathbf{A}}, \tilde{\mathbf{B}})$, roughly speaking.

What are the estimators estimating?

Using the vectorized representation,

$$(\hat{\mathbf{A}}, \hat{\mathbf{B}}) = \arg\min_{\mathbf{A},\mathbf{B}} \sum_i ||\mathbf{y}_i - (\mathbf{B} \otimes \mathbf{A})\mathbf{x}_i||^2 / n$$

$$= \arg\min_{\mathbf{A},\mathbf{B}} \operatorname{tr}\left((\mathbf{B}^T\mathbf{B} \otimes \mathbf{A}^T\mathbf{A}) \sum \mathbf{x}_i\mathbf{x}_i^T / n\right) - 2\operatorname{tr}\left((\mathbf{B} \otimes \mathbf{A}) \sum \mathbf{x}_i\mathbf{y}_i^T / n\right)$$

$$= \arg\min_{\mathbf{A},\mathbf{B}} \operatorname{tr}((\mathbf{B}^T\mathbf{B} \otimes \mathbf{A}^T\mathbf{A})\mathbf{S}_{xx}) - 2\operatorname{tr}((\mathbf{B} \otimes \mathbf{A})\mathbf{S}_{xy})$$

$$\approx \arg\min_{\mathbf{A},\mathbf{B}} \operatorname{tr}((\mathbf{B}^T\mathbf{B} \otimes \mathbf{A}^T\mathbf{A})\Sigma_{xx}) - 2\operatorname{tr}((\mathbf{B} \otimes \mathbf{A})\Sigma_{xy}) = (\tilde{\mathbf{A}}, \tilde{\mathbf{B}})$$

$(\tilde{\mathbf{A}}, \tilde{\mathbf{B}})$ are the *pseudotrue values* of the parameters.

$(\hat{\mathbf{A}}, \hat{\mathbf{B}})$ are estimating $(\tilde{\mathbf{A}}, \tilde{\mathbf{B}})$, roughly speaking.

What are the estimators estimating?

Using the vectorized representation,

$$
\begin{aligned}
(\hat{\mathbf{A}}, \hat{\mathbf{B}}) &= \arg\min_{\mathbf{A},\mathbf{B}} \sum_i ||\mathbf{y}_i - (\mathbf{B} \otimes \mathbf{A})\mathbf{x}_i||^2 / n \\
&= \arg\min_{\mathbf{A},\mathbf{B}} \operatorname{tr}\left((\mathbf{B}^T\mathbf{B} \otimes \mathbf{A}^T\mathbf{A}) \sum \mathbf{x}_i\mathbf{x}_i^T / n\right) - 2\operatorname{tr}\left((\mathbf{B} \otimes \mathbf{A}) \sum \mathbf{x}_i\mathbf{y}_i^T / n\right) \\
&= \arg\min_{\mathbf{A},\mathbf{B}} \operatorname{tr}((\mathbf{B}^T\mathbf{B} \otimes \mathbf{A}^T\mathbf{A})\mathbf{S}_{xx}) - 2\operatorname{tr}((\mathbf{B} \otimes \mathbf{A})\mathbf{S}_{xy}) \\
&\approx \arg\min_{\mathbf{A},\mathbf{B}} \operatorname{tr}((\mathbf{B}^T\mathbf{B} \otimes \mathbf{A}^T\mathbf{A})\Sigma_{xx}) - 2\operatorname{tr}((\mathbf{B} \otimes \mathbf{A})\Sigma_{xy}) = (\tilde{\mathbf{A}}, \tilde{\mathbf{B}})
\end{aligned}
$$

$(\tilde{\mathbf{A}}, \tilde{\mathbf{B}})$ are the *pseudotrue values* of the parameters.

$(\hat{\mathbf{A}}, \hat{\mathbf{B}})$ are estimating $(\tilde{\mathbf{A}}, \tilde{\mathbf{B}})$, roughly speaking.

What are the estimators estimating?

Using the vectorized representation,

$$
\begin{aligned}
(\hat{\mathbf{A}}, \hat{\mathbf{B}}) &= \arg\min_{\mathbf{A},\mathbf{B}} \sum_i ||\mathbf{y}_i - (\mathbf{B} \otimes \mathbf{A})\mathbf{x}_i||^2 / n \\
&= \arg\min_{\mathbf{A},\mathbf{B}} \operatorname{tr}\left((\mathbf{B}^T\mathbf{B} \otimes \mathbf{A}^T\mathbf{A}) \sum \mathbf{x}_i\mathbf{x}_i^T / n\right) - 2\operatorname{tr}\left((\mathbf{B} \otimes \mathbf{A}) \sum \mathbf{x}_i\mathbf{y}_i^T / n\right) \\
&= \arg\min_{\mathbf{A},\mathbf{B}} \operatorname{tr}((\mathbf{B}^T\mathbf{B} \otimes \mathbf{A}^T\mathbf{A})\mathbf{S}_{xx}) - 2\operatorname{tr}((\mathbf{B} \otimes \mathbf{A})\mathbf{S}_{xy}) \\
&\approx \arg\min_{\mathbf{A},\mathbf{B}} \operatorname{tr}((\mathbf{B}^T\mathbf{B} \otimes \mathbf{A}^T\mathbf{A})\Sigma_{xx}) - 2\operatorname{tr}((\mathbf{B} \otimes \mathbf{A})\Sigma_{xy}) = (\tilde{\mathbf{A}}, \tilde{\mathbf{B}})
\end{aligned}
$$

$(\tilde{\mathbf{A}}, \tilde{\mathbf{B}})$ are the *pseudotrue values* of the parameters.

$(\hat{\mathbf{A}}, \hat{\mathbf{B}})$ are estimating $(\tilde{\mathbf{A}}, \tilde{\mathbf{B}})$, roughly speaking.

What are the estimators estimating?

Using the vectorized representation,

$$
\begin{aligned}
(\hat{\mathbf{A}}, \hat{\mathbf{B}}) &= \arg\min_{\mathbf{A},\mathbf{B}} \sum_i ||\mathbf{y}_i - (\mathbf{B} \otimes \mathbf{A})\mathbf{x}_i||^2 / n \\
&= \arg\min_{\mathbf{A},\mathbf{B}} \operatorname{tr}\left((\mathbf{B}^T\mathbf{B} \otimes \mathbf{A}^T\mathbf{A}) \sum \mathbf{x}_i\mathbf{x}_i^T / n\right) - 2\operatorname{tr}\left((\mathbf{B} \otimes \mathbf{A}) \sum \mathbf{x}_i\mathbf{y}_i^T / n\right) \\
&= \arg\min_{\mathbf{A},\mathbf{B}} \operatorname{tr}((\mathbf{B}^T\mathbf{B} \otimes \mathbf{A}^T\mathbf{A})\mathbf{S}_{xx}) - 2\operatorname{tr}((\mathbf{B} \otimes \mathbf{A})\mathbf{S}_{xy}) \\
&\approx \arg\min_{\mathbf{A},\mathbf{B}} \operatorname{tr}((\mathbf{B}^T\mathbf{B} \otimes \mathbf{A}^T\mathbf{A})\Sigma_{xx}) - 2\operatorname{tr}((\mathbf{B} \otimes \mathbf{A})\Sigma_{xy}) = (\tilde{\mathbf{A}}, \tilde{\mathbf{B}})
\end{aligned}
$$

$(\tilde{\mathbf{A}}, \tilde{\mathbf{B}})$ are the *pseudotrue values* of the parameters.

$(\hat{\mathbf{A}}, \hat{\mathbf{B}})$ are estimating $(\tilde{\mathbf{A}}, \tilde{\mathbf{B}})$, roughly speaking.

What are the estimators estimating?

Using the vectorized representation,

$$
\begin{aligned}
(\hat{\mathbf{A}}, \hat{\mathbf{B}}) &= \arg\min_{\mathbf{A},\mathbf{B}} \sum_i ||\mathbf{y}_i - (\mathbf{B} \otimes \mathbf{A})\mathbf{x}_i||^2 / n \\
&= \arg\min_{\mathbf{A},\mathbf{B}} \operatorname{tr}\left((\mathbf{B}^T\mathbf{B} \otimes \mathbf{A}^T\mathbf{A}) \sum \mathbf{x}_i\mathbf{x}_i^T / n\right) - 2\operatorname{tr}\left((\mathbf{B} \otimes \mathbf{A}) \sum \mathbf{x}_i\mathbf{y}_i^T / n\right) \\
&= \arg\min_{\mathbf{A},\mathbf{B}} \operatorname{tr}((\mathbf{B}^T\mathbf{B} \otimes \mathbf{A}^T\mathbf{A})\mathbf{S}_{xx}) - 2\operatorname{tr}((\mathbf{B} \otimes \mathbf{A})\mathbf{S}_{xy}) \\
&\approx \arg\min_{\mathbf{A},\mathbf{B}} \operatorname{tr}((\mathbf{B}^T\mathbf{B} \otimes \mathbf{A}^T\mathbf{A})\Sigma_{xx}) - 2\operatorname{tr}((\mathbf{B} \otimes \mathbf{A})\Sigma_{xy}) = (\tilde{\mathbf{A}}, \tilde{\mathbf{B}})
\end{aligned}
$$

$(\tilde{\mathbf{A}}, \tilde{\mathbf{B}})$ are the *pseudotrue values* of the parameters.

$(\hat{\mathbf{A}}, \hat{\mathbf{B}})$ are estimating $(\tilde{\mathbf{A}}, \tilde{\mathbf{B}})$, roughly speaking.

What are the estimators estimating?

Using the vectorized representation,

$$
\begin{aligned}
(\hat{\mathbf{A}}, \hat{\mathbf{B}}) &= \arg\min_{\mathbf{A},\mathbf{B}} \sum_i ||\mathbf{y}_i - (\mathbf{B} \otimes \mathbf{A})\mathbf{x}_i||^2 / n \\
&= \arg\min_{\mathbf{A},\mathbf{B}} \operatorname{tr}\left((\mathbf{B}^T\mathbf{B} \otimes \mathbf{A}^T\mathbf{A}) \sum \mathbf{x}_i\mathbf{x}_i^T / n\right) - 2\operatorname{tr}\left((\mathbf{B} \otimes \mathbf{A}) \sum \mathbf{x}_i\mathbf{y}_i^T / n\right) \\
&= \arg\min_{\mathbf{A},\mathbf{B}} \operatorname{tr}((\mathbf{B}^T\mathbf{B} \otimes \mathbf{A}^T\mathbf{A})\mathbf{S}_{xx}) - 2\operatorname{tr}((\mathbf{B} \otimes \mathbf{A})\mathbf{S}_{xy}) \\
&\approx \arg\min_{\mathbf{A},\mathbf{B}} \operatorname{tr}((\mathbf{B}^T\mathbf{B} \otimes \mathbf{A}^T\mathbf{A})\Sigma_{xx}) - 2\operatorname{tr}((\mathbf{B} \otimes \mathbf{A})\Sigma_{xy}) = (\tilde{\mathbf{A}}, \tilde{\mathbf{B}})
\end{aligned}
$$

$(\tilde{\mathbf{A}}, \tilde{\mathbf{B}})$ are the *pseudotrue values* of the parameters.

$(\hat{\mathbf{A}}, \hat{\mathbf{B}})$ are estimating $(\tilde{\mathbf{A}}, \tilde{\mathbf{B}})$, roughly speaking.

What are the estimators estimating?

Using the vectorized representation,

$$
\begin{aligned}
(\hat{\mathbf{A}}, \hat{\mathbf{B}}) &= \arg\min_{\mathbf{A},\mathbf{B}} \sum_i ||\mathbf{y}_i - (\mathbf{B} \otimes \mathbf{A})\mathbf{x}_i||^2 / n \\
&= \arg\min_{\mathbf{A},\mathbf{B}} \operatorname{tr}\left((\mathbf{B}^T\mathbf{B} \otimes \mathbf{A}^T\mathbf{A}) \sum \mathbf{x}_i\mathbf{x}_i^T / n\right) - 2\operatorname{tr}\left((\mathbf{B} \otimes \mathbf{A}) \sum \mathbf{x}_i\mathbf{y}_i^T / n\right) \\
&= \arg\min_{\mathbf{A},\mathbf{B}} \operatorname{tr}((\mathbf{B}^T\mathbf{B} \otimes \mathbf{A}^T\mathbf{A})\mathbf{S}_{xx}) - 2\operatorname{tr}((\mathbf{B} \otimes \mathbf{A})\mathbf{S}_{xy}) \\
&\approx \arg\min_{\mathbf{A},\mathbf{B}} \operatorname{tr}((\mathbf{B}^T\mathbf{B} \otimes \mathbf{A}^T\mathbf{A})\Sigma_{xx}) - 2\operatorname{tr}((\mathbf{B} \otimes \mathbf{A})\Sigma_{xy}) = (\tilde{\mathbf{A}}, \tilde{\mathbf{B}})
\end{aligned}
$$

$(\tilde{\mathbf{A}}, \tilde{\mathbf{B}})$ are the *pseudotrue values* of the parameters.

$(\hat{\mathbf{A}}, \hat{\mathbf{B}})$ are estimating $(\tilde{\mathbf{A}}, \tilde{\mathbf{B}})$, roughly speaking.

What are the estimators estimating?

Using the vectorized representation,

$$
\begin{aligned}
(\hat{\mathbf{A}}, \hat{\mathbf{B}}) &= \arg\min_{\mathbf{A},\mathbf{B}} \sum_i ||\mathbf{y}_i - (\mathbf{B} \otimes \mathbf{A})\mathbf{x}_i||^2 / n \\
&= \arg\min_{\mathbf{A},\mathbf{B}} \operatorname{tr}\left((\mathbf{B}^T\mathbf{B} \otimes \mathbf{A}^T\mathbf{A}) \sum \mathbf{x}_i\mathbf{x}_i^T / n\right) - 2\operatorname{tr}\left((\mathbf{B} \otimes \mathbf{A}) \sum \mathbf{x}_i\mathbf{y}_i^T / n\right) \\
&= \arg\min_{\mathbf{A},\mathbf{B}} \operatorname{tr}((\mathbf{B}^T\mathbf{B} \otimes \mathbf{A}^T\mathbf{A})\mathbf{S}_{xx}) - 2\operatorname{tr}((\mathbf{B} \otimes \mathbf{A})\mathbf{S}_{xy}) \\
&\approx \arg\min_{\mathbf{A},\mathbf{B}} \operatorname{tr}((\mathbf{B}^T\mathbf{B} \otimes \mathbf{A}^T\mathbf{A})\Sigma_{xx}) - 2\operatorname{tr}((\mathbf{B} \otimes \mathbf{A})\Sigma_{xy}) = (\tilde{\mathbf{A}}, \tilde{\mathbf{B}})
\end{aligned}
$$

$(\tilde{\mathbf{A}}, \tilde{\mathbf{B}})$ are the *pseudotrue values* of the parameters.

$(\hat{\mathbf{A}}, \hat{\mathbf{B}})$ are estimating $(\tilde{\mathbf{A}}, \tilde{\mathbf{B}})$, roughly speaking.

What are the estimates estimating?

Correct mean model: If $\mathrm{E}[\mathbf{y}|\mathbf{x}] = (\mathbf{B}_0 \otimes \mathbf{A}_0)\mathbf{x}$, then

$$\tilde{\mathbf{B}} \otimes \tilde{\mathbf{A}} = \mathbf{B}_0 \otimes \mathbf{A}_0.$$

(A unique minimizer, assuming Σ_{xx} is positive definite.)

Incorrect mean model: If $\mathrm{E}[\mathbf{y}|\mathbf{x}] \neq (\mathbf{B}_0 \otimes \mathbf{A}_0)\mathbf{x}$, but $\mathbf{Y}_{[i,]}$, $\mathbf{X}_{[i',]}$ conditionally independent, then

$$\tilde{a}_{i,i'} = 0$$

(under some conditions on Σ_{xx}.)

What are the estimates estimating?

Correct mean model: If $\mathrm{E}[\mathbf{y}|\mathbf{x}] = (\mathbf{B}_0 \otimes \mathbf{A}_0)\mathbf{x}$, then

$$\tilde{\mathbf{B}} \otimes \tilde{\mathbf{A}} = \mathbf{B}_0 \otimes \mathbf{A}_0.$$

(A unique minimizer, assuming Σ_{xx} is positive definite.)

Incorrect mean model: If $\mathrm{E}[\mathbf{y}|\mathbf{x}] \neq (\mathbf{B}_0 \otimes \mathbf{A}_0)\mathbf{x}$, but $\mathbf{Y}_{[i,]}$, $\mathbf{X}_{[i',]}$ conditionally independent, then

$$\tilde{a}_{i,i'} = 0$$

(under some conditions on Σ_{xx}.)

Model comparison

Additive effects:

$$\mathrm{E}[\mathbf{Y}_t] = \mathbf{A}\mathbf{X}_t\mathbf{1}\mathbf{1}^T + \mathbf{1}\mathbf{1}^T\mathbf{X}_t\mathbf{B}^T$$
$$\mathrm{E}[\mathbf{y}_t] = (\mathbf{1}\mathbf{1}^T \otimes \mathbf{A} + \mathbf{B} \otimes \mathbf{1}\mathbf{1}^T)\,\mathbf{x}_t$$
$$\mathrm{E}[y_{i,j,t}] = \sum_{i'}\sum_{j'}(a_{i,i'} + b_{j,j'})x_{i',j',t}$$

Dyadic effects:

$$\mathrm{E}[\mathbf{Y}_t] = \mathbf{M} \circ \mathbf{X}_t$$
$$\mathrm{E}[\mathbf{y}_t] = \mathbf{D}(\mathbf{m})\,\mathbf{x}_t$$
$$\mathrm{E}[y_{i,j,t}] = m_{i,j}x_{i,j,t}$$

Model comparison

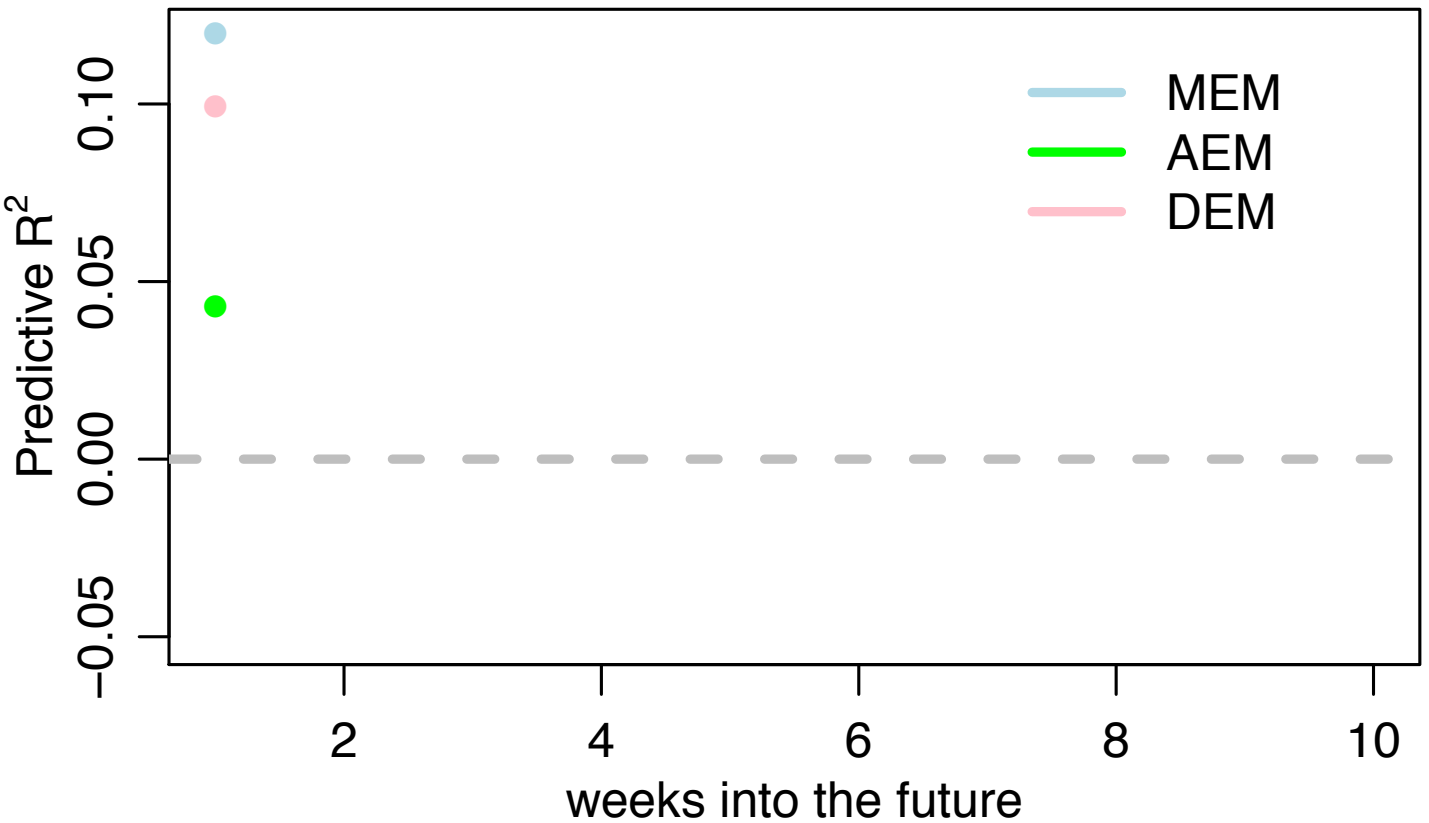
Additive effects:

$$\mathrm{E}[\mathbf{Y}_t] = \mathbf{A}\mathbf{X}_t\mathbf{1}\mathbf{1}^T + \mathbf{1}\mathbf{1}^T\mathbf{X}_t\mathbf{B}^T$$
$$\mathrm{E}[\mathbf{y}_t] = (\mathbf{1}\mathbf{1}^T \otimes \mathbf{A} + \mathbf{B} \otimes \mathbf{1}\mathbf{1}^T)\,\mathbf{x}_t$$
$$\mathrm{E}[y_{i,j,t}] = \sum_{i'}\sum_{j'}(a_{i,i'} + b_{j,j'})x_{i',j',t}$$

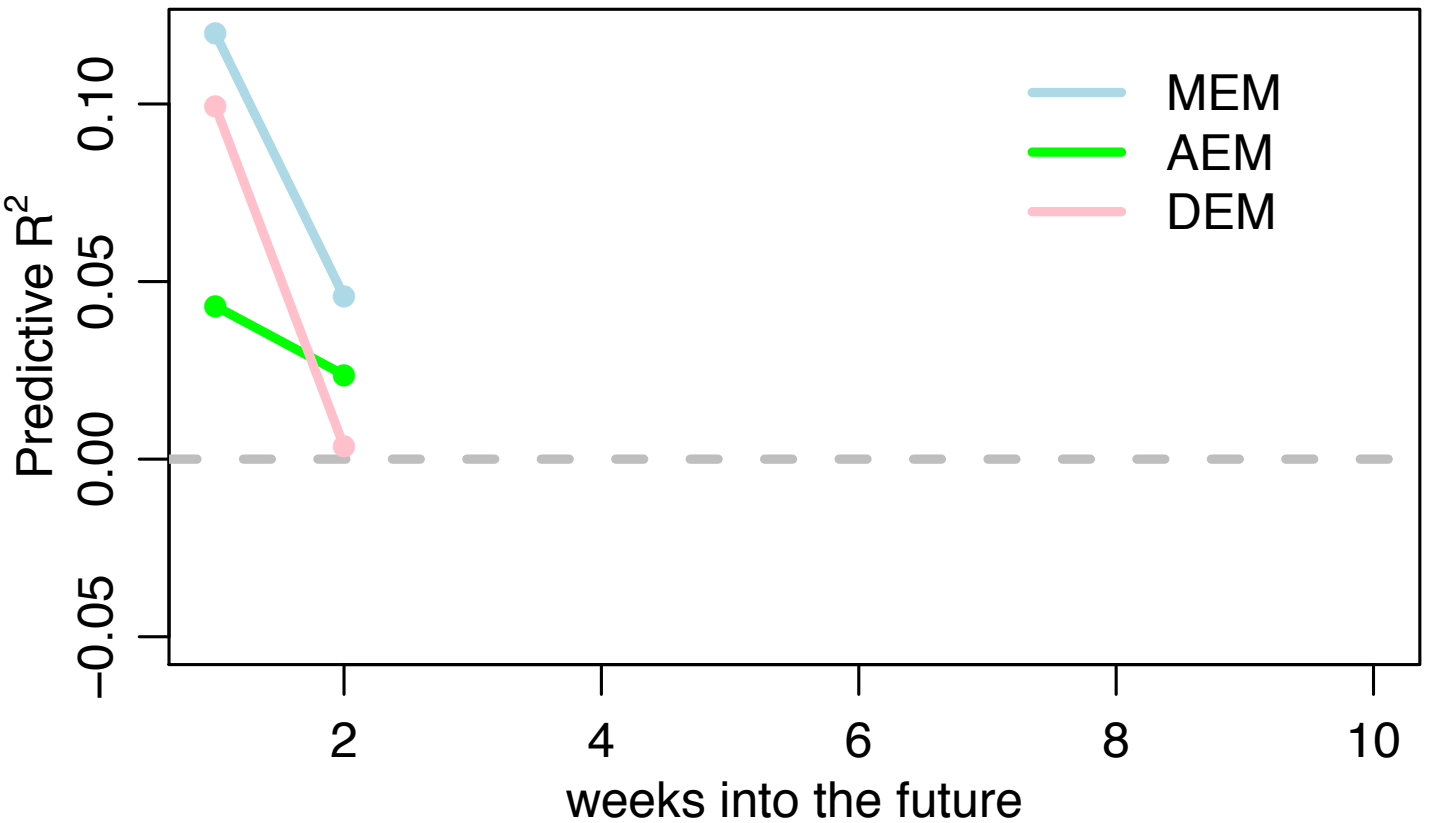
Dyadic effects:

$$\mathrm{E}[\mathbf{Y}_t] = \mathbf{M} \circ \mathbf{X}_t$$
$$\mathrm{E}[\mathbf{y}_t] = \mathbf{D}(\mathbf{m})\,\mathbf{x}_t$$
$$\mathrm{E}[y_{i,j,t}] = m_{i,j}x_{i,j,t}$$

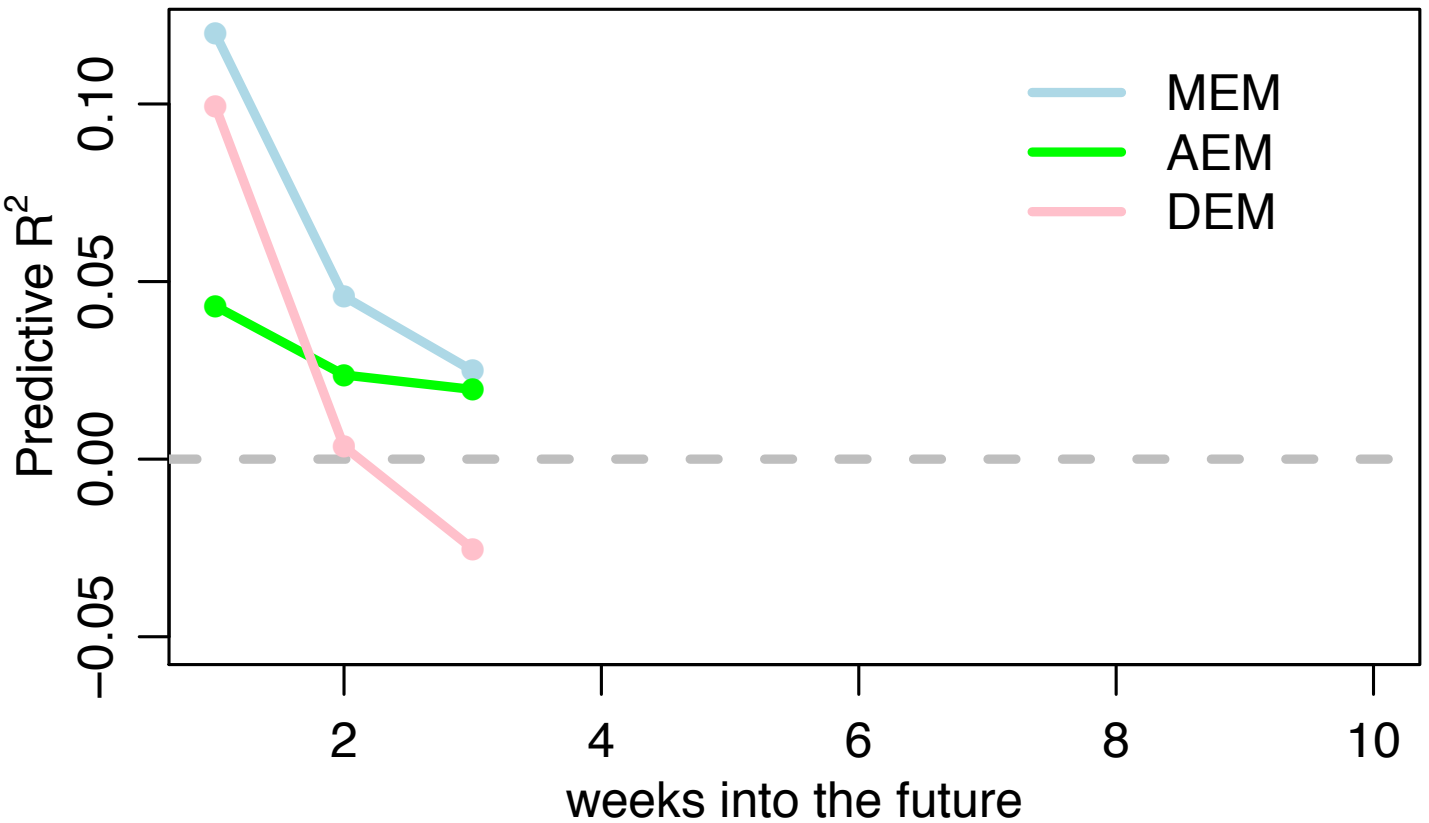
Out of sample forecasts



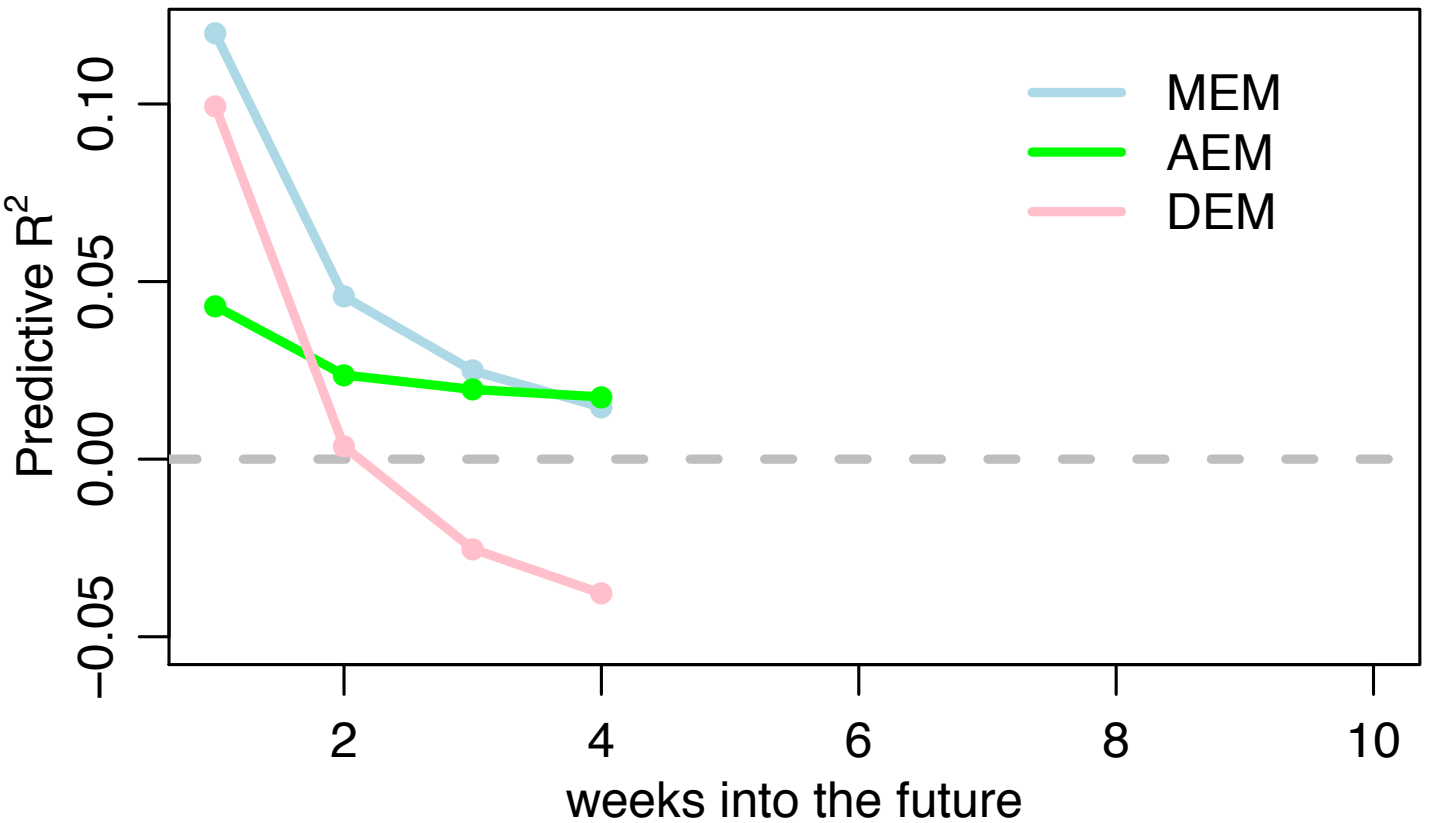
Out of sample forecasts



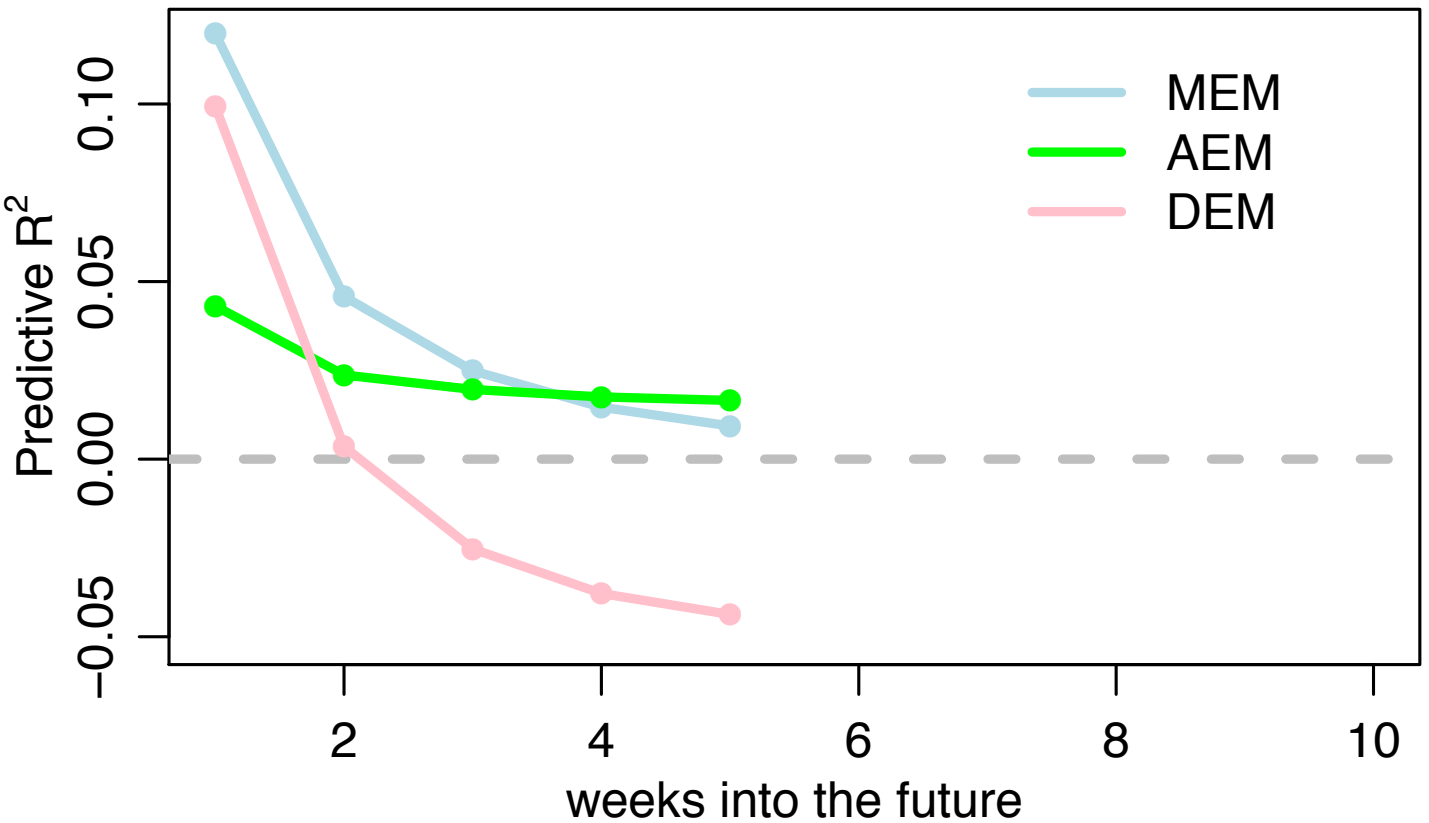
Out of sample forecasts



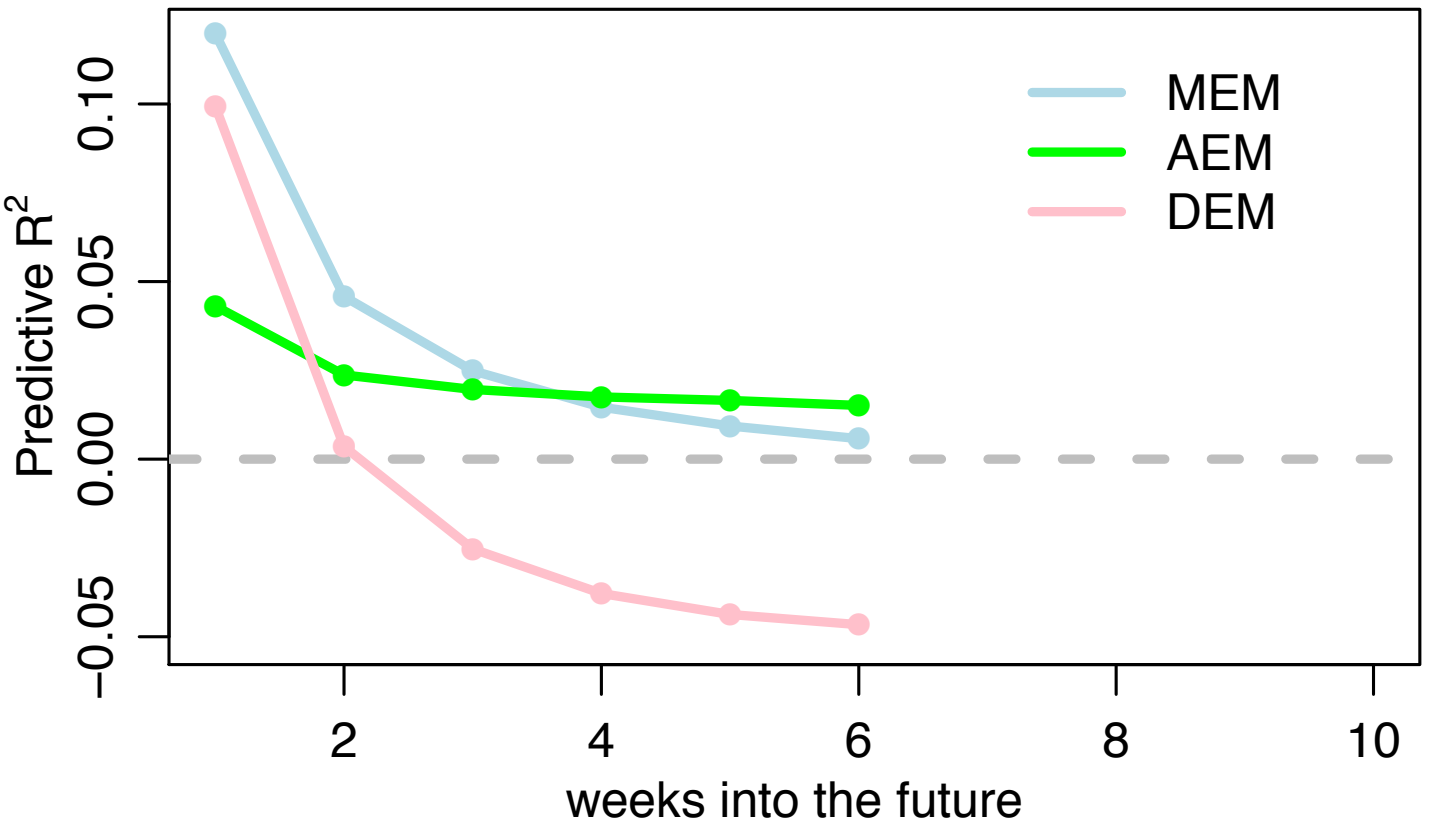
Out of sample forecasts



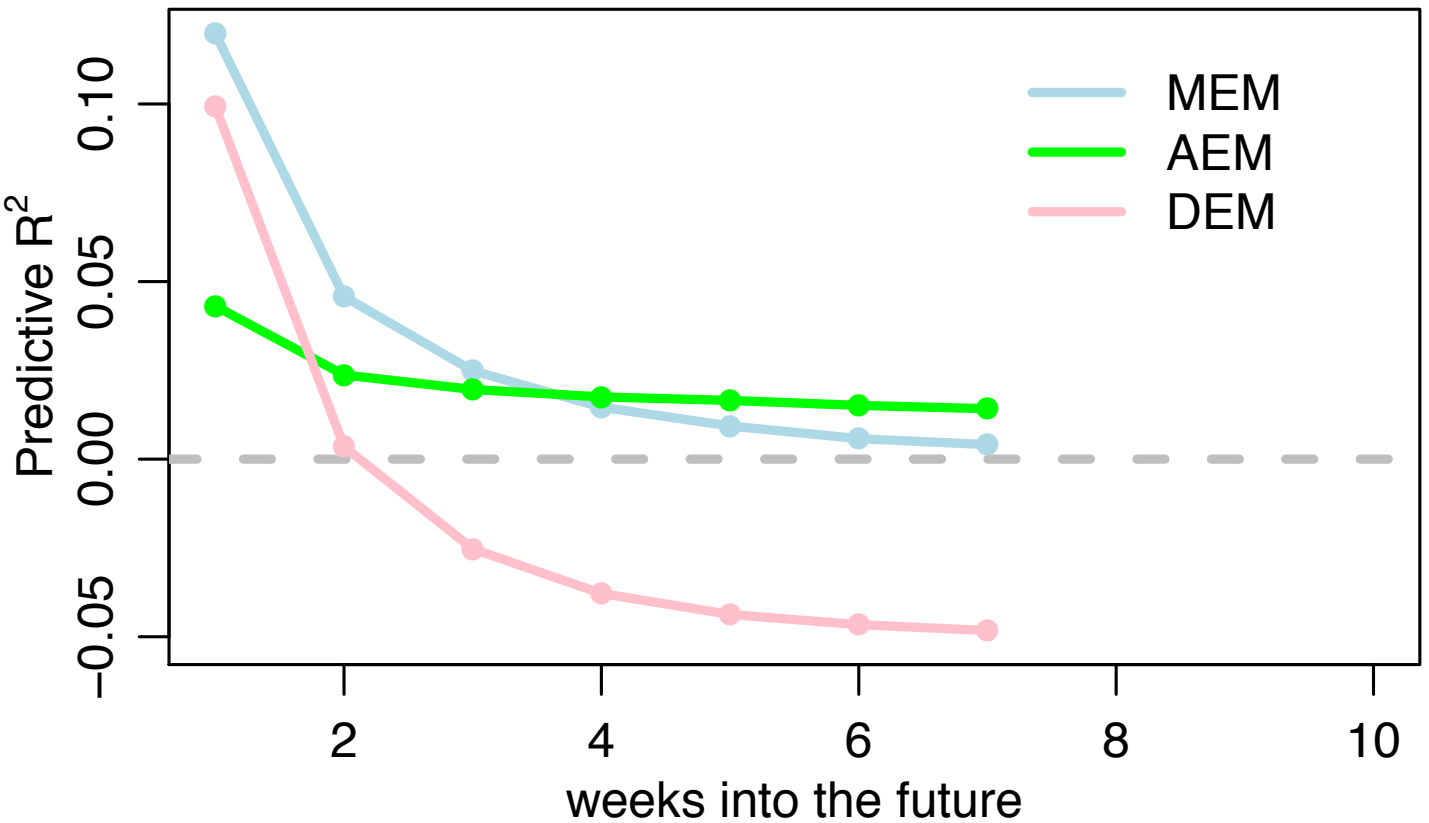
Out of sample forecasts



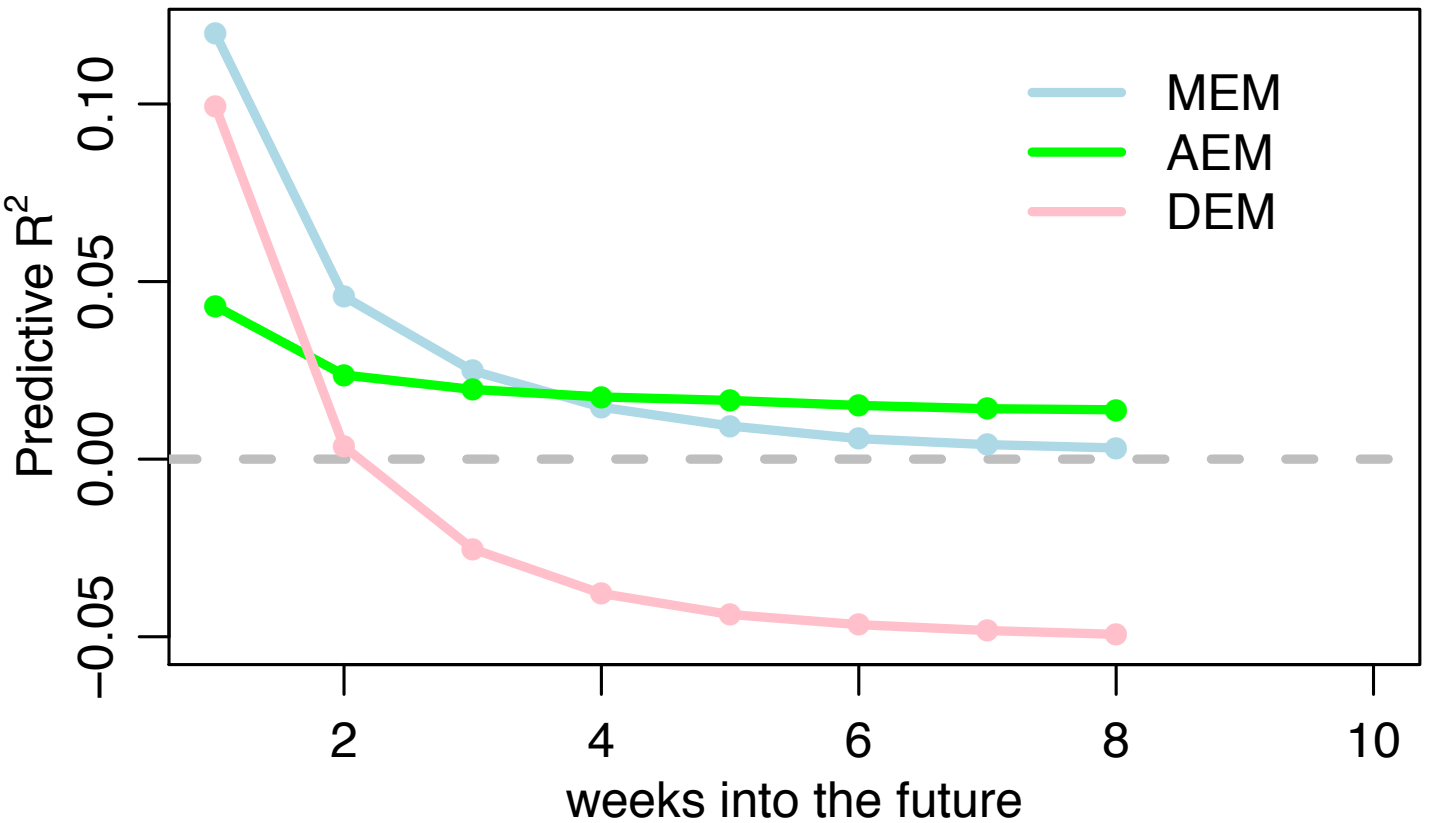
Out of sample forecasts



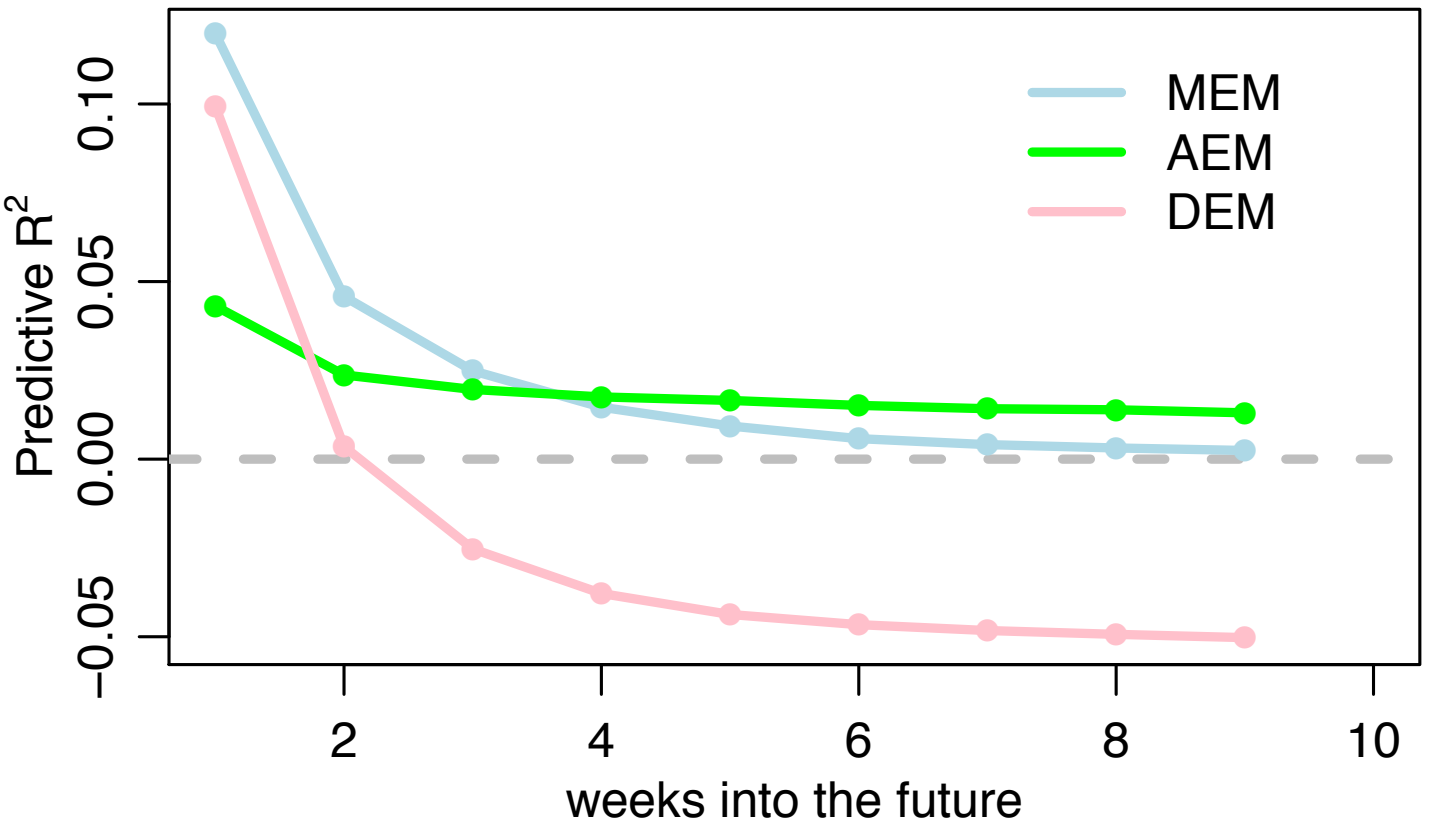
Out of sample forecasts



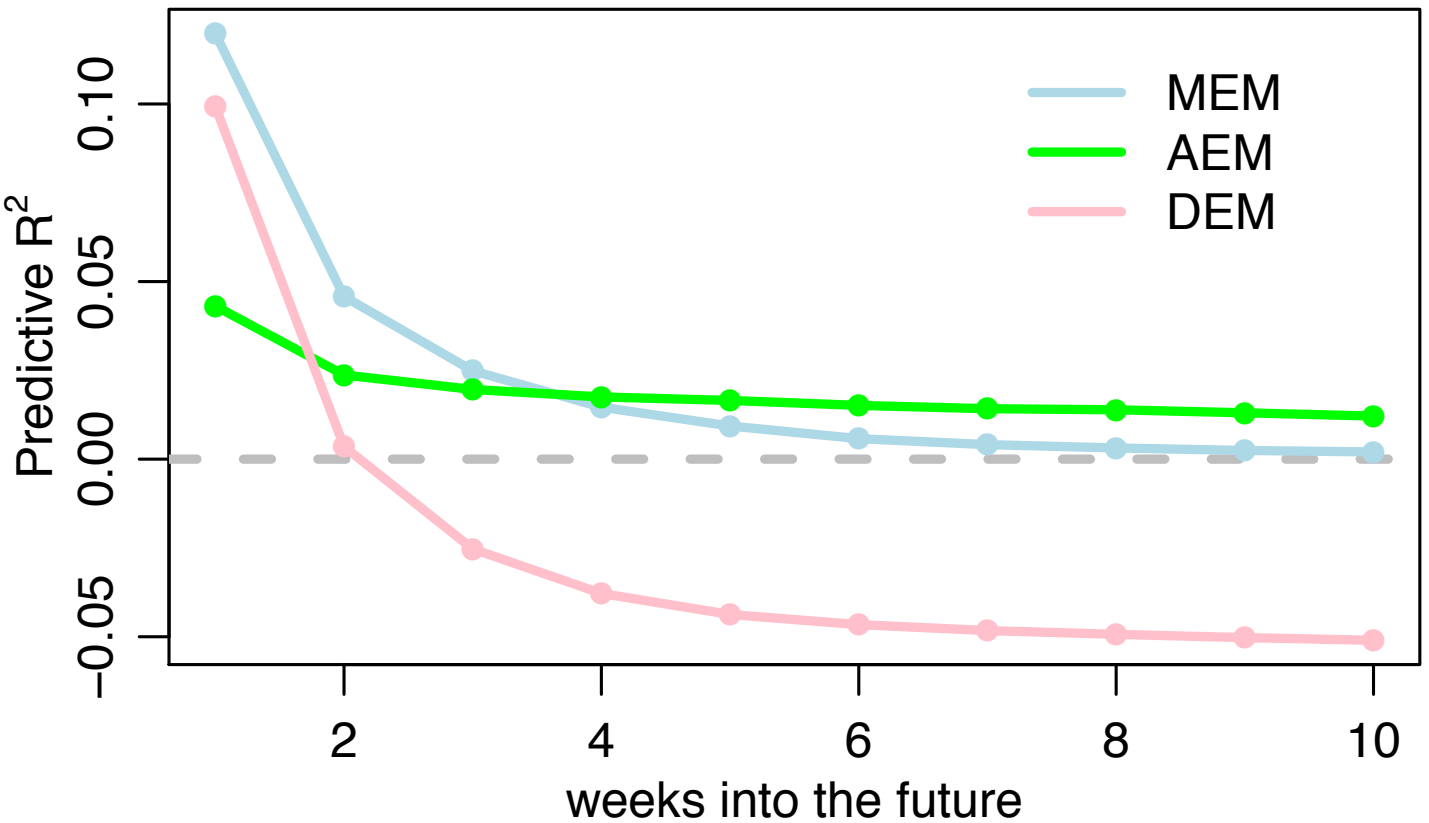
Out of sample forecasts



Out of sample forecasts



Out of sample forecasts



Tensor data

The ICEWS data includes categories of events, and so is **multivariate network data**.

Such data can be represented as a tensor $\{y_{i,j,k}\} = \mathbf{Y} \in \mathbb{R}^{m \times m \times p}$

$i \in \{1, \ldots, m\}$ indexes actors;

$j \in \{1, \ldots, m\}$ indexes targets;

$k \in \{1, \ldots, p\}$ indexes types of actions.

For this analysis, actions are categorized as follows:

$k = 1$: positive verbal

$k = 2$: positive material

$k = 3$: negative verbal

$k = 4$: negative material

Tensor data

The ICEWS data includes categories of events, and so is **multivariate network data**.

Such data can be represented as a tensor $\{y_{i,j,k}\} = \mathbf{Y} \in \mathbb{R}^{m \times m \times p}$

$i \in \{1, \ldots, m\}$ indexes actors;

$j \in \{1, \ldots, m\}$ indexes targets;

$k \in \{1, \ldots, p\}$ indexes types of actions.

For this analysis, actions are categorized as follows:

$k = 1$: positive verbal

$k = 2$: positive material

$k = 3$: negative verbal

$k = 4$: negative material

Tensor data

The ICEWS data includes categories of events, and so is **multivariate network data**.

Such data can be represented as a tensor $\{y_{i,j,k}\} = \mathbf{Y} \in \mathbb{R}^{m \times m \times p}$

$i \in \{1, \ldots, m\}$ indexes actors;

$j \in \{1, \ldots, m\}$ indexes targets;

$k \in \{1, \ldots, p\}$ indexes types of actions.

For this analysis, actions are categorized as follows:

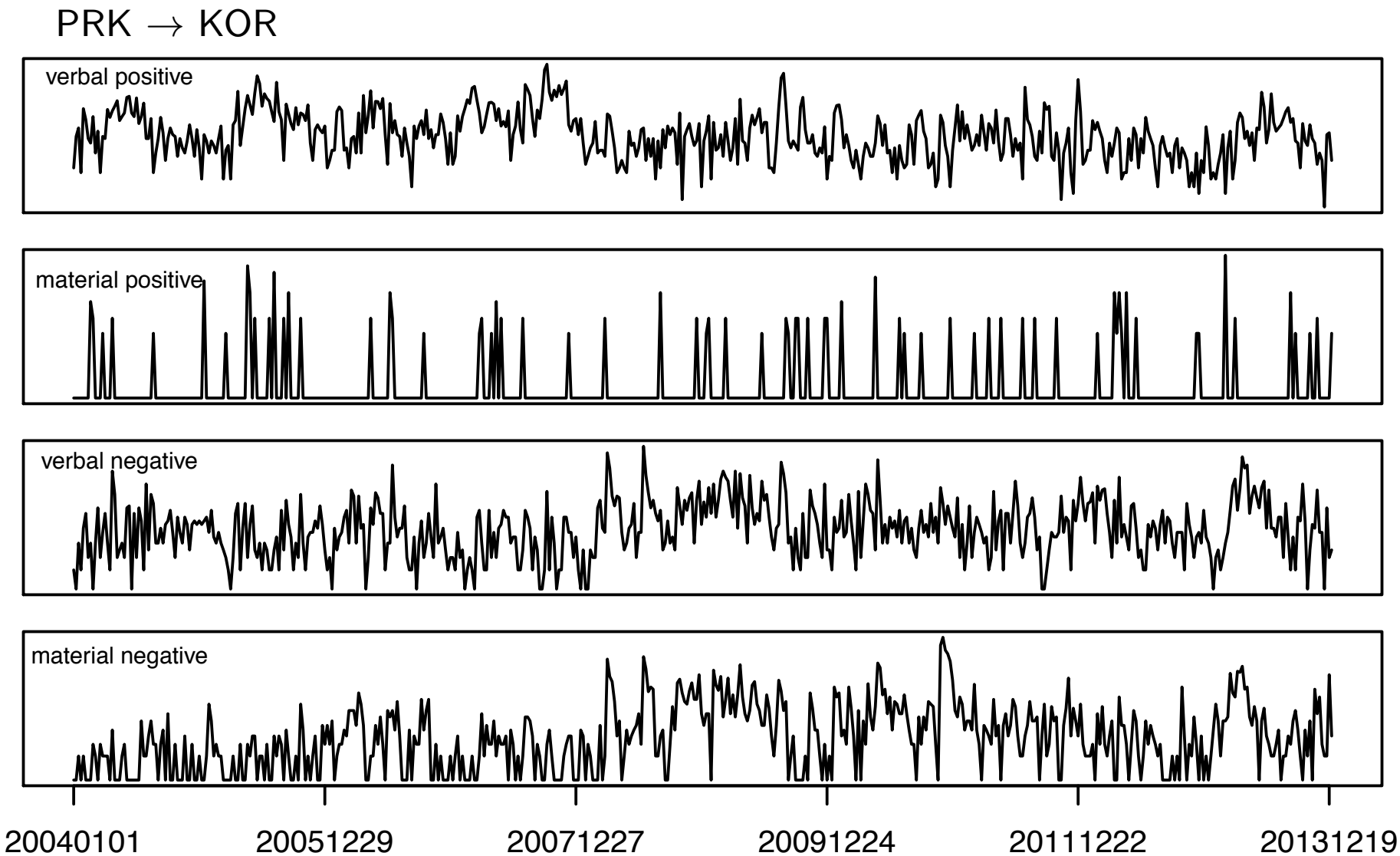
$k = 1$: positive verbal

$k = 2$: positive material

$k = 3$: negative verbal

$k = 4$: negative material

Multivariate multiway data



Higher-order generalizations

Bilinear regression: $\mathbf{Y} \in \mathbb{R}^{m_1 \times m_2}$, $\mathbf{X} \in \mathbb{R}^{p_1 \times p_2}$

$$\mathbf{Y} = \mathbf{A}\ \mathbf{X}\ \mathbf{B}^T + \mathbf{E}$$

Multilinear regression: $\mathbf{Y} \in \mathbb{R}^{m_1 \times m_2 \times m_3}$, $\mathbf{X} \in \mathbb{R}^{p_1 \times p_2 \times p_3}$

Higher-order generalizations

Bilinear regression: $\mathbf{Y} \in \mathbb{R}^{m_1 \times m_2}$, $\mathbf{X} \in \mathbb{R}^{p_1 \times p_2}$

$$\mathbf{Y} = \mathbf{A}\ \mathbf{X}\ \mathbf{B}^T + \mathbf{E}$$

Multilinear regression: $\mathbf{Y} \in \mathbb{R}^{m_1 \times m_2 \times m_3}$, $\mathbf{X} \in \mathbb{R}^{p_1 \times p_2 \times p_3}$

Higher-order generalizations

Bilinear regression: $\mathbf{Y} \in \mathbb{R}^{m_1 \times m_2}$, $\mathbf{X} \in \mathbb{R}^{p_1 \times p_2}$

$$\mathbf{Y} = \mathbf{A}\ \mathbf{X}\ \mathbf{B}^T + \mathbf{E}$$

Multilinear regression: $\mathbf{Y} \in \mathbb{R}^{m_1 \times m_2 \times m_3}$, $\mathbf{X} \in \mathbb{R}^{p_1 \times p_2 \times p_3}$

$$\mathbf{Y} = \left[\begin{array}{c} \mathbf{C} \\ \mathbf{A}\ \mathbf{X}\ \mathbf{B}^T \end{array} \right] + \mathbf{E}\ ?$$

Higher-order generalizations

Bilinear regression: $\mathbf{Y} \in \mathbb{R}^{m_1 \times m_2}$, $\mathbf{X} \in \mathbb{R}^{p_1 \times p_2}$

$$\mathbf{Y} = \mathbf{A}\ \mathbf{X}\ \mathbf{B}^T + \mathbf{E}$$

Multilinear regression: $\mathbf{Y} \in \mathbb{R}^{m_1 \times m_2 \times m_3}$, $\mathbf{X} \in \mathbb{R}^{p_1 \times p_2 \times p_3}$

$$\mathbf{Y} = \left[\begin{array}{c} \mathbf{C} \\ \mathbf{A}\ \mathbf{X}\ \mathbf{B}^T \end{array} \right] + \mathbf{E}\ ?$$

How can such a transformation be defined?

Higher-order generalizations

Bilinear regression: $\mathbf{Y} \in \mathbb{R}^{m_1 \times m_2}$, $\mathbf{X} \in \mathbb{R}^{p_1 \times p_2}$, $\mathbf{y} = \text{vec}(\mathbf{Y}), \mathbf{x} = \text{vec}(\mathbf{X})$

$$\mathbf{y} = (\mathbf{B} \otimes \mathbf{A})\,\mathbf{x} + \mathbf{e}$$

Multilinear regression: $\mathbf{Y} \in \mathbb{R}^{m_1 \times m_2 \times m_3}$, $\mathbf{X} \in \mathbb{R}^{p_1 \times p_2 \times p_3}$, $\mathbf{y} = \text{vec}(\mathbf{Y}), \mathbf{x} = \text{vec}(\mathbf{X})$

$$\mathbf{y} = (\mathbf{C} \otimes \mathbf{B} \otimes \mathbf{A})\,\mathbf{x} + \mathbf{e}$$

Higher-order generalizations

Bilinear regression: $\mathbf{Y} \in \mathbb{R}^{m_1 \times m_2}$, $\mathbf{X} \in \mathbb{R}^{p_1 \times p_2}$, $\mathbf{y} = \text{vec}(\mathbf{Y}), \mathbf{x} = \text{vec}(\mathbf{X})$

$$\mathbf{y} = (\mathbf{B} \otimes \mathbf{A})\,\mathbf{x} + \mathbf{e}$$

Multilinear regression: $\mathbf{Y} \in \mathbb{R}^{m_1 \times m_2 \times m_3}$, $\mathbf{X} \in \mathbb{R}^{p_1 \times p_2 \times p_3}$, $\mathbf{y} = \text{vec}(\mathbf{Y}), \mathbf{x} = \text{vec}(\mathbf{X})$

$$\mathbf{y} = (\mathbf{C} \otimes \mathbf{B} \otimes \mathbf{A})\,\mathbf{x} + \mathbf{e}$$

Higher-order generalizations

Bilinear regression: $\mathbf{Y} \in \mathbb{R}^{m_1 \times m_2}$, $\mathbf{X} \in \mathbb{R}^{p_1 \times p_2}$, $\mathbf{y} = \text{vec}(\mathbf{Y}), \mathbf{x} = \text{vec}(\mathbf{X})$

$$\mathbf{y} = (\mathbf{B} \otimes \mathbf{A})\,\mathbf{x} + \mathbf{e}$$

Multilinear regression: $\mathbf{Y} \in \mathbb{R}^{m_1 \times m_2 \times m_3}$, $\mathbf{X} \in \mathbb{R}^{p_1 \times p_2 \times p_3}$, $\mathbf{y} = \text{vec}(\mathbf{Y}), \mathbf{x} = \text{vec}(\mathbf{X})$

$$\mathbf{y} = (\mathbf{C} \otimes \mathbf{B} \otimes \mathbf{A})\,\mathbf{x} + \mathbf{e}$$

$$\mathbf{Y} = \mathbf{X} \times \{\mathbf{A}, \mathbf{B}, \mathbf{C}\} + \mathbf{E}$$

Higher-order generalizations

Bilinear regression: $\mathbf{Y} \in \mathbb{R}^{m_1 \times m_2}$, $\mathbf{X} \in \mathbb{R}^{p_1 \times p_2}$, $\mathbf{y} = \text{vec}(\mathbf{Y}), \mathbf{x} = \text{vec}(\mathbf{X})$

$$\mathbf{y} = (\mathbf{B} \otimes \mathbf{A})\,\mathbf{x} + \mathbf{e}$$

Multilinear regression: $\mathbf{Y} \in \mathbb{R}^{m_1 \times m_2 \times m_3}$, $\mathbf{X} \in \mathbb{R}^{p_1 \times p_2 \times p_3}$, $\mathbf{y} = \text{vec}(\mathbf{Y}), \mathbf{x} = \text{vec}(\mathbf{X})$

$$\mathbf{y} = (\mathbf{C} \otimes \mathbf{B} \otimes \mathbf{A})\,\mathbf{x} + \mathbf{e}$$

$$\mathbf{Y} = \mathbf{X} \times \{\mathbf{A}, \mathbf{B}, \mathbf{C}\} + \mathbf{E}$$

$$y_{i,j,k} = \sum_{i'=1}^{p_1} \sum_{j'=1}^{p_2} \sum_{k'=1}^{p_3} x_{i',j',k'}\, a_{i,i'}\, b_{j,j'}\, c_{k,k'} + e_{i,j,k}$$

The multilinear Tucker product

Bilinear model: $\text{vec}(\mathbf{Y}) = (\mathbf{B} \otimes \mathbf{A})\,\text{vec}(\mathbf{X}) + \text{vec}(\mathbf{E})$

Multilinear model: $\text{vec}(\mathbf{Y}) = (\mathbf{C} \otimes \mathbf{B} \otimes \mathbf{A})\,\text{vec}(\mathbf{X}) + \text{vec}(\mathbf{E})$

$$
\begin{aligned}
E[y_{i,j,k}] &= \sum_{i'=1}^{p_1}\sum_{j'=1}^{p_2}\sum_{k'=1}^{p_3} x_{i',j',k'}\, a_{i,i'}\, b_{j,j'}\, c_{k,k'} \\
E[\mathbf{Y}] &= \mathbf{X} \times_1 \mathbf{A} \times_2 \mathbf{B} \times_3 \mathbf{C} \\
&= \mathbf{X} \times \{\mathbf{A}, \mathbf{B}, \mathbf{C}\}
\end{aligned}
$$

Estimation: If $\mathbf{Y} = \mathbf{X} \times \{\mathbf{A}, \mathbf{B}, \mathbf{C}\} + \mathbf{E}$, then

$$
\begin{aligned}
\mathbf{Y}_{(1)} &= \mathbf{A}\mathbf{X}_{(1)}(\mathbf{C} \otimes \mathbf{B})^T + \mathbf{E}_{(1)} \\
\mathbf{Y}_{(2)} &= \mathbf{B}\mathbf{X}_{(2)}(\mathbf{C} \otimes \mathbf{A})^T + \mathbf{E}_{(2)} \\
\mathbf{Y}_{(3)} &= \mathbf{C}\mathbf{X}_{(3)}(\mathbf{B} \otimes \mathbf{A})^T + \mathbf{E}_{(3)}
\end{aligned}
$$

Estimate $\{\mathbf{A}, \mathbf{B}, \mathbf{C}\}$ with ALS.

The multilinear Tucker product

Bilinear model: $\text{vec}(\mathbf{Y}) = (\mathbf{B} \otimes \mathbf{A})\,\text{vec}(\mathbf{X}) + \text{vec}(\mathbf{E})$

Multilinear model: $\text{vec}(\mathbf{Y}) = (\mathbf{C} \otimes \mathbf{B} \otimes \mathbf{A})\,\text{vec}(\mathbf{X}) + \text{vec}(\mathbf{E})$

$$
\begin{aligned}
E[y_{i,j,k}] &= \sum_{i'=1}^{p_1}\sum_{j'=1}^{p_2}\sum_{k'=1}^{p_3} x_{i',j',k'}\, a_{i,i'}\, b_{j,j'}\, c_{k,k'} \\
E[\mathbf{Y}] &= \mathbf{X} \times_1 \mathbf{A} \times_2 \mathbf{B} \times_3 \mathbf{C} \\
&= \mathbf{X} \times \{\mathbf{A}, \mathbf{B}, \mathbf{C}\}
\end{aligned}
$$

Estimation: If $\mathbf{Y} = \mathbf{X} \times \{\mathbf{A}, \mathbf{B}, \mathbf{C}\} + \mathbf{E}$, then

$$
\begin{aligned}
\mathbf{Y}_{(1)} &= \mathbf{A}\mathbf{X}_{(1)}(\mathbf{C} \otimes \mathbf{B})^T + \mathbf{E}_{(1)} \\
\mathbf{Y}_{(2)} &= \mathbf{B}\mathbf{X}_{(2)}(\mathbf{C} \otimes \mathbf{A})^T + \mathbf{E}_{(2)} \\
\mathbf{Y}_{(3)} &= \mathbf{C}\mathbf{X}_{(3)}(\mathbf{B} \otimes \mathbf{A})^T + \mathbf{E}_{(3)}
\end{aligned}
$$

Estimate $\{\mathbf{A}, \mathbf{B}, \mathbf{C}\}$ with ALS.

The multilinear Tucker product

Bilinear model: $\text{vec}(\mathbf{Y}) = (\mathbf{B} \otimes \mathbf{A})\,\text{vec}(\mathbf{X}) + \text{vec}(\mathbf{E})$

Multilinear model: $\text{vec}(\mathbf{Y}) = (\mathbf{C} \otimes \mathbf{B} \otimes \mathbf{A})\,\text{vec}(\mathbf{X}) + \text{vec}(\mathbf{E})$

$$
\begin{aligned}
E[y_{i,j,k}] &= \sum_{i'=1}^{p_1}\sum_{j'=1}^{p_2}\sum_{k'=1}^{p_3} x_{i',j',k'}\, a_{i,i'}\, b_{j,j'}\, c_{k,k'} \\
E[\mathbf{Y}] &= \mathbf{X} \times_1 \mathbf{A} \times_2 \mathbf{B} \times_3 \mathbf{C} \\
&= \mathbf{X} \times \{\mathbf{A}, \mathbf{B}, \mathbf{C}\}
\end{aligned}
$$

Estimation: If $\mathbf{Y} = \mathbf{X} \times \{\mathbf{A}, \mathbf{B}, \mathbf{C}\} + \mathbf{E}$, then

$$
\begin{aligned}
\mathbf{Y}_{(1)} &= \mathbf{A}\mathbf{X}_{(1)}(\mathbf{C} \otimes \mathbf{B})^T + \mathbf{E}_{(1)} \\
\mathbf{Y}_{(2)} &= \mathbf{B}\mathbf{X}_{(2)}(\mathbf{C} \otimes \mathbf{A})^T + \mathbf{E}_{(2)} \\
\mathbf{Y}_{(3)} &= \mathbf{C}\mathbf{X}_{(3)}(\mathbf{B} \otimes \mathbf{A})^T + \mathbf{E}_{(3)}
\end{aligned}
$$

Estimate $\{\mathbf{A}, \mathbf{B}, \mathbf{C}\}$ with ALS.

The multilinear Tucker product

Bilinear model: $\text{vec}(\mathbf{Y}) = (\mathbf{B} \otimes \mathbf{A})\,\text{vec}(\mathbf{X}) + \text{vec}(\mathbf{E})$

Multilinear model: $\text{vec}(\mathbf{Y}) = (\mathbf{C} \otimes \mathbf{B} \otimes \mathbf{A})\,\text{vec}(\mathbf{X}) + \text{vec}(\mathbf{E})$

$$
\begin{aligned}
E[y_{i,j,k}] &= \sum_{i'=1}^{p_1}\sum_{j'=1}^{p_2}\sum_{k'=1}^{p_3} x_{i',j',k'}\, a_{i,i'}\, b_{j,j'}\, c_{k,k'} \\
E[\mathbf{Y}] &= \mathbf{X} \times_1 \mathbf{A} \times_2 \mathbf{B} \times_3 \mathbf{C} \\
&= \mathbf{X} \times \{\mathbf{A}, \mathbf{B}, \mathbf{C}\}
\end{aligned}
$$

Estimation: If $\mathbf{Y} = \mathbf{X} \times \{\mathbf{A}, \mathbf{B}, \mathbf{C}\} + \mathbf{E}$, then

$$
\begin{aligned}
\mathbf{Y}_{(1)} &= \mathbf{A}\mathbf{X}_{(1)}(\mathbf{C} \otimes \mathbf{B})^T + \mathbf{E}_{(1)} \\
\mathbf{Y}_{(2)} &= \mathbf{B}\mathbf{X}_{(2)}(\mathbf{C} \otimes \mathbf{A})^T + \mathbf{E}_{(2)} \\
\mathbf{Y}_{(3)} &= \mathbf{C}\mathbf{X}_{(3)}(\mathbf{B} \otimes \mathbf{A})^T + \mathbf{E}_{(3)}
\end{aligned}
$$

Estimate $\{\mathbf{A}, \mathbf{B}, \mathbf{C}\}$ with ALS.

The multilinear Tucker product

Bilinear model: $\text{vec}(\mathbf{Y}) = (\mathbf{B} \otimes \mathbf{A})\,\text{vec}(\mathbf{X}) + \text{vec}(\mathbf{E})$

Multilinear model: $\text{vec}(\mathbf{Y}) = (\mathbf{C} \otimes \mathbf{B} \otimes \mathbf{A})\,\text{vec}(\mathbf{X}) + \text{vec}(\mathbf{E})$

$$
\begin{aligned}
E[y_{i,j,k}] &= \sum_{i'=1}^{p_1}\sum_{j'=1}^{p_2}\sum_{k'=1}^{p_3} x_{i',j',k'}\, a_{i,i'}\, b_{j,j'}\, c_{k,k'} \\
E[\mathbf{Y}] &= \mathbf{X} \times_1 \mathbf{A} \times_2 \mathbf{B} \times_3 \mathbf{C} \\
&= \mathbf{X} \times \{\mathbf{A}, \mathbf{B}, \mathbf{C}\}
\end{aligned}
$$

Estimation: If $\mathbf{Y} = \mathbf{X} \times \{\mathbf{A}, \mathbf{B}, \mathbf{C}\} + \mathbf{E}$, then

$$
\begin{aligned}
\mathbf{Y}_{(1)} &= \mathbf{A}\mathbf{X}_{(1)}(\mathbf{C} \otimes \mathbf{B})^T + \mathbf{E}_{(1)} \\
\mathbf{Y}_{(2)} &= \mathbf{B}\mathbf{X}_{(2)}(\mathbf{C} \otimes \mathbf{A})^T + \mathbf{E}_{(2)} \\
\mathbf{Y}_{(3)} &= \mathbf{C}\mathbf{X}_{(3)}(\mathbf{B} \otimes \mathbf{A})^T + \mathbf{E}_{(3)}
\end{aligned}
$$

Estimate $\{\mathbf{A}, \mathbf{B}, \mathbf{C}\}$ with ALS.

The multilinear Tucker product

Bilinear model: $\text{vec}(\mathbf{Y}) = (\mathbf{B} \otimes \mathbf{A})\,\text{vec}(\mathbf{X}) + \text{vec}(\mathbf{E})$

Multilinear model: $\text{vec}(\mathbf{Y}) = (\mathbf{C} \otimes \mathbf{B} \otimes \mathbf{A})\,\text{vec}(\mathbf{X}) + \text{vec}(\mathbf{E})$

$$
\begin{aligned}
E[y_{i,j,k}] &= \sum_{i'=1}^{p_1}\sum_{j'=1}^{p_2}\sum_{k'=1}^{p_3} x_{i',j',k'}\, a_{i,i'}\, b_{j,j'}\, c_{k,k'} \\
E[\mathbf{Y}] &= \mathbf{X} \times_1 \mathbf{A} \times_2 \mathbf{B} \times_3 \mathbf{C} \\
&= \mathbf{X} \times \{\mathbf{A}, \mathbf{B}, \mathbf{C}\}
\end{aligned}
$$

Estimation: If $\mathbf{Y} = \mathbf{X} \times \{\mathbf{A}, \mathbf{B}, \mathbf{C}\} + \mathbf{E}$, then

$$
\begin{aligned}
\mathbf{Y}_{(1)} &= \mathbf{A}\mathbf{X}_{(1)}(\mathbf{C} \otimes \mathbf{B})^T + \mathbf{E}_{(1)} \\
\mathbf{Y}_{(2)} &= \mathbf{B}\mathbf{X}_{(2)}(\mathbf{C} \otimes \mathbf{A})^T + \mathbf{E}_{(2)} \\
\mathbf{Y}_{(3)} &= \mathbf{C}\mathbf{X}_{(3)}(\mathbf{B} \otimes \mathbf{A})^T + \mathbf{E}_{(3)}
\end{aligned}
$$

Estimate $\{\mathbf{A}, \mathbf{B}, \mathbf{C}\}$ with ALS.

Reciprocity and other effects

Reciprocity: $y_{i,j,k} \sim x_{i,j,k} \overset{?}{+} x_{j,i,k}$

Non-parsimonious approach:

$$\mathbf{Y}_t = \mathbf{X}_t \times \{\mathbf{A}, \mathbf{B}, \mathbf{C}\} + \tilde{\mathbf{X}}_t \times \{\tilde{\mathbf{A}}, \tilde{\mathbf{B}}, \tilde{\mathbf{C}}\} + \mathbf{E}$$

$(m^2 + m^2 + p^2) \times 2$ parameters.

Multilinear alternative: Construct $\mathbf{X}_t \in \mathbb{R}^{m \times m \times p \times 2}$ as

- $x_{i,j,k,1,t} = y_{i,j,k,t-1}$
- $x_{i,j,k,2,t} = y_{j,i,k,t-1}$

$$\mathbf{Y}_t = \mathbf{X}_t \times \{\mathbf{A}, \mathbf{B}, \mathbf{C}, \mathbf{D}\} + \mathbf{E}$$

$m^2 + m^2 + p^2 + 2$ parameters.

A similar approach can be used to parsimoniously model

- other network effects (transitivity);
- higher-order autoregressive parameters.

(Hoff, 2015)

Reciprocity and other effects

Reciprocity: $y_{i,j,k} \sim x_{i,j,k} \overset{?}{+} x_{j,i,k}$

Non-parsimonious approach:

$$\mathbf{Y}_t = \mathbf{X}_t \times \{\mathbf{A}, \mathbf{B}, \mathbf{C}\} + \tilde{\mathbf{X}}_t \times \{\tilde{\mathbf{A}}, \tilde{\mathbf{B}}, \tilde{\mathbf{C}}\} + \mathbf{E}$$

$(m^2 + m^2 + p^2) \times 2$ parameters.

Multilinear alternative: Construct $\mathbf{X}_t \in \mathbb{R}^{m \times m \times p \times 2}$ as

- $x_{i,j,k,1,t} = y_{i,j,k,t-1}$
- $x_{i,j,k,2,t} = y_{j,i,k,t-1}$

$$\mathbf{Y}_t = \mathbf{X}_t \times \{\mathbf{A}, \mathbf{B}, \mathbf{C}, \mathbf{D}\} + \mathbf{E}$$

$m^2 + m^2 + p^2 + 2$ parameters.

A similar approach can be used to parsimoniously model

- other network effects (transitivity);
- higher-order autoregressive parameters.

(Hoff, 2015)

Reciprocity and other effects

Reciprocity: $y_{i,j,k} \sim x_{i,j,k} \overset{?}{+} x_{j,i,k}$

Non-parsimonious approach:

$$\mathbf{Y}_t = \mathbf{X}_t \times \{\mathbf{A}, \mathbf{B}, \mathbf{C}\} + \tilde{\mathbf{X}}_t \times \{\tilde{\mathbf{A}}, \tilde{\mathbf{B}}, \tilde{\mathbf{C}}\} + \mathbf{E}$$

$(m^2 + m^2 + p^2) \times 2$ parameters.

Multilinear alternative: Construct $\mathbf{X}_t \in \mathbb{R}^{m \times m \times p \times 2}$ as

- $x_{i,j,k,1,t} = y_{i,j,k,t-1}$
- $x_{i,j,k,2,t} = y_{j,i,k,t-1}$

$$\mathbf{Y}_t = \mathbf{X}_t \times \{\mathbf{A}, \mathbf{B}, \mathbf{C}, \mathbf{D}\} + \mathbf{E}$$

$m^2 + m^2 + p^2 + 2$ parameters.

A similar approach can be used to parsimoniously model

- other network effects (transitivity);
- higher-order autoregressive parameters.

(Hoff, 2015)

Reciprocity and other effects

Reciprocity: $y_{i,j,k} \sim x_{i,j,k} \overset{?}{+} x_{j,i,k}$

Non-parsimonious approach:

$$\mathbf{Y}_t = \mathbf{X}_t \times \{\mathbf{A}, \mathbf{B}, \mathbf{C}\} + \tilde{\mathbf{X}}_t \times \{\tilde{\mathbf{A}}, \tilde{\mathbf{B}}, \tilde{\mathbf{C}}\} + \mathbf{E}$$

$(m^2 + m^2 + p^2) \times 2$ parameters.

Multilinear alternative: Construct $\mathbf{X}_t \in \mathbb{R}^{m \times m \times p \times 2}$ as

- $x_{i,j,k,1,t} = y_{i,j,k,t-1}$
- $x_{i,j,k,2,t} = y_{j,i,k,t-1}$

$$\mathbf{Y}_t = \mathbf{X}_t \times \{\mathbf{A}, \mathbf{B}, \mathbf{C}, \mathbf{D}\} + \mathbf{E}$$

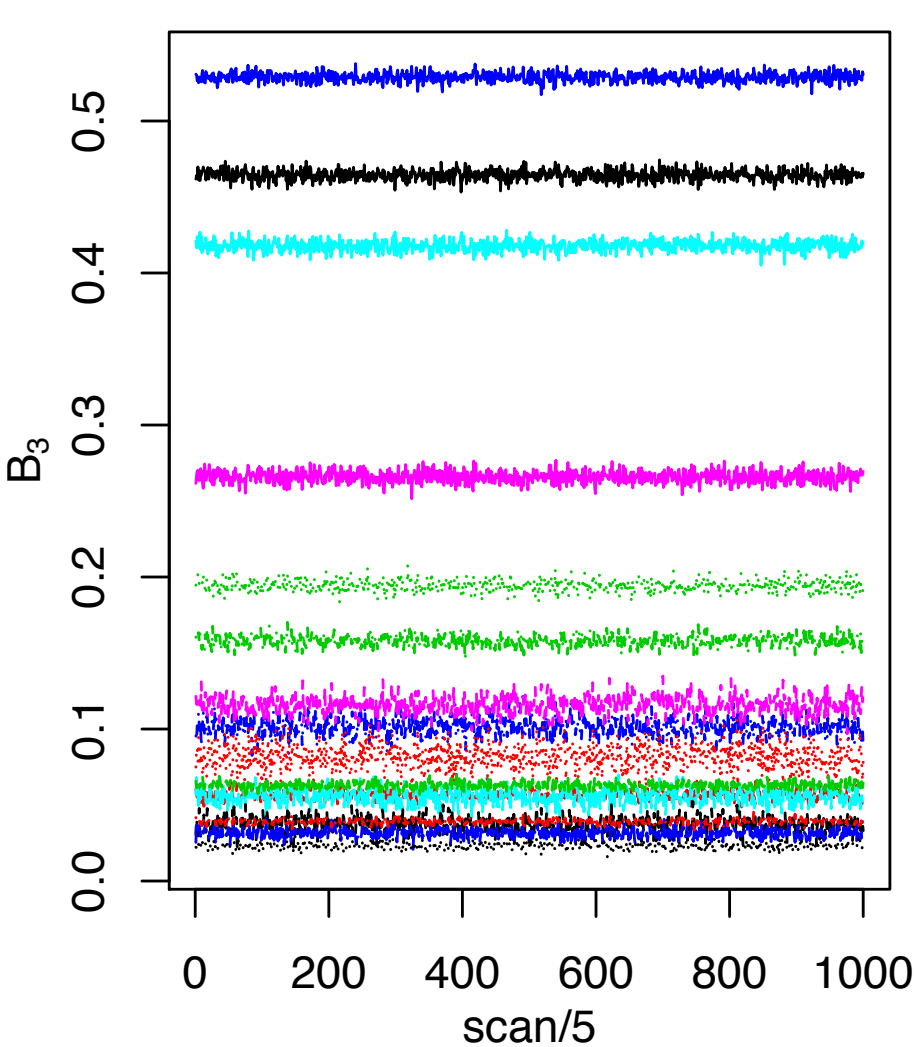
$m^2 + m^2 + p^2 + 2$ parameters.

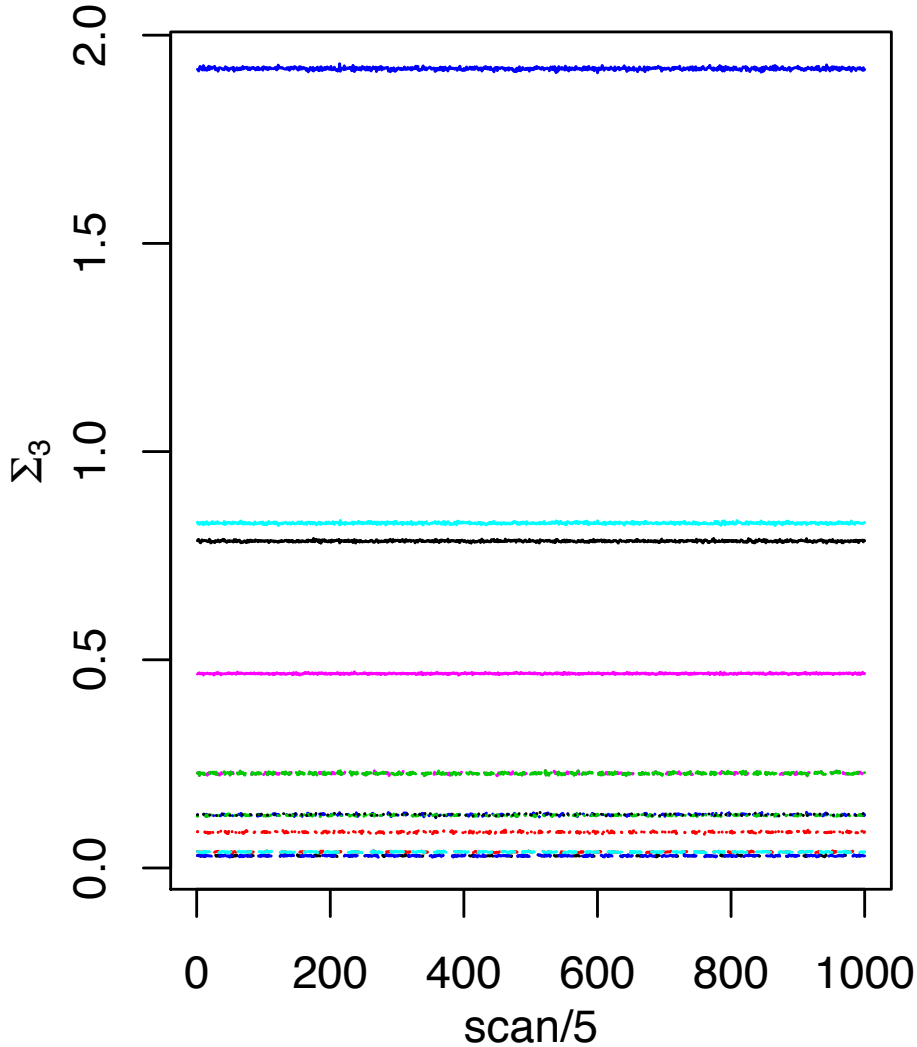
A similar approach can be used to parsimoniously model

- other network effects (transitivity);
- higher-order autoregressive parameters.

(Hoff, 2015)

Bayesian inference via MCMC





Large influence parameters

A		
i, i'	$\mathrm{E}[a_{i,i'}]$	$\mathrm{SD}[a_{i,i'}]$
GBR DEU	0.137	0.023
DEU FRA	0.121	0.018
TUR IRN	0.120	0.015
FRA DEU	0.120	0.021
JPN KOR	0.114	0.020
AUS GBR	0.097	0.016
GBR USA	0.096	0.012
LBN IRN	0.088	0.012
KOR CHN	0.088	0.015
UKR RUS	0.061	0.011

B		
j, j'	$\mathrm{E}[b_{j,j'}]$	$\mathrm{SD}[b_{j,j'}]$
GBR DEU	0.110	0.022
GBR AUS	0.101	0.024
ISR PSE	0.092	0.022
IRQ USA	0.067	0.012
AUS GBR	0.066	0.014
RUS USA	0.063	0.013
GBR USA	0.060	0.012
LBN ISR	0.060	0.014
PRK IRQ	0.054	0.011
SDN IRQ	0.047	0.011

All entries of **C** and **D** are nominally significant and positive.

Sparsity

Goal: Extend analysis to more countries.

Problem: Sparsity of relations.

- $m = 25$: one pair has no events, 3% of pairs have ten or fewer.
- $m = 150$: 18% of pairs have no events, 50% have 10 or fewer.

Sparsity

Goal: Extend analysis to more countries.

Problem: Sparsity of relations.

- $m = 25$: one pair has no events, 3% of pairs have ten or fewer.
- $m = 150$: 18% of pairs have no events, 50% have 10 or fewer.

Sparsity

Goal: Extend analysis to more countries.

Problem: Sparsity of relations.

- $m = 25$: one pair has no events, 3% of pairs have ten or fewer.
- $m = 150$: 18% of pairs have no events, 50% have 10 or fewer.

Sparsity

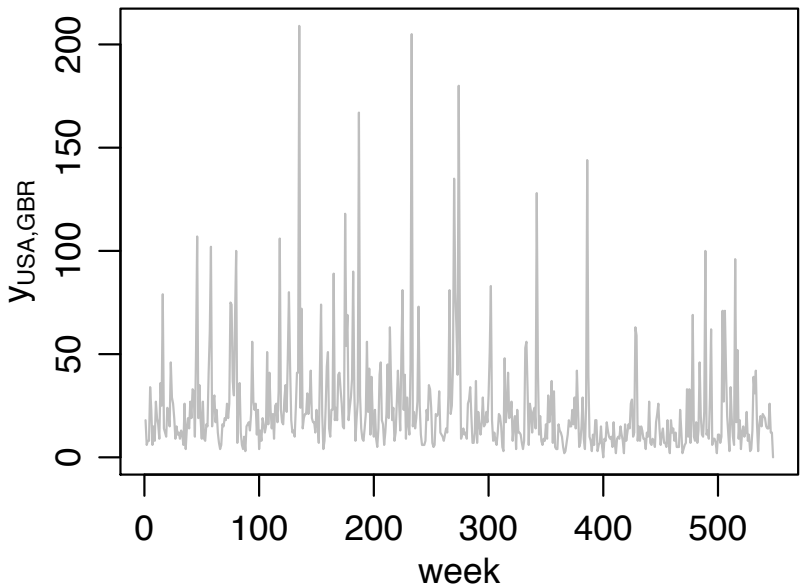
Goal: Extend analysis to more countries.

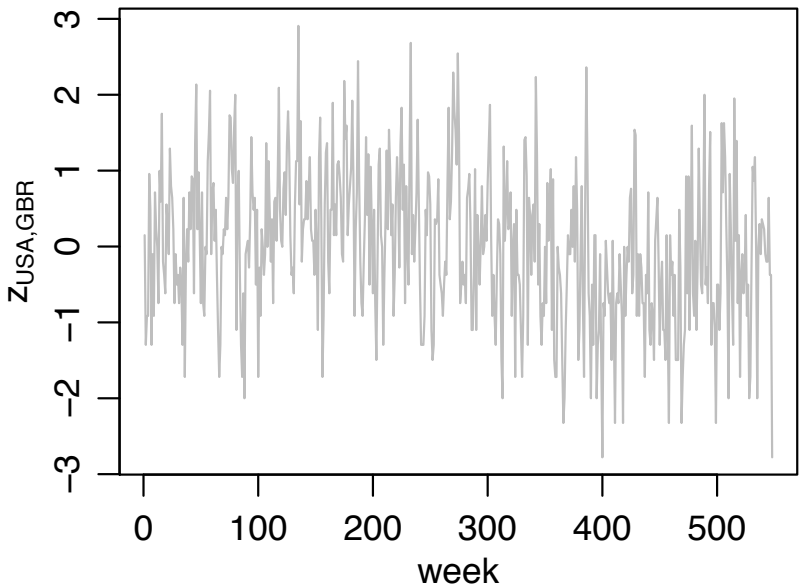
Problem: Sparsity of relations.

- $m = 25$: one pair has no events, 3% of pairs have ten or fewer.
- $m = 150$: 18% of pairs have no events, 50% have 10 or fewer.

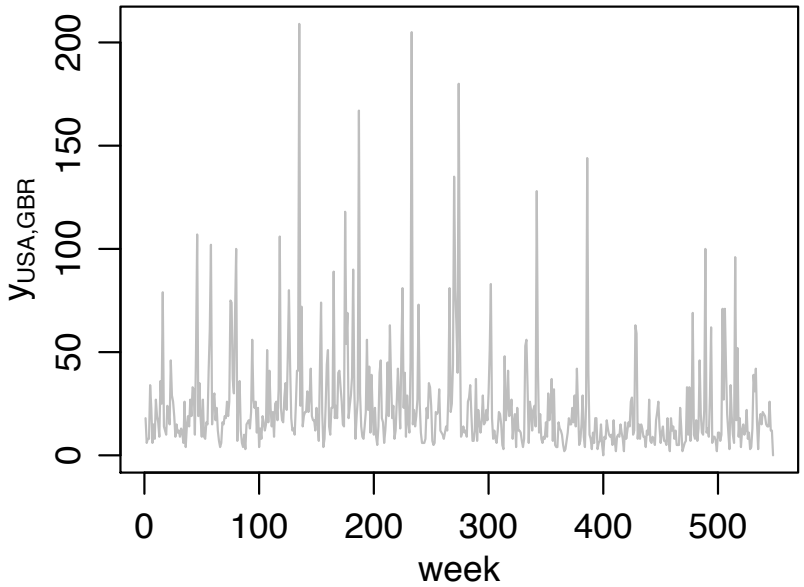
The bulk of the data are not transformably normal.

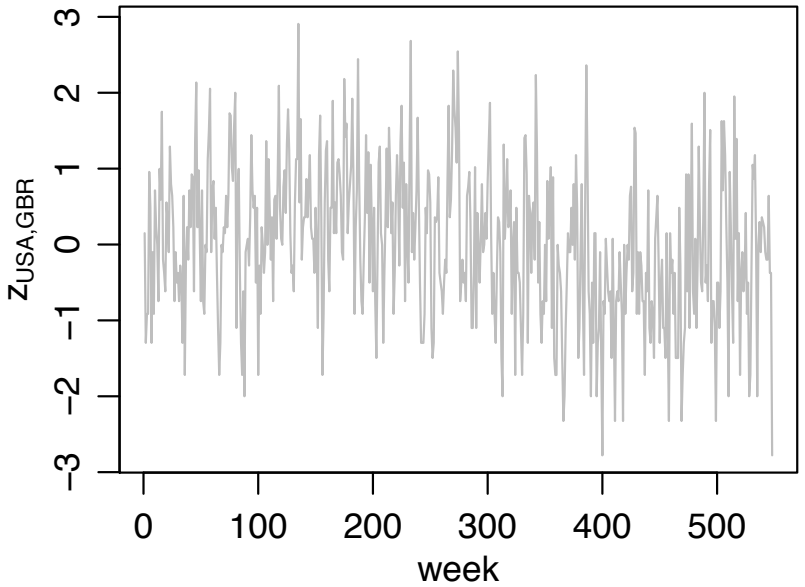
Sparsity

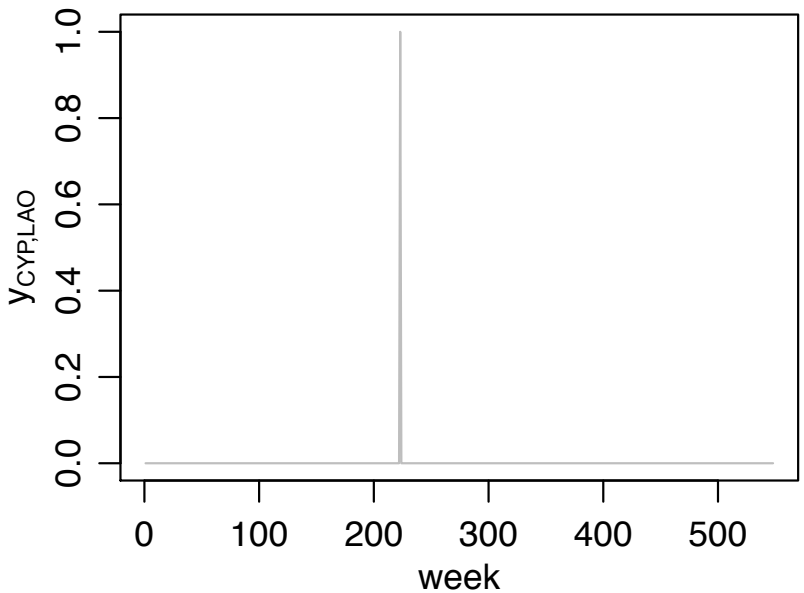

$y_{USA,GBR}$
0
50
100
150
200
0
100
200
300
400
500
week

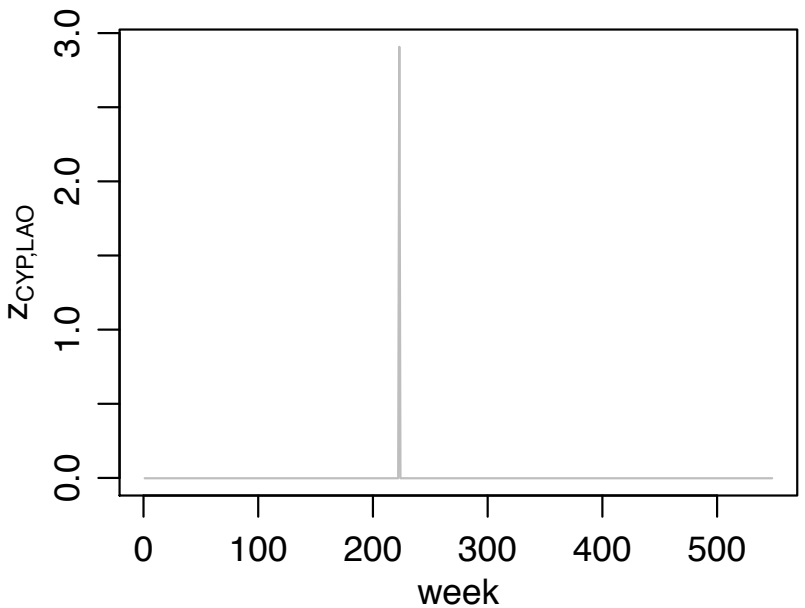

$z_{USA,GBR}$
−3
−2
−1
0
1
2
3
0
100
200
300
400
500
week

Sparsity


$y_{USA,GBR}$
0
50
100
150
200
0
100
200
300
400
500
week


$z_{USA,GBR}$
−3
−2
−1
0
1
2
3
0
100
200
300
400
500
week


$y_{CYP,LAO}$
0.0
0.2
0.4
0.6
0.8
1.0
0
100
200
300
400
500
week


$z_{CYP,LAO}$
0.0
1.0
2.0
3.0
0
100
200
300
400
500
week

Method 1: Ordinal HMM

Baseline model: $\Pr(y_{i,j,t} = 1|\mu_{i,j}) = \Phi(\mu_{i,j})$, or equivalently

$$z_{i,j,t} = \mu_{i,j} + \epsilon_{i,j,t}$$
$$y_{i,j,t} = 1(z_{i,j,t} > 0)$$

Probit HMM: $\Pr(y_{i,j,t} = 1|\mu_{i,j}, \theta_{i,j,t}) = \Phi(\mu_{i,j} + \theta_{i,j,t})$, or equivalently

$$z_{i,j,t} = \mu_{i,j} + \theta_{i,j,t} + \epsilon_{i,j,t}$$
$$y_{i,j,t} = 1(z_{i,j,t} > 0) \ ,$$

where

$$\mathbf{\Theta}_t = \mathbf{A}\mathbf{\Theta}_{t-1}\mathbf{B}^T + \sigma\mathbf{E}_t$$

Ordinal HMM: $y_{i,j,t} = f(z_{i,j,t})$, f non-decreasing.

(I treat f semiparametrically, using a rank likelihood as in Hoff (2007))

Method 1: Ordinal HMM

Baseline model: $\Pr(y_{i,j,t} = 1|\mu_{i,j}) = \Phi(\mu_{i,j})$, or equivalently

$$z_{i,j,t} = \mu_{i,j} + \epsilon_{i,j,t}$$
$$y_{i,j,t} = 1(z_{i,j,t} > 0)$$

Probit HMM: $\Pr(y_{i,j,t} = 1|\mu_{i,j}, \theta_{i,j,t}) = \Phi(\mu_{i,j} + \theta_{i,j,t})$, or equivalently

$$z_{i,j,t} = \mu_{i,j} + \theta_{i,j,t} + \epsilon_{i,j,t}$$
$$y_{i,j,t} = 1(z_{i,j,t} > 0) \ ,$$

where

$$\Theta_t = \mathbf{A}\Theta_{t-1}\mathbf{B}^T + \sigma\mathbf{E}_t$$

Ordinal HMM: $y_{i,j,t} = f(z_{i,j,t})$, f non-decreasing.

(I treat f semiparametrically, using a rank likelihood as in Hoff (2007))

Method 1: Ordinal HMM

Baseline model: $\Pr(y_{i,j,t} = 1|\mu_{i,j}) = \Phi(\mu_{i,j})$, or equivalently

$$z_{i,j,t} = \mu_{i,j} + \epsilon_{i,j,t}$$
$$y_{i,j,t} = 1(z_{i,j,t} > 0)$$

Probit HMM: $\Pr(y_{i,j,t} = 1|\mu_{i,j}, \theta_{i,j,t}) = \Phi(\mu_{i,j} + \theta_{i,j,t})$, or equivalently

$$z_{i,j,t} = \mu_{i,j} + \theta_{i,j,t} + \epsilon_{i,j,t}$$
$$y_{i,j,t} = 1(z_{i,j,t} > 0) \ ,$$

where

$$\mathbf{\Theta}_t = \mathbf{A}\mathbf{\Theta}_{t-1}\mathbf{B}^T + \sigma\mathbf{E}_t$$

Ordinal HMM: $y_{i,j,t} = f(z_{i,j,t})$, f non-decreasing.

(I treat f semiparametrically, using a rank likelihood as in Hoff (2007))

Method 1: Ordinal HMM

Baseline model: $\Pr(y_{i,j,t} = 1|\mu_{i,j}) = \Phi(\mu_{i,j})$, or equivalently

$$z_{i,j,t} = \mu_{i,j} + \epsilon_{i,j,t}$$
$$y_{i,j,t} = 1(z_{i,j,t} > 0)$$

Probit HMM: $\Pr(y_{i,j,t} = 1|\mu_{i,j}, \theta_{i,j,t}) = \Phi(\mu_{i,j} + \theta_{i,j,t})$, or equivalently

$$z_{i,j,t} = \mu_{i,j} + \theta_{i,j,t} + \epsilon_{i,j,t}$$
$$y_{i,j,t} = 1(z_{i,j,t} > 0) \ ,$$

where

$$\mathbf{\Theta}_t = \mathbf{A}\mathbf{\Theta}_{t-1}\mathbf{B}^T + \sigma\mathbf{E}_t$$

Ordinal HMM: $y_{i,j,t} = f(z_{i,j,t})$, f non-decreasing.

(I treat f semiparametrically, using a rank likelihood as in Hoff (2007))

Method 1: Ordinal HMM

Baseline model: $\Pr(y_{i,j,t} = 1|\mu_{i,j}) = \Phi(\mu_{i,j})$, or equivalently

$$z_{i,j,t} = \mu_{i,j} + \epsilon_{i,j,t}$$
$$y_{i,j,t} = 1(z_{i,j,t} > 0)$$

Probit HMM: $\Pr(y_{i,j,t} = 1|\mu_{i,j}, \theta_{i,j,t}) = \Phi(\mu_{i,j} + \theta_{i,j,t})$, or equivalently

$$z_{i,j,t} = \mu_{i,j} + \theta_{i,j,t} + \epsilon_{i,j,t}$$
$$y_{i,j,t} = 1(z_{i,j,t} > 0) \ ,$$

where

$$\mathbf{\Theta}_t = \mathbf{A}\mathbf{\Theta}_{t-1}\mathbf{B}^T + \sigma\mathbf{E}_t$$

Ordinal HMM: $y_{i,j,t} = f(z_{i,j,t})$, f non-decreasing.

(I treat f semiparametrically, using a rank likelihood as in Hoff (2007))

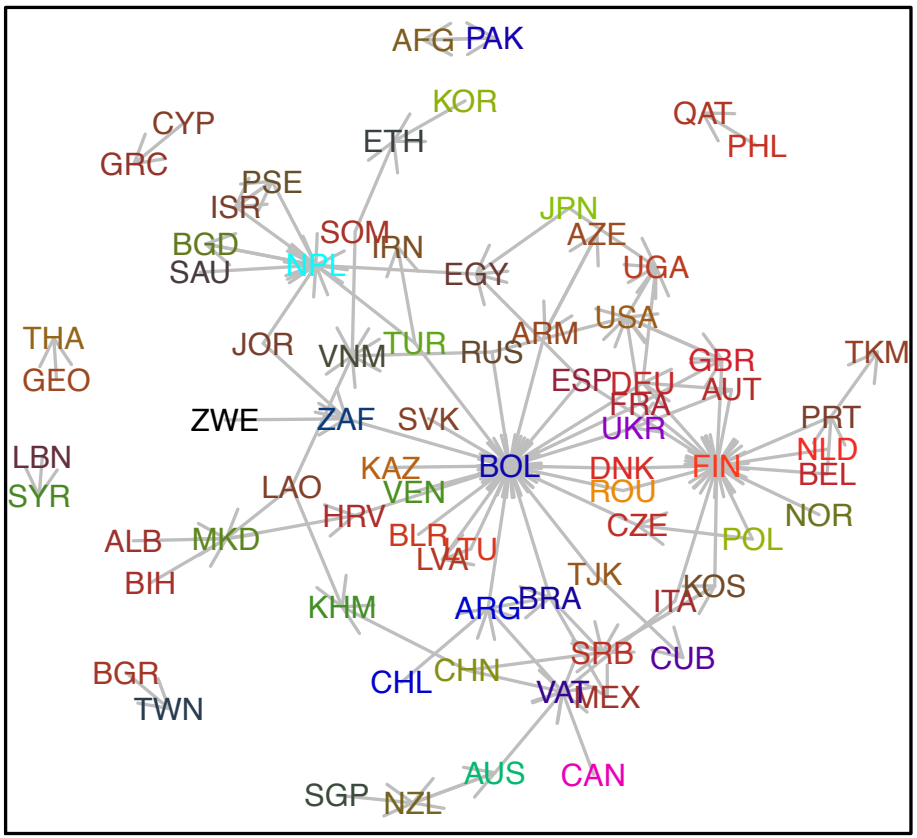
Method 1: Ordinal HMM

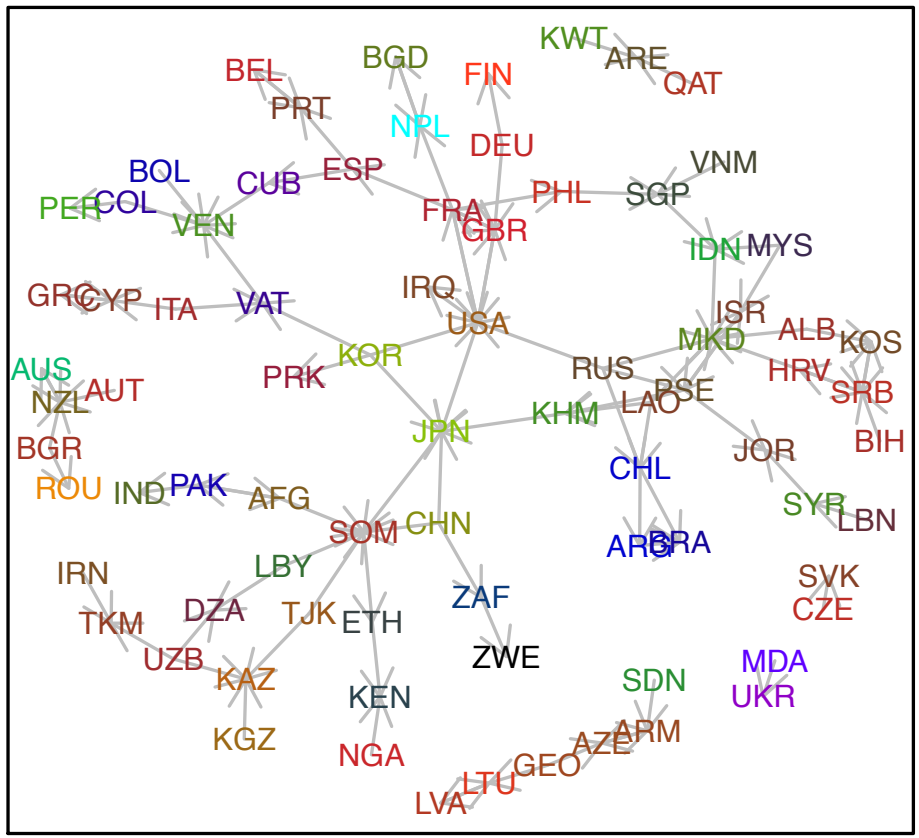
A		
i_1, i_2	$\mathrm{E}[a_{i_1,i_2}]$	$\mathrm{SD}[a_{i_1,i_2}]$
VEN BOL	0.109	0.020
DEU FIN	0.142	0.020
ESP FIN	0.121	0.020
POL FIN	0.106	0.018
PSE ISR	0.146	0.017
PSE NPL	0.177	0.036
AUS NZL	0.128	0.020
LBN SYR	0.106	0.016
GBR USA	0.074	0.012
ITA VAT	0.089	0.018

B		
i_1, i_2	$\mathrm{E}[b_{i_1,i_2}]$	$\mathrm{SD}[b_{2,i_1,i_2}]$
PSE ISR	0.178	0.018
SRB KOS	0.127	0.018
LTU LVA	0.141	0.020
ALB MKD	0.151	0.025
AUS NZL	0.149	0.023
BEL PRT	0.135	0.020
ISR PSE	0.137	0.016
KOS SRB	0.122	0.022
LBN SYR	0.086	0.014
GBR USA	0.070	0.012

Table: Posterior means and standard deviations of the top ten elements of **A** and **B**, in terms of the ratio of mean to standard deviation.

Method 1: Ordinal HMM





Method 2: Social influence regression

Multilinear GLM:

$$y_{i,j,t} \sim \text{Poisson}(e^{\eta_{i,j,t}})$$
$$y_{i,j,t} \sim \text{binary}(e^{\eta_{i,j,t}}/(1+e^{\eta_{i,j,t}}))$$
$$\eta_{i,j,t} = \mathbf{a_i}^T \mathbf{X}_t \mathbf{b_j}$$

Influence coefficients $a_{i,i'}, b_{j,j'}$ generally associated with exogenous covariates:

- distance between i and i';
- alliances, trade agreements between i and i';

Idea: Simplify the model by making this explicit.

$$a_{i,i'} = \alpha^T \mathbf{w}_{i,i'}$$
$$b_{j,j'} = \beta^T \mathbf{w}_{j,j'}$$

Method 2: Social influence regression

Multilinear GLM:

$$y_{i,j,t} \sim \text{Poisson}(e^{\eta_{i,j,t}})$$
$$y_{i,j,t} \sim \text{binary}(e^{\eta_{i,j,t}}/(1+e^{\eta_{i,j,t}}))$$
$$\eta_{i,j,t} = \mathbf{a_i}^T \mathbf{X}_t \mathbf{b_j}$$

Influence coefficients $a_{i,i'}, b_{j,j'}$ generally associated with exogenous covariates:

- distance between i and i';
- alliances, trade agreements between i and i';

Idea: Simplify the model by making this explicit.

$$a_{i,i'} = \alpha^T \mathbf{w}_{i,i'}$$
$$b_{j,j'} = \beta^T \mathbf{w}_{j,j'}$$

Method 2: Social influence regression

Multilinear GLM:

$$y_{i,j,t} \sim \text{Poisson}(e^{\eta_{i,j,t}})$$
$$y_{i,j,t} \sim \text{binary}(e^{\eta_{i,j,t}}/(1+e^{\eta_{i,j,t}}))$$
$$\eta_{i,j,t} = \mathbf{a_i}^T \mathbf{X}_t \mathbf{b_j}$$

Influence coefficients $a_{i,i'}, b_{j,j'}$ generally associated with exogenous covariates:

- distance between i and i';
- alliances, trade agreements between i and i';

Idea: Simplify the model by making this explicit.

$$a_{i,i'} = \alpha^T \mathbf{w}_{i,i'}$$
$$b_{j,j'} = \beta^T \mathbf{w}_{j,j'}$$

Method 2: Social influence regression

$$
\begin{aligned}
\eta_{i,j,t} = \mathbf{a_i}^T \mathbf{X}_t \mathbf{b_j} &= \sum_{i'j'} a_{ii'} b_{jj'} x_{i'j't} \\
&= \sum_{i'j'} \alpha^T w_{ii'} \beta^T w_{jj'} x_{i'j't} \\
&= \alpha^T \left(\sum_{i'j'} x_{i'j't} w_{ii'} w_{jj'}^T \right) \beta \\
&= \alpha^T \tilde{X}_{ijt} \beta
\end{aligned}
$$

The number of parameters has gone from $2 \times n \times (n-1)$ to $2 \times q$.

The model is basically of the type considered by Zhou et al. (2013).

Method 2: Social influence regression

$$
\begin{aligned}
\eta_{i,j,t} = \mathbf{a_i}^T \mathbf{X}_t \mathbf{b_j} &= \sum_{i'j'} a_{ii'} b_{jj'} x_{i'j't} \\
&= \sum_{i'j'} \alpha^T w_{ii'} \beta^T w_{jj'} x_{i'j't} \\
&= \alpha^T \left(\sum_{i'j'} x_{i'j't} w_{ii'} w_{jj'}^T \right) \beta \\
&= \alpha^T \tilde{X}_{ijt} \beta
\end{aligned}
$$

The number of parameters has gone from $2 \times n \times (n-1)$ to $2 \times q$.

The model is basically of the type considered by Zhou et al. (2013).

Method 2: Social influence regression

$$
\begin{aligned}
\eta_{i,j,t} = \mathbf{a_i}^T \mathbf{X}_t \mathbf{b_j} &= \sum_{i'j'} a_{ii'} b_{jj'} x_{i'j't} \\
&= \sum_{i'j'} \alpha^T w_{ii'} \beta^T w_{jj'} x_{i'j't} \\
&= \alpha^T \left(\sum_{i'j'} x_{i'j't} w_{ii'} w_{jj'}^T \right) \beta \\
&= \alpha^T \tilde{X}_{ijt} \beta
\end{aligned}
$$

The number of parameters has gone from $2 \times n \times (n-1)$ to $2 \times q$.

The model is basically of the type considered by Zhou et al. (2013).

Method 2: Social influence regression

Generalized tensor regression:

$$\mathrm{E}[y_{ij,t}] = f(\eta_{ij,t})$$
$$\eta_{ij,t} = \theta^T z_{ij,t} + \alpha^T \left(\sum x_{i'j't} \mathbf{w}_{ii'} \mathbf{w}_{jj'}^T \right) \beta$$

Covariates:

- $\mathbf{z}_{ij,t} = (\mathrm{ldist}_{ij}, \mathrm{ally}_{ij}, \mathrm{pta}_{ij}, y_{ij,t-1}, y_{ji,t-1})$
- $\mathbf{w}_{ii'} = (\mathrm{ldist}_{ii'}, \mathrm{ally}_{ii'}, \mathrm{pta}_{ii'})$

Method 2: Social influence regression

Dyadic component $\hat{\theta}$:

coef	est	se	z
ldist	-0.1313	0.0006	-219.8854
ally	0.0057	0.0006	8.7728
pta	0.0486	0.0008	64.2236
lag.int	0.8279	0.0019	426.1852
tlag.int	0.2366	0.0019	121.8315

Initiator social influence $\hat{\alpha}$:

coef	est	se	z
ldist	-0.0034	0e+00	-82.5877
ally	0.0048	0e+00	110.5489
pta	-0.0057	1e-04	-100.8460

Target social influence $\hat{\beta}$:

coef	est	se	z
ldist	-0.0092	1e-04	-95.7684
ally	0.0089	1e-04	96.6667
pta	-0.0140	1e-04	-108.2068

Method 2: Social influence regression

Dyadic component $\hat{\theta}$:

coef	est	rse	z
ldist	-0.1313	0.0043	-30.3751
ally	0.0057	0.0055	1.0317
pta	0.0486	0.0056	8.6624
lag.int	0.8279	0.0228	36.3424
tlag.int	0.2366	0.0228	10.3780

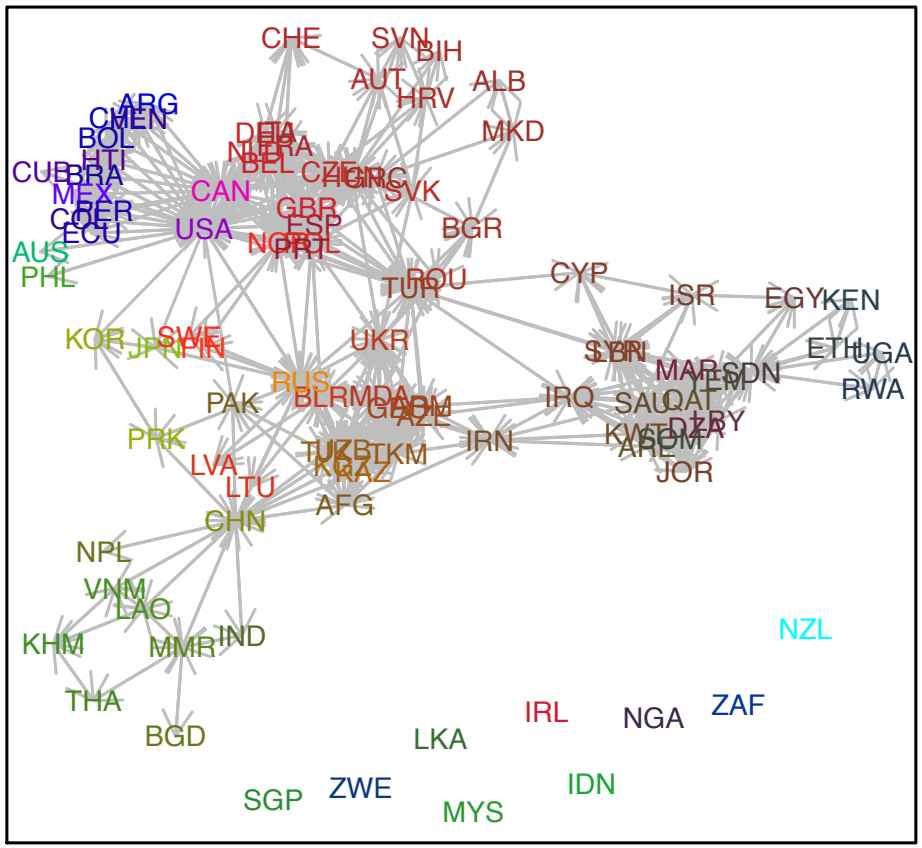
Initiator social influence $\hat{\alpha}$:

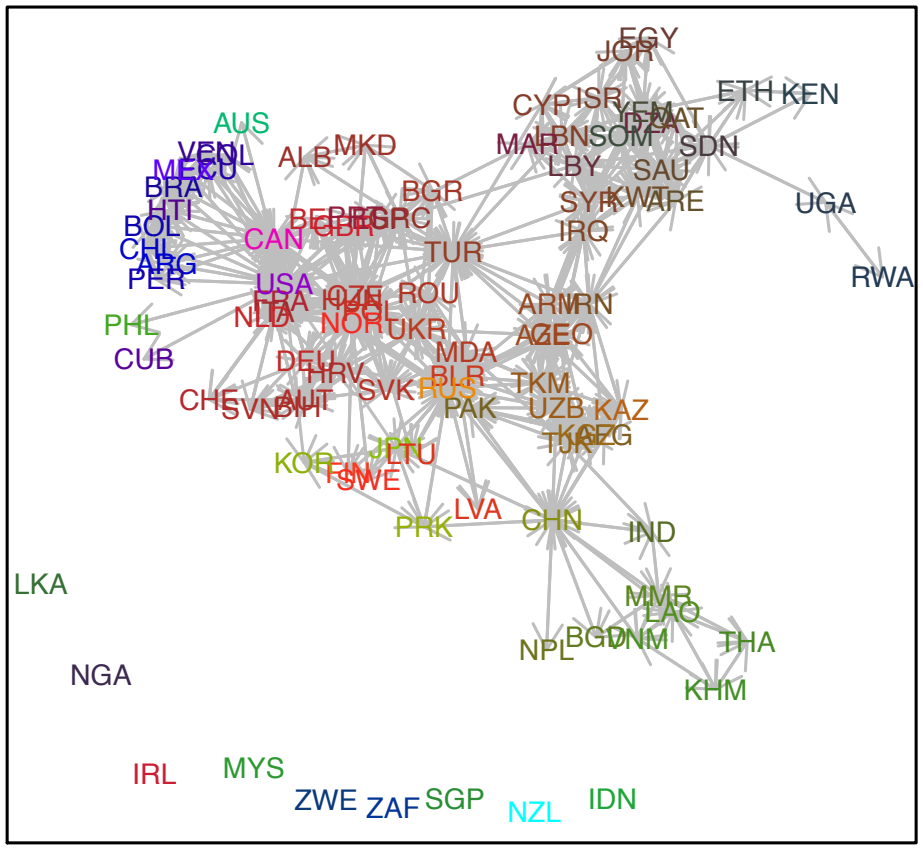
coef	est	rse	z
ldist	-0.0034	3e-04	-10.0735
ally	0.0048	3e-04	15.5362
pta	-0.0057	4e-04	-15.5470

Target social influence $\hat{\beta}$:

coef	est	rse	z
ldist	-0.0092	0.0009	-10.6345
ally	0.0089	0.0008	10.9789
pta	-0.0140	0.0010	-14.5538

Method 2: Social influence regression





$$a_{ii'} = \alpha^T w_{ii'} \quad , \quad b_{jj'} = \beta^T w_{jj'}$$

Discussion

- Multivariate relational data can be represented as a tensor.
- A "simple" way to regress one tensor on another is via a multilinear model.
- For IR data, multilinear models fit better than analogous linear models.
- Parameters give a summary of countries' political relevance to each other.

- Current work and some open questions:
  - Should the model be larger (eBayes/hBayes)?
  - Should the model be smaller (sparsification)?
  - Can prior information/exogenous covariates be used to inform/model $\mathbf{A}$, $\mathbf{B}$?

Discussion

- Multivariate relational data can be represented as a tensor.
- A "simple" way to regress one tensor on another is via a multilinear model.
- For IR data, multilinear models fit better than analogous linear models.
- Parameters give a summary of countries' political relevance to each other.
- Current work and some open questions:
  - Should the model be larger (eBayes/hBayes)?
  - Should the model be smaller (sparsification)?
  - Can prior information/exogenous covariates be used to inform/model $\mathbf{A}$, $\mathbf{B}$?

Discussion

- Multivariate relational data can be represented as a tensor.
- A "simple" way to regress one tensor on another is via a multilinear model.
- For IR data, multilinear models fit better than analogous linear models.
- Parameters give a summary of countries' political relevance to each other.
- Current work and some open questions:
  - Should the model be larger (eBayes/hBayes)?
  - Should the model be smaller (sparsification)?
  - Can prior information/exogenous covariates be used to inform/model $\mathbf{A}$, $\mathbf{B}$?

Discussion

- Multivariate relational data can be represented as a tensor.
- A "simple" way to regress one tensor on another is via a multilinear model.
- For IR data, multilinear models fit better than analogous linear models.
- Parameters give a summary of countries' political relevance to each other.

- Current work and some open questions:
  - Should the model be larger (eBayes/hBayes)?
  - Should the model be smaller (sparsification)?
  - Can prior information/exogenous covariates be used to inform/model $\mathbf{A}$, $\mathbf{B}$?

Discussion

- Multivariate relational data can be represented as a tensor.
- A "simple" way to regress one tensor on another is via a multilinear model.
- For IR data, multilinear models fit better than analogous linear models.
- Parameters give a summary of countries' political relevance to each other.

- Current work and some open questions:
  - Should the model be larger (eBayes/hBayes)?
  - Should the model be smaller (sparsification)?
  - Can prior information/exogenous covariates be used to inform/model $\mathbf{A}$, $\mathbf{B}$?

Discussion

- Multivariate relational data can be represented as a tensor.
- A "simple" way to regress one tensor on another is via a multilinear model.
- For IR data, multilinear models fit better than analogous linear models.
- Parameters give a summary of countries' political relevance to each other.

- Current work and some open questions:
  - Should the model be larger (eBayes/hBayes)?
  - Should the model be smaller (sparsification)?
  - Can prior information/exogenous covariates be used to inform/model **A**, **B**?

Discussion

- Multivariate relational data can be represented as a tensor.
- A "simple" way to regress one tensor on another is via a multilinear model.
- For IR data, multilinear models fit better than analogous linear models.
- Parameters give a summary of countries' political relevance to each other.

- Current work and some open questions:
  - Should the model be larger (eBayes/hBayes)?
  - Should the model be smaller (sparsification)?
  - Can prior information/exogenous covariates be used to inform/model **A**, **B**?

Discussion

- Multivariate relational data can be represented as a tensor.
- A "simple" way to regress one tensor on another is via a multilinear model.
- For IR data, multilinear models fit better than analogous linear models.
- Parameters give a summary of countries' political relevance to each other.

- Current work and some open questions:
  - Should the model be larger (eBayes/hBayes)?
  - Should the model be smaller (sparsification)?
  - Can prior information/exogenous covariates be used to inform/model **A**, **B**?

Discussion

- Multivariate relational data can be represented as a tensor.
- A "simple" way to regress one tensor on another is via a multilinear model.
- For IR data, multilinear models fit better than analogous linear models.
- Parameters give a summary of countries' political relevance to each other.

- Current work and some open questions:
  - Should the model be larger (eBayes/hBayes)?
  - Should the model be smaller (sparsification)?
  - Can prior information/exogenous covariates be used to inform/model **A**, **B**?

Discussion

- Multivariate relational data can be represented as a tensor.
- A "simple" way to regress one tensor on another is via a multilinear model.
- For IR data, multilinear models fit better than analogous linear models.
- Parameters give a summary of countries' political relevance to each other.

- Current work and some open questions:
  - Should the model be larger (eBayes/hBayes)?
  - Should the model be smaller (sparsification)?
  - Can prior information/exogenous covariates be used to inform/model $\mathbf{A}$, $\mathbf{B}$?

Discussion

- Multivariate relational data can be represented as a tensor.
- A "simple" way to regress one tensor on another is via a multilinear model.
- For IR data, multilinear models fit better than analogous linear models.
- Parameters give a summary of countries' political relevance to each other.

- Current work and some open questions:
  - Should the model be larger (eBayes/hBayes)?
  - Should the model be smaller (sparsification)?
  - Can prior information/exogenous covariates be used to inform/model $\mathbf{A}$, $\mathbf{B}$?

References

Basu, S., J. Dunagan, K. Duh, and K.-K. Muniswamy-Reddy (2012). Blr-d: applying bilinear logistic regression to factored diagnosis problems. *ACM SIGOPS Operating Systems Review 45*(3), 31–38.

Gabriel, K. R. (1998). Generalised bilinear regression. *Biometrika 85*(3), 689–700.

Hoff, P. D. (2007). Extending the rank likelihood for semiparametric copula estimation. *Ann. Appl. Stat. 1*(1), 265–283.

Hoff, P. D. (2015). Multilinear tensor regression for longitudinal relational data. *Ann. Appl. Stat. 9*(3), 1169–1193.

Minhas, S., P. D. Hoff, and M. D. Ward (2016). A new approach to analyzing coevolving longitudinal networks in international relations. *Journal of Peace Research*, 0022343316630783.

Potthoff, R. F. and S. N. Roy (1964). A generalized multivariate analysis of variance model useful especially for growth curve problems. *Biometrika 51*, 313–326.

Shi, J. V., Y. Xu, and R. G. Baraniuk (2014). Sparse bilinear logistic regression. *arXiv eprint 1404.4104*.

Zhou, H., L. Li, and H. Zhu (2013). Tensor regression with applications in neuroimaging data analysis. *J. Amer. Statist. Assoc. 108*(502), 540–552.

Model-based TSVD

SVD: For $\mathbf{Y} \in \mathbb{R}^{m_1 \times m_2}$

$$\mathbf{Y} = \mathbf{U}\mathbf{D}\mathbf{V}^T$$

$$\mathbf{Y} \approx \mathbf{U}_{[,1:r]}\mathbf{D}_{[1:r,1:r]}\mathbf{V}^T_{[,1:r]}$$

SVD model:

$$\mathbf{Y} = \mathbf{U}\mathbf{D}\mathbf{V}^T + \mathbf{E}$$

where $\mathbf{U}^T\mathbf{U} = \mathbf{V}^T\mathbf{V} = \mathbf{I}_r$, $\mathbf{D} = \text{diag}(d_1, \ldots, d_r)$, $\{e_{i,j}\} \sim \text{iid } N(0, \sigma^2)$.

TSVD: For $\mathbf{Y} \in \mathbb{R}^{m_1 \times m_2 \times m_3}$

$$\mathbf{Y} = \mathbf{S} \times \{\mathbf{U}, \mathbf{V}, \mathbf{W}\}$$

$$\mathbf{Y} \approx \mathbf{S}_{[1:r_1,1:r_2,1:r_3]} \times \{\mathbf{U}_{[,1:r_1]}, \mathbf{V}_{[,1:r_2]}, \mathbf{W}_{[,1:r_3]}\}$$

TSVD model:

$$\mathbf{Y} = \mathbf{S} \times \{\mathbf{U}, \mathbf{V}, \mathbf{W}\} + \mathbf{E}$$

where $\mathbf{U}^T\mathbf{U} = \mathbf{I}_{r_1}$, $\mathbf{V}^T\mathbf{V} = \mathbf{I}_{r_2}$, $\mathbf{W}^T\mathbf{W} = \mathbf{I}_{r_3}$, $\mathbf{S} \in \mathbf{R}^{r_1 \times r_2 \times r_3}$, $\{e_{i,j}\} \sim \text{iid } N(0, \sigma^2)$.

Model-based TSVD

SVD: For $\mathbf{Y} \in \mathbb{R}^{m_1 \times m_2}$

$$\mathbf{Y} = \mathbf{U}\mathbf{D}\mathbf{V}^T$$

$$\mathbf{Y} \approx \mathbf{U}_{[,1:r]}\mathbf{D}_{[1:r,1:r]}\mathbf{V}^T_{[,1:r]}$$

SVD model:

$$\mathbf{Y} = \mathbf{U}\mathbf{D}\mathbf{V}^T + \mathbf{E}$$

where $\mathbf{U}^T\mathbf{U} = \mathbf{V}^T\mathbf{V} = \mathbf{I}_r$, $\mathbf{D} = \text{diag}(d_1, \ldots, d_r)$, $\{e_{i,j}\} \sim \text{iid } N(0, \sigma^2)$.

TSVD: For $\mathbf{Y} \in \mathbb{R}^{m_1 \times m_2 \times m_3}$

$$\mathbf{Y} = \mathbf{S} \times \{\mathbf{U}, \mathbf{V}, \mathbf{W}\}$$

$$\mathbf{Y} \approx \mathbf{S}_{[1:r_1,1:r_2,1:r_3]} \times \{\mathbf{U}_{[,1:r_1]}, \mathbf{V}_{[,1:r_2]}, \mathbf{W}_{[,1:r_3]}\}$$

TSVD model:

$$\mathbf{Y} = \mathbf{S} \times \{\mathbf{U}, \mathbf{V}, \mathbf{W}\} + \mathbf{E}$$

where $\mathbf{U}^T\mathbf{U} = \mathbf{I}_{r_1}$, $\mathbf{V}^T\mathbf{V} = \mathbf{I}_{r_2}$, $\mathbf{W}^T\mathbf{W} = \mathbf{I}_{r_3}$, $\mathbf{S} \in \mathbf{R}^{r_1 \times r_2 \times r_3}$, $\{e_{i,j}\} \sim \text{iid } N(0, \sigma^2)$.

Model-based TSVD

SVD: For $\mathbf{Y} \in \mathbb{R}^{m_1 \times m_2}$

$$\mathbf{Y} = \mathbf{U}\mathbf{D}\mathbf{V}^T$$

$$\mathbf{Y} \approx \mathbf{U}_{[,1:r]}\mathbf{D}_{[1:r,1:r]}\mathbf{V}^T_{[,1:r]}$$

SVD model:

$$\mathbf{Y} = \mathbf{U}\mathbf{D}\mathbf{V}^T + \mathbf{E}$$

where $\mathbf{U}^T\mathbf{U} = \mathbf{V}^T\mathbf{V} = \mathbf{I}_r$, $\mathbf{D} = \text{diag}(d_1, \ldots, d_r)$, $\{e_{i,j}\} \sim \text{iid } N(0, \sigma^2)$.

TSVD: For $\mathbf{Y} \in \mathbb{R}^{m_1 \times m_2 \times m_3}$

$$\mathbf{Y} = \mathbf{S} \times \{\mathbf{U}, \mathbf{V}, \mathbf{W}\}$$

$$\mathbf{Y} \approx \mathbf{S}_{[1:r_1,1:r_2,1:r_3]} \times \{\mathbf{U}_{[,1:r_1]}, \mathbf{V}_{[,1:r_2]}, \mathbf{W}_{[,1:r_3]}\}$$

TSVD model:

$$\mathbf{Y} = \mathbf{S} \times \{\mathbf{U}, \mathbf{V}, \mathbf{W}\} + \mathbf{E}$$

where $\mathbf{U}^T\mathbf{U} = \mathbf{I}_{r_1}$, $\mathbf{V}^T\mathbf{V} = \mathbf{I}_{r_2}$, $\mathbf{W}^T\mathbf{W} = \mathbf{I}_{r_3}$, $\mathbf{S} \in \mathbf{R}^{r_1 \times r_2 \times r_3}$, $\{e_{i,j}\} \sim \text{iid } N(0, \sigma^2)$.

Model-based TSVD

SVD: For $\mathbf{Y} \in \mathbb{R}^{m_1 \times m_2}$

$$\mathbf{Y} = \mathbf{U}\mathbf{D}\mathbf{V}^T$$

$$\mathbf{Y} \approx \mathbf{U}_{[,1:r]}\mathbf{D}_{[1:r,1:r]}\mathbf{V}^T_{[,1:r]}$$

SVD model:

$$\mathbf{Y} = \mathbf{U}\mathbf{D}\mathbf{V}^T + \mathbf{E}$$

where $\mathbf{U}^T\mathbf{U} = \mathbf{V}^T\mathbf{V} = \mathbf{I}_r$, $\mathbf{D} = \text{diag}(d_1, \ldots, d_r)$, $\{e_{i,j}\} \sim \text{iid } N(0, \sigma^2)$.

TSVD: For $\mathbf{Y} \in \mathbb{R}^{m_1 \times m_2 \times m_3}$

$$\mathbf{Y} = \mathbf{S} \times \{\mathbf{U}, \mathbf{V}, \mathbf{W}\}$$

$$\mathbf{Y} \approx \mathbf{S}_{[1:r_1,1:r_2,1:r_3]} \times \{\mathbf{U}_{[,1:r_1]}, \mathbf{V}_{[,1:r_2]}, \mathbf{W}_{[,1:r_3]}\}$$

TSVD model:

$$\mathbf{Y} = \mathbf{S} \times \{\mathbf{U}, \mathbf{V}, \mathbf{W}\} + \mathbf{E}$$

where $\mathbf{U}^T\mathbf{U} = \mathbf{I}_{r_1}$, $\mathbf{V}^T\mathbf{V} = \mathbf{I}_{r_2}$, $\mathbf{W}^T\mathbf{W} = \mathbf{I}_{r_3}$, $\mathbf{S} \in \mathbf{R}^{r_1 \times r_2 \times r_3}$, $\{e_{i,j}\} \sim \text{iid } N(0, \sigma^2)$.

Model-based TSVD (Hoff [2013])

$$\mathbf{Y} = \sigma \mathbf{S} \times \{\mathbf{U}, \mathbf{V}, \mathbf{W}\} + \sigma \mathbf{E}$$

This model is invariant under transformations of the form

$$g : \mathbf{Y} \to a\mathbf{Y} \times \{\mathbf{A}, \mathbf{B}, \mathbf{C}\}$$
$$\bar{g} : (\sigma, \mathbf{U}, \mathbf{V}, \mathbf{W}, \mathbf{S}) \to (a\sigma, \mathbf{AU}, \mathbf{BV}, \mathbf{CW}, \mathbf{S})$$

for $a > 0$, $\mathbf{A}^T\mathbf{A} = \mathbf{I}_{r_1}$, $\mathbf{B}^T\mathbf{B} = \mathbf{I}_{r_2}$, $\mathbf{C}^T\mathbf{C} = \mathbf{I}_{r_3}$.

For known $\mathbf{S}$, a best equivariant estimator exists and is a Bayes rule under

$$\pi(\sigma, \mathbf{U}, \mathbf{V}, \mathbf{W}) = \frac{1}{\sigma}.$$

For unknown $\mathbf{S}$, what is a reasonable prior?

Model-based TSVD (Hoff [2013])

$$\mathbf{Y} = \sigma \mathbf{S} \times \{\mathbf{U}, \mathbf{V}, \mathbf{W}\} + \sigma \mathbf{E}$$

This model is invariant under transformations of the form

$$g : \mathbf{Y} \to a\mathbf{Y} \times \{\mathbf{A}, \mathbf{B}, \mathbf{C}\}$$
$$\bar{g} : (\sigma, \mathbf{U}, \mathbf{V}, \mathbf{W}, \mathbf{S}) \to (a\sigma, \mathbf{AU}, \mathbf{BV}, \mathbf{CW}, \mathbf{S})$$

for $a > 0$, $\mathbf{A}^T\mathbf{A} = \mathbf{I}_{r_1}$, $\mathbf{B}^T\mathbf{B} = \mathbf{I}_{r_2}$, $\mathbf{C}^T\mathbf{C} = \mathbf{I}_{r_3}$.

For known $\mathbf{S}$, a best equivariant estimator exists and is a Bayes rule under

$$\pi(\sigma, \mathbf{U}, \mathbf{V}, \mathbf{W}) = \frac{1}{\sigma}.$$

For unknown $\mathbf{S}$, what is a reasonable prior?

Model-based TSVD (Hoff [2013])

$$\mathbf{Y} = \sigma \mathbf{S} \times \{\mathbf{U}, \mathbf{V}, \mathbf{W}\} + \sigma \mathbf{E}$$

This model is invariant under transformations of the form

$$g : \mathbf{Y} \to a\mathbf{Y} \times \{\mathbf{A}, \mathbf{B}, \mathbf{C}\}$$
$$\bar{g} : (\sigma, \mathbf{U}, \mathbf{V}, \mathbf{W}, \mathbf{S}) \to (a\sigma, \mathbf{AU}, \mathbf{BV}, \mathbf{CW}, \mathbf{S})$$

for $a > 0$, $\mathbf{A}^T\mathbf{A} = \mathbf{I}_{r_1}$, $\mathbf{B}^T\mathbf{B} = \mathbf{I}_{r_2}$, $\mathbf{C}^T\mathbf{C} = \mathbf{I}_{r_3}$.

For known $\mathbf{S}$, a best equivariant estimator exists and is a Bayes rule under

$$\pi(\sigma, \mathbf{U}, \mathbf{V}, \mathbf{W}) = \frac{1}{\sigma}.$$

For unknown $\mathbf{S}$, what is a reasonable prior?

Model-based TSVD (Hoff [2013])

$$\mathbf{Y} = \sigma \mathbf{S} \times \{\mathbf{U}, \mathbf{V}, \mathbf{W}\} + \sigma \mathbf{E}$$

This model is invariant under transformations of the form

$$g : \mathbf{Y} \to a\mathbf{Y} \times \{\mathbf{A}, \mathbf{B}, \mathbf{C}\}$$
$$\bar{g} : (\sigma, \mathbf{U}, \mathbf{V}, \mathbf{W}, \mathbf{S}) \to (a\sigma, \mathbf{AU}, \mathbf{BV}, \mathbf{CW}, \mathbf{S})$$

for $a > 0$, $\mathbf{A}^T\mathbf{A} = \mathbf{I}_{r_1}$, $\mathbf{B}^T\mathbf{B} = \mathbf{I}_{r_2}$, $\mathbf{C}^T\mathbf{C} = \mathbf{I}_{r_3}$.

For known $\mathbf{S}$, a best equivariant estimator exists and is a Bayes rule under

$$\pi(\sigma, \mathbf{U}, \mathbf{V}, \mathbf{W}) = \frac{1}{\sigma}.$$

For unknown $\mathbf{S}$, what is a reasonable prior?

Prior over the core array (Hoff [2013])

Hierarchical normal prior:

$$\mathbf{s} = \text{vec}(\mathbf{S}) \sim N(\mathbf{0}, \boldsymbol{\Psi}_3 \otimes \boldsymbol{\Psi}_2 \otimes \boldsymbol{\Psi}_1) \ , \ (\boldsymbol{\Psi}_1, \boldsymbol{\Psi}_2, \boldsymbol{\Psi}_3) \sim \pi$$

Marginal distribution: $\mathbf{y} = \text{vec}(\mathbf{Y})$ is normal mean zero with

$$\text{Cov}[\mathbf{y}] = (\mathbf{W}\boldsymbol{\Psi}_3\mathbf{W}^T \otimes \mathbf{V}\boldsymbol{\Psi}_2\mathbf{V}^T \otimes \mathbf{U}\boldsymbol{\Psi}_1\mathbf{U}^T) + \mathbf{I}$$

- Eigenvectors of $\boldsymbol{\Psi}$'s are confounded with those of $\mathbf{U}, \mathbf{V}, \mathbf{W}$;
- scales of the $\boldsymbol{\Psi}$'s are confounded with each other.

Identifiable parametrization:

$$\boldsymbol{\Psi}_3 \otimes \boldsymbol{\Psi}_2 \otimes \boldsymbol{\Psi}_1 = \tau^2(\boldsymbol{\Lambda}_3 \otimes \boldsymbol{\Lambda}_2 \otimes \boldsymbol{\Lambda}_1),$$

where $\tau^2 > 0$, $\boldsymbol{\Lambda}_k$ is diagonal with $\text{tr}(\boldsymbol{\Lambda}_k) = 1$.

Prior over the core array (Hoff [2013])

Hierarchical normal prior:

$$\mathbf{s} = \text{vec}(\mathbf{S}) \sim N(\mathbf{0}, \boldsymbol{\Psi}_3 \otimes \boldsymbol{\Psi}_2 \otimes \boldsymbol{\Psi}_1) \ , \ (\boldsymbol{\Psi}_1, \boldsymbol{\Psi}_2, \boldsymbol{\Psi}_3) \sim \pi$$

Marginal distribution: $\mathbf{y} = \text{vec}(\mathbf{Y})$ is normal mean zero with

$$\text{Cov}[\mathbf{y}] = (\mathbf{W}\boldsymbol{\Psi}_3\mathbf{W}^T \otimes \mathbf{V}\boldsymbol{\Psi}_2\mathbf{V}^T \otimes \mathbf{U}\boldsymbol{\Psi}_1\mathbf{U}^T) + \mathbf{I}$$

- Eigenvectors of $\boldsymbol{\Psi}$'s are confounded with those of $\mathbf{U}, \mathbf{V}, \mathbf{W}$;
- scales of the $\boldsymbol{\Psi}$'s are confounded with each other.

Identifiable parametrization:

$$\boldsymbol{\Psi}_3 \otimes \boldsymbol{\Psi}_2 \otimes \boldsymbol{\Psi}_1 = \tau^2(\boldsymbol{\Lambda}_3 \otimes \boldsymbol{\Lambda}_2 \otimes \boldsymbol{\Lambda}_1),$$

where $\tau^2 > 0$, $\boldsymbol{\Lambda}_k$ is diagonal with $\text{tr}(\boldsymbol{\Lambda}_k) = 1$.

Prior over the core array (Hoff [2013])

Hierarchical normal prior:

$$\mathbf{s} = \text{vec}(\mathbf{S}) \sim N(\mathbf{0}, \boldsymbol{\Psi}_3 \otimes \boldsymbol{\Psi}_2 \otimes \boldsymbol{\Psi}_1) \ , \ (\boldsymbol{\Psi}_1, \boldsymbol{\Psi}_2, \boldsymbol{\Psi}_3) \sim \pi$$

Marginal distribution: $\mathbf{y} = \text{vec}(\mathbf{Y})$ is normal mean zero with

$$\text{Cov}[\mathbf{y}] = (\mathbf{W}\boldsymbol{\Psi}_3\mathbf{W}^T \otimes \mathbf{V}\boldsymbol{\Psi}_2\mathbf{V}^T \otimes \mathbf{U}\boldsymbol{\Psi}_1\mathbf{U}^T) + \mathbf{I}$$

- Eigenvectors of $\boldsymbol{\Psi}$'s are confounded with those of $\mathbf{U}, \mathbf{V}, \mathbf{W}$;
- scales of the $\boldsymbol{\Psi}$'s are confounded with each other.

Identifiable parametrization:

$$\boldsymbol{\Psi}_3 \otimes \boldsymbol{\Psi}_2 \otimes \boldsymbol{\Psi}_1 = \tau^2(\boldsymbol{\Lambda}_3 \otimes \boldsymbol{\Lambda}_2 \otimes \boldsymbol{\Lambda}_1),$$

where $\tau^2 > 0$, $\boldsymbol{\Lambda}_k$ is diagonal with $\text{tr}(\boldsymbol{\Lambda}_k) = 1$.

Prior over the core array (Hoff [2013])

Hierarchical normal prior:

$$\mathbf{s} = \text{vec}(\mathbf{S}) \sim N(\mathbf{0}, \boldsymbol{\Psi}_3 \otimes \boldsymbol{\Psi}_2 \otimes \boldsymbol{\Psi}_1) \ , \ (\boldsymbol{\Psi}_1, \boldsymbol{\Psi}_2, \boldsymbol{\Psi}_3) \sim \pi$$

Marginal distribution: $\mathbf{y} = \text{vec}(\mathbf{Y})$ is normal mean zero with

$$\text{Cov}[\mathbf{y}] = (\mathbf{W}\boldsymbol{\Psi}_3\mathbf{W}^T \otimes \mathbf{V}\boldsymbol{\Psi}_2\mathbf{V}^T \otimes \mathbf{U}\boldsymbol{\Psi}_1\mathbf{U}^T) + \mathbf{I}$$

- Eigenvectors of $\boldsymbol{\Psi}$'s are confounded with those of $\mathbf{U}, \mathbf{V}, \mathbf{W}$;
- scales of the $\boldsymbol{\Psi}$'s are confounded with each other.

Identifiable parametrization:

$$\boldsymbol{\Psi}_3 \otimes \boldsymbol{\Psi}_2 \otimes \boldsymbol{\Psi}_1 = \tau^2(\boldsymbol{\Lambda}_3 \otimes \boldsymbol{\Lambda}_2 \otimes \boldsymbol{\Lambda}_1),$$

where $\tau^2 > 0$, $\boldsymbol{\Lambda}_k$ is diagonal with $\text{tr}(\boldsymbol{\Lambda}_k) = 1$.

Prior over the core array (Hoff [2013])

Hierarchical normal prior:

$$\mathbf{s} = \text{vec}(\mathbf{S}) \sim N(\mathbf{0}, \boldsymbol{\Psi}_3 \otimes \boldsymbol{\Psi}_2 \otimes \boldsymbol{\Psi}_1) \ , \ (\boldsymbol{\Psi}_1, \boldsymbol{\Psi}_2, \boldsymbol{\Psi}_3) \sim \pi$$

Marginal distribution: $\mathbf{y} = \text{vec}(\mathbf{Y})$ is normal mean zero with

$$\text{Cov}[\mathbf{y}] = (\mathbf{W}\boldsymbol{\Psi}_3\mathbf{W}^T \otimes \mathbf{V}\boldsymbol{\Psi}_2\mathbf{V}^T \otimes \mathbf{U}\boldsymbol{\Psi}_1\mathbf{U}^T) + \mathbf{I}$$

- Eigenvectors of $\boldsymbol{\Psi}$'s are confounded with those of $\mathbf{U}, \mathbf{V}, \mathbf{W}$;
- scales of the $\boldsymbol{\Psi}$'s are confounded with each other.

Identifiable parametrization:

$$\boldsymbol{\Psi}_3 \otimes \boldsymbol{\Psi}_2 \otimes \boldsymbol{\Psi}_1 = \tau^2(\boldsymbol{\Lambda}_3 \otimes \boldsymbol{\Lambda}_2 \otimes \boldsymbol{\Lambda}_1),$$

where $\tau^2 > 0$, $\boldsymbol{\Lambda}_k$ is diagonal with $\text{tr}(\boldsymbol{\Lambda}_k) = 1$.

Model and prior (Hoff [2013])

Mean model

$$\mathbf{Y} = \sigma \mathbf{S} \times \{\mathbf{U}, \mathbf{V}, \mathbf{W}\} + \sigma \mathbf{E}$$

$$\text{Cov}[\text{vec}(\mathbf{S})] = \tau^2 \mathbf{\Lambda}_3 \otimes \mathbf{\Lambda}_2 \otimes \mathbf{\Lambda}_1$$

Covariance model

$$\mathbf{y} \sim N(\mathbf{0}, \mathbf{\Sigma})$$

$$\mathbf{\Sigma} = \sigma^2(\tau^2 \mathbf{W}\mathbf{\Lambda}_3\mathbf{W}^T \otimes \mathbf{V}\mathbf{\Lambda}_2\mathbf{V}^T \otimes \mathbf{U}\mathbf{\Lambda}_1\mathbf{U}^T + \mathbf{I})$$

Priors

- $\pi(\sigma) = 1/\sigma$
- $\mathbf{U}, \mathbf{V}, \mathbf{W}, \mathbf{\Lambda}_1, \mathbf{\Lambda}_2, \mathbf{\Lambda}_3$ uniform
- $1/\tau^2 \sim \text{gamma}(\nu_0/2, \nu_0\tau_0^2/2)$

Model and prior (Hoff [2013])

Mean model

$$\mathbf{Y} = \sigma \mathbf{S} \times \{\mathbf{U}, \mathbf{V}, \mathbf{W}\} + \sigma \mathbf{E}$$

$$\text{Cov}[\text{vec}(\mathbf{S})] = \tau^2 \mathbf{\Lambda}_3 \otimes \mathbf{\Lambda}_2 \otimes \mathbf{\Lambda}_1$$

Covariance model

$$\mathbf{y} \sim N(\mathbf{0}, \mathbf{\Sigma})$$

$$\mathbf{\Sigma} = \sigma^2(\tau^2 \mathbf{W}\mathbf{\Lambda}_3\mathbf{W}^T \otimes \mathbf{V}\mathbf{\Lambda}_2\mathbf{V}^T \otimes \mathbf{U}\mathbf{\Lambda}_1\mathbf{U}^T + \mathbf{I})$$

Priors

- $\pi(\sigma) = 1/\sigma$
- $\mathbf{U}, \mathbf{V}, \mathbf{W}, \mathbf{\Lambda}_1, \mathbf{\Lambda}_2, \mathbf{\Lambda}_3$ uniform
- $1/\tau^2 \sim \text{gamma}(\nu_0/2, \nu_0\tau_0^2/2)$

Model and prior (Hoff [2013])

Mean model

$$\mathbf{Y} = \sigma \mathbf{S} \times \{\mathbf{U}, \mathbf{V}, \mathbf{W}\} + \sigma \mathbf{E}$$

$$\text{Cov}[\text{vec}(\mathbf{S})] = \tau^2 \mathbf{\Lambda}_3 \otimes \mathbf{\Lambda}_2 \otimes \mathbf{\Lambda}_1$$

Covariance model

$$\mathbf{y} \sim N(\mathbf{0}, \mathbf{\Sigma})$$

$$\mathbf{\Sigma} = \sigma^2(\tau^2 \mathbf{W}\mathbf{\Lambda}_3\mathbf{W}^T \otimes \mathbf{V}\mathbf{\Lambda}_2\mathbf{V}^T \otimes \mathbf{U}\mathbf{\Lambda}_1\mathbf{U}^T + \mathbf{I})$$

Priors

- $\pi(\sigma) = 1/\sigma$
- $\mathbf{U}, \mathbf{V}, \mathbf{W}, \mathbf{\Lambda}_1, \mathbf{\Lambda}_2, \mathbf{\Lambda}_3$ uniform
- $1/\tau^2 \sim \text{gamma}(\nu_0/2, \nu_0\tau_0^2/2)$

Simulation study (Hoff [2013])

$$\mathbf{Y} = \mathbf{M} + \sigma\mathbf{E}, \quad \mathbf{Y} \in \mathbb{R}^{60\times 50\times 40}$$

Task: Estimate $\mathbf{M}$ with $\text{rank}(\mathbf{M}) = (r_1, r_2, r_3)$

Estimate $\mathbf{M} = \sigma\mathbf{S} \times \{\mathbf{U}, \mathbf{V}, \mathbf{W}\}$, $\mathbf{S} \in \mathbb{R}^{r_1\times r_2\times r_3}$

Conditions: $\mathbf{s} \sim N_{r_1 r_2 r_3}(\mathbf{0}, \tau_0^2\mathbf{I})$

- low or high rank: $\mathbf{r} = (6, 5, 4)$ or $\mathbf{r} = (30, 25, 20)$;
- low or high noise: τ_0^2 large or small.

Estimators:

$\hat{\mathbf{M}}_{OLS}$: available via ALS

$\hat{\mathbf{M}}_{HOM}$: homoscedastic Bayes estimator, $\mathbf{s} \sim N(\mathbf{0}, \tau^2\mathbf{I})$ ("oracle")

$\hat{\mathbf{M}}_{HET}$: heteroscedastic Bayes estimator, $\mathbf{s} \sim N(\mathbf{0}, \tau^2\mathbf{\Lambda}_3 \otimes \mathbf{\Lambda}_2 \otimes \mathbf{\Lambda}_1)$

Simulation study (Hoff [2013])

$$\mathbf{Y} = \mathbf{M} + \sigma\mathbf{E}, \quad \mathbf{Y} \in \mathbb{R}^{60\times 50\times 40}$$

Task: Estimate $\mathbf{M}$ with $\text{rank}(\mathbf{M}) = (r_1, r_2, r_3)$

Estimate $\mathbf{M} = \sigma\mathbf{S} \times \{\mathbf{U}, \mathbf{V}, \mathbf{W}\}$, $\mathbf{S} \in \mathbb{R}^{r_1\times r_2\times r_3}$

Conditions: $\mathbf{s} \sim N_{r_1 r_2 r_3}(\mathbf{0}, \tau_0^2\mathbf{I})$

- low or high rank: $\mathbf{r} = (6, 5, 4)$ or $\mathbf{r} = (30, 25, 20)$;
- low or high noise: τ_0^2 large or small.

Estimators:

$\hat{\mathbf{M}}_{OLS}$: available via ALS

$\hat{\mathbf{M}}_{HOM}$: homoscedastic Bayes estimator, $\mathbf{s} \sim N(\mathbf{0}, \tau^2\mathbf{I})$ ("oracle")

$\hat{\mathbf{M}}_{HET}$: heteroscedastic Bayes estimator, $\mathbf{s} \sim N(\mathbf{0}, \tau^2\mathbf{\Lambda}_3 \otimes \mathbf{\Lambda}_2 \otimes \mathbf{\Lambda}_1)$

Simulation study (Hoff [2013])

$$\mathbf{Y} = \mathbf{M} + \sigma\mathbf{E}, \quad \mathbf{Y} \in \mathbb{R}^{60\times 50\times 40}$$

Task: Estimate $\mathbf{M}$ with $\text{rank}(\mathbf{M}) = (r_1, r_2, r_3)$

Estimate $\mathbf{M} = \sigma\mathbf{S} \times \{\mathbf{U}, \mathbf{V}, \mathbf{W}\}$, $\mathbf{S} \in \mathbb{R}^{r_1\times r_2\times r_3}$

Conditions: $\mathbf{s} \sim N_{r_1 r_2 r_3}(\mathbf{0}, \tau_0^2\mathbf{I})$

- low or high rank: $\mathbf{r} = (6, 5, 4)$ or $\mathbf{r} = (30, 25, 20)$;
- low or high noise: τ_0^2 large or small.

Estimators:

$\hat{\mathbf{M}}_{OLS}$: available via ALS

$\hat{\mathbf{M}}_{HOM}$: homoscedastic Bayes estimator, $\mathbf{s} \sim N(\mathbf{0}, \tau^2\mathbf{I})$ ("oracle")

$\hat{\mathbf{M}}_{HET}$: heteroscedastic Bayes estimator, $\mathbf{s} \sim N(\mathbf{0}, \tau^2\mathbf{\Lambda}_3 \otimes \mathbf{\Lambda}_2 \otimes \mathbf{\Lambda}_1)$

Simulation study (Hoff [2013])

$$\mathbf{Y} = \mathbf{M} + \sigma\mathbf{E}, \quad \mathbf{Y} \in \mathbb{R}^{60\times 50\times 40}$$

Task: Estimate $\mathbf{M}$ with $\text{rank}(\mathbf{M}) = (r_1, r_2, r_3)$

Estimate $\mathbf{M} = \sigma\mathbf{S} \times \{\mathbf{U}, \mathbf{V}, \mathbf{W}\}$, $\mathbf{S} \in \mathbb{R}^{r_1\times r_2\times r_3}$

Conditions: $\mathbf{s} \sim N_{r_1 r_2 r_3}(\mathbf{0}, \tau_0^2\mathbf{I})$

- low or high rank: $\mathbf{r} = (6, 5, 4)$ or $\mathbf{r} = (30, 25, 20)$;
- low or high noise: τ_0^2 large or small.

Estimators:

$\hat{\mathbf{M}}_{OLS}$: available via ALS

$\hat{\mathbf{M}}_{HOM}$: homoscedastic Bayes estimator, $\mathbf{s} \sim N(\mathbf{0}, \tau^2\mathbf{I})$ ("oracle")

$\hat{\mathbf{M}}_{HET}$: heteroscedastic Bayes estimator, $\mathbf{s} \sim N(\mathbf{0}, \tau^2\mathbf{\Lambda}_3 \otimes \mathbf{\Lambda}_2 \otimes \mathbf{\Lambda}_1)$

Simulation study (Hoff [2013])

$$\mathbf{Y} = \mathbf{M} + \sigma\mathbf{E}, \quad \mathbf{Y} \in \mathbb{R}^{60\times 50\times 40}$$

Task: Estimate $\mathbf{M}$ with $\text{rank}(\mathbf{M}) = (r_1, r_2, r_3)$

Estimate $\mathbf{M} = \sigma\mathbf{S} \times \{\mathbf{U}, \mathbf{V}, \mathbf{W}\}$, $\mathbf{S} \in \mathbb{R}^{r_1\times r_2\times r_3}$

Conditions: $\mathbf{s} \sim N_{r_1 r_2 r_3}(\mathbf{0}, \tau_0^2\mathbf{I})$

- low or high rank: $\mathbf{r} = (6, 5, 4)$ or $\mathbf{r} = (30, 25, 20)$;
- low or high noise: τ_0^2 large or small.

Estimators:

$\hat{\mathbf{M}}_{OLS}$: available via ALS

$\hat{\mathbf{M}}_{HOM}$: homoscedastic Bayes estimator, $\mathbf{s} \sim N(\mathbf{0}, \tau^2\mathbf{I})$ ("oracle")

$\hat{\mathbf{M}}_{HET}$: heteroscedastic Bayes estimator, $\mathbf{s} \sim N(\mathbf{0}, \tau^2\mathbf{\Lambda}_3 \otimes \mathbf{\Lambda}_2 \otimes \mathbf{\Lambda}_1)$

Simulation study (Hoff [2013])

$$\mathbf{Y} = \mathbf{M} + \sigma\mathbf{E}, \quad \mathbf{Y} \in \mathbb{R}^{60\times 50\times 40}$$

Task: Estimate $\mathbf{M}$ with $\text{rank}(\mathbf{M}) = (r_1, r_2, r_3)$

Estimate $\mathbf{M} = \sigma\mathbf{S} \times \{\mathbf{U}, \mathbf{V}, \mathbf{W}\}$, $\mathbf{S} \in \mathbb{R}^{r_1\times r_2\times r_3}$

Conditions: $\mathbf{s} \sim N_{r_1 r_2 r_3}(\mathbf{0}, \tau_0^2\mathbf{I})$

- low or high rank: $\mathbf{r} = (6, 5, 4)$ or $\mathbf{r} = (30, 25, 20)$;
- low or high noise: τ_0^2 large or small.

Estimators:

$\hat{\mathbf{M}}_{OLS}$: available via ALS

$\hat{\mathbf{M}}_{HOM}$: homoscedastic Bayes estimator, $\mathbf{s} \sim N(\mathbf{0}, \tau^2\mathbf{I})$ ("oracle")

$\hat{\mathbf{M}}_{HET}$: heteroscedastic Bayes estimator, $\mathbf{s} \sim N(\mathbf{0}, \tau^2\boldsymbol{\Lambda}_3 \otimes \boldsymbol{\Lambda}_2 \otimes \boldsymbol{\Lambda}_1)$

Simulation study (Hoff [2013])

$$\mathbf{Y} = \mathbf{M} + \sigma\mathbf{E}, \quad \mathbf{Y} \in \mathbb{R}^{60\times 50\times 40}$$

Task: Estimate $\mathbf{M}$ with $\text{rank}(\mathbf{M}) = (r_1, r_2, r_3)$

Estimate $\mathbf{M} = \sigma\mathbf{S} \times \{\mathbf{U}, \mathbf{V}, \mathbf{W}\}$, $\mathbf{S} \in \mathbb{R}^{r_1\times r_2\times r_3}$

Conditions: $\mathbf{s} \sim N_{r_1 r_2 r_3}(\mathbf{0}, \tau_0^2\mathbf{I})$

- low or high rank: $\mathbf{r} = (6, 5, 4)$ or $\mathbf{r} = (30, 25, 20)$;
- low or high noise: τ_0^2 large or small.

Estimators:

$\hat{\mathbf{M}}_{OLS}$: available via ALS

$\hat{\mathbf{M}}_{HOM}$: homoscedastic Bayes estimator, $\mathbf{s} \sim N(\mathbf{0}, \tau^2\mathbf{I})$ ("oracle")

$\hat{\mathbf{M}}_{HET}$: heteroscedastic Bayes estimator, $\mathbf{s} \sim N(\mathbf{0}, \tau^2\mathbf{\Lambda}_3 \otimes \mathbf{\Lambda}_2 \otimes \mathbf{\Lambda}_1)$

Simulation study (Hoff [2013])

$$\mathbf{Y} = \mathbf{M} + \sigma\mathbf{E}, \quad \mathbf{Y} \in \mathbb{R}^{60\times 50\times 40}$$

Task: Estimate $\mathbf{M}$ with $\text{rank}(\mathbf{M}) = (r_1, r_2, r_3)$

Estimate $\mathbf{M} = \sigma\mathbf{S} \times \{\mathbf{U}, \mathbf{V}, \mathbf{W}\}$, $\mathbf{S} \in \mathbb{R}^{r_1\times r_2\times r_3}$

Conditions: $\mathbf{s} \sim N_{r_1 r_2 r_3}(\mathbf{0}, \tau_0^2\mathbf{I})$

- low or high rank: $\mathbf{r} = (6, 5, 4)$ or $\mathbf{r} = (30, 25, 20)$;
- low or high noise: τ_0^2 large or small.

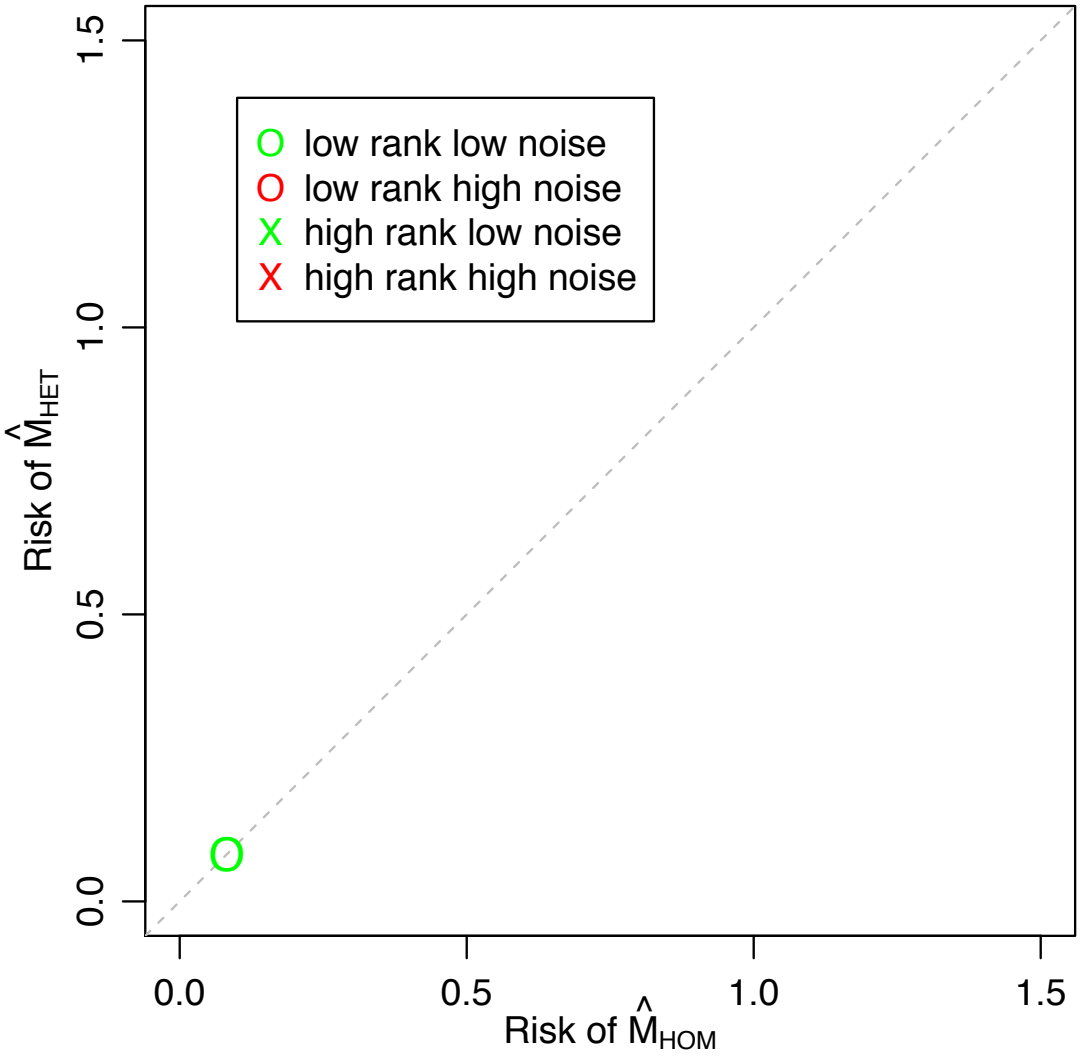
Estimators:

$\hat{\mathbf{M}}_{OLS}$: available via ALS

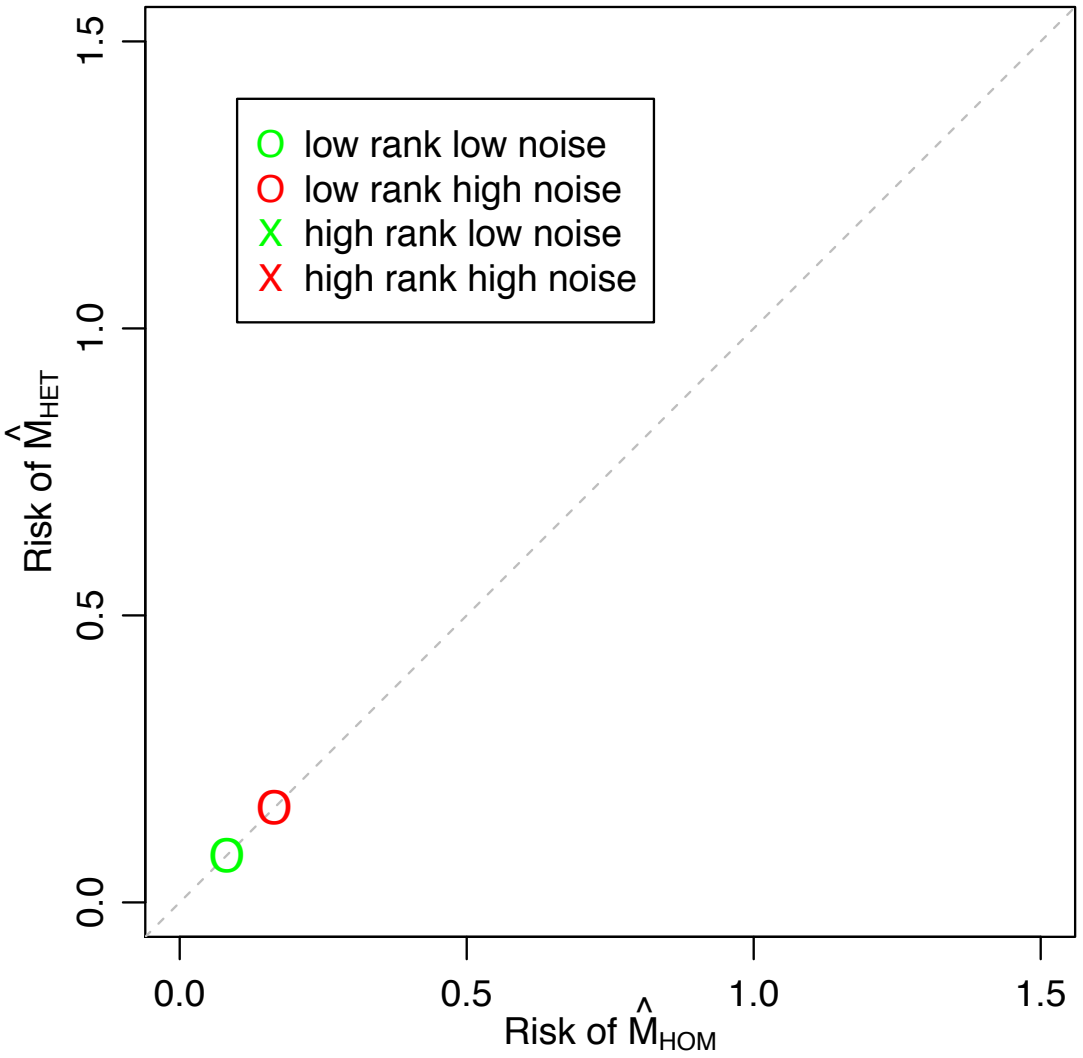
$\hat{\mathbf{M}}_{HOM}$: homoscedastic Bayes estimator, $\mathbf{s} \sim N(\mathbf{0}, \tau^2\mathbf{I})$ ("oracle")

$\hat{\mathbf{M}}_{HET}$: heteroscedastic Bayes estimator, $\mathbf{s} \sim N(\mathbf{0}, \tau^2\mathbf{\Lambda}_3 \otimes \mathbf{\Lambda}_2 \otimes \mathbf{\Lambda}_1)$

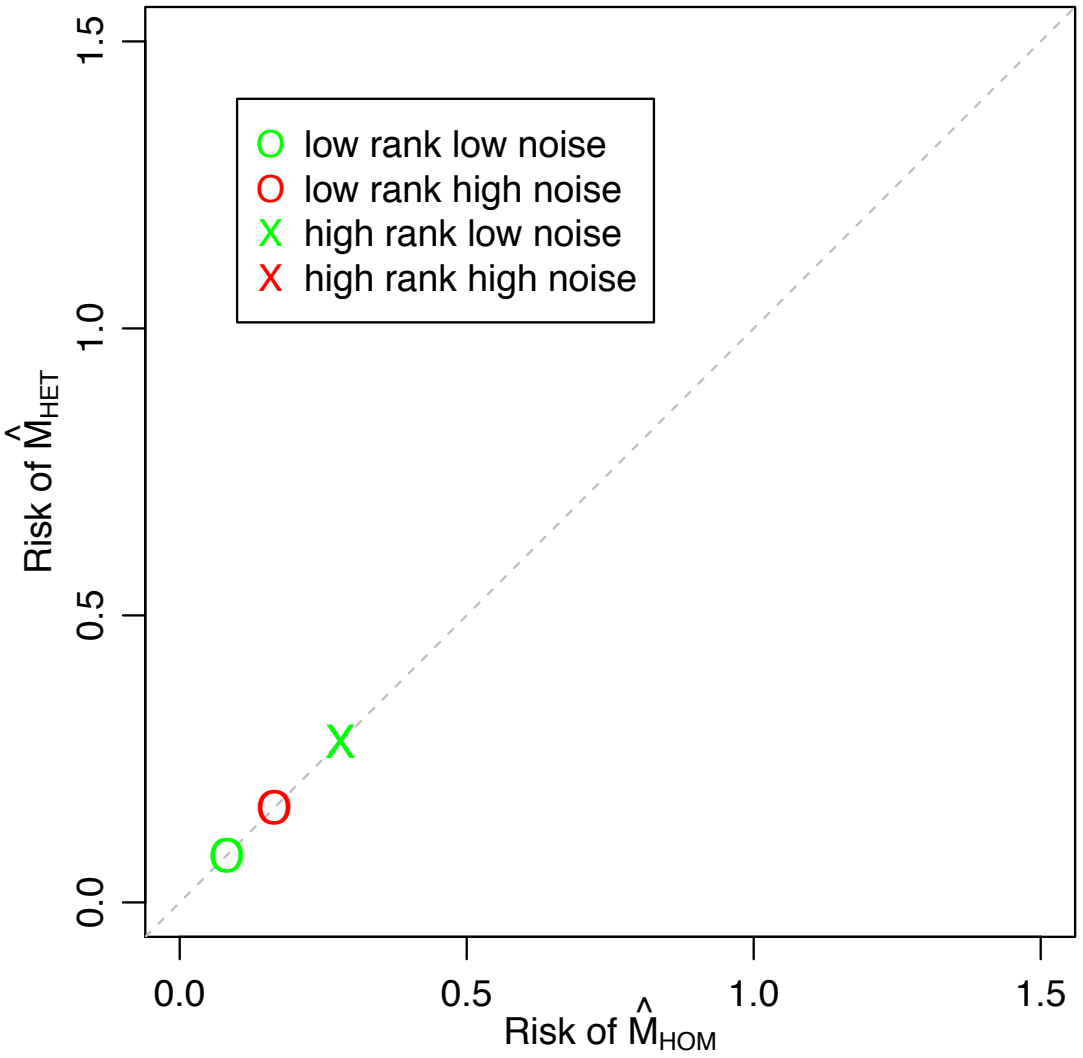
Risks


low rank low noise
low rank high noise
high rank low noise
high rank high noise
Risk of $\hat{M}_{HET}$
Risk of $\hat{M}_{HOM}$
0.0
0.5
1.0
1.5

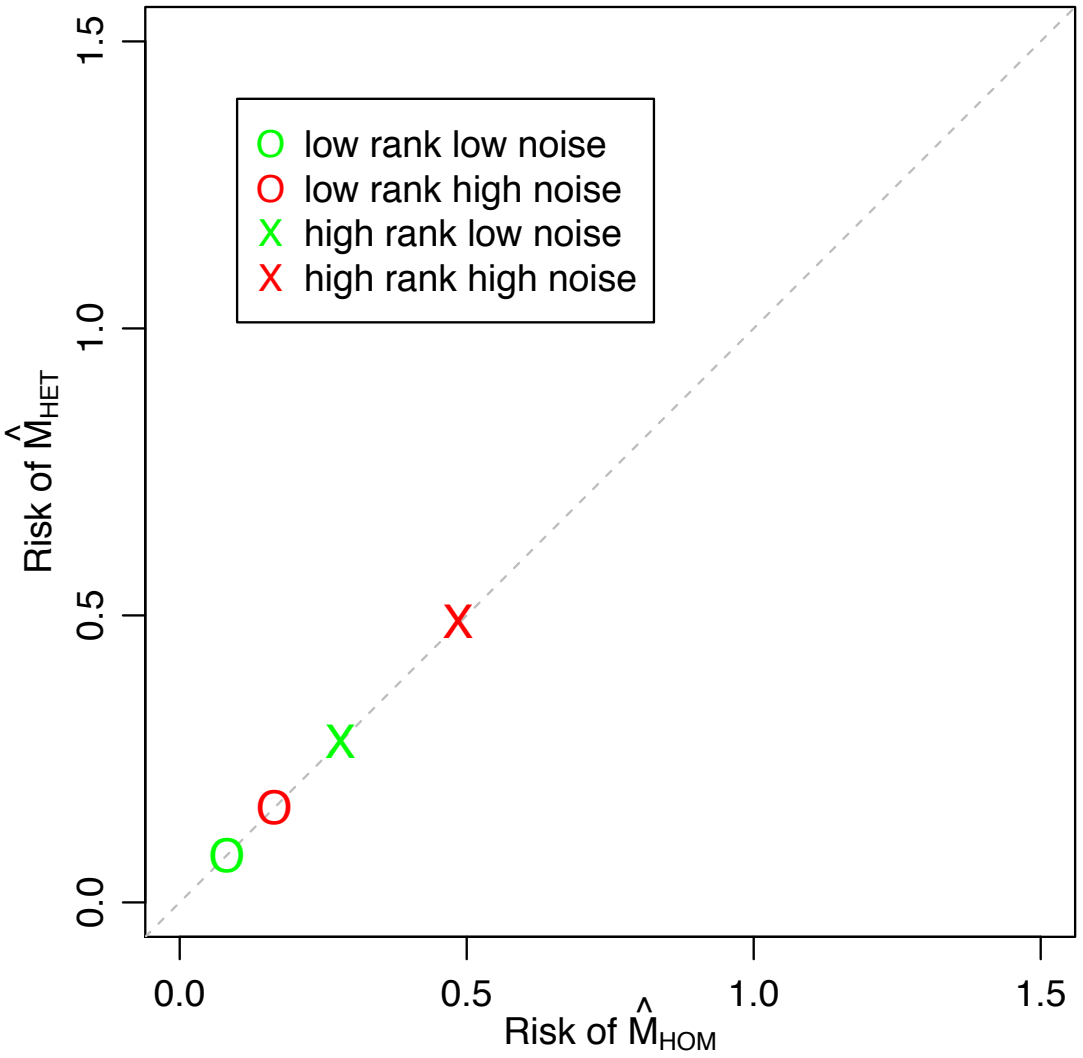
Risks


low rank low noise
low rank high noise
high rank low noise
high rank high noise
Risk of $\hat{M}_{HET}$
Risk of $\hat{M}_{HOM}$
0.0
0.5
1.0
1.5

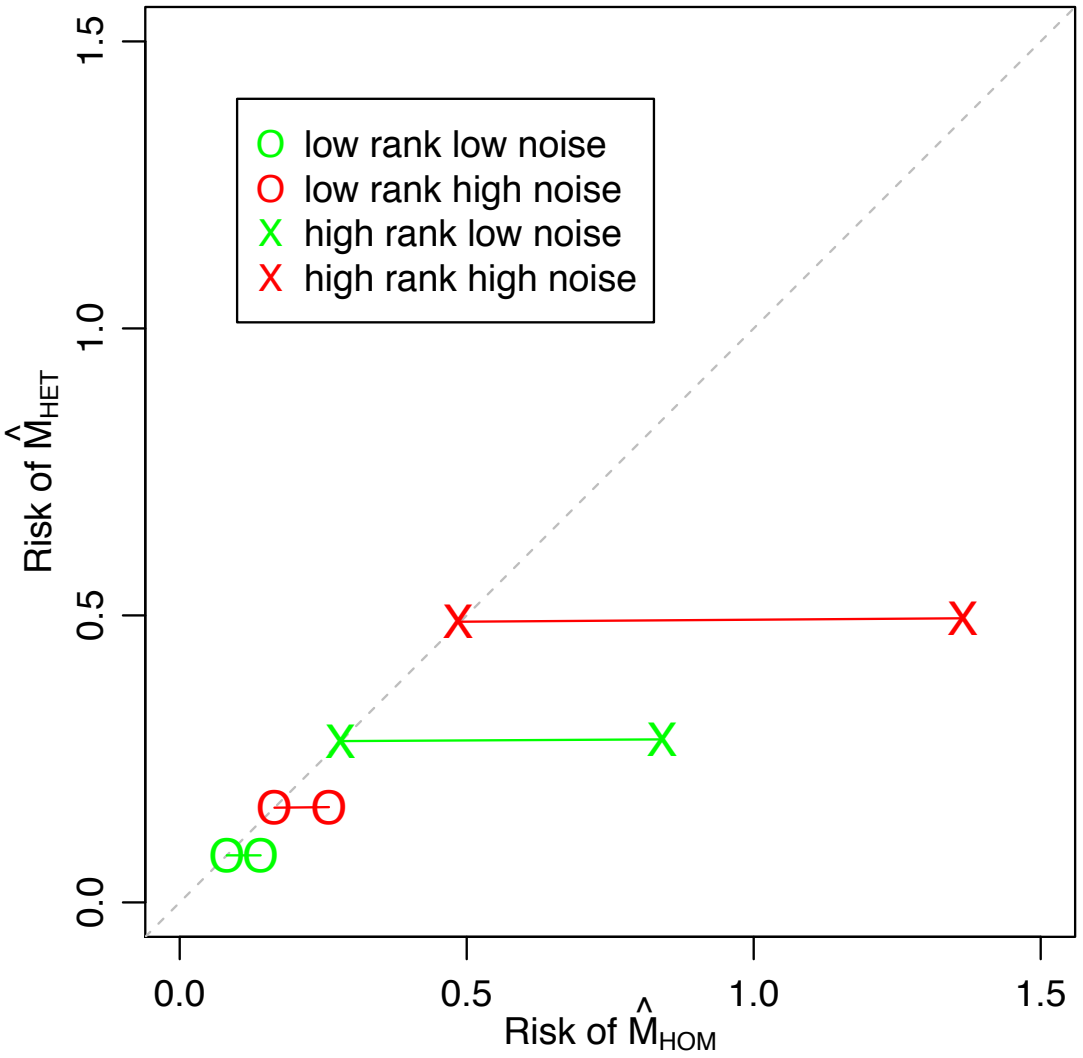
Risks


low rank low noise
low rank high noise
high rank low noise
high rank high noise
Risk of $\hat{M}_{HET}$
Risk of $\hat{M}_{HOM}$
0.0
0.5
1.0
1.5

Risks


O low rank low noise
O low rank high noise
X high rank low noise
X high rank high noise
Risk of $\hat{M}_{HET}$
Risk of $\hat{M}_{HOM}$
0.0
0.5
1.0
1.5

Risks - misspecified rank



Extension and application (Hoff [2013])

Longitudinal multivariate IR data : $\mathbf{Y} = \{y_{i,j,k,l}\} \in \mathbb{R}^{30\times 30\times 20\times 52}$

- i and j index initiator and target of action
- k indexes the action category
- l indexes week

Extension and application (Hoff [2013])

Longitudinal multivariate IR data : $\mathbf{Y} = \{y_{i,j,k,l}\} \in \mathbb{R}^{30\times 30\times 20\times 52}$

- i and j index initiator and target of action
- k indexes the action category
- l indexes week

Extension and application (Hoff [2013])

Longitudinal multivariate IR data : $\mathbf{Y} = \{y_{i,j,k,l}\} \in \mathbb{R}^{30\times 30\times 20\times 52}$

- i and j index initiator and target of action
- k indexes the action category
- l indexes week

Extension and application (Hoff [2013])

Longitudinal multivariate IR data : $\mathbf{Y} = \{y_{i,j,k,l}\} \in \mathbb{R}^{30\times 30\times 20\times 52}$

- i and j index initiator and target of action
- k indexes the action category
- l indexes week

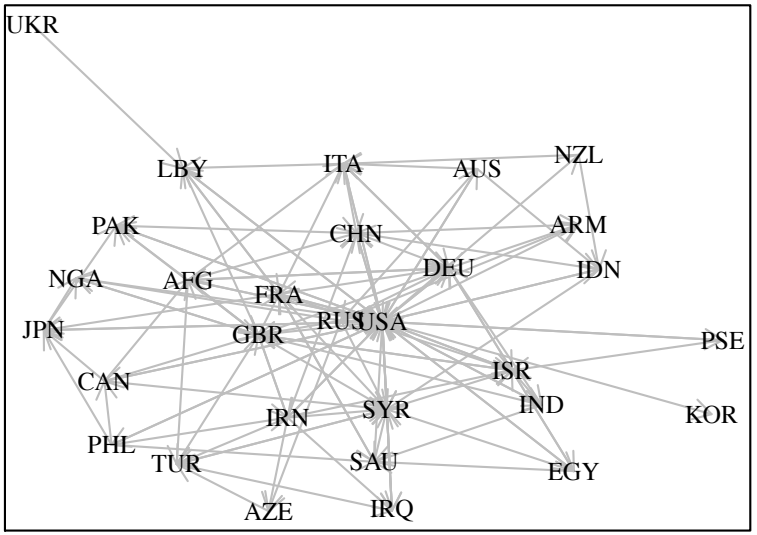
Extension and application (Hoff [2013])

Longitudinal multivariate IR data : $\mathbf{Y} = \{y_{i,j,k,l}\} \in \mathbb{R}^{30\times 30\times 20\times 52}$

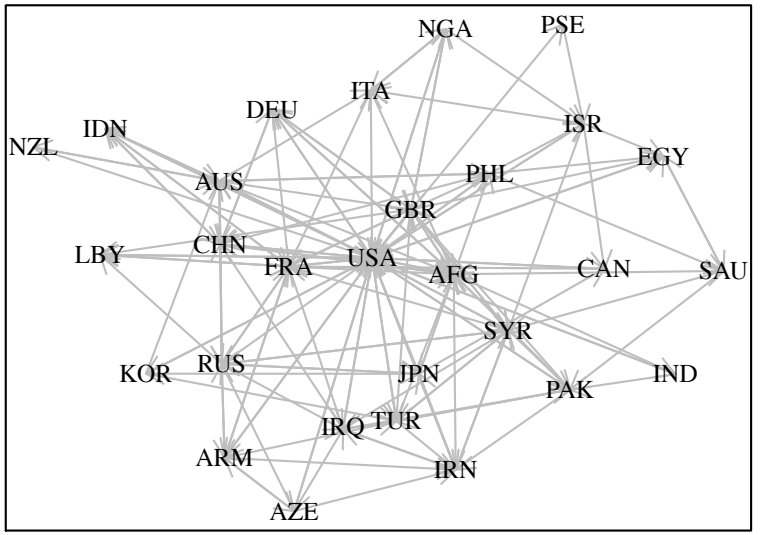
- i and j index initiator and target of action
- k indexes the action category
- l indexes week

Extension and application (Hoff [2013])

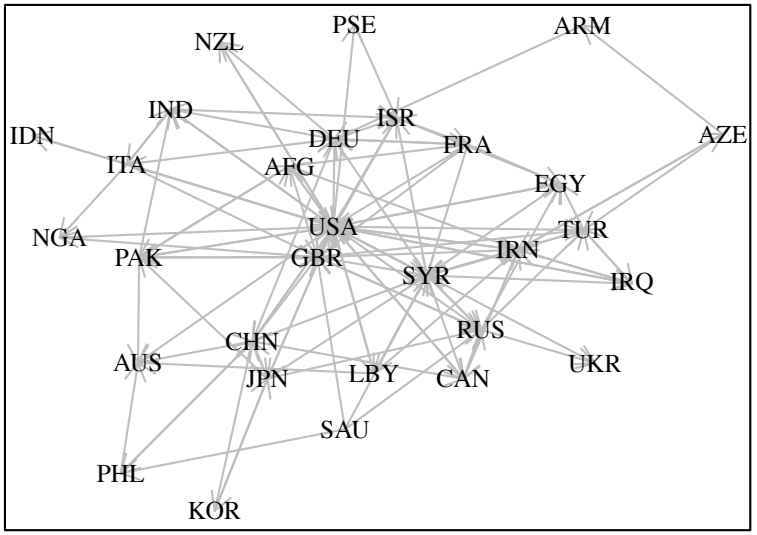
provide aid



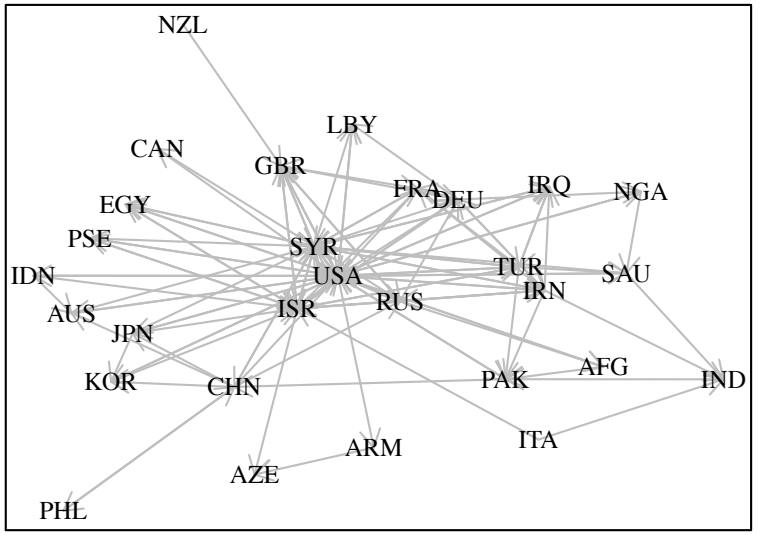
engage in material cooperation



demand

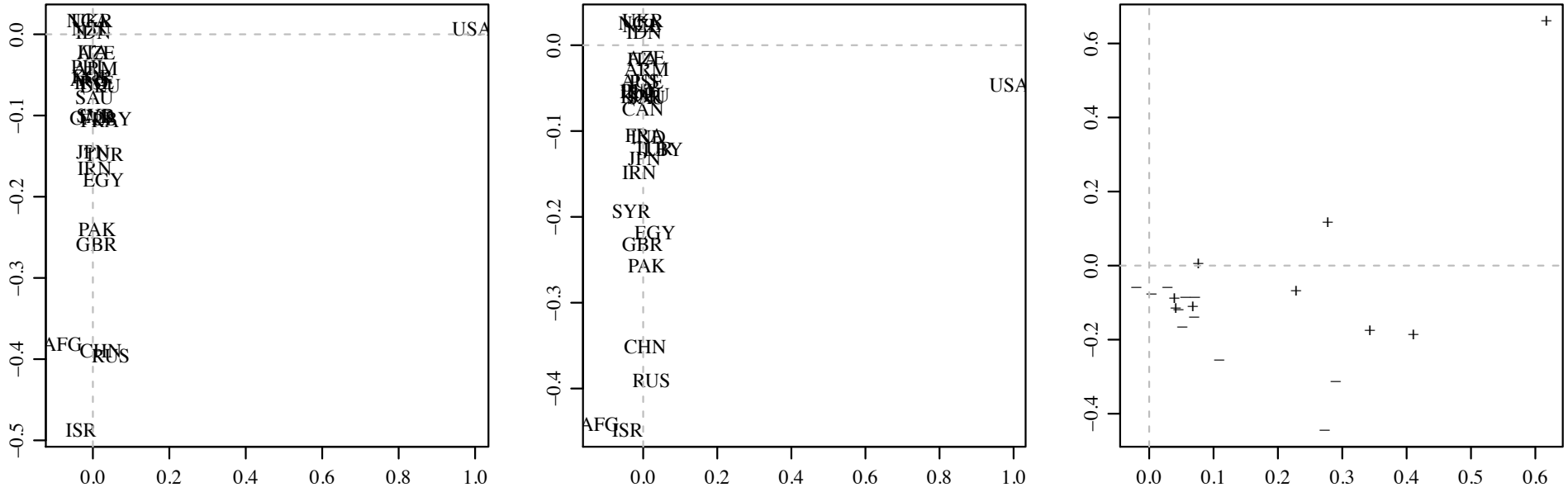


threaten

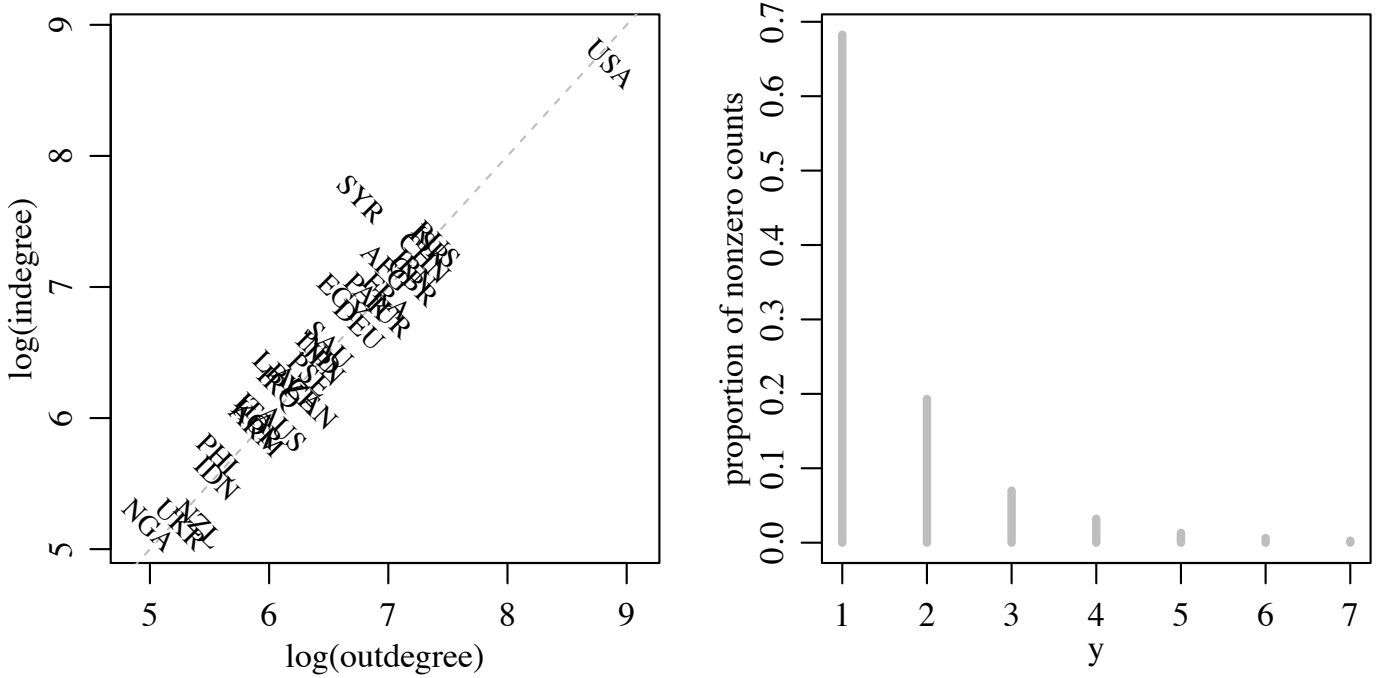


Extension and application

Eigenvectors of least-squares reduced rank approximation (HOSVD):



Extension and application (Hoff [2013])



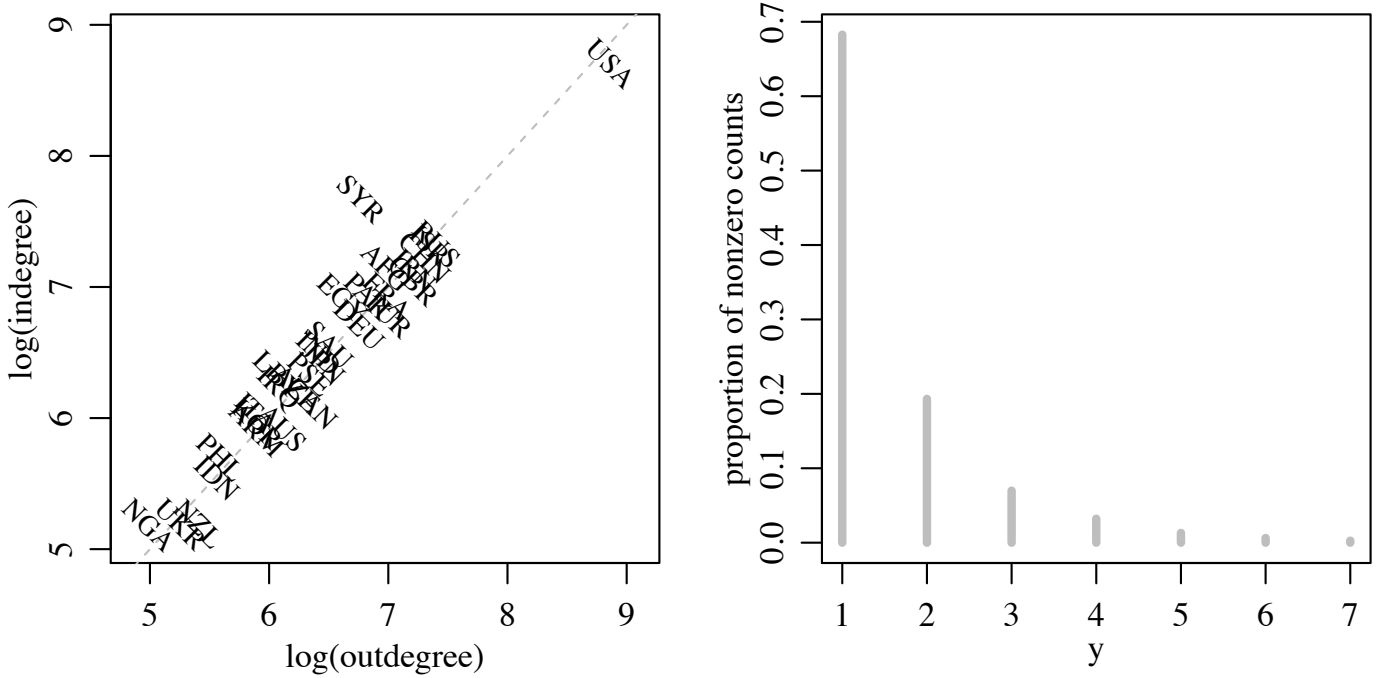
Data are highly skewed counts - normal model/least squares not appropriate.

Scale-free TDM:

$$\mathbf{Z} = \mathbf{S} \times \{\mathbf{U}, \mathbf{V}, \mathbf{W}, \mathbf{X}\} + \mathbf{E}$$

$$\mathbf{Y}_k = g_k(\mathbf{Z}_k)$$

Extension and application (Hoff [2013])



Data are highly skewed counts - normal model/least squares not appropriate.

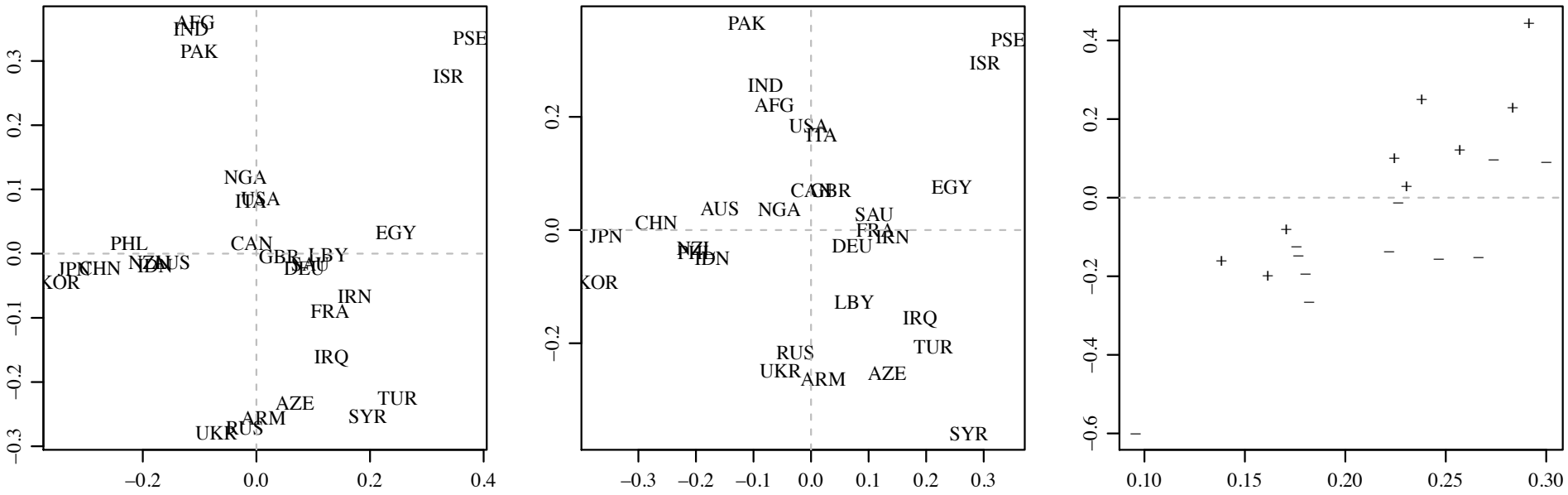
Scale-free TDM:

$$\mathbf{Z} = \mathbf{S} \times \{\mathbf{U}, \mathbf{V}, \mathbf{W}, \mathbf{X}\} + \mathbf{E}$$

$$\mathbf{Y}_k = g_k(\mathbf{Z}_k)$$

Extension and application (Hoff [2013])

Eigenvectors of posterior mean array:

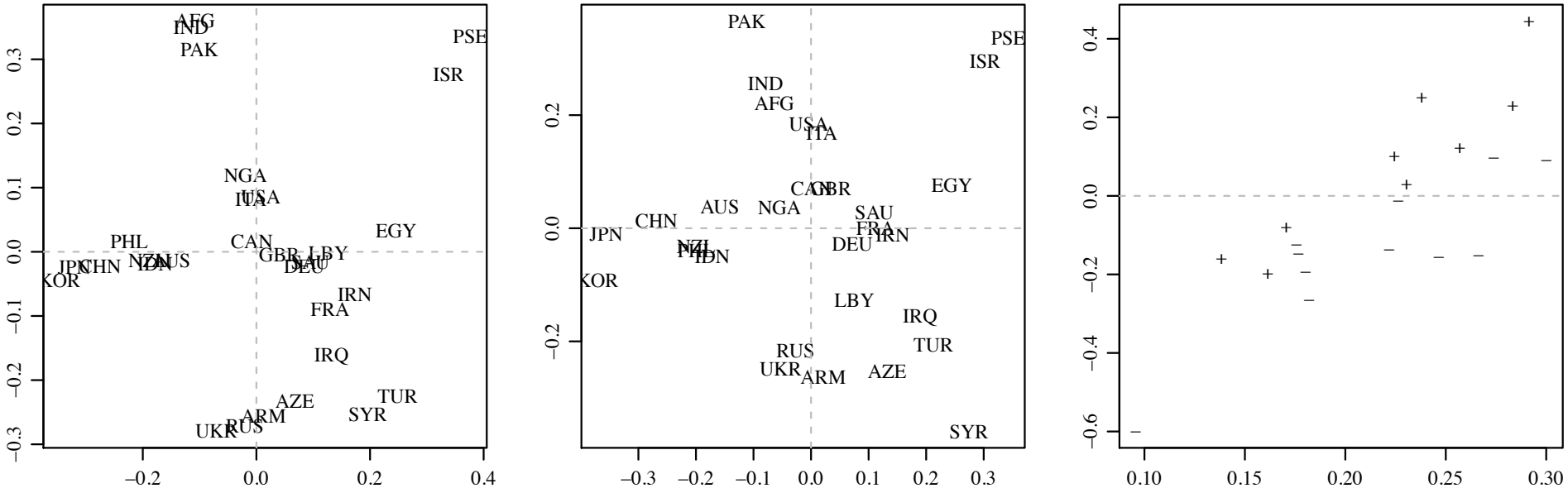


The SF-TSVD provides

- a different (and more interesting) description of heterogeneity;
- a better fit in terms of scale-free measures of association.

Extension and application (Hoff [2013])

Eigenvectors of posterior mean array:

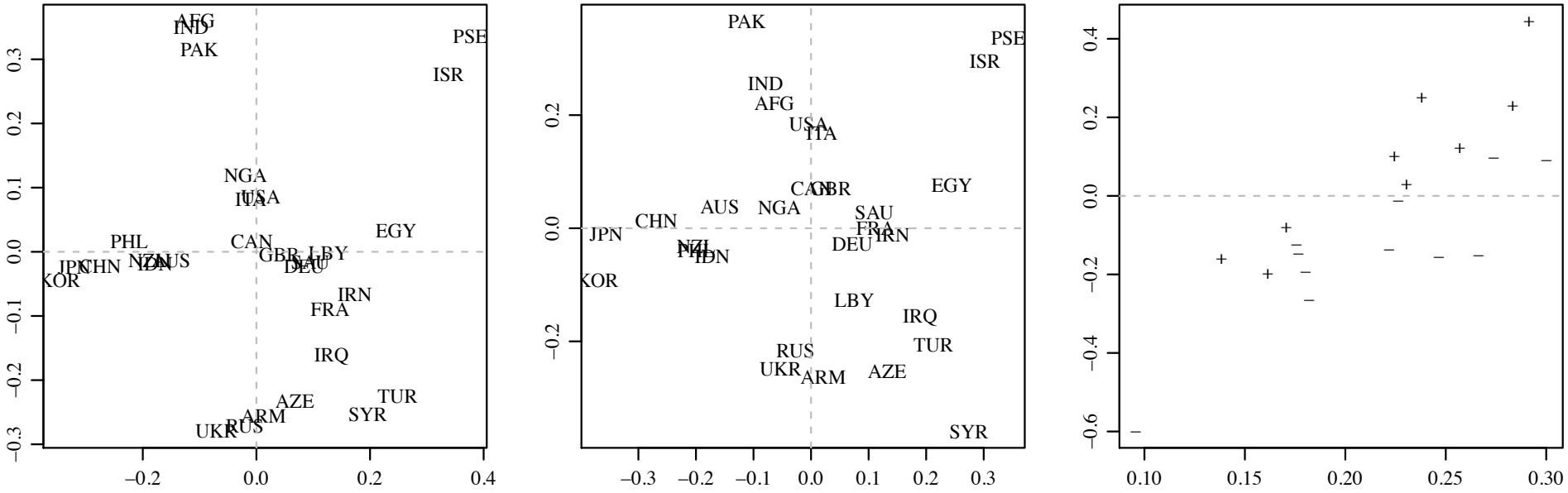


The SF-TSVD provides

- a different (and more interesting) description of heterogeneity;
- a better fit in terms of scale-free measures of association.

Extension and application (Hoff [2013])

Eigenvectors of posterior mean array:

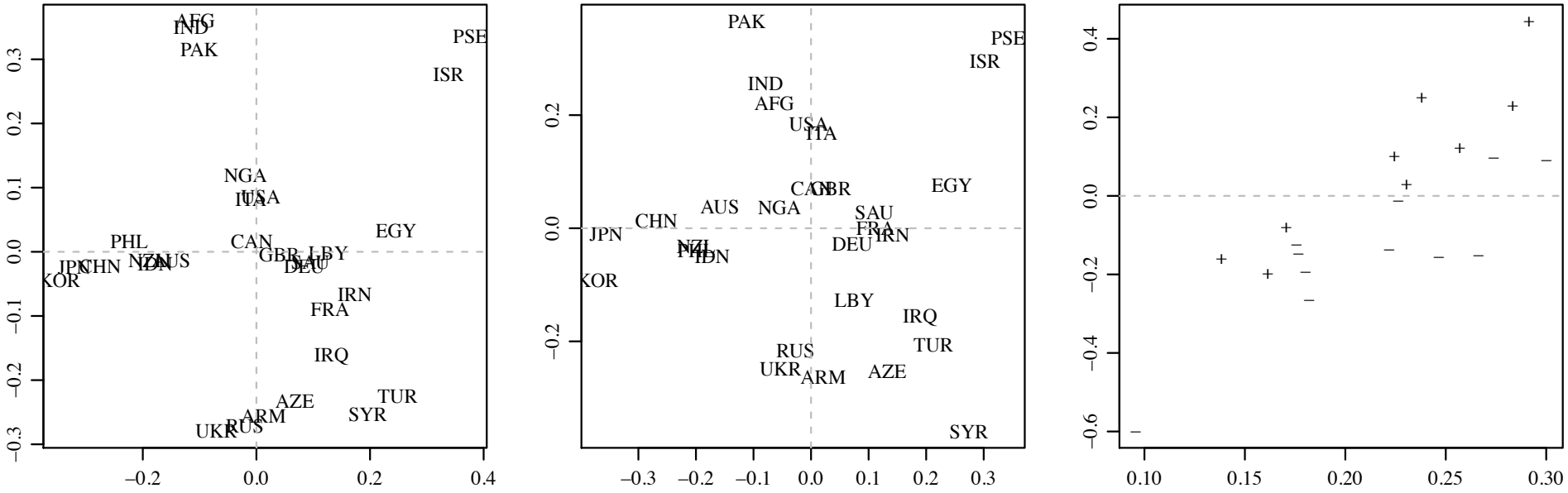


The SF-TSVD provides

- a different (and more interesting) description of heterogeneity;
- a better fit in terms of scale-free measures of association.

Extension and application (Hoff [2013])

Eigenvectors of posterior mean array:



The SF-TSVD provides

- a different (and more interesting) description of heterogeneity;
- a better fit in terms of scale-free measures of association.