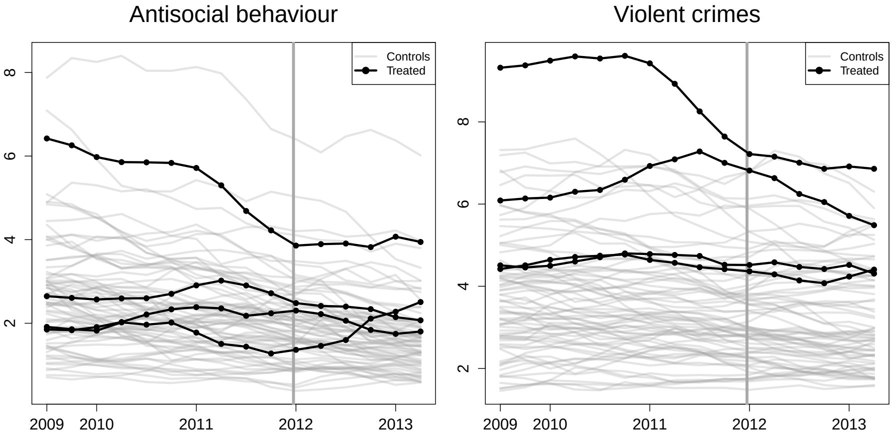
A Bayesian multivariate factor analysis model for evaluating an intervention using observational time-series data on multiple outcomes

Pantelis Samartsidis
MRC Biostatistics Unit, University of Cambridge

BayesPharma 2019
Lyon, 21-24 May

Motivating example

- Application: impact of stricter alcohol licensing policies on alcohol related crimes



Motivating example: details

- Intervention: adaptation of *cumulative impact policies*
- **Objective:** estimate the **causal effect**
- Outcomes (per 10,000)
  1. Antisocial behaviour incidence
  2. Hospital admissions
  3. Sexual crimes
  4. Violent crimes
- Units: 91 local councils in England & Wales (4 treated)
- Study period: mid-2009 to 2015 (quarterly data)
- See de Vocht *et al.* (2017)

Notation

Notation:

(i) n units (index i): n_1 controls and n_2 treated

(ii) T times (index t): T_1 pre- and T_2 post-intervention

(iii) K outcomes (index k)

(iv) Data y_{itk} and covariates $\boldsymbol{x}_{it}$

Rubin causal model:

- Treatment-free outcomes $y_{itk}^{(0)}$
- Outcomes under intervention $y_{itk}^{(1)}$ $(i > n_1$ and $t > T_1)$
- **Counterfactual** estimates $\hat{y}_{itk}^{(0)}$ $(i > n_1$ and $t > T_1)$
- Causal effect estimates $\hat{\theta}_{itk} = y_{itk}^{(1)} - \hat{y}_{itk}^{(0)}$

Challenges for causal inference

Main challenges:

- Observational data
  - (i) Adjust for effect of covariates
  - (ii) Potential of **unobserved confounding**
- Few treated units: propensity score methodologies not suitable

Our contribution:

- Explicit temporal modelling (efficiency)
- Using **multivariate outcomes** (efficiency)
- Quantification of uncertainty

Factor analysis

The **factor analysis** (FA, see e.g. Xu (2017)) model assumes

$$y_{it}^{(0)} =$$

Factor analysis

The **factor analysis** (FA, see e.g. Xu (2017)) model assumes

$$y_{it}^{(0)} = \boldsymbol{x}_{it}^{\top}\boldsymbol{\beta}$$

- Coefficients $\boldsymbol{\beta} \sim \mathrm{N}_d\left(\mathbf{0}, 10^3\boldsymbol{I}\right)$

Factor analysis

The **factor analysis** (FA, see e.g. Xu (2017)) model assumes

$$y_{it}^{(0)} = \boldsymbol{x}_{it}^\top \boldsymbol{\beta} + \boldsymbol{\lambda}_i^\top \boldsymbol{f}_t$$

- Coefficients $\boldsymbol{\beta} \sim \mathrm{N}_d \left(\mathbf{0}, 10^3 \boldsymbol{I}\right)$
- **Loadings** $\boldsymbol{\lambda}_i \sim \mathrm{N}_p \left(\mathbf{0}, \boldsymbol{I}\right)$ for all units i (latent)
- **Factors** $\boldsymbol{f}_t \sim \mathrm{N}_p \left(\mathbf{0}, \boldsymbol{I}\right)$ for all times t (latent)

Factor analysis

The **factor analysis** (FA, see e.g. Xu (2017)) model assumes

$$y_{it}^{(0)} = \boldsymbol{x}_{it}^\top \boldsymbol{\beta} + \boldsymbol{\lambda}_i^\top \boldsymbol{f}_t + \varepsilon_{it}$$

- Coefficients $\boldsymbol{\beta} \sim \mathrm{N}_d\left(\mathbf{0}, 10^3 \boldsymbol{I}\right)$
- **Loadings** $\boldsymbol{\lambda}_i \sim \mathrm{N}_p\left(\mathbf{0}, \boldsymbol{I}\right)$ for all units i (latent)
- **Factors** $\boldsymbol{f}_t \sim \mathrm{N}_p\left(\mathbf{0}, \boldsymbol{I}\right)$ for all times t (latent)
- Errors $\varepsilon_{it} \sim \mathrm{N}\left(0, \psi_i^2\right)$, where $\psi_i^2 \sim \mathrm{IG}\left(0.01, 0.01\right)$

Factor analysis

The **factor analysis** (FA, see e.g. Xu (2017)) model assumes

$$y_{it}^{(0)} = \boldsymbol{x}_{it}^\top \boldsymbol{\beta} + \boldsymbol{\lambda}_i^\top \boldsymbol{f}_t + \varepsilon_{it}$$

- Coefficients $\boldsymbol{\beta} \sim \mathrm{N}_d\left(\mathbf{0}, 10^3 \boldsymbol{I}\right)$
- **Loadings** $\boldsymbol{\lambda}_i \sim \mathrm{N}_p\left(\mathbf{0}, \boldsymbol{I}\right)$ for all units i (latent)
- **Factors** $\boldsymbol{f}_t \sim \mathrm{N}_p\left(\mathbf{0}, \boldsymbol{I}\right)$ for all times t (latent)
- Errors $\varepsilon_{it} \sim \mathrm{N}\left(0, \psi_i^2\right)$, where $\psi_i^2 \sim \mathrm{IG}\left(0.01, 0.01\right)$

Some comments:

- Loadings/factors can account for unobserved confounding
- Implied assumption: intervention as good as randomized conditional on $\boldsymbol{x}_{it}$ and $\boldsymbol{\lambda}_i$
- Relates to *difference-in-differences* and *synthetic controls*

Inducing temporal dependence

- **Motivation**: autocorrelations likely in time-series data
- For each $j = 1, \ldots, p$, we assume factors are **AR(1)** i.e.

$$f_{tj} = \rho_j f_{t-1,j} + \eta_{tj},$$

  where $\rho_j \in (-1, 1)$ and $\eta_{tj} \sim \mathrm{N}(0, 1)$
- Let $\boldsymbol{\Lambda}$ be the $n \times p$ matrix with rows $\boldsymbol{\lambda_i}$ and let $\boldsymbol{y}_{\cdot t} = (y_{1t}, \ldots, y_{nt})^\top$
- Marginally, we have that

$$\mathbb{C}\mathrm{ov}\left(\boldsymbol{y}_{\cdot t}, \boldsymbol{y}_{\cdot s}\right) = \boldsymbol{\Lambda}\mathbb{C}\mathrm{ov}\left(\boldsymbol{f}_t, \boldsymbol{f}_s\right)\boldsymbol{\Lambda}^\top$$

- Standard FA treats different time points as independent

Choosing p

- **Motivation**: credible intervals should reflect uncertainty in p
- Let $\boldsymbol{\lambda}_i$ be of dimension ∞ for each i
- The **multiplicative Gamma shrinkage process** prior assumes

$$\lambda_{ij} \sim \mathrm{N}\left(0, \frac{1}{\phi_{ij}\tau_j}\right),$$

  where $\tau_j = \prod_{\ell=1}^{j} \delta_j$
- Priors: $\phi_{ij} \sim \mathrm{Gamma}\left(\frac{3}{2}, \frac{3}{2}\right)$, $\delta_1 \sim \mathrm{Gamma}\,(2.1, 1)$ and $\delta_j \sim \mathrm{Gamma}\,(3.1, 1)$ $(j > 1)$
- Elements of $\boldsymbol{\lambda}_i$ will progressively **shrink to zero**
- See Bhattacharya and Dunson (2011)

Multivariate factor analysis

- **Motivation**: the K outcomes potentially share some variability
- The multivariate FA (MVFA) assumes that

$$y_{itk}^{(0)} = \boldsymbol{x}_{it}^{\top}\boldsymbol{\beta}_k + \boldsymbol{\lambda}_i^{\top}\boldsymbol{f}_{tk} + \boldsymbol{\gamma}_{ik}^{\top}\boldsymbol{s}_{tk} + \varepsilon_{itk}$$

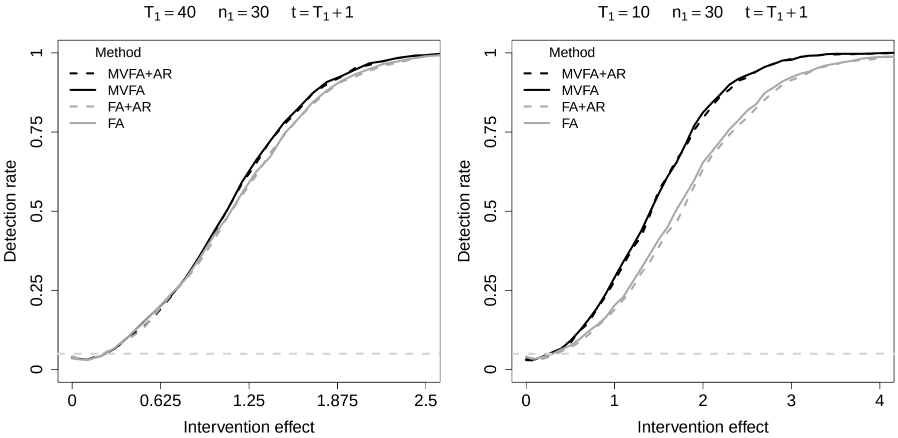
- Think of $\boldsymbol{\gamma}_{ik}$ as unobserved variables that affect k-th outcome **only**
- MVFA learns $\boldsymbol{\lambda}_i$ using data on all K outcomes
- MVFA can make use of covariates that are affected by treatment
- Similar approach by De Vito *et al.* (2018) in meta-analysis

Simulation study

- **Goal**: to assess gains (if any) of using MVFA instead of FA
- **Setting**:
  1. $K = 3$ outcomes
  2. $p_1 = 4$ shared and $p_2 = 2$ outcome-specific loadings
  3. $n_1 \in \{30, 15, 5\}$ and $T_1 \in \{40, 20, 10\}$
  4. $n_2 = 5$ and $t_2 = 5$
  5. $\rho_{jk} = 0.9$ for all j and k
- $\mathrm{logit}\left\{\mathbb{P}\left(\text{i is treated}\right)\right\} = \alpha \sum_{t=T_1+1}^{T} \left[\boldsymbol{\lambda}_i^\top \boldsymbol{f}_{t1} + \boldsymbol{\gamma}_{i1}^\top \boldsymbol{s}_{t1}\right]$
- Models compared: MVFA+AR, MVFA, FA+AR and FA
- Data generated from MVFA+AR (*all models are correctly specified*)
- Interested in **power to detect an intervention effect**

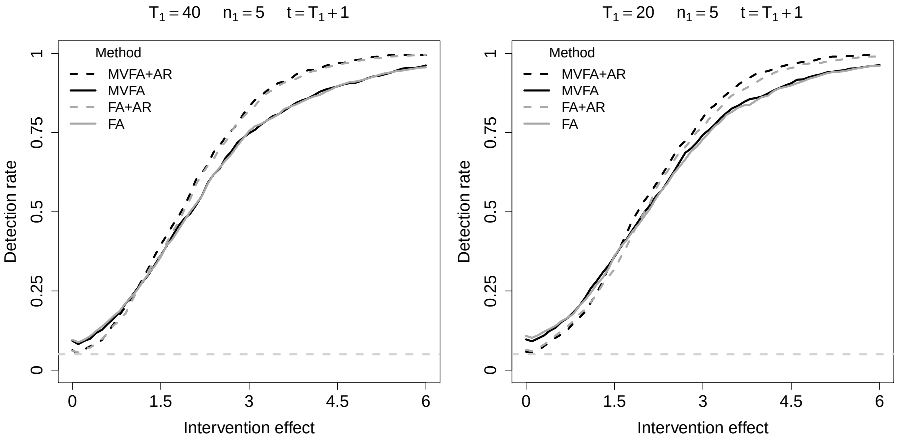
Simulation results: joint outcome modelling ($k = 1$)

- Gains in power when T_1 is small
- T_1 large: $\boldsymbol{\lambda}_i$ can be accurately estimated



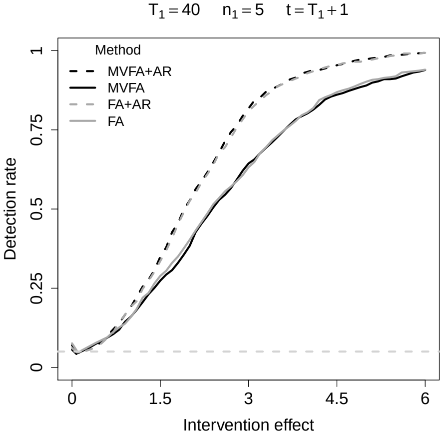
Simulation results: AR ($k = 1$)

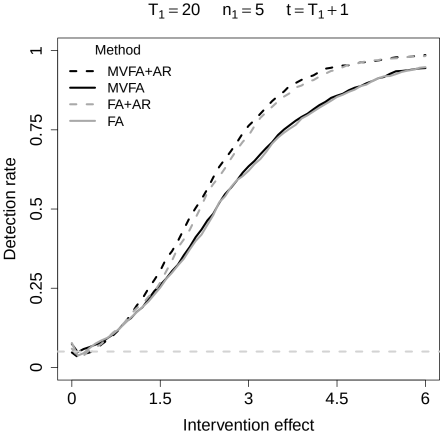
- Gains in power when n_1 is small
- n_1 large: $\boldsymbol{f}_t$ can be accurately estimated



Simulation results: AR ($k = 2$)

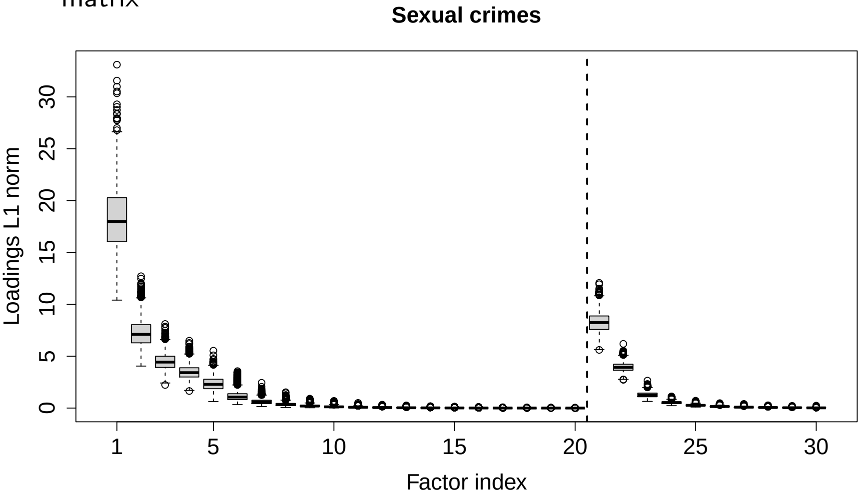
- Gains in power when n_1 is small
- n_1 large: $\boldsymbol{f}_t$ can be accurately estimated





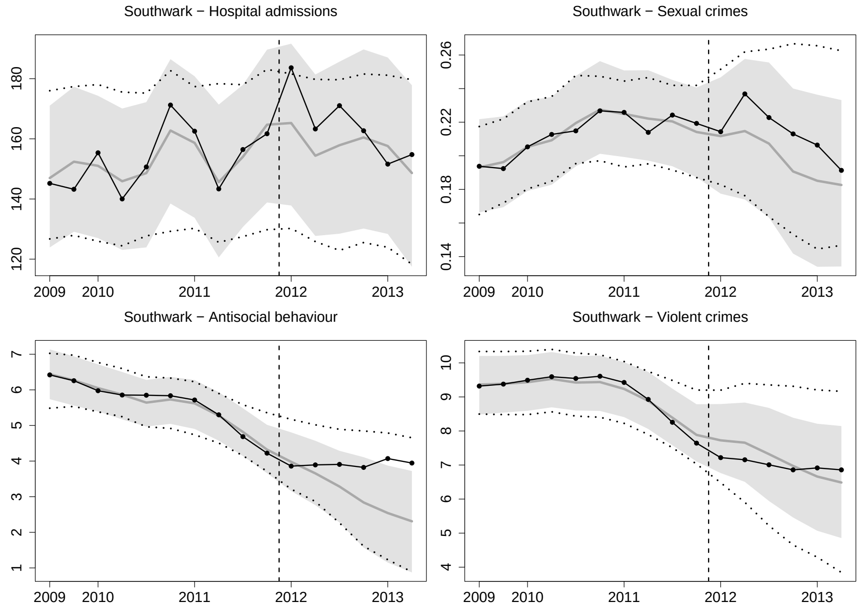
Real data results: evidence for common factors

- Posterior of $\sum_{i=1}^{n} |\gamma_{ikj}|$, L_1-norm of the j-th column of loadings matrix



* *Left: outcome-specific loadings. Right: shared loadings*

Real data results: Southwark



Discussion

Conclusions:

- Proposed extensions can improve quality of causal estimates
  1. AR: large T_1 and small n_1
  2. Joint outcome modelling: small T_1
- MVFA also reduced bias/false positive rates of the FA model
- Empirical application: we did not detect a significant intervention effect

Future work:

- Shared factors (rather than loadings)
- Model temporal/between-outcome correlation of the error terms
- Spatial MVFA models

Acknowledgements

Jointly with:

- Daniela De Angelis, Shaun Seaman (Cambridge)
- Silvia Montagna (Torino)
- Matthew Hickman (Bristol)
- Andre Charlett (Public Health England)

Funding:

- NIHR Health Protection Unit on Evaluation of Interventions
- Medical Research Council
- Public Health England

Acknowledgements

Jointly with:

- Daniela De Angelis, Shaun Seaman (Cambridge)
- Silvia Montagna (Torino)
- Matthew Hickman (Bristol)
- Andre Charlett (Public Health England)

Funding:

- NIHR Health Protection Unit on Evaluation of Interventions
- Medical Research Council
- Public Health England

THANK YOU!!!

References I

Bhattacharya, A. and Dunson, D. B. (2011). Sparse bayesian infinite factor models. *Biometrika*, **98**(2), 291–306.

De Vito, R., Bellio, R., Trippa, L., and Parmigiani, G. (2018). Multi-study factor analysis. *Biometrics*.

de Vocht, F., Tilling, K., Pliakas, T., Angus, C., Egan, M., Brennan, A., Campbell, R., and Hickman, M. (2017). The intervention effect of local alcohol licensing policies on hospital admission and crime: a natural experiment using a novel bayesian synthetictime-series method. *Journal of Epidemiology & Community Health*, **71**(9), 912–918.

Xu, Y. (2017). Generalized synthetic control method: Causal inference with interactive fixed effects models. *Political Analysis*, **25**(1), 57–76.