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

Antisocial behaviour

Violent 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

(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 \( x_{it} \)

Rubin causal model:

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

Main challenges:

- Observational data
  - Adjust for effect of covariates
  - 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
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

\[
y_{it}^{(0)} = x_{it} \beta
\]

- Coefficients \( \beta \sim N_d \left( 0, 10^3 I \right) \)
Factor analysis

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

\[ y_{it}^{(0)} = \mathbf{x}_{it}^\top \beta + \lambda_i^\top f_t \]

- **Coefficients** \( \beta \sim N_d \left( \mathbf{0}, 10^3 \mathbf{I} \right) \)
- **Loadings** \( \lambda_i \sim N_p \left( \mathbf{0}, \mathbf{I} \right) \) for all units \( i \) (latent)
- **Factors** \( f_t \sim N_p \left( \mathbf{0}, \mathbf{I} \right) \) for all times \( t \) (latent)

Some comments:

- Loadings/factors can account for unobserved confounding
- Implied assumption: intervention as good as randomized conditional on \( x_{it} \) and \( \lambda_i \)
- Relates to difference-in-differences and synthetic controls
Factor analysis

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

\[ y_{it}^{(0)} = x_{it}^\top \beta + \lambda_i^\top f_t + \varepsilon_{it} \]

- **Coefficients** \( \beta \sim N_d \left( \mathbf{0}, 10^3 I \right) \)
- **Loadings** \( \lambda_{i} \sim N_p \left( \mathbf{0}, I \right) \) for all units \( i \) (latent)
- **Factors** \( f_t \sim N_p \left( \mathbf{0}, I \right) \) for all times \( t \) (latent)
- **Errors** \( \varepsilon_{it} \sim N \left( \mathbf{0}, \psi_i^2 \right) \), where \( \psi_i^2 \sim IG \left( 0.01, 0.01 \right) \)
Factor analysis

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

\[ y_{it}^{(0)} = x_{it}^\top \beta + \lambda_i^\top f_t + \varepsilon_{it} \]

- **Coefficients** $\beta \sim N_d(0, 10^3 I)$
- **Loadings** $\lambda_i \sim N_p(0, I)$ for all units $i$ (latent)
- **Factors** $f_t \sim N_p(0, I)$ for all times $t$ (latent)
- **Errors** $\varepsilon_{it} \sim N(0, \psi_i^2)$, where $\psi_i^2 \sim IG(0.01, 0.01)$

Some comments:
- Loadings/factors can account for unobserved confounding
- Implied assumption: intervention as good as randomized conditional on $x_{it}$ and $\lambda_i$
- 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 $\eta_{tj} \sim N(0, 1)$
- Let $\Lambda$ be the $n \times p$ matrix with rows $\lambda_i$ and let $y_t = (y_{1t}, \ldots, y_{nt})^\top$
- Marginally, we have that
  \[ \text{Cov} (y_t, y_s) = \Lambda \text{Cov} (f_t, f_s) \Lambda^\top \]
- Standard FA treats different time points as independent
Motivation: credible intervals should reflect uncertainty in $p$

Let $\lambda_i$ be of dimension $\infty$ for each $i$

The multiplicative Gamma shrinkage process prior assumes

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

where $\tau_j = \prod_{\ell=1}^{j} \delta_j$

Priors: $\phi_{ij} \sim \text{Gamma}\left(\frac{3}{2}, \frac{3}{2}\right)$, $\delta_1 \sim \text{Gamma}\left(2.1, 1\right)$ and $\delta_{j} \sim \text{Gamma}\left(3.1, 1\right)$ ($j > 1$)

Elements of $\lambda_i$ will progressively shrink to zero

See Bhattacharya and Dunson (2011)
Motivation: the \( K \) outcomes potentially share some variability

The multivariate FA (MVFA) assumes that

\[
y_{itk}^{(0)} = x_{it}^\top \beta_k + \lambda_i^\top f_{tk} + \gamma_{ik}^\top s_{tk} + \varepsilon_{itk}
\]

Think of \( \gamma_{ik} \) as unobserved variables that affect \( k \)-th outcome only

MVFA learns \( \lambda_i \) 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 $k$

- $\text{logit}\{\mathbb{P}(i \text{ is treated})\} = \alpha \sum_{t=T_1+1}^{T} [\lambda_i^T f_{t1} + \gamma_{i1}^T s_{t1}]$

- 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: $\lambda_i$ can be accurately estimated

<table>
<thead>
<tr>
<th>$T_1$</th>
<th>$n_1$</th>
<th>$t$</th>
<th>Intervention effect</th>
<th>Detection rate</th>
</tr>
</thead>
<tbody>
<tr>
<td>40</td>
<td>30</td>
<td>$T_1+1$</td>
<td>0</td>
<td>0.625</td>
</tr>
<tr>
<td></td>
<td></td>
<td></td>
<td>1.25</td>
<td>0.25</td>
</tr>
<tr>
<td></td>
<td></td>
<td></td>
<td>1.875</td>
<td>0.5</td>
</tr>
<tr>
<td></td>
<td></td>
<td></td>
<td>2.5</td>
<td>0.75</td>
</tr>
</tbody>
</table>

Method
- MVFA+AR
- MVFA
- FA+AR
- FA

<table>
<thead>
<tr>
<th>$T_1$</th>
<th>$n_1$</th>
<th>$t$</th>
<th>Intervention effect</th>
<th>Detection rate</th>
</tr>
</thead>
<tbody>
<tr>
<td>10</td>
<td>30</td>
<td>$T_1+1$</td>
<td>0</td>
<td>1</td>
</tr>
<tr>
<td></td>
<td></td>
<td></td>
<td>1.25</td>
<td>0.25</td>
</tr>
<tr>
<td></td>
<td></td>
<td></td>
<td>2</td>
<td>0.5</td>
</tr>
<tr>
<td></td>
<td></td>
<td></td>
<td>3</td>
<td>0.75</td>
</tr>
<tr>
<td></td>
<td></td>
<td></td>
<td>4</td>
<td>1</td>
</tr>
</tbody>
</table>

Method
- MVFA+AR
- MVFA
- FA+AR
- FA
Simulation results: AR \((k = 1)\)

- Gains in power when \(n_1\) is small
- \(n_1\) large: \(f_t\) can be accurately estimated

\[
T_1 = 40 \quad n_1 = 5 \quad t = T_1 + 1
\]

\[
T_1 = 20 \quad n_1 = 5 \quad t = T_1 + 1
\]
Simulation results: AR \((k = 2)\)

- Gains in power when \(n_1\) is small
- \(n_1\) large: \(f_t\) can be accurately estimated

\[
T_1 = 40 \quad n_1 = 5 \quad t = T_1 + 1
\]

\[
T_1 = 20 \quad n_1 = 5 \quad t = T_1 + 1
\]
Real data results: evidence for common factors

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

Sexual crimes

*Left: outcome-specific loadings. Right: shared loadings
Real data results: Southwark

Southwark – Antisocial behaviour

Southwark – Violent crimes

Southwark – Hospital admissions

Southwark – Sexual crimes
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!!!

