A Bayesian multivariate factor analysis model for causal inference using time-series observational data on mixed outcomes

Angelos Alexopoulos

Statistical modelling of epidemic outbreaks,

May 5, 2023

Co-authors

- MRC Biostatistics Unit
    - Pantelis Samartsidis
    - Shaun R. Seaman
    - Daniela De Angelis
- UK Health Security Agency
    - Abbie Harrison
    - Gareth J. Hughes
    - Charlotte Anderson
    - Andre Charlet
    - Isabel Oliver
    - Daniela De Angelis

Introduction: The problem

- **Motivation:**
  - ✓ **Quick** evaluation of the *local tracing partnerships* (LTPs) introduced by England's NHS *Test & Trace* (TT) programme to improve tracing of **Covid-19** cases and their contacts.
  - ✓ LTPs **main task**: trace cases resident in local authorities not successfully traced by TT.
  - ✓ **Question**: impact of LTPs on the effectiveness of TT.

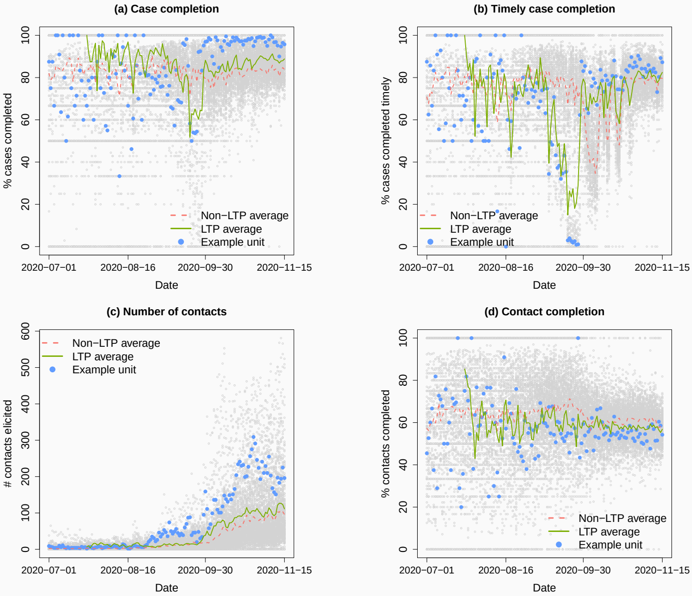
Introduction: The problem

- **Motivation:**
  - ✓ **Quick** evaluation of the *local tracing partnerships* (LTPs) introduced by England's NHS *Test & Trace* (TT) programme to improve tracing of **Covid-19** cases and their contacts.
  - ✓ LTPs **main task**: trace cases resident in local authorities not successfully traced by TT.
  - ✓ **Question**: impact of LTPs on the effectiveness of TT.
- **The statistical problem:**
  - ✓ Estimation of causal effects of **intervention** (LTPs introduction) based on **observational** data; **judge** the effectiveness of TT in terms of four outcomes measured daily.

Introduction: The problem

- **Motivation:**
  - ✓ **Quick** evaluation of the *local tracing partnerships* (LTPs) introduced by England's NHS *Test & Trace* (TT) programme to improve tracing of **Covid-19** cases and their contacts.
  - ✓ LTPs **main task**: trace cases resident in local authorities not successfully traced by TT.
  - ✓ **Question**: impact of LTPs on the effectiveness of TT.
- **The statistical problem:**
  - ✓ Estimation of causal effects of **intervention** (LTPs introduction) based on **observational** data; **judge** the effectiveness of TT in terms of four outcomes measured daily.
- **Challenges:**
  - ✓ Measurements for multiple sample units (local authorities) at multiple time points.
  - ✓ Mixed (continuous and discrete) outcomes; some of those with limited amount of information (e.g. low counts).
  - ✓ Absence of randomisation in intervention wrt time and units.

Introduction: Outcomes of interest



Introduction: Contribution

- Tackle limitations of **causal factor analysis** ("matrix completion"):
  - ✓ Extension to a **multivariate factor analysis** model.
  - ✓ **Joint modelling** of mixed outcomes to increase statistical efficiency.
  - ✓ Use **Bayesian** methods to quantify uncertainty for the causal effects.
- Construction of a bespoke **MCMC** algorithm.
  - ✓ Dealing **efficiently** with problems caused by non-identifiability of factor models by customising modern samplers.
  - ✓ Rely on **data augmentation** to facilitate sampling from the full conditionals.

Notation and assumptions

- Assume D_1 **continuous**, D_2 **binomial** and D_3 count outcomes.
- Notation for **binomial** outcomes; similarly for remaining outcomes.
  - n_{it}: # of new cases in unit i and day t
  - k_{it}: completed cases out of n_{it}
  - p_{it}: probability of case completion
  - N units total, N_1 controls. For $i > N_1$, T_i is the last day before LTP

Notation and assumptions

- Assume D_1 **continuous**, D_2 **binomial** and D_3 count outcomes.
- Notation for **binomial** outcomes; similarly for remaining outcomes.
  - n_{it}: # of new cases in unit i and day t
  - k_{it}: completed cases out of n_{it}
  - p_{it}: probability of case completion
  - N units total, N_1 controls. For $i > N_1$, T_i is the last day before LTP
- **Potential outcomes framework** (Holland, 1986)
  - Treatment free outcomes $p_{it}^{(0)}$, $n_{it}^{(0)}$ and $k_{it}^{(0)}$ (all i and t)
  - Outcomes under LTP $p_{it}^{(1)}$, $n_{it}^{(1)}$ and $k_{it}^{(1)}$ ($i > N_1$ and $t > T_i$)
  - Hence, data are: $$\{n_{it}, k_{it}\} = \begin{cases} \{n_{it}^{(1)}, k_{it}^{(1)}\} & i > N_1 \text{ and } t > T_i \\ \{n_{it}^{(0)}, k_{it}^{(0)}\} & \text{otherwise} \end{cases}$$
- y_{it} and z_{it} denote continuous and count outcomes respectively.

Causal effects

For $i > N_1$ and $t > T_i$, we are interested in estimating

- Effect of LTP on case completion probability

$$\beta_{it} = p_{it}^{(1)} - p_{it}^{(0)}$$

- The total number of **additional cases completed** thanks to LTP

$$\gamma_{it} = k_{it}^{(1)} - \tilde{k}_{it}^{(0)},$$

where $\tilde{k}_{it}^{(0)} \sim \mathrm{Bin}(n_{it}^{(1)}, p_{it}^{(0)})$.

Causal effects

For $i > N_1$ and $t > T_i$, we are interested in estimating

- Effect of LTP on case completion probability

$$\beta_{it} = p_{it}^{(1)} - p_{it}^{(0)}$$

- The total number of **additional cases completed** thanks to LTP

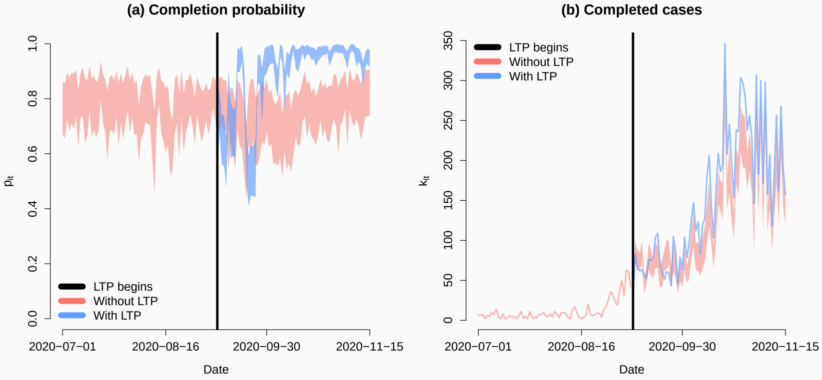
$$\gamma_{it} = k_{it}^{(1)} - \tilde{k}_{it}^{(0)},$$

where $\tilde{k}_{it}^{(0)} \sim \mathrm{Bin}(n_{it}^{(1)}, p_{it}^{(0)})$.

For a **count outcome** (e.g. # contacts), we take $\delta_{it} = z_{it}^{(1)} - z_{it}^{(0)}$, unless there is an offset; $\alpha_{it} = y_{it}^{(1)} - y_{it}^{(0)}$ for **continuous** outcomes.

Causal inference as a missing data problem

To estimate **causal** effects we need to **impute** the counterfactuals $p_{it}^{(0)}$ and $\tilde{k}_{it}^{(0)} \sim \mathrm{Bin}(n_{it}^{(1)}, p_{it}^{(0)})$ for $i > N_1$ and $t > T_i$



Imputation through latent factors

Normal outcomes as an **illustration**:

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

Some comments:

- The loadings $\boldsymbol{\lambda}_i \in \mathbb{R}^J$ and factors $\boldsymbol{f}_t \in \mathbb{R}^J$ are unobserved.
- We fit this model **discarding all post-intervention** data; for $i > N_1$ and $t > T_i$, we estimate

$$\hat{y}_{it}^{(0)} = \hat{\boldsymbol{\eta}}^\top \boldsymbol{x}_{it} + \hat{\boldsymbol{\lambda}}_i^\top \hat{\boldsymbol{f}}_t$$

- **Implicit causal assumption**: $\boldsymbol{x}_{it}$ and $\boldsymbol{\lambda}_i$ account for (observed/unobserved) confounding of the causal effects (Xu, 2017).

Imputation through latent factors

Normal outcomes as an **illustration**:

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

Some comments:

- The loadings $\boldsymbol{\lambda}_i \in \mathbb{R}^J$ and factors $\boldsymbol{f}_t \in \mathbb{R}^J$ are unobserved.
- We fit this model **discarding all post-intervention** data; for $i > N_1$ and $t > T_i$, we estimate

$$\hat{y}_{it}^{(0)} = \hat{\boldsymbol{\eta}}^\top \boldsymbol{x}_{it} + \hat{\boldsymbol{\lambda}}_i^\top \hat{\boldsymbol{f}}_t$$

- **Implicit causal assumption**: $\boldsymbol{x}_{it}$ and $\boldsymbol{\lambda}_i$ account for (observed/unobserved) confounding of the causal effects (Xu, 2017).

- Now we can estimate the causal effect $\alpha_{it} = y_{it}^{(1)} - \hat{y}_{it}^{(0)}$.

A multivariate factor analysis model

For unit i, day t, and outcome d, we assume a **factor analysis** model

$$
\begin{aligned}
y_{itd}^{(0)} &\sim \mathrm{N}\left(\mu_{itd}, \sigma_i^2\right), \;\; \mu_{itd} = \boldsymbol{\lambda}_i^\top \boldsymbol{f}_{td} + \boldsymbol{\eta}_{1,d}^\top \boldsymbol{x}_{it}, \\
k_{itd}^{(0)} &\sim \mathrm{Bin}\left(n_{itd}, p_{itd}\right), \;\; \mathrm{logit}(p_{itd}) = \boldsymbol{\lambda}_i^\top \boldsymbol{g}_{td} + \boldsymbol{\eta}_{2,d}^\top \boldsymbol{x}_{it}, \\
z_{itd}^{(0)} &\sim \mathrm{NegBin}\left(w_{itd} q_{itd} \xi_d^{-1}, (1+\xi_d)^{-1}\right), \;\; \log(q_{itd}) = \boldsymbol{\lambda}_i^\top \boldsymbol{h}_{td} + \boldsymbol{\eta}_{3,d}^\top \boldsymbol{x}_{it}.
\end{aligned}
$$

- $\boldsymbol{f}_{td}, \boldsymbol{g}_{td}, \boldsymbol{h}_{td} \in \mathbb{R}^J$ unobserved factors; $\boldsymbol{\lambda}_i \in \mathbb{R}^J$ common across outcomes factor loadings for unit i; $\boldsymbol{x}_{it} \in \mathbb{R}^P$ covariates not affected by the intervention; $\boldsymbol{\eta}_{1,d}, \boldsymbol{\eta}_{2,d}, \boldsymbol{\eta}_{3,d}$ regression coefficients.
- $\boldsymbol{x}_{it}$ and $\boldsymbol{\lambda}_i$ control for potential observed and unobserved confounding respectively.

Prior on factors

- For normal outcome d, we *a priori* assume that $f_{tdj} \sim \mathrm{N}(0, \psi_{jd})$, $j = 1, \ldots, J$.
- Analogous for binomial and NB outcomes.
- To allow loadings to affect **any subset** of outcomes, we introduce variables M_j for the variance of factors f_{tdj}, g_{tdj} and h_{tdj}, e.g.

Loading	Normal	Binomial	Neg. Binomial	M_j
$j = 1$	$\psi_{11} \sim \mathrm{Uni}[0,1]$	$\psi_{12} = 1$	$\psi_{13} \sim \mathrm{Uni}[0,1]$	2
$j = 2$	$\psi_{21} \sim \mathrm{Uni}[0,1]$	$\psi_{22} \sim \mathrm{Uni}[0,1]$	$\psi_{23} = 1$	3
$j = 3$	$\psi_{31} = 1$	$\psi_{32} \sim \mathrm{Uni}[0,1]$	$\psi_{33} \sim \mathrm{Uni}[0,1]$	1
$j = 4$	$\psi_{41} = 1$	$\psi_{42} \sim \mathrm{Uni}[0,1]$	$\psi_{43} \sim \mathrm{Uni}[0,1]$	1

...

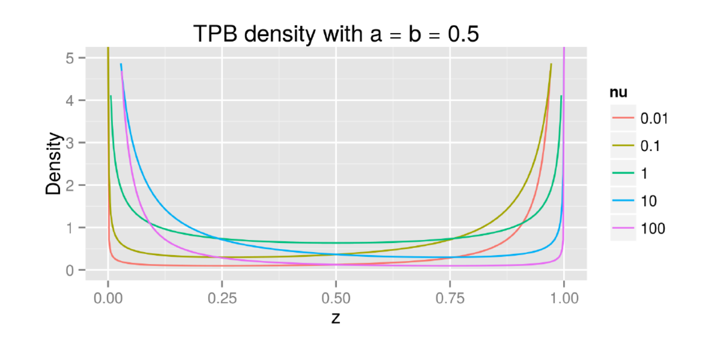
Prior on loadings

Uncertainty in J: start with J large, let data determine how many are needed. We follow Gao et al, (2016):

$$\lambda_{ij} \sim \mathrm{N}(0, \frac{1}{\phi_{ij}} - 1), \qquad \phi_{ij} \sim \mathrm{TPB}(0.5, 0.5, \frac{1}{\zeta_j} - 1),$$

$$\zeta_j \sim \mathrm{TPB}(0.5, 0.5, \frac{1}{\rho} - 1), \qquad \rho \sim \mathrm{TPB}(0.5, 0.5, \nu)$$

TPB: three parameter beta distribution



Bayesian estimation

We draw samples from the **posterior** of the causal effects $\alpha_{itd}, \beta_{itd}, \gamma_{itd}$ by using **Metropolis-within-Gibbs** that targets

$$\pi\left(\left\{\boldsymbol{\lambda}_i, \{\xi_{id}\}_{d=1}^{D_3}\right\}_{i=1}^{N}, \left\{\{\boldsymbol{g}_{td}\}_{t=1}^{T}, \boldsymbol{\eta}_{2,d}\right\}_{d=1}^{D_2}, \left\{\{\boldsymbol{h}_{td}\}_{t=1}^{T}, \boldsymbol{\eta}_{3,d}\right\}_{d=1}^{D_3}, \boldsymbol{\theta} \mid \mathrm{data}\right),$$

where

$$\boldsymbol{\theta} = \left\{\left\{\{\boldsymbol{f}_{td}\}_{t=1}^{T}, \left\{\sigma_{id}^2\right\}_{i=1}^{N}\right\}_{d=1}^{D_1}, \{\xi_d\}_{d=1}^{D_3}, \left\{\{\phi_{ij}\}_{i=1}^{N}, \zeta_j, \{v_{j,l}\}_{\ell=1}^{D}, M_j\right\}_{j=1}^{J^*}, \rho\right\}$$

and

$$\mathrm{data} = \left\{\left\{\{y_{itd}\}_{d=1}^{D_1}, \{k_{itd}, n_{itd}\}_{d=1}^{D_2}, \{z_{itd}, w_{itd}\}_{d=1}^{D_3}, \boldsymbol{x}_{it}\right\}_{t=1}^{T_i}\right\}_{i=1}^{N}.$$

- **Problem**: Non identifiability of latent factors, $\boldsymbol{\Lambda}\boldsymbol{F}^\top = \boldsymbol{\Lambda}\boldsymbol{Q}\boldsymbol{Q}^{-1}\boldsymbol{F}^\top$ ($\boldsymbol{Q}$ orthonormal) as well as sign/label-switching problems make popular schemes (e.g. HMC/MALA) to fail.

Bayesian estimation

We draw samples from the **posterior** of the causal effects $\alpha_{itd}, \beta_{itd}, \gamma_{itd}$ by using **Metropolis-within-Gibbs** that targets

$$\pi\left(\left\{\boldsymbol{\lambda}_i, \{\xi_{id}\}_{d=1}^{D_3}\right\}_{i=1}^{N}, \left\{\{\boldsymbol{g}_{td}\}_{t=1}^{T}, \boldsymbol{\eta}_{2,d}\right\}_{d=1}^{D_2}, \left\{\{\boldsymbol{h}_{td}\}_{t=1}^{T}, \boldsymbol{\eta}_{3,d}\right\}_{d=1}^{D_3}, \boldsymbol{\theta} \mid \mathrm{data}\right),$$

where

$$\boldsymbol{\theta} = \left\{\left\{\{\boldsymbol{f}_{td}\}_{t=1}^{T}, \left\{\sigma_{id}^2\right\}_{i=1}^{N}\right\}_{d=1}^{D_1}, \{\xi_d\}_{d=1}^{D_3}, \left\{\{\phi_{ij}\}_{i=1}^{N}, \zeta_j, \{v_{j,l}\}_{\ell=1}^{D}, M_j\right\}_{j=1}^{J^*}, \rho\right\}$$

and

$$\mathrm{data} = \left\{\left\{\{y_{itd}\}_{d=1}^{D_1}, \{k_{itd}, n_{itd}\}_{d=1}^{D_2}, \{z_{itd}, w_{itd}\}_{d=1}^{D_3}, \boldsymbol{x}_{it}\right\}_{t=1}^{T_i}\right\}_{i=1}^{N}.$$

- **Problem**: Non identifiability of latent factors, $\boldsymbol{\Lambda}\boldsymbol{F}^\top = \boldsymbol{\Lambda}\boldsymbol{Q}\boldsymbol{Q}^{-1}\boldsymbol{F}^\top$ ($\boldsymbol{Q}$ orthonormal) as well as sign/label-switching problems make popular schemes (e.g. HMC/MALA) to fail.
- **Solution**: We employ the (simplified) manifold MALA with a state dependent proposal covariance matrix and we focus on facilitating computations.

Data augmentation to facilitate computations

- We **introduce** $\omega_{itd} \sim \mathrm{PG}(n_{itd}, 0)$; the binomial likelihood writes
$$\pi(k_{itd} \mid n_{itd}, \boldsymbol{\lambda}_i, \boldsymbol{g}_{td}, \boldsymbol{\eta}_{2,d}, \boldsymbol{x}_{it}, \omega_{itd}) \propto \exp\left\{-\frac{\omega_{itd}}{2}\left(\frac{\kappa_{itd}}{\omega_{itd}} - \boldsymbol{\lambda}_i^\top \boldsymbol{g}_{td} - \boldsymbol{\eta}_{2,d}^\top \boldsymbol{x}_{it}\right)^2\right\},$$
where $\kappa_{itd} = k_{itd} - n_{itd}/2$ (Polson et al., 2013); drawing $\boldsymbol{g}_{td}$ and $\boldsymbol{\eta}_{2,d}$ is performed with Gibbs steps.

Data augmentation to facilitate computations

- We **introduce** $\omega_{itd} \sim \mathrm{PG}(n_{itd}, 0)$; the binomial likelihood writes

$$\pi(k_{itd} \mid n_{itd}, \boldsymbol{\lambda}_i, \boldsymbol{g}_{td}, \boldsymbol{\eta}_{2,d}, \boldsymbol{x}_{it}, \omega_{itd}) \propto \exp\left\{-\frac{\omega_{itd}}{2}\left(\frac{\kappa_{itd}}{\omega_{itd}} - \boldsymbol{\lambda}_i^\top \boldsymbol{g}_{td} - \boldsymbol{\eta}_{2,d}^\top \boldsymbol{x}_{it}\right)^2\right\},$$

  where $\kappa_{itd} = k_{itd} - n_{itd}/2$ (Polson et al., 2013); drawing $\boldsymbol{g}_{td}$ and $\boldsymbol{\eta}_{2,d}$ is performed with Gibbs steps.

- For count outcomes we **introduce** $L_{itd} \sim \mathrm{CRT}(z_{itd}, w_{itd} q_{itd}/\xi_d)$; Zhou and Carin (2015) show that

$$\pi\left(z_{itd}, L_{itd} \mid w_{itd}, q_{itd}, \boldsymbol{x}_{it}, \xi_d\right) = \frac{\xi_d^{z_{itd}} L_{itd}! |S(z_{itd}, L_{itd})|}{(1+\xi_d)^{z_{itd}} (\log(1+\xi_d))^{L_{itd}}} \times \mathrm{Pois}\left(L_{itd}; \frac{w_{itd} q_{itd}}{\xi_d} \log(1+\xi_d)\right).$$

  ✓ Derivatives wrt $q_{itd} = \exp\left(\boldsymbol{\lambda}_i^\top \boldsymbol{h}_{td} + \boldsymbol{\eta}_{3,d}^\top \boldsymbol{x}_{it}\right)$ require **less computational cost** compared to the (non-aumented) NB likelihood.

  ✓ Gibbs step to update L_{itd}.

  - Check **Samartsidis et al., 2021** for technical details.

Real data analysis

The data

***Our simulation study shows that joint modelling outperforms univariate models in detecting the intervention.**

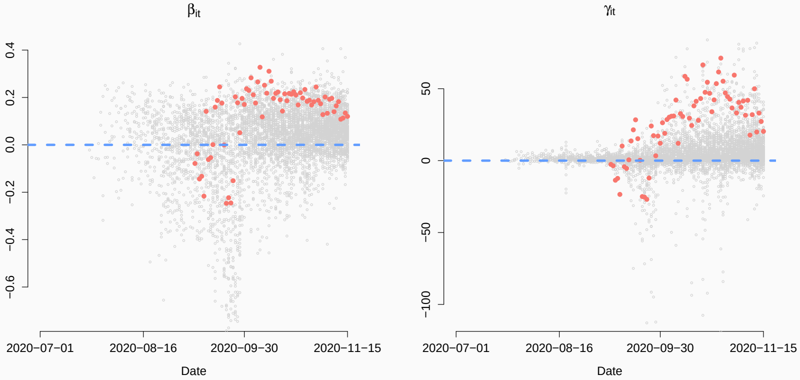
The data

***Our simulation study shows that joint modelling outperforms univariate models in detecting the intervention.**

- $N = 181$ units, $T = 138$ time points, $N_1 = 63$ units did not introduce LTP during the study period.
- Three binomial outcomes:
  - ✓ *Case completion*: proportion of cases completed out of new cases.
  - ✓ *Timeliness*: as above, within 48 hours.
  - ✓ *Contact completion*: proportion of contacts completed.
- One count outcome:
  - ✓ *Number of contacts* elicited from the completed cases

Case completion: causal effects

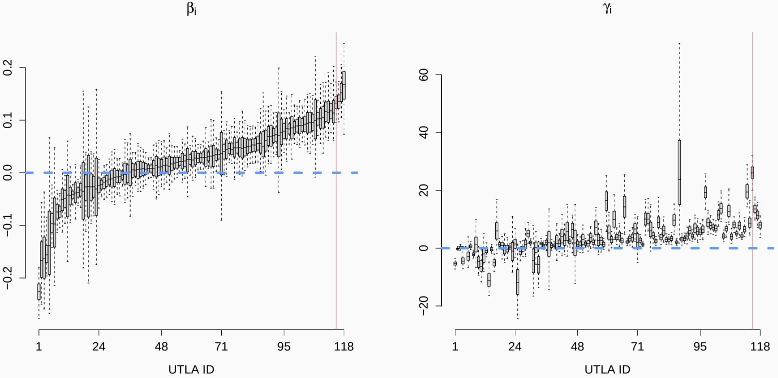
Point estimates (posterior means) of β_{it} and γ_{it} are shown below:



Both β_{it} and γ_{it} are positive on average. However, there is substantial heterogeneity. This is also true for the remaining outcomes

Case completion: average unit effects

For treated units, we define the average effects as $\beta_i = \frac{1}{T-T_i}\sum_{t=T_i+1}^{T}\beta_{it}$ and $\gamma_i = \frac{1}{T-T_i}\sum_{t=T_i+1}^{T}\gamma_{it}$. Posterior summaries are shown below:



Discussion

New method for causal inference with time-series observational data

- Can deal with outcomes of mixed type.
- Increases efficiency by jointly modelling multiple outcomes.
- Uncertainty quantification building efficient Bayesian estimation techniques.

Evaluation of LTPs

- On average, LTPs improved case completion and timely case completion.
- LTPs might have had an adverse effect on # of contacts elicited.
- Considerable heterogeneity in the estimates of the causal effects.