Structural parameters estimation in multi-partite networks.

Lucas Resende, **Guillaume Lecué**, Lionel Wilner, Philippe Choné

The Graph4Health project (ENSAE, ESSEC, INSERM, CASD)

Some of our aims:

1. understand the **impact of medical deserts** on health outcomes
2. construct tools to **evaluate health policy**
3. construct a **recommendation system** to help 'Médecin traitant' to recommend specialists, medical procedures, medicines, etc. to patients
4. prediction tools at the patient level: 'predictive medicine'

Based on the SNDS database (système national de santé).

The SNDS database

The SNDS collects and stores health data from various sources:

1. public health agency (CNAM): SNIIRAM from 'carte vitale' (système national d’information interrégimes de l’Assurance Maladie) (1998)
2. hospitals: PMSI programme de médicalisation des systèmes d’information
3. INSERM's databases on medical causes of death: CépiDc (Centre d’épidémiologie sur les causes médicales de décès, depuis 2017)
4. disability data: MDPH and CNSA (2018) (maisons départementales des personnes handicapées and Caisse nationale de solidarité pour l’autonomie, 2018)
5. a sample of data from complementary health insurance (2019)

One of the largest health database over 66 millions of French peoples since 1998.

The SNDS database

The purpose of the SNDS is to make these data available for studies, research or evaluation contributing to one of the following purposes:

1. information on health;
2. implementation of public health policies;
3. knowledge of health expenditure;
4. informing professionals and institutions about their activities;
5. innovation in the fields of health and medico-social care;
6. surveillance, monitoring and health security.

Accessing the SNDS

Since April 2017, any person or structure, public or private, for-profit or not-for-profit, can access SNDS data with the authorization of the Cnil , in order to carry out a study, research or evaluation of public interest.

Access to and use of SNDS data can only be made under conditions that respect the SNDS security guidelines , aimed at guaranteeing the confidentiality and integrity of the data and the traceability of access and other processing.

Graph4Health's dataset

We have been given access to 11 years of SNDS (2012--2022). So far we have data from 2016 to 2019 (**50 TB** of data).

Our aim in this presentation

Set of questions 1 (structural / homophily parameters estimation):

1. What is the impact of distance on our visits to doctors?
2. Do women (resp. men) visit more female (resp. male) doctors?
3. Do young (resp. old) people visit more often young (resp. old) doctors?

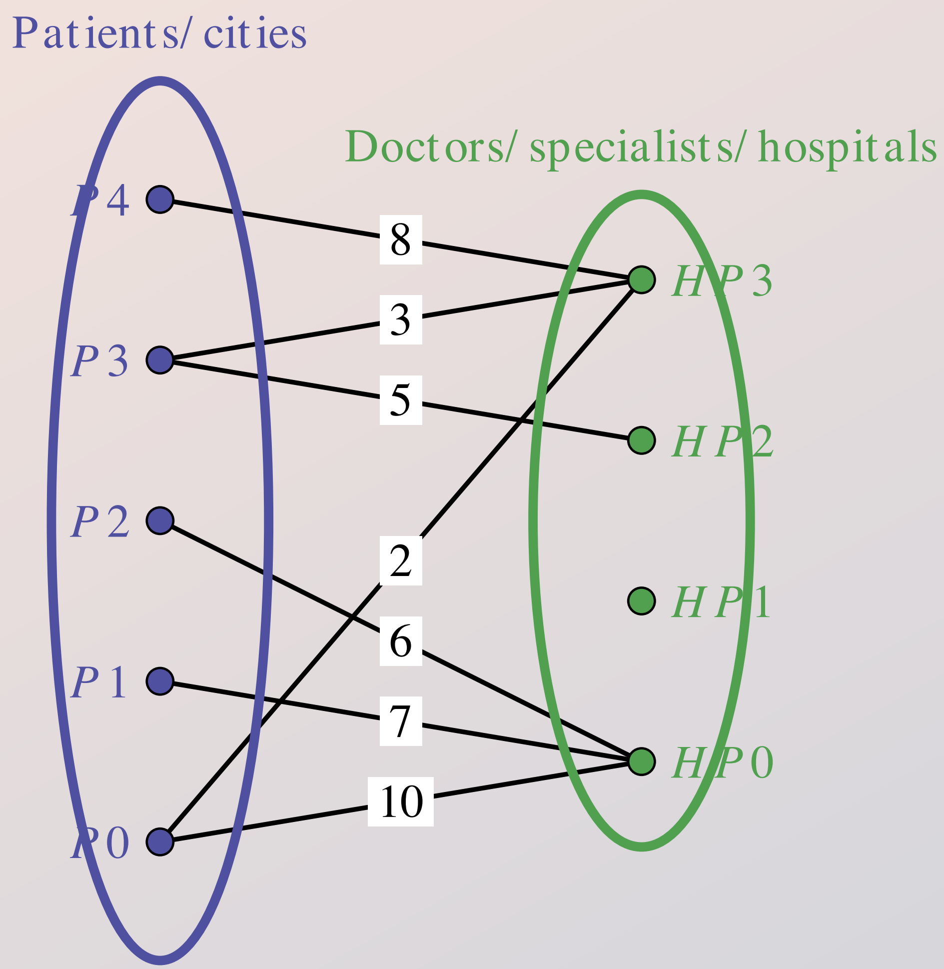
Based on **Cross-section data**: look at all consults in 2016.

Set of question 2 (causality):

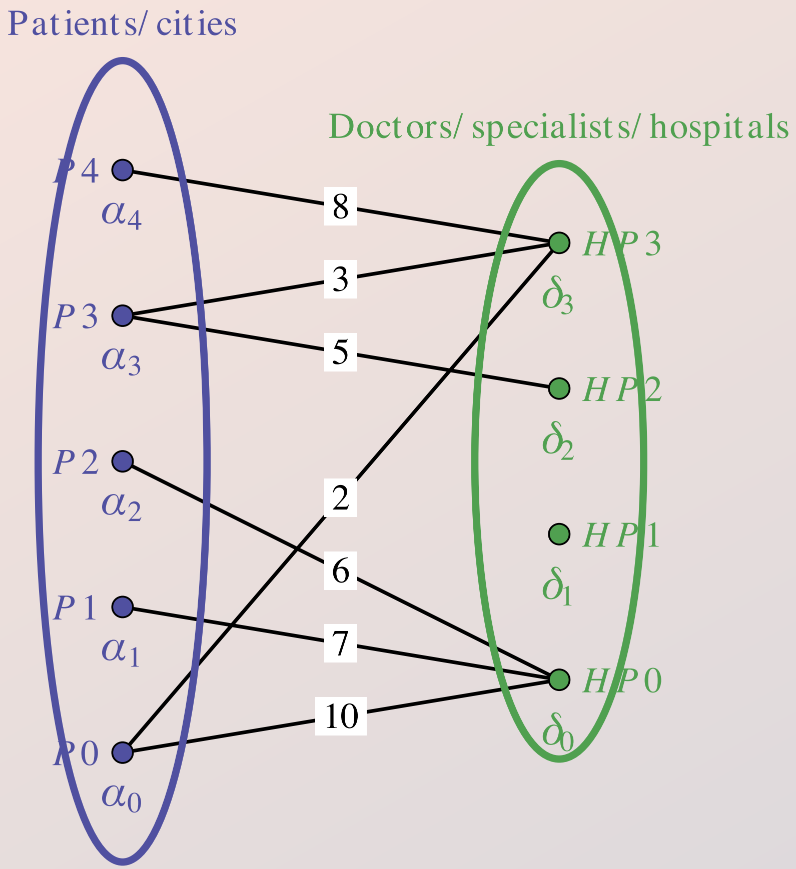
1. what is the causal effect of a policy reform in May 2017 (Sector~1 GPs' fees have increased by 8.7%) on spatial accessibility to health care in France?

Based on **longitudinal data**: all consults from 2016 to 2018.

Cross-section data from year 2016



Statistical modelling: Poisson regression with fixed effects



The Model: Edges' weight are independent and distributed according to (patient i and doctor j)

$$Y_{ij} \sim Poisson(\lambda_{ij})$$

where:

$$\lambda_{ij} = \exp(\alpha_i + \delta_j + < X_{ij}, \beta >)$$

1. α_i and δ_j are the fixed effects (incidental parameters)
2. β is the structural parameter
3. X_{ij} a vector of features of the dyads (i, j) (distance, same sex, same age bin, etc.)

Maximum likelihood estimation (1/2)

Our aim: estimate β, the structural parameter. But there are incidental parameters of no interest = nuisance parameters...many of them...

First approach = MLE

Step 1: The probability density function under $\mathbb{P}_{(\beta,(\alpha_i)_i,(\delta_j)_j)}$ is

$$f_{(\beta,(\alpha_i),(\delta_j))} : y \in \mathbb{N}^{PD} \to \mathbb{P}_{(\beta,(\alpha_i),(\delta_j))}[Y = y]$$

$$= \prod_{i,j} \mathbb{P}_{(\beta,(\alpha_i),(\delta_j))}[Y_{ij} = y_{ij}] = \prod_{i,j} \frac{\lambda_{ij}^{y_{ij}}}{y_{ij}!} \exp(-\lambda_{ij})$$

Step 2: the log-likelihood function is

$$\mathcal{L}(\beta, (\alpha_i), (\delta_j)) = \log f_{(\beta,(\alpha_i),(\delta_j))}(Y)$$

Maximum likelihood estimation (2/2)

Step 3: The MLE $(\hat{\beta}, (\hat{\alpha}_i), (\hat{\delta}_j))$ maximizes

$$(\beta, (\alpha_i), (\delta_j)) \to \sum_{ij} Y_{ij}(\alpha_i + \delta_j + < X_{ij}, \beta >) - \exp(\alpha_i + \delta_j + < X_{ij}, \beta >).$$

Hence, to estimate β (that is less than 10 real numbers in general) we end up estimating 70 millions coefficients!

The impact of the incidental parameters (ie $\hat{\alpha}_i$'s and $\hat{\delta}_j$'s) on the estimation quality of $\hat{\beta}$ is known as the Incidental Parameter Problem = IPP (NS '48)

An example of IPP in the Gaussian model (Neyman and Scott'48)

Gaussian model with incidental parameter (i.e. FE $(\alpha_i)_i$) and a structural paremeter (unknow variance σ^2): $i \in [n]$ and $j \in [T]$

$$X_{ij} = \alpha_i + \sigma g_{ij}, \text{ where } g_{ij} \overset{i.i.d.}{\sim} \mathcal{N}(0,1)$$

- MLE is $\hat{\alpha}_i = \bar{X}_{i\cdot}, i \in [n]$ and

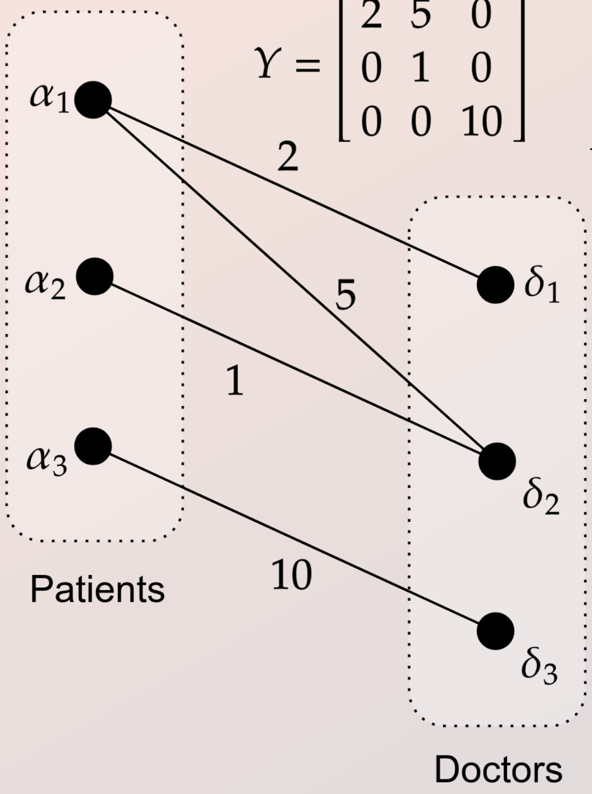
$$\hat{\sigma}^2 = \frac{1}{nT} \sum_{i=1}^{n} \sum_{j=1}^{T} \left(X_{ij} - \bar{X}_{i\cdot}\right)^2 \sim \sigma^2 \frac{\chi^2[n(T-1)]}{nT}.$$

- $\hat{\sigma}^2$ is not consistent as $n \to \infty$ (and fixed T): indeed, for all n, T,

$$\mathbb{E}[\hat{\sigma}^2] = \sigma^2 \frac{T-1}{T} \text{ and } \mathbb{V}[\hat{\sigma}^2] = 2\frac{(T-1)}{nT^2}.$$

(*Note:* A consistent estimator with a bounded variance is asymptotical unbiased)

The IPP in the Poisson model for bi-partite graphs (FW '18)



$$Y = \begin{bmatrix} 2 & 5 & 0 \\ 0 & 1 & 0 \\ 0 & 0 & 10 \end{bmatrix}$$

$$X = \begin{bmatrix} 0 & 1 & 7 \\ 5 & 1.5 & 0 \\ 10 & 10 & 0 \end{bmatrix}$$

$$\mathbb{P}(i \text{ visited } j \text{ a total of } k \text{ times}) = e^{-\lambda_{ij}} \frac{\lambda_{ij}^k}{k!}$$

where the Poisson itensity is $\lambda_{ij} = \exp(\beta_\star X_{ij} + \alpha_i + \delta_j)$

$\widehat{\beta}_{PPML} \not\to \beta_\star$ (in distribution under weak exogeneity)

A key idea: the diff-in-diff (DiD) - example in linear regression

Problem: how to get rid of the fixed effects so that we can just estimate β?

The DiD in the linear regression model with fixed effects

$$Z_{ij} = \alpha_i + \delta_j + < X_{ij}, \beta > + \epsilon_{ij}.$$

We observe that

$$Z_\xi = (Z_{ij} - Z_{ij'}) - (Z_{i'j} - Z_{i'j'}) =< \tilde{X}_\xi, \beta > + \epsilon_\xi$$

where

$$\xi = \begin{pmatrix} i & j \\ i' & j' \end{pmatrix} \text{ and } \tilde{X}_\xi = (X_{ij} - X_{ij'}) - (X_{i'j} - X_{i'j'})$$

This can be done for any tetrads ξ.

The DiD strategy in Poisson regression: sufficient statistics for the FE

How can we proceed the diff-in-diff trick in our Poisson regression model?
Go for CMLE (conditional MLE).

The log-likelihood function is (up to a constant independent of the parameters)

$$(\beta, (\alpha_i), (\delta_j)) \to \sum_i d_{i.} \alpha_i + \sum_j d_{.j} \delta_j + \sum_{ij} Y_{ij} < X_{ij}, \beta > - \exp(\alpha_i + \delta_j + < X_{ij}, \beta >)$$

where

$$d_{i.} = \sum_j Y_{ij} \text{ is the degree of } i \text{ and } d_{.j} = \sum_i Y_{ij} \text{ is the degree of } j.$$

Hence, **the degrees are sufficient statistics for the FE**

Recap on sufficient statistics

Definition

If the likelihood function can be written as

$$\ell(\theta) = \exp\left(h(Y).\, g(\theta, S(Y))\right)$$

then $S(Y)$ is a sufficient statistics for parameter θ.

Definition

If the likelihood function for $\theta = (\beta, \gamma)$ can be written as

$$\ell(\theta) = \exp\left(h(Y, \beta).\, g(S(Y), \beta, \gamma)\right)$$

then $S(Y)$ is a sufficient statistics for parameter γ.

CMLE = conditional Maximum likelihood Estimation

Idea

Conditionning on a sufficient statistics for γ (a nuisance parameter) removes γ from the likelihood function.

$$\mathbb{P}_\theta[Y = y | S(Y) = S(y)] = \frac{\mathbb{P}_\theta[Y = y \text{ and } S(Y) = S(y)]}{\mathbb{P}_\theta[S(Y) = S(y)]} = \frac{\mathbb{P}_\theta[Y = y]}{\mathbb{P}_\theta[S(Y) = S(y)]}$$

$$= \frac{e^{h(y,\beta)} e^{g(S(y),\theta)}}{\sum_{z:S(z)=S(y)} e^{h(z,\beta)} e^{g(S(z),\theta)}} = \frac{e^{h(y,\beta)}}{\sum_{z:S(z)=S(y)} e^{h(z,\beta)}} := \ell(\beta, y)$$

The idea of CMLE is to maximize $\beta \to \log \ell(\beta, Y)$ (where Y is my data).

Why CMLE works?

Theorem (CMLE principle)

Let $S(Y)$ be a sufficient statistics for γ. Denote

$$\mathcal{L}(\beta, y) := \log \mathbb{P}_{(\beta,\gamma)}[Y = y | S(Y) = S(y)] \text{ (which is independent of } \gamma)$$

Let β^* and assume that $Y \sim \mathbb{P}_{\theta^*}$ for $\theta^* = (\beta^*, \gamma^*)$ - whatever γ^* is.
Then

$$\beta \to \mathbb{E}_{\theta^*} \mathcal{L}(\beta, Y)$$

is maximal at $\beta = \beta^*$.

Proof of the CMLE principle

Assume that the pdf of Y is of the shape $p_\theta(y) = g(y, \beta)h(S(y), \theta)$ under $\mathbb{P}_\theta$.

Step 1: Law of iterated expectation:

$$\mathbb{E}_{\theta^*}\mathcal{L}(\beta, Y) = \mathbb{E}_{\theta^*}\left[\mathbb{E}_{\theta^*}\left[\mathcal{L}(\beta, Y)|S(Y)\right]\right].$$

Step 2: For all s, we have

$$\mathbb{E}_{\theta^*}\left[\mathcal{L}(\beta, Y)|S(Y) = s\right] = \mathbb{E}_{\beta^*}\left[\log f_\beta(Z)\right].$$

where $f_\beta(z) = \frac{g(z,\beta)}{\sum_{z:S(z)=s} g(z,\beta)}$ is the pdf of $Z = (Y|S(Y) = s)$ under $\mathbb{P}_\theta$

Step 3: From Jensen's inequality, we get

$$\mathbb{E}_{\beta^*}\left[\log f_\beta(Z)\right] \leq \mathbb{E}_{\beta^*}\left[\log f_{\beta^*}(Z)\right]. \qquad \square$$

CMLE in our Poisson model

Let $S(y) = \left(\left(\sum_j y_{ij}\right)_i, \left(\sum_i y_{ij}\right)_j\right)$ be the vector of all degrees of graph y: we know that $S(Y)$ is a sufficient statistics for all FEs. Let $\theta = (\beta, (\alpha_i), (\delta_j))$

$$\mathbb{P}_\theta[Y = y | S(Y) = S(y)] = \frac{\mathbb{P}_\theta[Y = y]}{\mathbb{P}_\theta[S(Y) = S(y)]} = \frac{\prod_{ij} \frac{\lambda_{ij}^{y_{ij}}}{y_{ij}!} e^{-\lambda_{ij}}}{\sum_{z:S(z)=S(y)} \prod_{ij} \frac{\lambda_{ij}^{z_{ij}}}{z_{ij}!} e^{-\lambda_{ij}}}$$

$$= \left(\sum_{z:S(z)=S(y)} \left(\prod_{ij} \lambda_{ij}^{z_{ij}-y_{ij}} \right) \left(\prod_{ij} \frac{y_{ij}!}{z_{ij}!} \right) \right)^{-1}$$

and for all z and y having the same degrees (i.e. $S(y) = S(z)$):

$$\left(\prod_{ij} \lambda_{ij}^{z_{ij}-y_{ij}} \right) = \exp\left(\sum_{ij} (\alpha_i + \delta_j + < X_{ij}, \beta >)(z_{ij} - y_{ij}) \right) = \exp\left(< \sum_{ij} (z_{ij} - y_{ij}) X_{ij}, \beta > \right)$$

CMLE in our Poisson model

All FEs diseappered from the conditional likelihood function:

$$\log \mathbb{P}_\theta[Y = y | S(Y) = S(y)] = -\log\left(\sum_{z:S(z)=S(y)} \exp\left(< \sum_{ij} (z_{ij} - y_{ij}) X_{ij}, \beta > \right) \left(\prod_{ij} \frac{y_{ij}!}{z_{ij}!} \right) \right)$$

hence, the conditional maximum likelihood estimator is $\hat{\beta}$ minimizing

$$\beta \to \log\left(\sum_{z:S(z)=S(Y)} \exp\left(< \sum_{ij} (z_{ij} - Y_{ij}) X_{ij}, \beta > \right) \left(\prod_{ij} \frac{Y_{ij}!}{z_{ij}!} \right) \right)$$

But the sum $\sum_{z:S(z)=S(Y)}$ is not computationally feasible!

Problem

It is not possible to construct the conditional log-likelihood function

$$\beta \to \log\left(\sum_{z:S(z)=S(Y)} \exp\left(< \sum_{ij}(z_{ij} - Y_{ij})X_{ij}, \beta > \right)\left(\prod_{ij} \frac{Y_{ij}!}{z_{ij}!}\right)\right)$$

However, we don't have to take all z such that $S(z) = S(Y)$, we may restrict ourself to a subset of them: look at tetrads and use a Diff-in-Diff argument on $\log(\lambda_{ij})$!

The diff-in-diff strategy in our Poisson regression model

Let $\xi = \begin{pmatrix} i & j \\ i' & j' \end{pmatrix}$ be a tetrad. We have

$$\underbrace{\left(\ln\lambda_{ij} - \ln\lambda_{ij'}\right) - \left(\ln\lambda_{i'j} - \ln\lambda_{i'j'}\right)}_{\text{cancel } \alpha_i,\alpha_{i'},\delta_j,\delta_{j'}} =< \beta, \tilde{X}_\xi >$$

where $\tilde{X}_\xi$ is the tetrad feature: $\tilde{X}_\xi = (X_{ij} - X_{ij'}) - (X_{i'j} - X_{i'j'})$

We define the <u>DiD sign matrix s_ξ w.r.t. tetrad</u> ξ at edge (i_0, j_0) by

$$s_\xi(i_0 j_0) = \begin{cases} 0 \text{ if } (i_0, j_0) \text{ is not an edge of } \xi \\ (\mathbf{1}(i_0 = i) - \mathbf{1}(i_0 = i'))(\mathbf{1}(j_0 = j) - \mathbf{1}(j_0 = j')) & \text{otherwise.} \end{cases}$$

$$\sum_{\text{edge in } \xi} s_\xi(edge) \ln\lambda_{edge} = \left(\ln\lambda_{ij} - \ln\lambda_{ij'}\right) - \left(\ln\lambda_{i'j} - \ln\lambda_{i'j'}\right) =< \beta, \tilde{X}_\xi >$$

The DiD in (conditional) likelihood for a given tetrad

- *Question:* Given a graph y, how to construct a new graph z such that

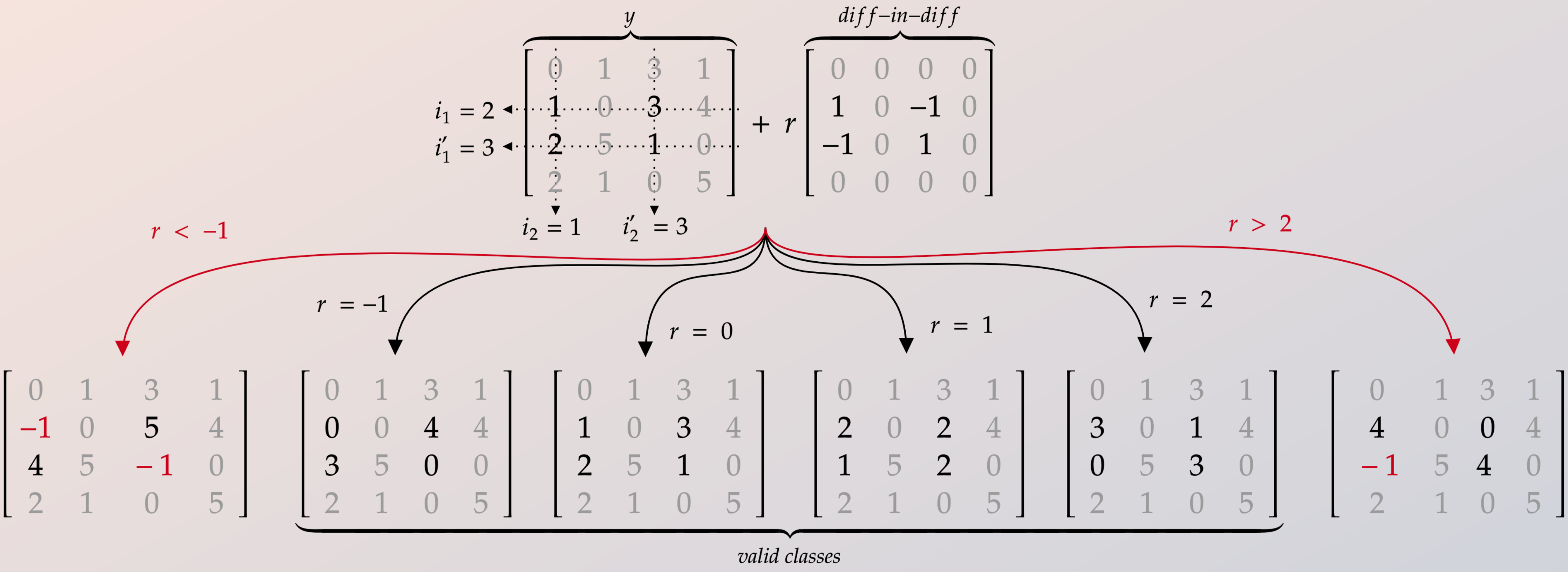
$$\ln \mathbb{P}_\theta (Y = z) - \ln \mathbb{P}_\theta (Y = y) = \sum_{i,j} (z_{ij} - y_{ij}) \ln \lambda_{ij}(\theta) - (\ln z_{ij}! - \ln y_{ij}!)$$

where $\theta = (\beta, (\alpha_i)_i, (\delta_j)_j)$ and $\ln \lambda_{ij}(\theta) = \alpha_i + \psi_j + < X_{ij}, \beta >$, is **not** a function of the fixed effects $(\alpha_i)_i, (\delta_j)_j$?

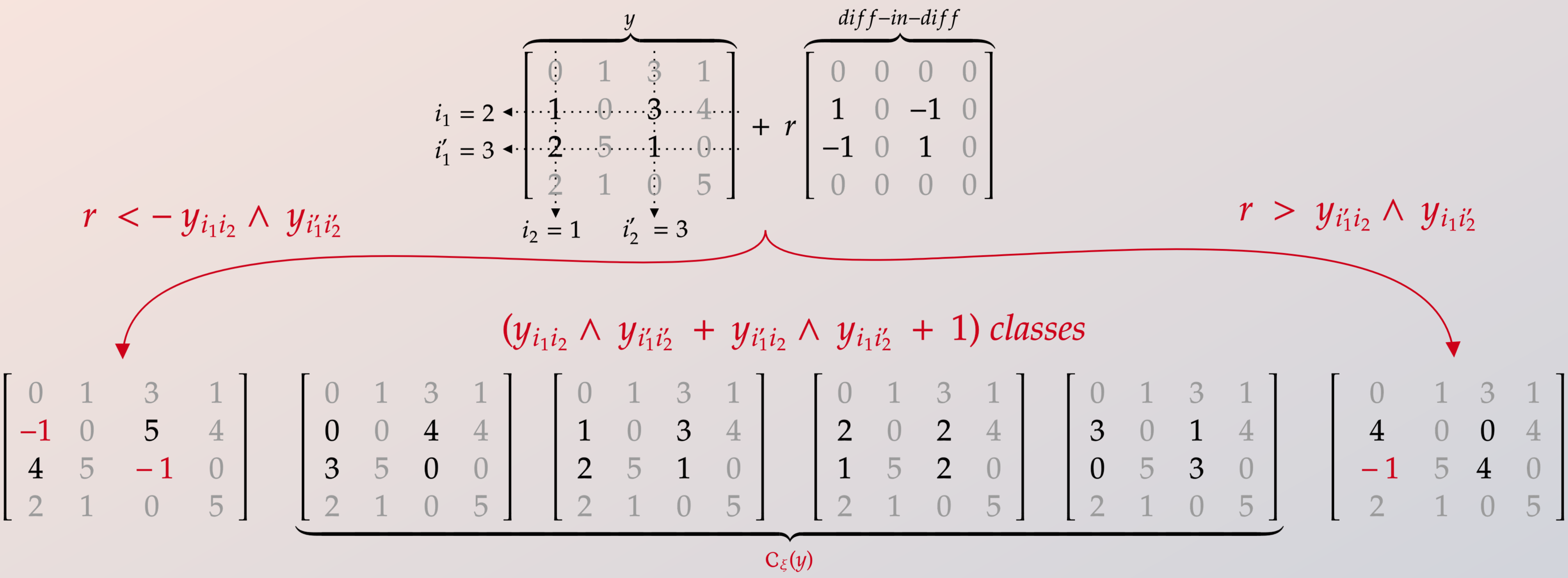
- *Answer:* Take a polyad ξ and let $z_{ij} - y_{ij}$ be a multiple of the DiD sign $s_\xi(ij)$:

$$z := T_\xi^r(y) := y + r s_\xi, \text{ i.e. } z_{ij} = y_{ij} + r s_\xi(ij), \text{ for } r \in \mathbb{Z}$$

The set of valid r's defines an orbit of graphs $\mathcal{O}_\xi(y)$ for $\xi = \begin{pmatrix} i_1 & i_2 \\ i_1' & i_2' \end{pmatrix}$



A multiclass classification problem for each tetrad ξ



A multiclass classification problem for each tetrad ξ

Definition

For every tetrad ξ we let the orbit of y be:

$$\mathcal{O}_\xi(y) := \left\{T_\xi^r(y) : -m_\xi(y) \leq r \leq M_\xi(y)\right\}$$

where

$$m_\xi(y) = \min(y_{ij} : s_\xi(ij) = 1) \text{ and } M_\xi(y) = \min(y_{ij} : s_\xi(ij) = -1).$$

- <u>To design a loss function:</u> for each ξ, we look at our data as a multiclass classification problem. The number of classes is $1 + m_\xi(y) + M_\xi(y)$
- A multiple multi-class classification problems dataset $(\tilde{X}_\xi, Y_\xi = 0)_\xi$ where $Y_\xi = 0$ is the class 0 observed among all classes $\{-m_\xi(Y) \leq r \leq M_\xi(Y)\}$

Designing a loss function from the tetrads' classification problems

- The set of *active tetrads* is

$$\Xi_a := \{\xi : m_\xi(y) + M_\xi(y) \geq 1\}$$

(If $m_\xi(y) + M_\xi(y) = 0$, there is no classification problem)

Loss function and the tetrads estimator

Our loss function is $\beta \to \widehat{L}(\beta) = \sum_{\xi \in \Xi_a} \ell_\xi(\beta|Y)$ where for all ξ and y

$$\ell_\xi(\beta|y) = -\ln \mathbb{P}_\beta(Y = y | Y \in \mathcal{O}_\xi(y)).$$

The **tetrads estimator** is $\widehat{\beta}_{\Xi_a} = \arg\min_\beta \widehat{L}(\beta)$.

Computational properties of the loss function

Newton's method to construct the tetrads estimator $\widehat{\beta}_{\Xi_a}$: $\beta_{k+1} = \beta_k - \widehat{H}_k^{-1}\widehat{S}_k$ where $\widehat{H}_k = \nabla^2\widehat{L}(\beta_k)$ and $\widehat{S}_k = \nabla\widehat{L}(\beta_k)$.

We have for all tetrads ξ and graph y:

$$\ell_\xi(\beta|y) = \ln\left(\sum_{r=-m_\xi(y)}^{M_\xi(y)} \exp\left(r < \tilde{X}_\xi, \beta > + \sum_{ij} \ln\frac{y_{ij}!}{(y_{ij} + r s_\xi(ij))!}\right)\right)$$

$$\nabla\ell_\xi(\beta|y) = (\mathbb{E}_\beta[m_\xi(Y)|Y \in \mathcal{O}_\xi(y)] - m_\xi(y))\widetilde{X}_\xi$$

$$\nabla^2\ell_\xi(\beta|y) = (\mathbb{V}_\beta[m_\xi(Y)|Y \in \mathcal{O}_\xi(y)])\widetilde{X}_\xi\widetilde{X}_\xi^\top$$

- It is strictly convex as long as $\widetilde{X}_\xi\widetilde{X}_\xi^T \succ 0$.
- Computational tricks to sum over all active tetrads.

Try it on Colab

We use Lucas Resende's package:

- first on synthetic data for a two-way model:

$$\lambda_{ij} = \exp(\alpha_i + \delta_j + \beta_0 X_{ij}^{(0)} + \beta_1 X_{ij}^{(1)})$$

- then on the trade data from Santos Silva & Tenreyro (2006) -- for only 20 countries
- then we will go to synthetic data from a three-ways model.

link to Colab

Extension to multi-ways models: A 'trade data' example

- **Trade data**: in the literature on gravity models for trade there is interest in learning from 4-dimensional data where
  - i_1 is an exporter country;
  - i_2 is an importer country;
  - i_3 is the industry;
  - i_4 is the time.
- Measuring $Y_{i_1 i_2 i_3 i_3}$ as the volume of trade and taking $X_{i_1 i_2 i_3 i_4}$ as the presence (or amount) of tariffs from i_1 to i_2 in industry i_3 at time i_4 we can measure the impact of tariffs.

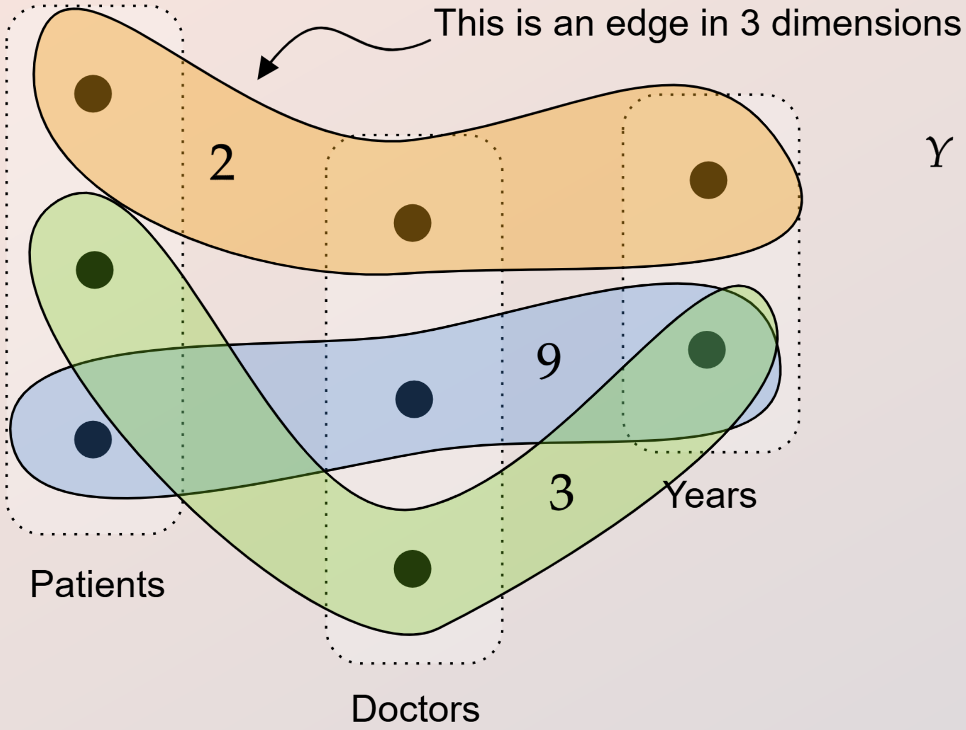
Extension to the multi-ways model

- $i_1 \in [n_1]$, $i_2 \in [n_2]$, ..., $i_D \in [n_D]$, $n = \prod_{d=1}^{D} n_d$
- $Y_{i_1 i_2 \ldots i_D} \geq 0$ and $X_{i_1 i_2 \ldots i_D} \in \mathbb{R}^p$
- $\mathbf{i} = (i_1, i_2, \ldots, i_d) \in \prod_{d=1}^{D} [n_d]$

$$\mathbb{P}_{\beta_\star, \theta_\star}(Y = y | X) = \prod_{\mathbf{i}} e^{-\lambda_{\mathbf{i}}(\beta_\star, \theta_\star)} \frac{\lambda_{\mathbf{i}}(\beta_\star, \theta_\star)^{y_{\mathbf{i}}}}{y_{\mathbf{i}}!}$$

- $\lambda_{\mathbf{i}}(\beta, \theta) = \exp\left\{\beta^T X_{\mathbf{i}} + \theta^{(1)}_{i_2 i_3 \ldots i_D} + \theta^{(2)}_{i_1 i_3 \ldots i_D} + \cdots + \theta^{(D)}_{i_1 i_2 \ldots i_{D-1}}\right\}$
- $\beta \in \mathbb{R}^p$ is the parameter of interest. θ is a vector containing the fixed effects (incidental parameters).

The IPP in Poisson ML for 3-way model (WZ '21)



$$Y = \begin{bmatrix} 2 & 0 & 0 \\ 0 & 0 & 0 \\ 0 & 0 & 0 \end{bmatrix}, \begin{bmatrix} 0 & 0 & 0 \\ 0 & 0 & 3 \\ 0 & 9 & 0 \end{bmatrix}$$

$$X = \begin{bmatrix} 0 & 9 & 3 \\ 3 & 1 & 0 \\ 5 & 9 & 7 \end{bmatrix}, \begin{bmatrix} 8 & 9 & 8 \\ 7 & 0 & 0 \\ 4 & 0 & 9 \end{bmatrix}$$

$$\mathbb{P}(i \text{ visited } j\ k \text{ times in in year } t) = e^{-\lambda_{ijt}} \frac{\lambda_{ijt}^{k}}{k!}$$

$$\text{where } \lambda_{ijt} = \exp(\beta_\star X_{ijt} + \alpha_{ij} + \delta_{it} + \gamma_{jt})$$

$\widehat{\beta}_{PPML} \not\to \beta_\star$ (in distribution)

Introducing polyads in multi-ways models

- A *polyad* is a pair of edges $\mathbf{i} = (i_d), \mathbf{i}' = (i'_d)$ such that $i_d \neq i'_d$ for all $d \in [D]$:

$$\xi = \begin{bmatrix} i_1 & i_2 & \cdots & i_D \\ i'_1 & i'_2 & \cdots & i'_D \end{bmatrix}$$

Note: we can form 2^D edges in a polyads.

- The *diff-in-diff sign tensor* s_ξ is defined for all edge $\mathbf{j} = (j_d)_d$ as

$$s_\xi(\mathbf{j}) = \prod_{d=1}^{D} \left(\mathbf{1}(i_d = j_d) - \mathbf{1}(i'_d = j'_d) \right)$$

The diff-in-diff strategy in the conditional likelihood for each polyad

- diff-in-diff works in the multi-way case:

$$\underbrace{\sum_{\mathbf{j}\in\prod_{d=1}^{D}\{i_d,i'_d\}} s_\xi(\mathbf{j})\ln\lambda_{\mathbf{j}}}_{\text{diff-in-diff of } \ln\lambda} =< \beta^T, \widetilde{X}_\xi > \text{ where } \widetilde{X}_\xi := \underbrace{\sum_{\mathbf{j}\in\prod_{d=1}^{D}\{i_d,i'_d\}} s_\xi(\mathbf{j})X_{\mathbf{j}}}_{\text{diff-in-diff of X}}$$

Loss function and the polyads estimator (same definition as $D = 2$)

Our loss function is $\beta \to \widehat{L}(\beta) = \sum_{\xi\in\Xi_a} \ell_\xi(\beta|Y)$ where for all ξ and y

$$\ell_\xi(\beta|y) = -\ln\mathbb{P}_\beta(Y = y|Y \in \mathcal{O}_\xi(y)),$$

$\mathcal{O}_\xi(y) := \{y + rs_\xi : -m_\xi(y) \le r \le M_\xi(y)\}$, $m_\xi(y) = \min(y_{\mathbf{i}} : s_\xi(\mathbf{i}) = 1)$, $M_\xi(y) = \min(y_{\mathbf{i}} : s_\xi(\mathbf{i}) = -1)$ and the **polyads estimator** is $\widehat{\beta}_{\Xi_a} = \arg\min_\beta \widehat{L}(\beta)$.

Definitions of Density and sparsity of a graph

- The *density of a graph* is given by
$$\rho = \frac{|E|}{n} \text{ where } E = \{\mathbf{i} : y_{\mathbf{i}} > 0\}.$$
- We say that a sequence of graphs $(y_{(k)})_{k>0}$ is *sparse* if $\rho_{(k)} \to 0$ and $n_{(k)} \to \infty$ as $k \to \infty$.

Ex.: Social networks typically have $\rho_{(k)} = O\left(\frac{1}{\sqrt{n_{(k)}}}\right)$.

Statistical properties

Consistency

Let $M_\lambda = \max_{\mathbf{i}} \left| \ln \lambda_{\mathbf{i}} \right|$. If Y follows the Poisson multi-way model and some technical conditions hold (boundness of the risk and separability of the minimum), then:

- Conditionaly on X, $\widehat{\beta}_{\Xi_a} \to \beta_\star$ in probability as $\rho \gg \frac{M_\lambda}{n}$.

Statistical properties

Asymptotic normality

- Adding other technical conditions, it holds $\forall v \in \mathbb{R}^p$, $||v|| = 1$, that

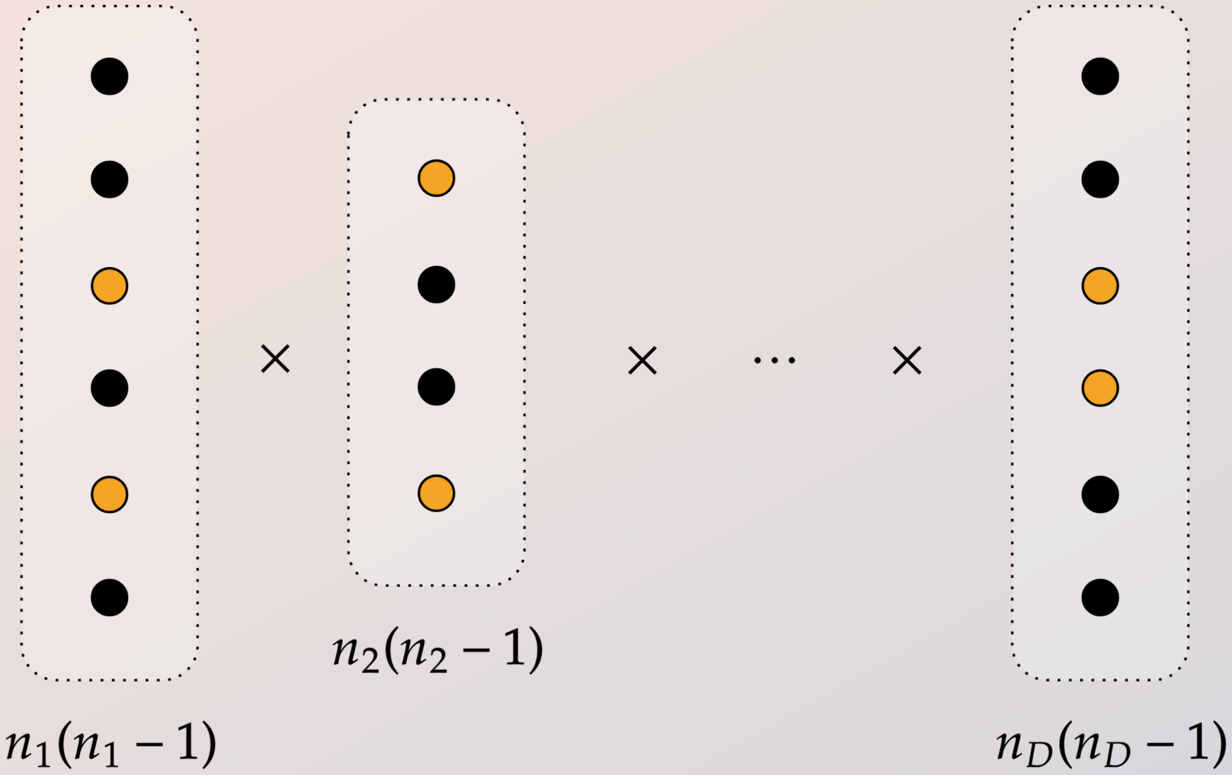
$$\frac{\langle \widehat{\beta}_{\Xi_a} - \beta_\star, v \rangle}{\sqrt{v^T (\nabla^2 \widehat{L}_{\Xi_a}^{-1}) \widehat{\Sigma}_{\Xi_a} (\nabla^2 \widehat{L}_{\Xi_a}^{-1}) v}} \to \mathcal{N}(0, 1) \text{ as } \rho \gg \frac{M_\lambda}{n}$$

where $\widehat{\Sigma}_{\Xi_a} = \sum_{\xi, \xi' \in \Xi_a} (\nabla \ell_\xi)(\nabla \ell_\xi)^T \mathbf{1}_{\xi \text{ and } \xi' \text{ share at least one edge}}.$

- In the worst case, $\widehat{\Sigma}_{\Xi_a}$ can be evaluated in $O(|E|^3)$. Typically it requires $O(|E|^2)$.

Looping over active polyads in Newton's method

The total number of polyads



$$\prod_{d=1}^{D} \binom{n_d}{2} = O(n^2)$$

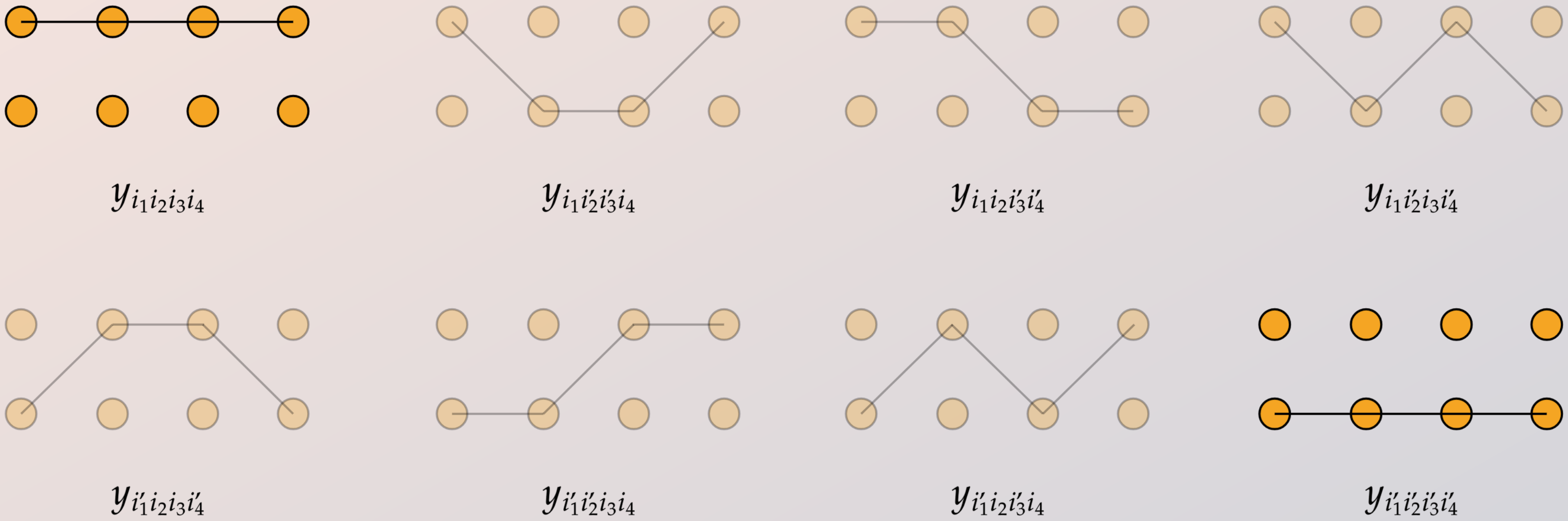
complexity for checking all polyads

$$O(n)$$

complexity for running PPML

Looping over active polyads in Newton's method

A computational trick



- If a ξ is active, then $\min_{\mathbf{i}:s_\xi(\mathbf{i})} y_{\mathbf{i}} > 0$, thus $y_{\mathbf{i}} > 0$, $y_{\mathbf{i}'} > 0$ (up to permutation).

Looping over <u>active</u> polyads in Newton's method

A computational trick

- Instead of looking at all $O(n^2)$ possible polyads we can search over all pairs $(\mathbf{i}, \mathbf{i}')$ such that $y_{\mathbf{i}} > 0$ and $y_{\mathbf{i}'} > 0$.
- This can be done in $O(|E|^2)$ where $E = \{\mathbf{i} : y_{\mathbf{i}} > 0\}$.
- It is better than $O(n^2)$ and can even be better than $O(n)$ (PPML).

Data generating setup (from WZ '21)

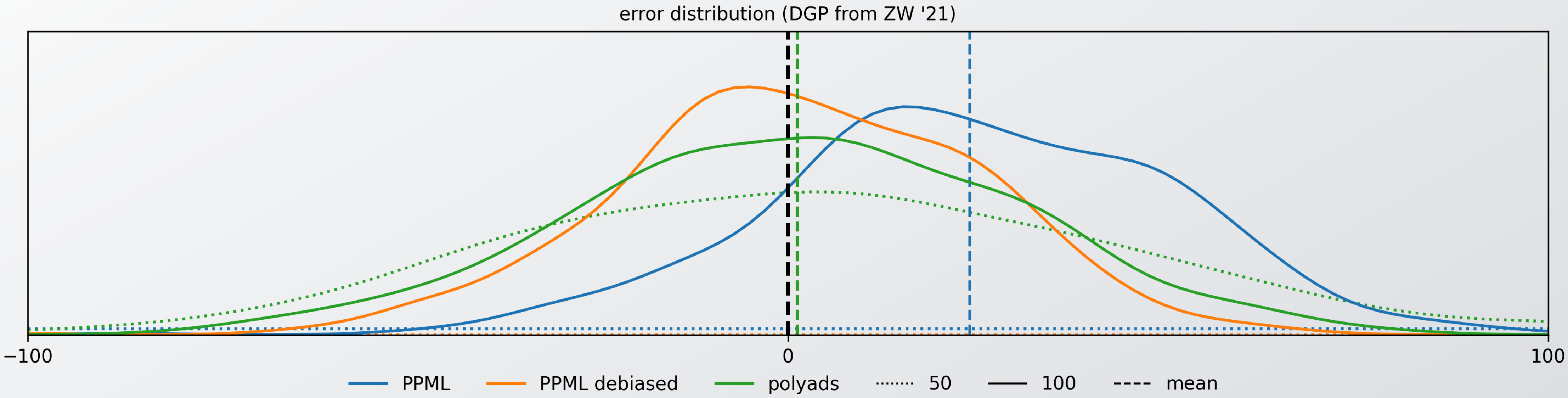
- 3-way DGP: $n_1 = n_2$ vary and $n_3 = 5$ is fixed.
- $\theta_{ij}, \theta_{it}, \theta_{jt}$ are i.i.d. normal with zero mean and variance $1/16$. $\beta_\star = 1$.
- The features are correlated with θ and with previous features in the 3rd axis:

$$X_{ijt} = \begin{cases} \frac{1}{2} X_{ij(t-1)} + \theta_{it} + \theta_{jt} + \frac{1}{4}\mathcal{N}(0,1), \text{ if } t > 1 \\ \theta_{it} + \theta_{jt} + \frac{1}{4}\mathcal{N}(0,1), \text{ if } t = 1 \end{cases}$$

- We let $\mathbb{E}(Y_{ijt}) = \exp\left(c + \beta_\star X_{ijt} + \theta_{ij} + \theta_{it} + \theta_{jt}\right)$.
- The constant c let us control the density. We consider Y_{ijt} Poisson and NB.

Computational experiments

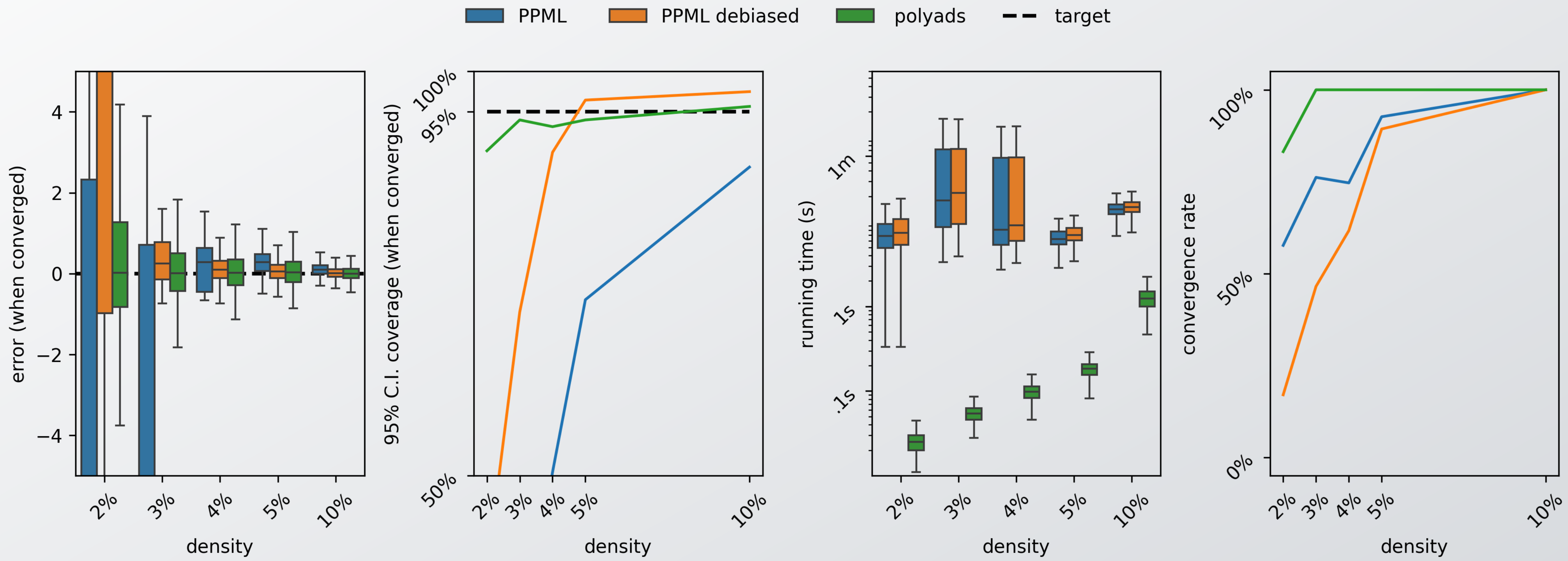
The initial example



- Coverage of the 95% C.I. ($n_1 = 100$): 81%(PPML), 97%(PPML debiased), 94% (Polyads). **We have a good concentration even when $n_1 = 50$.**

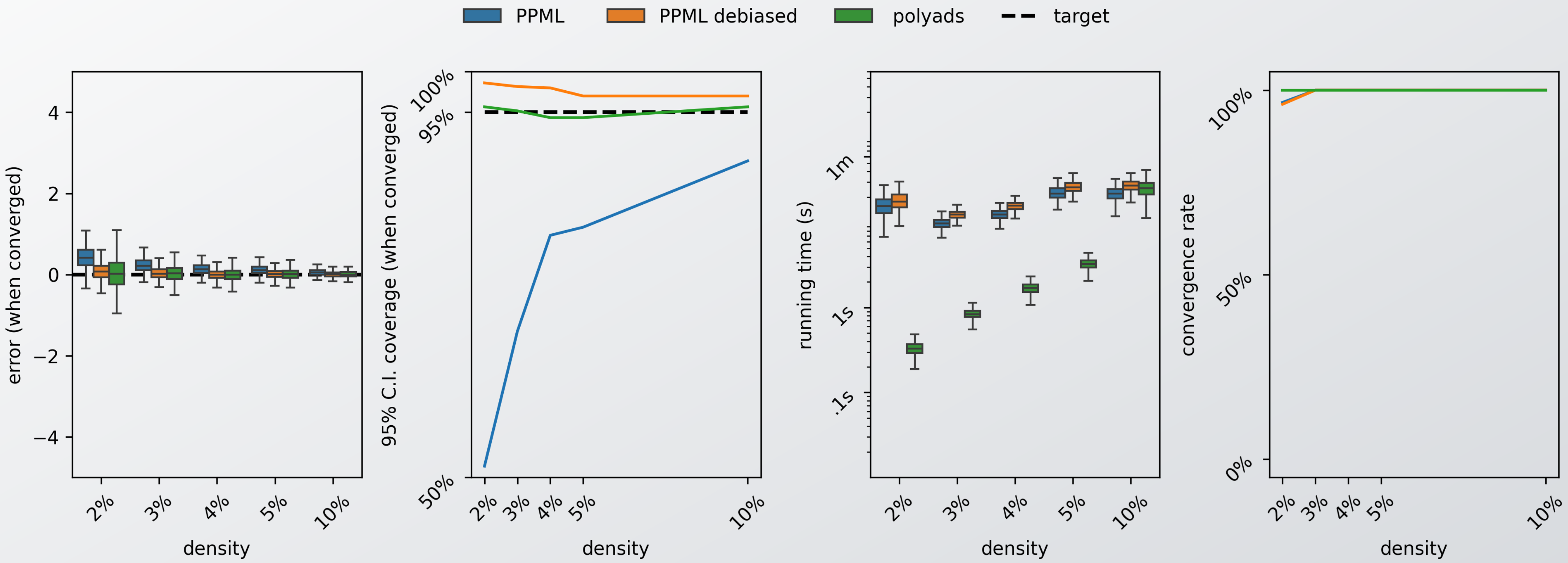
Computational experiments

Short panel ($n_1 = n_2 = 50, n_3 = 5$)



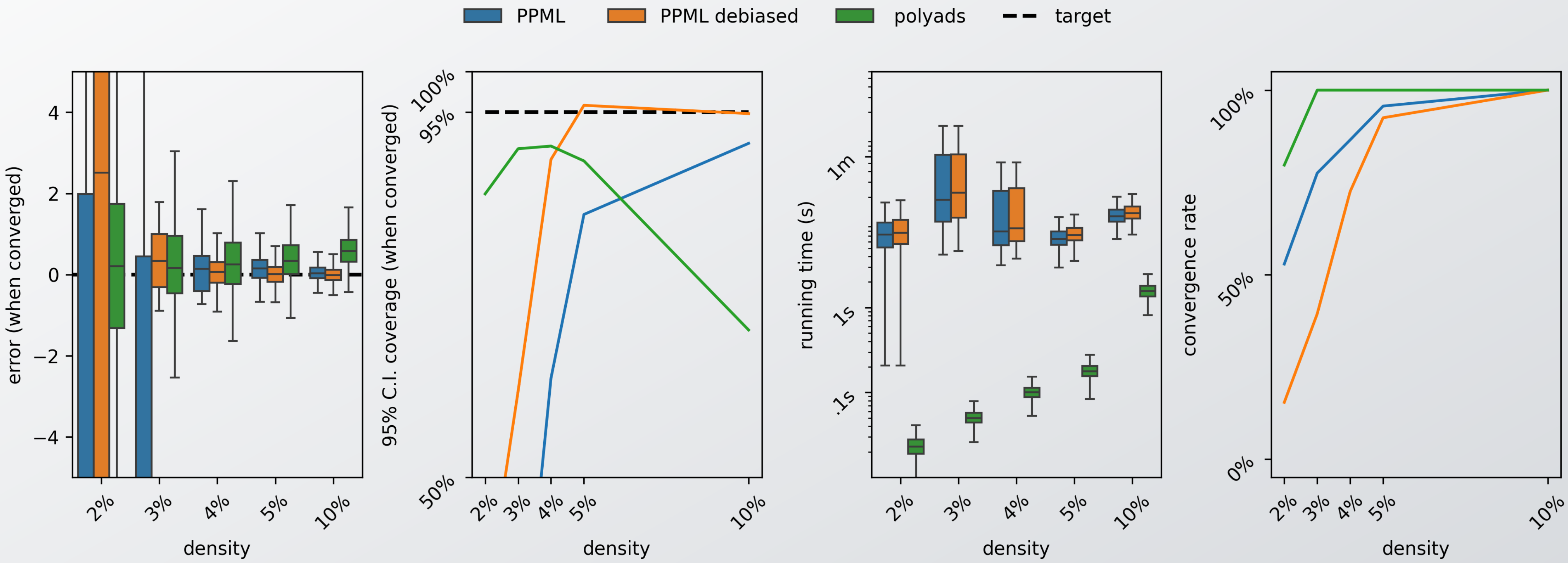
Computational experiments

Short panel ($n_1 = n_2 = 100, n_3 = 5$)



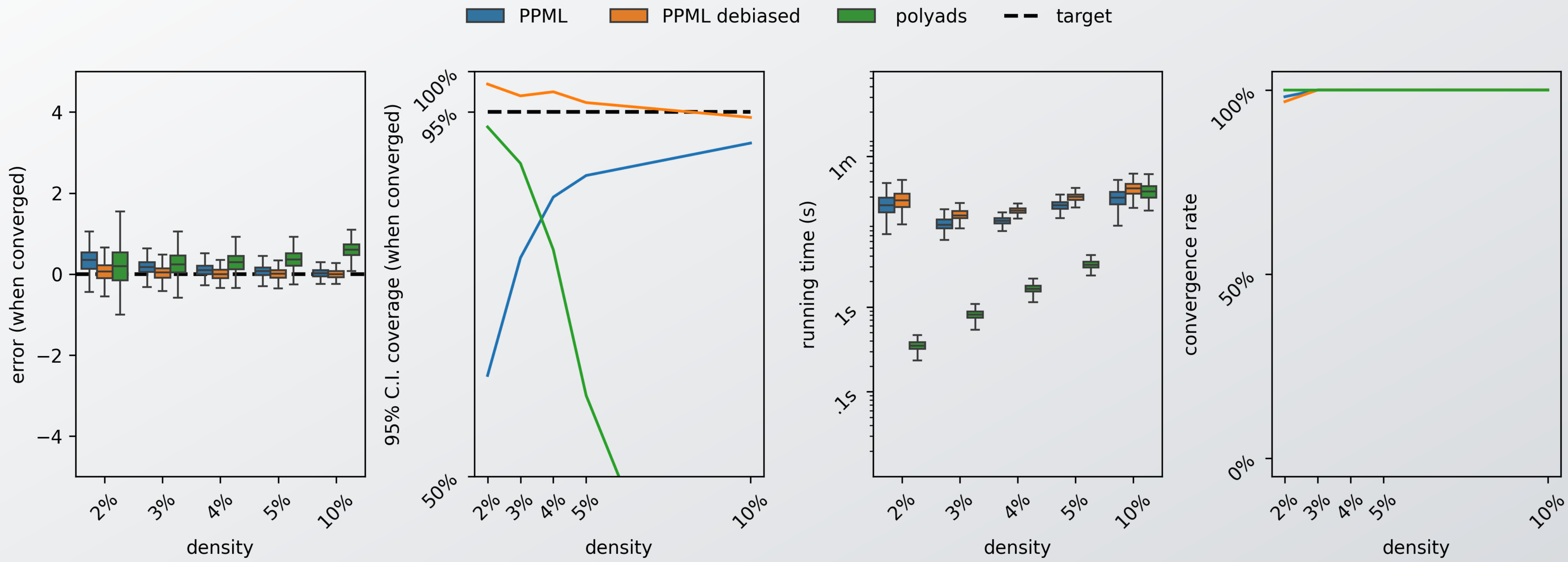
Computational experiments

Short panel ($n_1 = n_2 = 50, n_3 = 5$, NB)



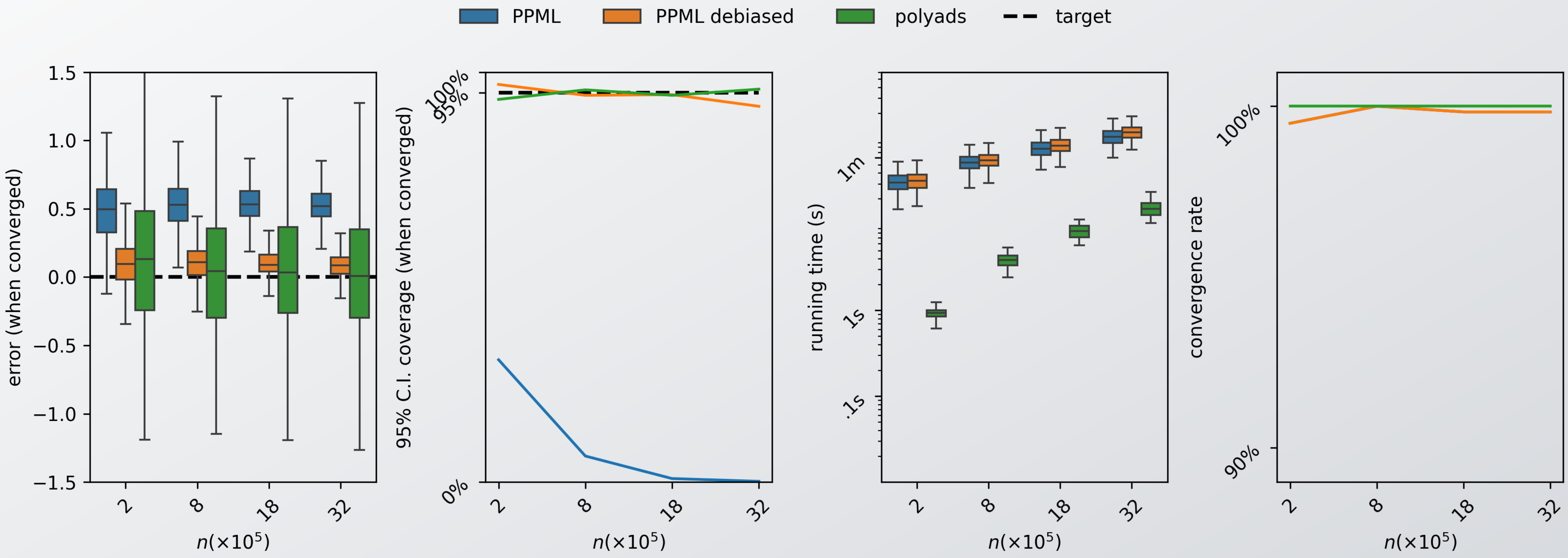
Computational experiments

Short panel ($n_1 = n_2 = 100, n_3 = 5$, NB)



Computational experiments

Sparse case ($\rho = \frac{4}{\sqrt{n}}$, NB)



Causal inference via structural parameter estimation on panel data

Context: We study health insurance claims data over the years 2016 to 2018.

In May 2017, the fees charged by general practitioners (GPs) belonging to the regulated sector (sector~1) have increased by 8.7%.

We find that the stronger financial incentives have caused physician activity (as measured by number of visits) to rise by approximately 10%

Questions about accessibility: Did this policy induce any change in gender homophily between patients and general practitioners (GPs)?

Did patient prefer more medical offices that are located in the same municipality?

Solution Using a three-way model and controlling for dyads fixed effects allows to estimate how the reform has changed the doctor-patient connections.

Causal inference via structural parameter estimation on panel data

Data aggregation: We aggregate data at the city-sex level: index i (resp. j) stands thus for the set of patients (resp. doctors) in a given municipality with given gender.

Outcome: Y_{ijt} is the number of visits by patient i to doctor j on month t.

Features construction (1/2):
The treatment T_j is a binary variable equal to 1 for sector 1 GPs, and to 0 for direct access specialists.

The reform has been implemented from May 2017 onward, hence the definition of Post_t, a dummy variable equal to 1 after that date.

Causal inference via structural parameter estimation on panel data

Features construction (2/2):

- the interactions between $\mathrm{Post}_t \times T_j$
- a dummy variable equals to 1 when patients and doctors have the same sex,
- a dummy variable that is equal to 1 when the medical office chosen is located in the same municipality as the patient's home,
- travel time (measured in minutes between the centroids of municipalities).

Model: A three-way Poisson model $Y_{ijt} \sim \mathcal{P}(\lambda_{ijt})$ with intensity given by

$$\ln \lambda_{ijt} = \left(\beta_{\mathtt{d}} d_{ij} + \beta_{\mathtt{sc}} \mathbf{1}\{\mathrm{city}_i = \mathrm{city}_j\} + \beta_{\mathtt{ss}} \mathbf{1}\{\mathrm{sex}_i = \mathrm{sex}_j\}\right) \times \mathrm{Post}_t \times T_j + u_{ij} + v_{jt} + w_{it}.$$

Causal inference via structural parameter estimation on panel data

$$\ln \lambda_{ijt} = \left(\beta_{\mathtt{d}} d_{ij} + \beta_{\mathtt{sc}} \mathbf{1}\{\text{city}_i = \text{city}_j\} + \beta_{\mathtt{ss}} \mathbf{1}\{\text{sex}_i = \text{sex}_j\}\right) \times \text{Post}_t \times T_j + u_{ij} + v_{jt} + w_{it}.$$

Parameter interpretation:

- $\beta_{\mathtt{d}}$: additional effect of the distance after the reform
- $\beta_{\mathtt{sc}}$: additional effect of living in the same city after the reform
- $\beta_{\mathtt{ss}}$: additional effect of homophily gender after the reform

Causal inference via structural parameter estimation on panel data

Result on SNDS: 95% confidence intervals (in parentheses) by subsample proportion and method (100 replications). $\beta_{\mathtt{d}}$ estimates are scaled by 10^4; $\beta_{\mathtt{sc}}$ and $\beta_{\mathtt{ss}}$ by 10^2.

Parameter	2% Subsample	3% Subsample	4% Subsample
$\beta_{\mathtt{d}}(\times 10^4)$	1.23 (0.26, 2.20)	1.73 (1.18, 2.28)	1.45 (1.03, 1.86)
$\beta_{\mathtt{sc}}(\times 10^2)$	-7.60 (-13.10, -2.10)	-3.91 (-6.55, -1.27)	-4.52 (-6.28, -2.75)
$\beta_{\mathtt{ss}}(\times 10^2)$	-2.48 (-4.45, -0.50)	-0.35 (-1.61, 0.92)	-0.20 (-1.13, 0.73)

Interpretation:

- a negative coefficient means that after the reform patients are more likely to visit a doctor outside their own city
- Having 0 in the confidence interval means we cannot tell about the effect of the reform

Conclusion

Our contributions

- Multi-way.
- Faster than PPML on sparse networks.
- No IPP problem by construction.
- Convex loss with good convergence.
- Reliable confidence intervals under model assumptions.