Matrix Completion Methods for Causal Panel Data Models

Susan Athey et al. (2021), JASA

Naoki Eguchi

June 5, 2025

Faculty of Medicine, Kyoto University

Introduction

Today's Agenda ; Keyword : Imputation

-

Imputation

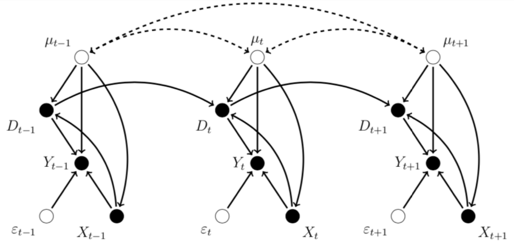
- As many panel data methods, we want to know ATT : $\mathbb{E}[Y_{it}(1) - Y_{it}(0)|W_i = 1]$.
- Thus, it boils down to estimate (impute) the counterfactual $Y_{it}(0)$.
  - Horizontal : Under unconfoundedness, we can impute counterfactual PO using observed outcomes for control units.
  - Vertical : By SCM, we can also impute it using weighted average outcomes for control units with most predictive weights trained with pre-treatment datas.

$$\mathbf{Y} = \begin{pmatrix} \checkmark & \checkmark & \checkmark \\ \vdots & \vdots & \vdots \\ \checkmark & \checkmark & \checkmark \\ \checkmark & \checkmark & ? \\ \vdots & \vdots & \vdots \\ \checkmark & \checkmark & ? \end{pmatrix}.$$

$$\mathbf{Y} = \begin{pmatrix} \checkmark & \checkmark & \dots & \checkmark & \checkmark & \dots & \checkmark \\ \checkmark & \checkmark & \dots & \checkmark & \checkmark & \dots & \checkmark \\ \checkmark & \checkmark & \dots & \checkmark & ? & \dots & ? \end{pmatrix}$$

Xu (2024): Counterfactual estimation

- functional form: $Y_{it}(0) = f(\mathbf{X_{it}}) + h(\mathbf{U_{it}}) + \epsilon_{it}$
  $\rightarrow$No anticipation, carryover, feedback, LDV
- strict exogeneity: $\forall i, j \in \{1, \ldots, N\}$, $\forall s, t \in \{1, \ldots, T\}, \epsilon_{it} \perp\!\!\!\perp \{D_{js}, \mathbf{X_{js}}, \mathbf{U_{js}}\}$



- low-dimensional decomposition: $h(\mathbf{U_{it}}) = \{L_{it}\}, \text{rank}(\mathbf{L}_{N \times T}) \ll \min\{N, T\}$
  $\rightarrow$ The rank (= number of factors) is FIXED !!

Set Up

Notation and Estimand

- Consider a setting with N units observed over T periods characterized by a binary treatment W_{it} and hence two POs $Y_{it}(1), Y_{it}(0)$.
    - $\mathbf{X} \in \mathbb{R}^{N \times P}$, $\mathbf{Z} \in \mathbb{R}^{T \times Q}$: observe (unit / time)-specific covariance matrix
- Estimand : $\mathbf{Y} = \{Y_{it}(0)^1\} = \begin{pmatrix} Y_{11}(0) & \cdots & Y_{1T}(0) \\ \vdots & \ddots & \vdots \\ Y_{N1}(0) & \cdots & Y_{NT}(0) \end{pmatrix}$ ($\leftarrow$ Matrix!!)
- $W_{it} = \begin{cases} 1 & \text{if } (i,t) \in \mathcal{M} \\ 0 & \text{if } (i,t) \in \mathcal{O} \end{cases}$

[1] 以降は簡単のため，$Y_{it}(0) = Y_{it}$ とし，“(0)” を省略して表記する．

Patterns of data matrix

- Ordinary case (rich data wrt. units and times)

$$\mathbf{Y}_{N\times T} = \begin{pmatrix} \checkmark & \checkmark & \checkmark & \checkmark & \dots & \checkmark \\ \checkmark & \checkmark & \checkmark & \checkmark & \dots & \checkmark \\ \checkmark & \checkmark & \checkmark & \checkmark & \dots & \checkmark \\ \checkmark & \checkmark & \checkmark & ? & \dots & ? \\ \checkmark & \checkmark & \checkmark & ? & \dots & ? \\ \vdots & \vdots & \vdots & \vdots & \ddots & \vdots \\ \checkmark & \checkmark & \checkmark & ? & \dots & ? \end{pmatrix}$$

- Staggered adoption

$$\mathbf{Y}_{N\times T} = \begin{array}{c} \begin{pmatrix} \checkmark & \checkmark & \checkmark & \checkmark & \dots & \checkmark \\ \checkmark & \checkmark & \checkmark & \checkmark & \dots & ? \\ \checkmark & \checkmark & ? & ? & \dots & ? \\ \checkmark & \checkmark & ? & ? & \dots & ? \\ \vdots & \vdots & \vdots & \vdots & \ddots & \vdots \\ \checkmark & ? & ? & ? & \dots & ? \end{pmatrix} \end{array} \begin{array}{r} \text{(never adopter)} \\ \text{(late adopter)} \\ \\ \text{(medium adopter)} \\ \\ \text{(early adopter)} \end{array}$$

$$\mathbf{Y} = \begin{pmatrix} \checkmark & \checkmark & \checkmark \\ \vdots & \vdots & \vdots \\ \checkmark & \checkmark & \checkmark \\ \checkmark & \checkmark & ? \\ \vdots & \vdots & \vdots \\ \checkmark & \checkmark & ? \end{pmatrix}.$$

1. Regress the last period outcome on the lagged outcomes. (among untreated)
2. Predict the missing POs using the estimated regression.

$$\forall (i,T) \in \mathcal{M},\ \hat{Y}_{iT} = \hat{\beta}_0 + \sum_{t=1}^{T-1} \hat{\beta}_t Y_{it}, \text{ where } \hat{\beta} = \arg\min_{\beta} \sum_{i:(i,T)\in\mathcal{O}} (Y_{iT} - \beta_0 - \sum_{t=1}^{T-1} \beta_t Y_{it})^2.$$

$\rightarrow$ Nonparametrically,

Vertical regression and synthesis control : fat matrix ($T \gg N$)

$$\mathbf{Y} = \begin{pmatrix} \checkmark & \checkmark & \dots & \checkmark & \checkmark & \dots & \checkmark \\ \checkmark & \checkmark & \dots & \checkmark & \checkmark & \dots & \checkmark \\ \checkmark & \checkmark & \dots & \checkmark & ? & \dots & ? \end{pmatrix}$$

1. Regress the outcomes for treated unit prior to the treatment on the outcomes for the control units in the same periods.
2. Predct the missing POs using the estimated regression.

$$\forall (N,t) \in \mathcal{M},\ \hat{Y}_{Nt} = \hat{\gamma}_0 + \sum_{i=1}^{N-1} \hat{\gamma}_i Y_{it}, \text{ where } \hat{\gamma} = \arg\min_{\gamma} \sum_{t:(N,t)\in\mathcal{O}} (Y_{Nt} - \gamma_0 - \sum_{i=1}^{N-1} \gamma_i Y_{it})^2.$$

$\rightarrow$ Vertical regression is generalization of ADH(2010) in that it relaxes two restrictions :

- the coefficients $\hat{\gamma}$ are nonnegative. (Interpretability ; What is a negative weight?)
- the intercept in this regression is 0. (This is seen to be plausible in recent literatures.)

Matrix Completion

Model

- Under no covariates, we model the $N \times T$ matrix of complete matrix $\mathbf{Y}$ as

$$\mathbf{Y} = \mathbf{L}^* + \epsilon, \text{ where } \mathbb{E}[\epsilon|\mathbf{L}^*] = 0.$$

Assumption 1

- ϵ is independent of $\mathbf{L}^*$
- The element of ϵ are $\sigma - sub - Gaussian$ and independent each other.
  $\Leftrightarrow \forall t,\ \mathbb{E}[\exp(t\epsilon)] \leq \exp(\frac{\sigma^2 t^2}{2})$.

- The goal is to estimate the matrix $\mathbf{L}^*$. (low-rank assumption)
  $\rightarrow$ Note that two types[2] of fixed effects are included.

[2] Interactive fixed effect も低ランクとしていいが, rank が明示的でないため具体的には考えない.

Theoritical Bounds
for the Estimation Error

Two illustrations

References

References

- Susan Athey, Mohsen Bayati, Nikolay Doudchenko, Guido Imbens, and Khashayar Khosravi (2021), *Matrix Completion Methods for Causal Panel Data Models*, Journal of the American Statistical Association.
- Licheng Liu, Ye Wang, Yiqing Xu (2024), *A Practical Guide to Counterfactual Estimators for Causal Inference with Time-Series Cross-Sectional Data*, American Journal of Political Science.
- Martin J. Wainwright (2019), *High-Dimensional Statistics*, Cambridge University Press. (mainly Chapter 10)
- 冨岡亮太 (2015), スパース性に基づく機械学習, 講談社.
- 川口康平, 澤田真行 (2024), 因果推論の計量経済学, 日本評論社.