Matrix Completion (Athey et al. 2021)

Matrix Completion

To estimate average causal effects in panel data with units exposed to treatment intermittently, two literatures are pivotal:

- **Unconfoundedness** [@imbens2015causal]: Imputes missing potential control outcomes for treated units using observed outcomes from similar control units in previous periods.
- **Synthetic Control** [@abadie2010synthetic]: Imputes missing control outcomes for treated units using weighted averages from control units, matching lagged outcomes between treated and control units.

Both exploit missing potential outcomes under different assumptions:

- Unconfoundedness assumes time patterns are stable across units.
- Synthetic control assumes unit patterns are stable over time.

Overview

Once regularization is applied, both approaches are applicable in similar settings

Matrix Completion method, nesting both, is based on matrix factorization, focusing on imputing missing matrix elements assuming:

1. Complete matrix = low-rank matrix + noise.
2. Missingness is completely at random.

It's distinguished by not imposing factorization restrictions but utilizing regularization to define the estimator, particularly effective with the nuclear norm as a regularizer for complex missing patterns [@athey2021matrix].

Contributions

1. Recognizing structured missing patterns allowing time correlation, enabling staggered adoption.
2. Modifying estimators for unregularized unit and time fixed effects.
3. Performing well across various T and N sizes, unlike unconfoundedness and synthetic control, which falter when $T >> N$ or $N >> T$, respectively.

Identifying Assumptions

1. SUTVA: Potential outcomes indexed only by the unit's contemporaneous treatment.
2. No dynamic effects (it's okay under staggered adoption, it gives a different interpretation of estimand).

Setup

- $Y_{it}(0)$ and $Y_{it}(1)$ represent potential outcomes of Y_{it}.
- W_{it} is a binary treatment indicator.

Aim to estimate the average effect for the treated:

$$\tau = \frac{\sum_{(i,t):W_{it}=1} [Y_{it}(1) - Y_{it}(0)]}{\sum_{i,t} W_{it}}$$

Setup

We observe all relevant values for $Y_{it}(1)$

We want to impute missing entries in the $Y(0)$ matrix for treated units with $W_{it} = 1$.

Define $\mathcal{M}$ as the set of pairs of indices (i, t), where $i \in N$ and $t \in T$, corresponding to missing entries with $W_{it} = 1$;

Define $\mathcal{O}$ as the set of pairs of indices corresponding to observed entries in $Y(0)$ with $W_{it} = 0$.

Data is conceptualized as two $N \times T$ matrices,

Setup

one incomplete

$$Y = \begin{pmatrix} Y_{11} & Y_{12} & ? & \cdots & Y_{1T} \\ ? & ? & Y_{23} & \cdots & ? \\ Y_{31} & ? & Y_{33} & \cdots & ? \\ \vdots & \vdots & \vdots & \ddots & \vdots \\ Y_{N1} & ? & Y_{N3} & \cdots & ? \end{pmatrix},$$

Setup

and one complete:

$$W = \begin{pmatrix} 0 & 0 & 1 & \cdots & 0 \\ 1 & 1 & 0 & \cdots & 1 \\ 0 & 1 & 0 & \cdots & 1 \\ \vdots & \vdots & \vdots & \ddots & \vdots \\ 0 & 1 & 0 & \cdots & 1 \end{pmatrix},$$

where

$$W_{it} = \begin{cases} 1 & \text{if } (i,t) \in \mathcal{M}, \\ 0 & \text{if } (i,t) \in \mathcal{O}, \end{cases}$$

is an indicator for the event that the corresponding component of Y, that is Y_{it}, is missing.

Setup

Patterns of missing data in $\mathbf{Y}$:

- Block (treatment) structure with 2 special cases
  - Single-treated-period block structure [@imbens2015causal]
  - Single-treated-unit block structure [@abadie2010synthetic]
- Staggered Adoption

Shape of matrix $\mathbf{Y}$:

- Thin ($N >> T$)
- Fat ($T >> N$)
- Square ($N \approx T$)

Combinations

Combinations of patterns of missingness and shape create different literatures:

- Horizontal Regression = Thin matrix + single-treated-period block (focusing on cross-section correlation patterns)
- Vertical Regression = Fat matrix + single-treated-unit block (focusing on time-series correlation patterns)
- TWFE = Square matrix

To combine, we can exploit both stable patterns over time, and across units (e.g., TWFE, interactive FEs or matrix completion).

Alternatively

For the same factor model

$$\mathbf{Y} = \mathbf{U}\mathbf{V}^T +$$

where $\mathbf{U}$ is $N \times R$ and $\mathbf{V}$ is $T \times R$

The interactive FE literature focuses on a fixed number of factors R in $\mathbf{U}, \mathbf{V}$, while matrix completion focuses on impute $\mathbf{Y}$ using some forms regularization (e.g., nuclear norm).

- We can also estimate the number of factors R [@bai2002determining, @moon2015linear]

To use the nuclear norm minimization estimator, we must add a penalty term to regularize the objective function. However, before doing so, we need to explicitly estimate the time (λ_t) and unit (μ_i) fixed effects implicitly embedded in the missing data matrix to reduce the bias of the regularization term.

Specifically,

$$Y_{it} = L_{it} + \sum_{p=1}^{P} \sum_{q=1}^{Q} X_{ip} H_{pq} Z_{qt} + \mu_i + \lambda_t + V_{it}\beta + \epsilon_{it}$$

where

- X_{ip} is a matrix of p variables for unit i
- Z_{qt} is a matrix of q variables for time t
- V_{it} is a matrix of time-varying variables.

Lasso-type l_1 norm ($||H|| = \sum_{p=1}^{p} \sum_{q=1}^{Q} |H_{pq}|$) is used shrink $H \to 0$

Regularize L

There are several options to regularize L:

1. Frobenius (i.e., Ridge): not informative since it imputes missing values as 0.
2. Nuclear Norm (i.e., Lasso): computationally feasible (using SOFT-IMPUTE algorithm [@Mazumder2010SpectralRA]).
3. Rank (i.e., Subset selection): not computationally feasible

Extension

This methods allows to

- use more covariates
- leverage data from treated units (can be used when treatment effect is constant and pattern of missing is not complex).
- have autocorrelated errors
- have weighted loss function (i.e., take into account the probability of outcomes for a unit being missing)