Matrix Completion Methods and Causal Panel Models

American Economic Association,

Continuing Education Program 2018

Machine Learning and Econometrics, Lecture 7-8

Guido Imbens

Motivating Example I

• California's anti-smoking legislation (Proposition 99) took effect in 1989.

• **What is the causal effect of the legislation on smoking rates in California in 1989?**

• We **observe** smoking rates in California in 1989 given the legislation. We need to **impute** the **counterfactual** smoking rates in California in 1989 had the legislation not been enacted.

• We have data in the absence of smoking legislation in California prior to 1989, and for other states both before and after 1989. (and other variables, but not of essence)

Motivating Example II

In US, on any day, 1 in 25 patients suffers at least one Hospital Acquired Infection (HAI).

• Hospital acquired infections cause 75,000 deaths per year, cost 35 billion dollars per year

• 13 states have adopted a reporting policy (at different times during 2000-2010) that rquires hospitals to report HAIs to the state.

What is (average) causal effect of reporting policy on deaths or costs?

Set Up: we observe (in addition to covariates):

$$\mathbf{Y} = \begin{pmatrix} Y_{11} & Y_{12} & Y_{13} & \dots & Y_{1T} \\ Y_{21} & Y_{22} & Y_{23} & \dots & Y_{2T} \\ Y_{31} & Y_{32} & Y_{33} & \dots & Y_{3T} \\ \vdots & \vdots & \vdots & \ddots & \vdots \\ Y_{N1} & Y_{N2} & Y_{N3} & \dots & Y_{NT} \end{pmatrix} \quad \text{(realized outcome).}$$

$$\mathbf{W} = \begin{pmatrix} 1 & 1 & 0 & \dots & 1 \\ 0 & 0 & 1 & \dots & 0 \\ 1 & 0 & 1 & \dots & 0 \\ \vdots & \vdots & \vdots & \ddots & \vdots \\ 1 & 0 & 1 & \dots & 0 \end{pmatrix} \quad \text{(binary treatment).}$$

• rows of $\mathbf{Y}$ and $\mathbf{W}$ correspond to physical units, columns correspond to time periods.

In terms of potential outcome matrices $\mathbf{Y}(0)$ and $\mathbf{Y}(1)$:

$$\mathbf{Y}(0) = \begin{pmatrix} ? & ? & \checkmark & \dots & ? \\ \checkmark & \checkmark & ? & \dots & \checkmark \\ ? & \checkmark & ? & \dots & \checkmark \\ \vdots & \vdots & \vdots & \ddots & \vdots \\ ? & \checkmark & ? & \dots & \checkmark \end{pmatrix} \qquad \mathbf{Y}(1) = \begin{pmatrix} \checkmark & \checkmark & ? & \dots & \checkmark \\ ? & ? & \checkmark & \dots & ? \\ \checkmark & ? & \checkmark & \dots & ? \\ \vdots & \vdots & \vdots & \ddots & \vdots \\ \checkmark & ? & \checkmark & \dots & ? \end{pmatrix}.$$

$Y_{it}(0)$ observed iff $W_{it} = 0$, $Y_{it}(1)$ observed iff $W_{it} = 1$.

In order to estimate the average treatment effect for the treated, (or other average, e.g., overall average effect)

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

We need to **impute** the missing potential outcomes in $\mathbf{Y}(0)$ (and in $\mathbf{Y}(1)$ for other estimands).

Focus on problem of imputing missing in $N \times T$ matrix $\mathbf{Y} = \mathbf{Y}(0)$

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

$\mathcal{O}$ and $\mathcal{M}$ are sets of indices (i,t) with $Y_{i,t}$ observed and missing, with cardinalities $|\mathcal{O}|$ and $|\mathcal{M}|$. Covariates, time-specific, unit-specific, time/unit-specific.

- Now the problem is a **Matrix Completion Problem**.

At times we focus on the special case with $W_{it} = 1$ iff $(i,t) = (N,T)$, so that $|\mathcal{O}| = N \times T - 1$ and $|\mathcal{M}| = 1$:

$$\mathbf{Y} = \begin{pmatrix} Y_{11} & Y_{12} & Y_{13} & \dots & Y_{1T} \\ Y_{21} & Y_{22} & Y_{23} & \dots & Y_{2T} \\ Y_{31} & Y_{32} & Y_{33} & \dots & Y_{3T} \\ \vdots & \vdots & \vdots & \ddots & \vdots \\ Y_{N1} & Y_{N2} & Y_{N3} & \dots & ? \end{pmatrix} \quad \text{(realized outcome).}$$

$$\mathbf{W} = \begin{pmatrix} 1 & 1 & 1 & \dots & 1 \\ 1 & 1 & 1 & \dots & 1 \\ 1 & 1 & 1 & \dots & 1 \\ \vdots & \vdots & \vdots & \ddots & \vdots \\ 1 & 1 & 1 & \dots & 0 \end{pmatrix} \quad \text{(binary treatment).}$$

Here single entry Y_{NT} needs to be imputed.

But also important **Staggered Adoption** (e.g., adoption of technology, Athey and Stern, 1998)

– Here each unit is characterized by an adoption date $T_i \in \{1, \ldots, T, \infty\}$ that is the first date they are exposed to the treatment.

– Once exposed a unit will subsequently be exposed, $W_{it+1} \geq W_{it}$.

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

Netflix Problem

- $N \approx 500,000$ (individuals), raises computational issues
- Large $T \approx 20,000$ (movies),
- General missing data pattern,
- Fraction of observed data is close to zero, $|\mathcal{O}| \ll N \times T$

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

How is this problem treated in the various literatures?

1. Program evaluation literature under unconfoundedness / selection-on-observables (Imbens-Rubin 2015)

2. Synthetic Control literature (Abadie-Diamond-Hainmueller, JASA 2010)

3. Differences-In-Differences Literature (Bertrand-Dulfo-Mullainath 2003) and Econometric Panel Data literature (Bai 2003, 2009; Bai and Ng 2002)

4. Machine Learning literature on matrix completion / net-flix problem (Candés-Recht, 2009; Keshavan, Montanari & Oh, 2010).

1. Program Evaluation Literature under unconfoundedness / selection-on-observables (Imbens-Rubin 2015)

Focuses on special case:

– **Thin** Matrix (N large, T small),

– Y_{iT} is missing for some i (N_t "treated units"), and no missing entries for other units ($N_c = N - N_t$ "control units").

– Y_{it} is always observed for $t < T$.

$$\mathbf{Y}_{N\times T} = \begin{pmatrix} \checkmark & \checkmark & \checkmark \\ \checkmark & \checkmark & \checkmark \\ \checkmark & \checkmark & ? \\ \checkmark & \checkmark & ? \\ \checkmark & \checkmark & \checkmark \\ \vdots & \vdots & \vdots \\ \checkmark & \checkmark & ? \end{pmatrix} \qquad \mathbf{W}_{N\times T} = \begin{pmatrix} 1 & 1 & 1 \\ 1 & 1 & 1 \\ 1 & 1 & 0 \\ 1 & 1 & 0 \\ 1 & 1 & 1 \\ \vdots & \vdots & \vdots \\ 1 & 1 & 0 \end{pmatrix}$$

Parametric version: **Horizontal** Regression

Specification:

$$Y_{iT} = \beta_0 + \sum_{t=1}^{T-1} \beta_t Y_{it} + \varepsilon_i$$

estimated on N_c control units, with T regressors.

Prediction for N_t treated units:

$$\widehat{Y}_{iT} = \widehat{\beta}_0 + \sum_{t=1}^{T-1} \widehat{\beta}_t Y_{Nt}$$

• Nonparametric version: for each treated unit i find a control unit j with $Y_{it} \approx Y_{jt}$, $t < T$.

• If N large relative to T_0 use regularized regression (lasso, ridge, elastic net)

2. Synthetic Control Literature (Abadie-Diamond-Hainmueller, JASA 2010)

Focuses on special case:

– **Fat** Matrix (N small, T large)

– Y_{Nt} is missing for $t \geq T_0$ and no missing entries for other units.

$$\mathbf{Y}_{N\times T} = \begin{pmatrix} \checkmark & \checkmark & \checkmark & \dots & \checkmark \\ \checkmark & \checkmark & \checkmark & \dots & \checkmark \\ \checkmark & \checkmark & ? & \dots & ? \end{pmatrix} \quad \mathbf{W}_{N\times T} = \begin{pmatrix} 1 & 1 & 1 & \dots & 1 \\ 1 & 1 & 1 & \dots & 1 \\ 1 & 1 & 0 & \dots & 0 \end{pmatrix}$$

Parametric version: **Vertical** Regression:

$$Y_{Nt} = \alpha_0 + \sum_{i=1}^{N-1} \alpha_i Y_{it} + \eta_t$$

estimated on T_0 pre-treatment periods, with N regressors (in ADH with restrictions $\alpha_0 = 0$, $\alpha_i \geq 0$).

Prediction for $T - T_0$ treated periods:

$$\widehat{Y}_{Nt} = \widehat{\alpha}_0 + \sum_{i=1}^{N-1} \widehat{\alpha}_i Y_{it}$$

• Nonparametric version: for each post-treatment period t find a pre-treatment period s with $Y_{is} \approx Y_{it}$, $i = 1, \ldots, N-1$.

• If T large relative to N_c use regularized regression (lasso, ridge, elastic net), following Doudchenko and Imbens (2016).

3. Differences-In-Differences Literature and Econometric Panel Data literature (Bertrand-Dulfo-Mullainathan, 2003; Bai 2003, 2009; Bai-Ng 2002)

Note: we model only $\mathbf{Y} = \mathbf{Y}(0)$ here, because $Y_{it}(1)$ is observed for the relevant units/time-periods.

Models developed here for complete-data matrices:
DID:

$$Y_{it} = \delta_i + \gamma_t + \varepsilon_{it}$$

Generalized Fixed Effects (Bai 2003, 2009)

$$Y_{it} = \sum_{r=1}^{R} \delta_{ir}\gamma_{rt} + \varepsilon_{it}$$

Estimate δ and γ by least squares and use to impute missing values.

4. Machine Learning literature on matrix completion / netflix problem (Candés-Recht, 2009; Keshavan, Montanari & Oh, 2010)

– Focus on setting with many missing entries: $|\mathcal{O}|/|\mathcal{M}| \approx 0$. E.g., netflix problem, with units corresponding to individuals, and time periods corresponding to movie titles, or image recovery from limited information, with i and t corresponding to different dimensions.

– Focus on computational feasibility, with both N and T large.

– Focus on randomly missing entries: $\mathbf{W} \perp\!\!\!\perp \mathbf{Y}$.

Like interactive fixed effect, focus on low-rank structure underlying observations.

Set Up

General Case: $N >> T$, $N << T$, $N \approx T$, possibly stable patterns over time, possibly stable patterns within units.

Set up without covariates, connecting to the interactive fixed effect literature (Bai, 2003), and the matrix completion literature (Cands, and Recht, 2009):

$$\mathbf{Y}_{N\times T} = \mathbf{L}_{N\times T} + \varepsilon_{N\times T}$$

- Key assumption:

$$\mathbf{W}_{N\times T} \perp\!\!\!\perp \varepsilon_{N\times T} \quad \text{(but } \mathbf{W} \text{ may depend on } \mathbf{L})$$

- Staggered entry, $W_{it+1} \geq W_{it}$

- $\mathbf{L}$ has low rank relative to N and T.

More general case, with unit-specific P-component covariate X_i, time-specific Q-component covariate Z_t, and unit-time-specific covariate V_{it}:

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

• We do not necessarily need the fixed effects γ_i and δ_t, these can be subsumed into $\mathbf{L}$. It is convenient to include the fixed effects given that we regularize $\mathbf{L}$.

Too many parameters (especially $N \times T$ matrix $\mathbf{L}$), so we need regularization:

We shrink $\mathbf{L}$ and $\mathbf{H}$ towards zero.

For $\mathbf{H}$ we use Lasso-type element-wise ℓ_1 norm: defined as $\|\mathbf{H}\|_{1,e} = \sum_{p=1}^{P} \sum_{q=1}^{Q} |H_{pq}|$.

How do we regularize $\mathbf{L}$?

$$\mathbf{L}_{N\times T} = \mathbf{S}_{N\times N}\mathbf{\Sigma}_{N\times T}\mathbf{R}_{T\times T}$$

$\mathbf{S}$, $\mathbf{R}$ unitary, $\mathbf{\Sigma}$ is rectangular diagonal with entries $\sigma_i(\mathbf{L})$ that are the **singular values**. Rank of $\mathbf{L}$ is number of non-zero $\sigma_i(\mathbf{L})$.

$$\|\mathbf{L}\|_F^2 = \sum_{i,t} |L_{it}|^2 = \sum_{j=1}^{\min(N,T)} \sigma_i^2(\mathbf{L}) \quad \text{(Frobenius, like ridge)}$$

$$\Longrightarrow \quad \|\mathbf{L}\|_* = \sum_{j=1}^{\min(N,T)} \sigma_i(\mathbf{L}) \quad \textbf{(nuclear norm, like LASSO)}$$

$$\|\mathbf{L}\|_R = \sum_{j=1}^{\min(N,T)} \mathbf{1}_{\sigma_i(\mathbf{L})>0} \quad \text{(Rank, like subset selection)}$$

Frobenius norm imputes missing values as 0.

Rank norm is computationally not feasible for general missing data patterns.

The preferred **Nuclear** norm leads to low-rank matrix but is computationally feasible.

Our proposed method: regularize using using nuclear norm:

$$\min_{\mathbf{L}} \frac{1}{|\mathcal{O}|} \sum_{(i,t)\in\mathcal{O}} (Y_{it} - L_{it})^2 + \lambda_L \|\mathbf{L}\|_*$$

For the general case we estimate $\mathbf{H}$, $\mathbf{L}$, δ, γ, and β as

$$\min_{\mathbf{H},\mathbf{L},\delta,\gamma} \frac{1}{|\mathcal{O}|} \sum_{(i,t)\in\mathcal{O}} \left(Y_{it} - L_{it} - \sum_{p=1}^{P} \sum_{q=1}^{Q} X_{ip} H_{pq} Z_{qt} - \gamma_i - \delta_t - V_{it}\beta \right)^2$$

$$+\lambda_L \|\mathbf{L}\|_* + \lambda_H \|\mathbf{H}\|_{1,e}$$

We choose λ_L and λ_H through cross-validation.

Algorithm (Mazumder, Hastie, & Tibshirani 2010)

Given any $N \times T$ matrix $\mathbf{A}$, define the two $N \times T$ matrices $\mathbf{P}_{\mathcal{O}}(\mathbf{A})$ and $\mathbf{P}^{\perp}_{\mathcal{O}}(\mathbf{A})$ with typical elements:

$$\mathbf{P}_{\mathcal{O}}(\mathbf{A})_{it} = \begin{cases} A_{it} & \text{if } (i,t) \in \mathcal{O}\,, \\ 0 & \text{if } (i,t) \notin \mathcal{O}\,, \end{cases}$$

and

$$\mathbf{P}^{\perp}_{\mathcal{O}}(\mathbf{A})_{it} = \begin{cases} 0 & \text{if } (i,t) \in \mathcal{O}\,, \\ A_{it} & \text{if } (i,t) \notin \mathcal{O}\,. \end{cases}$$

Let $\mathbf{A} = \mathbf{S}\mathbf{\Sigma}\mathbf{R}^\top$ be the Singular Value Decomposition for $\mathbf{A}$, with $\sigma_1(\mathbf{A}), \ldots, \sigma_{\min(N,T)}(\mathbf{A})$, denoting the singular values.

Then define the matrix shrinkage operator

$$\text{shrink}_\lambda(\mathbf{A}) = \mathbf{S}\tilde{\mathbf{\Sigma}}\mathbf{R}^\top,$$

where $\tilde{\mathbf{\Sigma}}$ is equal to $\mathbf{\Sigma}$ with the i-th singular value $\sigma_i(\mathbf{A})$ replaced by $\max(\sigma_i(\mathbf{A}) - \lambda, 0)$.

Given these definitions, the algorithm proceeds as follows.

– Start with the initial choice $\mathbf{L}_1(\lambda) = \mathbf{P}_{\mathcal{O}}(\mathbf{Y})$, with zeros for the missing values.

– Then for $k = 1, 2, \ldots,$ define,

$$\mathbf{L}_{k+1}(\lambda) = \text{shrink}_{\lambda}\Big\{\mathbf{P}_{\mathcal{O}}(\mathbf{Y}) + \mathbf{P}_{\mathcal{O}}^{\perp}(\mathbf{L}_k(\lambda))\Big\},$$

until the sequence $\{\mathbf{L}_k(\lambda)\}_{k\geq 1}$ converges.

– The limiting matrix $\mathbf{L}^*$ is our estimator for the regularization parameter λ, denoted by $\widehat{\mathbf{L}}(\lambda, \mathcal{O})$.

Results I (for case without covariates, and just $\mathbf{L}$, $\mathcal{O}$ is sufficiently random, and $\varepsilon_{it} = Y_{it} - L_{it}$ are iid with variance σ^2.

$\|\mathbf{Y}\|_F = \sqrt{\sum_{i,t} Y_{i,t}^2}$ is Frobenius norm. $\|\mathbf{Y}\|_\infty = \max_{i,t} |Y_{i,t}|$.

Let $\mathbf{Y}^*$ be the matrix including all the missing values (e.g., $\mathbf{Y}(1)$.

The estimated matrix $\hat{\mathbf{L}}$ is close to $\mathbf{L}^*$ in the following sense:

$$\frac{\left\|\hat{\mathbf{L}} - \mathbf{L}\right\|_F}{\|\mathbf{L}\|_F} \leq C \max\left(\sigma, \frac{\|\mathbf{L}\|_\infty}{\|\mathbf{L}\|_F}\right) \frac{\text{rank}(\mathbf{L})(N+T)\ln(N+T)}{|\mathcal{O}|}$$

In many cases the number of observed entries $|\mathcal{O}|$ is of order $N \times T$ so if $\text{rank}(\mathbf{L}) \ll (N+T)$ the error goes to 0 as $N+T$ grows.

Adaptive Properties of Matrix Regression

Suppose N is large, $T = 2$, $W_{i1} = 0$, many control units (treatment effect setting)

– In that case the natural imputation is $L_{i2} = Y_{i1} \times \rho$, where ρ is the within unit correlation between periods for the control units.

– The program-evaluation / horizontal-regression approach would lead to $L_{i2} = Y_{i1} \times \rho$

– The synthetic-control / vertical-regression approach would lead to $L_{i2} = 0$.

– How does the matrix completion estimator impute the missing values in this case?

Suppose for control units

$$\begin{pmatrix} Y_{i1} \\ Y_{i2} \end{pmatrix} \sim \mathcal{N}\left(\begin{pmatrix} 0 \\ 0 \end{pmatrix}, \begin{pmatrix} 1 & \rho \\ \rho & 1 \end{pmatrix} \right),$$

and suppose the pairs (Y_{i1}, Y_{i2}) are jointly independent.

Then for small λ, for the treated units, the imputed value is

$$\widehat{L}_{i2} = Y_{i1} \times \frac{\rho}{1 + \sqrt{(1 - \rho^2)}}$$

The regularization leads to a small amount of shrinkage towards zero relative to optimal imputation.

Matrix Completion method adapts well to shape of matrix and correlation structure.

Illustrations

• We take complete matrices $\mathbf{Y}_{N\times T}$,

• We pretend some entries are missing.

• We use different estimators to impute the "missing" entries and compare them to actual values, and calculate the mean-squared-error.

We compare three estimators:

1. Matrix Completion Nuclear Norm, MC-NNM

2. Horizontal Regr with Elastic Net Regularization, EN-H (Program evaluation type regression)

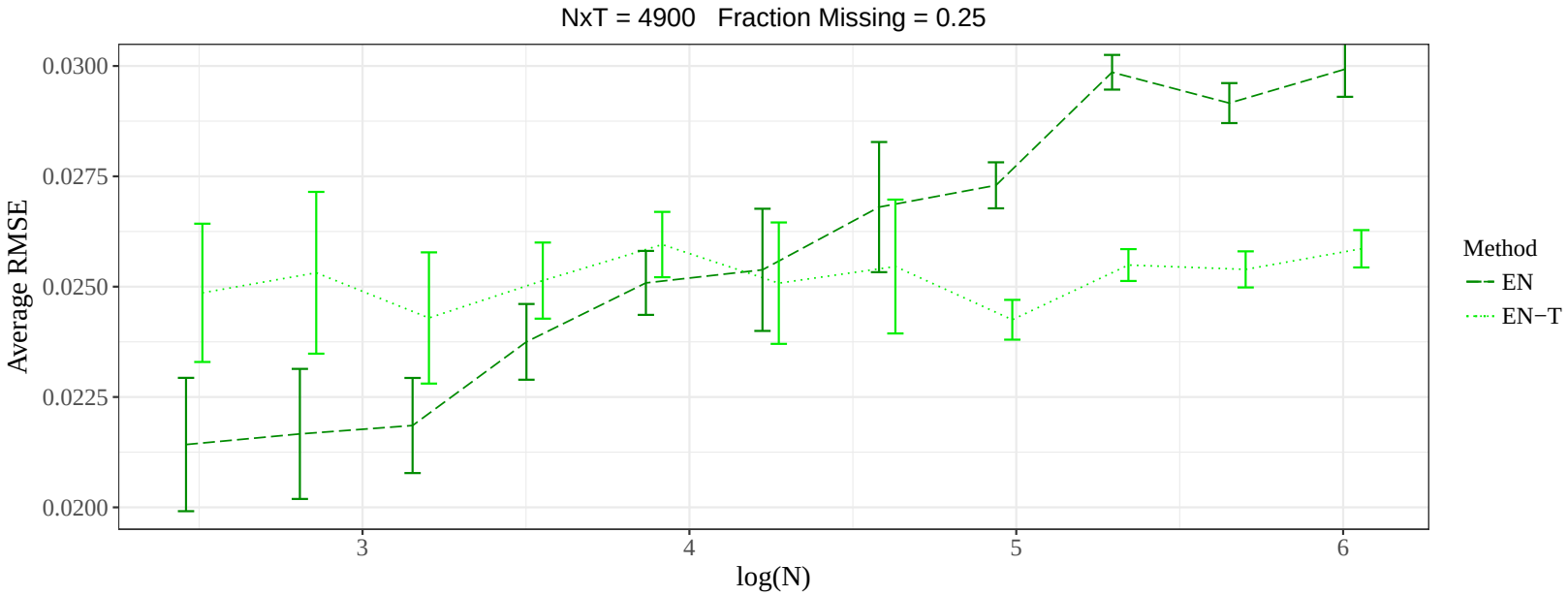
3. Vertical Regr with Elastic Net Regularization, EN-V (Synthetic Control type regression)

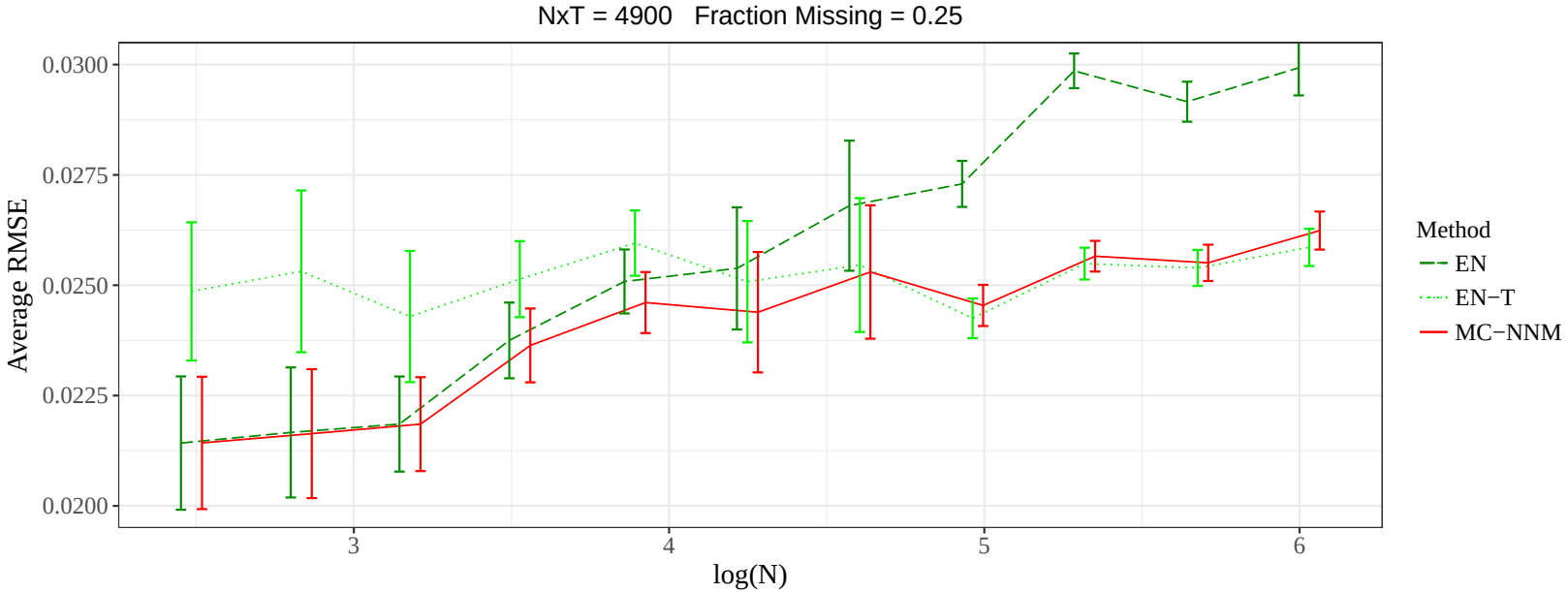
Illustrations I: Stock Market Data

We use daily returns for 2453 stocks over 10 years (3082 days). We create sub-samples by looking at the first T daily returns of N randomly sampled stocks for pairs of (N, T) such that $N \times T = 4900$, ranging from fat to thin:
$(N, T) = (10, 490), \ldots, (70, 70), \ldots, (490, 10)$.

Given the sample, we pretend that half the stocks are treated at the mid point over time, so that 25% of the entries in the matrix are missing.

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


NxT = 4900 Fraction Missing = 0.25
Average RMSE
0.0300
0.0275
0.0250
0.0225
0.0200
3
4
5
6
log(N)
Method
EN
EN−T


NxT = 4900 Fraction Missing = 0.25
Average RMSE
0.0300
0.0275
0.0250
0.0225
0.0200
3
4
5
6
log(N)
Method
EN
EN−T
MC−NNM

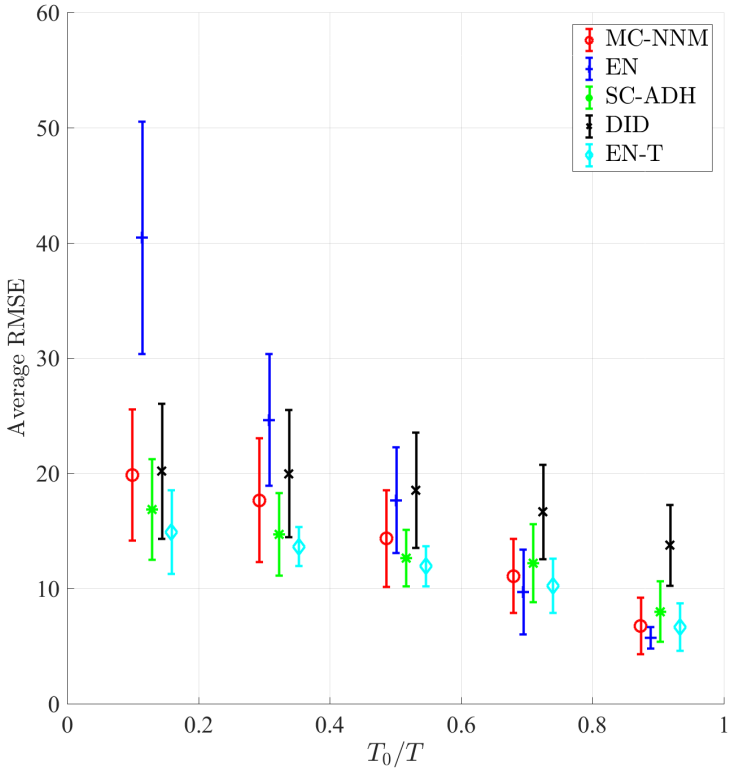
Illustrations II: California Smoking Rate Data

The outcome here is per capita smoking rates by state, for 38 states, 31 years.

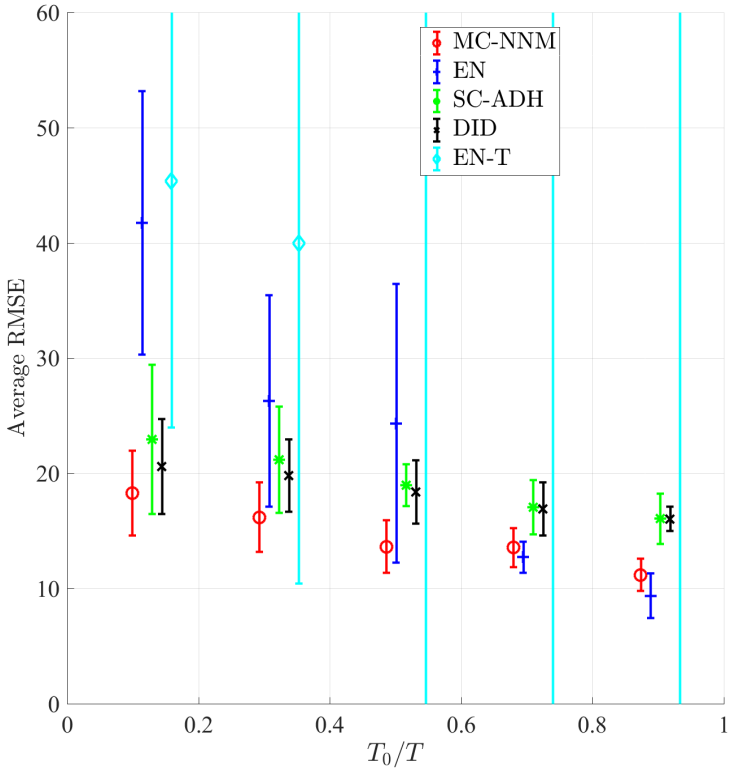
We compare both simultaneous adoption and staggered adoption.

Illustration I: California Smoking Example ($N = 38, T = 31$)

Simultaneous adoption, $N_t = 8$



Staggered adoption, $N_t = 35$



Generalizations I:

• Allow for propensity score weighting to focus on fit where it matters:

Model propensity score $E_{it} = \text{pr}(W_{it} = 1|X_i, Z_t, V_{it})$, $\mathbf{E}$ is $N \times T$ matrix with typical element E_{it}

Possibly using matrix completion:

$$\hat{\mathbf{E}} = \arg\min_{\mathbf{E}} \frac{1}{NT} \sum_{(i,t)} (W_{it} - E_{it})^2 + \lambda_L \|\mathbf{E}\|_*$$

and then

$$\min_{\mathbf{L}} \frac{1}{|\mathcal{O}|} \sum_{(i,t)\in\mathcal{O}} \frac{\hat{E}_{it}}{1 - \hat{E}_{it}} (Y_{it} - L_{it})^2 + \lambda_L \|\mathbf{L}\|_*$$

Generalizations II:

- Take account of of time series correlation in $\varepsilon_{it} = Y_{it} - L_{it}$

Modify objective function from logarithm of Gaussian likelihood based on independence to have autoregressive structure.

References

Abadie, Alberto, Alexis Diamond, and Jens Hainmueller. "Synthetic control methods for comparative case studies: Estimating the effect of Californias tobacco control program." *Journal of the American statistical Association* 105.490 (2010): 493-505.

Abadie, Alberto, Alexis Diamond, and Jens Hainmueller. "Comparative politics and the synthetic control method." *American Journal of Political Science* 59.2 (2015): 495-510.

Abadie, Alberto, and Jeremy L'Hour "A Penalized Synthetic Control Estimator for Disaggregated Data"

Athey, Susan, Guido W. Imbens, and Stefan Wager. Efficient inference of average treatment effects in high dimensions via

approximate residual balancing. arXiv preprint arXiv:1604.07125v3 (2016).

Bai, Jushan. "Inferential theory for factor models of large dimensions." *Econometrica* 71.1 (2003): 135-171.

Bai, Jushan. "Panel data models with interactive fixed effects." *Econometrica* 77.4 (2009): 1229-1279.

Bai, Jushan, and Serena Ng. "Determining the number of factors in approximate factor models." *Econometrica* 70.1 (2002): 191-221.

Candes, Emmanuel J., and Yaniv Plan. "Matrix completion with noise." *Proceedings of the IEEE* 98.6 (2010): 925-936.

Cands, Emmanuel J., and Benjamin Recht. "Exact matrix completion via convex optimization." *Foundations of Computational mathematics* 9.6 (2009): 717.

Chamberlain, G., and M. Rothschild. "Arbitrage, factor structure, and mean-variance analysis on large asset markets. *Econometrica* 51 12811304, 1983.

Doudchenko, Nikolay, and Guido W. Imbens. Balancing, regression, difference-in-differences and synthetic control methods: A synthesis. No. w22791. National Bureau of Economic Research, 2016.

Gobillon, Laurent, and Thierry Magnac. "Regional policy evaluation: Interactive fixed effects and synthetic controls." *Review of Economics and Statistics* 98.3 (2016): 535-551.

Imbens, G., and D. Rubin *Causal Inference* Cambridge University Press.

Keshavan, Raghunandan H., Andrea Montanari, and Sewoong Oh. "Matrix Completion from a Few Entries." *IEEE Transactions on Information Theory*, vol. 56,no. 6, pp.2980-2998, June 2010

Keshavan, Raghunandan H., Andrea Montanari, and Sewoong Oh. "Matrix completion from noisy entries." *Journal of Machine Learning Research* 11.Jul (2010): 2057-2078.

Liang, Dawen, et al. "Modeling user exposure in recommendation." *Proceedings of the 25th International Conference on World Wide Web.* International World Wide Web Conferences Steering Committee, 2016.

Mazumder, Rahul, Trevor Hastie, and Robert Tibshirani, (2010) "Spectral regularization algorithms for learning large incomplete matrices," *Journal of machine learning research*, 11(Aug):2287–2322.

Benjamin Recht, "A Simpler Approach to Matrix Completion", *Journal of Machine Learning Research* 12:3413-3430, 2011

Xu, Yiqing. "Generalized Synthetic Control Method: Causal Inference with Interactive Fixed Effects Models." *Political Analysis* 25.1 (2017): 57-76.