Time Series Analysis via Matrix Estimation

Devavrat Shah

Anish Agarwal Muhammad J Amjad Dennis Shen

Massachusetts Institute of Technology

https://arxiv.org/pdf/1802.09064.pdf https://arxiv.org/pdf/1711.06940.pdf

Questions from Retail

Question 1:

Estimate *demand (rate)* for Umbrellas at Target store in Sunnyvale

Question 2:

Estimate *future demand* for Umbrellas at Target store in Sunnyvale

Question 3:

What would *demand* for Umbrellas at Target store in Sunnyvale be if we did (not) introduce the mobile checkout

Questions from Retail and Time Series Analysis

Question 1: demand rate estimation

estimating latent state of a time-series with missing values

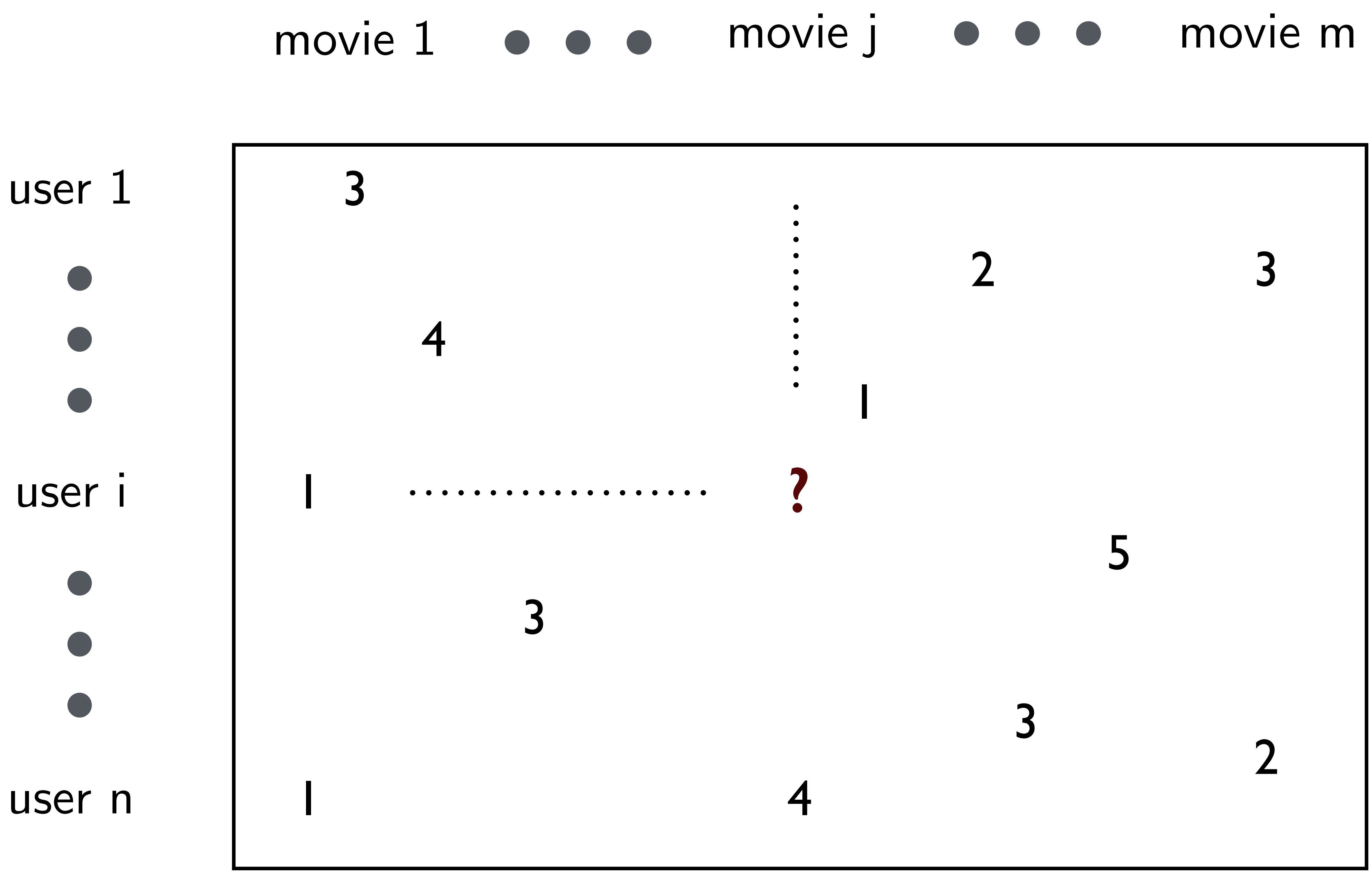
Question 2: future demand

forecasting state of time-series using historical (+ other time-series)

Question 3: demand with(out) intervention

comparing with synthetic control for time-series of interest
using other time-series

Matrix Estimation (ME)



Rating Matrix A

Matrix Estimation (ME)

Observation

ground truth: $A_{ij}, \ \forall (i,j) \in [n] \times [m]$

noisy observation for a subset E of entries: $Y_{ij}, \ \text{for } (i,j) \in E$

subject to some `noise' model: $Y_{ij} \sim A_{ij}, \ \forall \ i,j.$

Goal

produce an estimation $\hat{A}_{ij}$ for all $(i,j) \in [n] \times [m]$

so that the *prediction error*

$$\text{MSE}(\hat{A}) = \frac{1}{nm}\mathbb{E}\Big[\sum_{i,j}(\hat{A}_{ij} - A_{ij})^2\Big]$$

is *small*

Matrix Estimation (ME)

Latent Variable Model

row i has associated *latent* features $x_1(i) \in \mathcal{X}_1$

column j has associated *latent* features $x_2(j) \in \mathcal{X}_2$

entry corresponding to user i and movie j in A

$A_{ij} = f(x_1(i), x_2(j))$ where $f : \mathcal{X}_1 \times \mathcal{X}_2 \to \mathbb{R}$

That is, Y_{ij} is such that

$$\mathbb{E}\big[Y_{ij} | x_1(i), x_2(j)\big] = A_{ij} = f(x_1(i), x_2(j))$$

Canonical representation due to Row-Column Exchangeability

[Hoover 79, 82], [Aldous 81, 82, 85], [Lovasz-Szegedy 08], ...

Matrix Estimation (ME)

Goal

given (partial) observation of matrix $Y = [Y_{ij}]$

produce an estimation $\hat{A} = [\hat{A}_{ij}]$

so that the prediction error $\mathrm{MSE}(\hat{A})$ is small

Performance Metric

(random) fraction p of matrix $Y = [Y_{ij}]$ that needs to be observed

so that estimator is consistent, i.e.

$$\lim_{n,m\to\infty} \mathrm{MSE}(\hat{A}) = 0$$

Matrix Estimation (ME)

Model complexity

Feature space: $[0,1]^d$

Function space: bilinear, Lipschitz continuous

Noise model

Additive: $Y_{ij} = f(x_1(i), x_2(j)) + \eta_{ij}, \quad \mathbb{E}[\eta_{ij}] = 0$

Generic: $\mathbb{E}[Y_{ij}] = f(x_1(i), x_2(j)) \quad Y_{ij} \in [-B, B]$

Sample complexity

Number of samples observed

Matrix Estimation (ME)

[very large number of remarkable results are not reported here]

Result	Sample Complexity	Noise Model	Function Class	Guarantee
KMO10	$\Omega(nd \log n)$	Additive	bilinear(rank d)	MSE to 0
C15	$\Omega(n^{2-\frac{2}{d+2}})$	Generic	Lipschitz	MSE to 0
LLSS16	$\tilde{\Omega}(n^{3/2})$	Additive	Lipschitz	MSE to 0
BCLS17	$\omega(nd^5)$	Generic	bilinear(rank d)	MSE to 0

This Talk

Answer to all three time-series questions

- estimating latent state of a time-series with missing values
- forecasting state of time-series using historical (+ other time-series)
- comparing with synthetic control for time-series of interest

Via Matrix Estimation (ME)

- we'll assume access to Matrix Estimation (ME) as a *black-box* (BB-ME)
- transform all three questions to Matrix Estimation
  - and some post-processing

Answer 1: Time Series Imputation

"Ground Truth" of interest: $f(t), \ t \in \mathbb{R}$, for example

$$f(t) = \sum_{k=1}^{K} \alpha_k \sin(\omega_k t) + \beta_k \cos(\omega_k t)$$

$$f(t) = \sum_{k=1}^{K} \alpha_k t^{\beta_k} \quad \text{or} \quad f(t) = \sum_{k=1}^{K} \alpha_k f(t-k) \quad \text{or, their combination....}$$

Observation: $X(t), \ t \in \{0, 1, \ldots, T\}$ s. t. if observed (w.p. p)

$$\mathbb{E}[X(t)] = f(t) \text{ (+ independence, conditions on "noise")}$$

Goal: produce estimate $\hat{X}(t)$ so that

$$\mathrm{MSE}(\hat{X}, f) \text{ is small}$$

Answer 1: Time Series Imputation

An Example:

"Ground Truth" of interest: $f(t),\ t \in \mathbb{R}$ as described before

Observation: w.p. 0.1, observe X(t) where for some

$$X(t) \ \sim \ \text{Poisson}(f(t))$$

Goal: produce estimate $\hat{X}(t)$ so that

$$\text{MSE}(\hat{X}, f) \text{ is small}$$

Answer 1: Time Series Imputation

An Example:

"Ground Truth" of interest: $f(t), \ t \in \mathbb{R}$ as described before

Observation: w.p. 0.1, observe X(t) where for some $C \geq 1$

$$X(t) \sim \min\Big(C, \mathrm{Poisson}(f(t))\Big)$$

$$g(t) = \mathbb{E}\big[\min(C, \mathrm{Poisson}(f(t))\big]$$

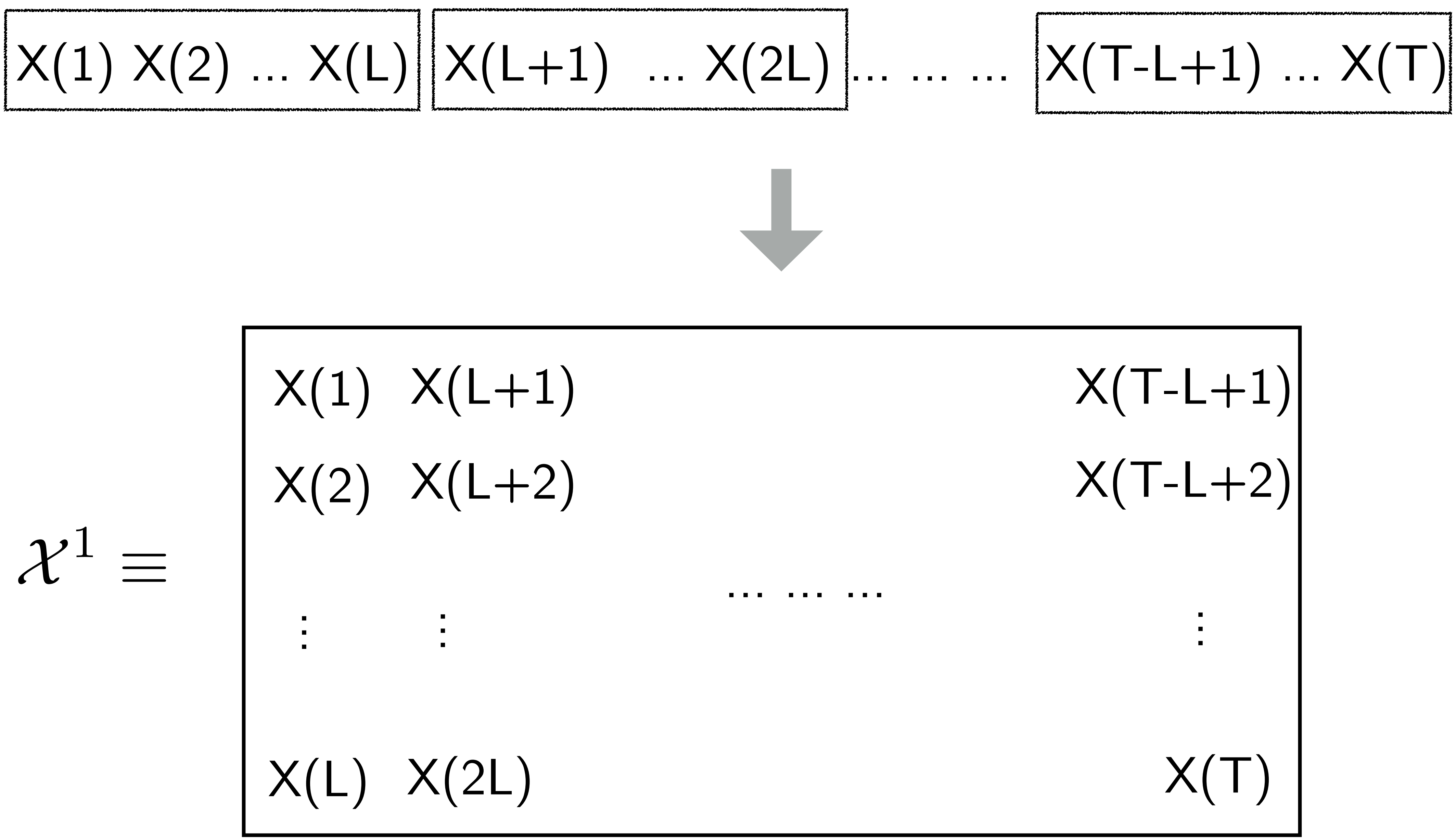
Goal: produce estimate $\hat{X}(t)$ so that

$$\mathrm{MSE}(\hat{X}, g) \text{ is small}$$

Answer 1: Time Series Imputation

Algorithm:

Transform to Matrix, Do Matrix Estimation, Undo Transformation



Answer 1: Time Series Imputation

Theorem (Informal):

The matrix satisfies Latent Variable Model with Lipschitz function.

For L *large enough* (depending upon model params) with $L^2 \ll T$

the resulting estimator is *consistent* as long as the fraction of observed

data, *p*, is *large enough* (+ good Matrix Estimation Black-Box).

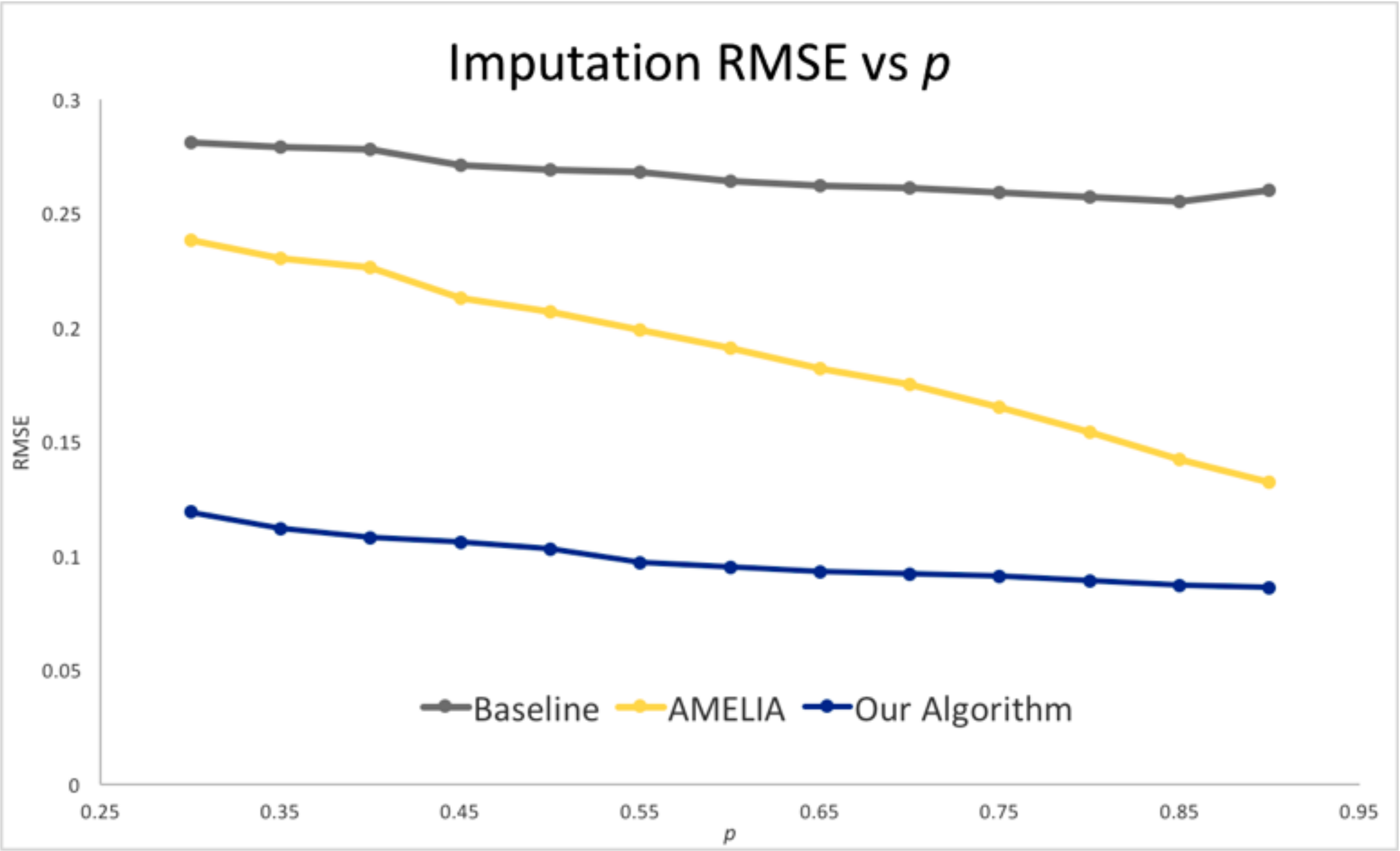
For example:

Sum of harmonics with period in {1,...,n}, $T = \omega(n^2)$ is sufficient

Contrast with $T = \omega(n)$ in the best case for additive noise

"Quadratic" loss may be min'l cost of "university" w.r.t. model/noise (?!)

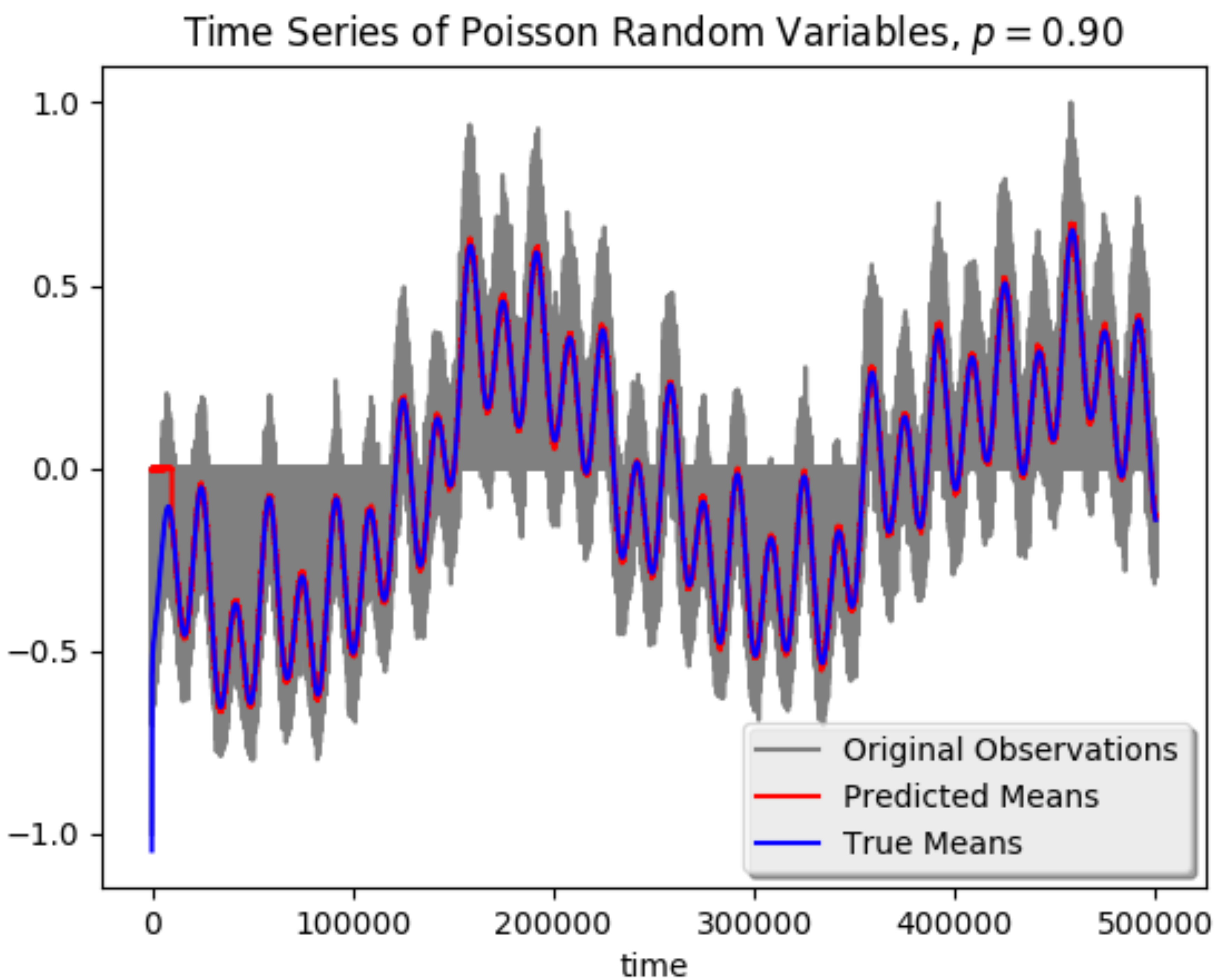
Answer 1: Time Series Imputation



mixture of periodic, trend and auto-regressive

with additive zero-mean noise and randomly missing values

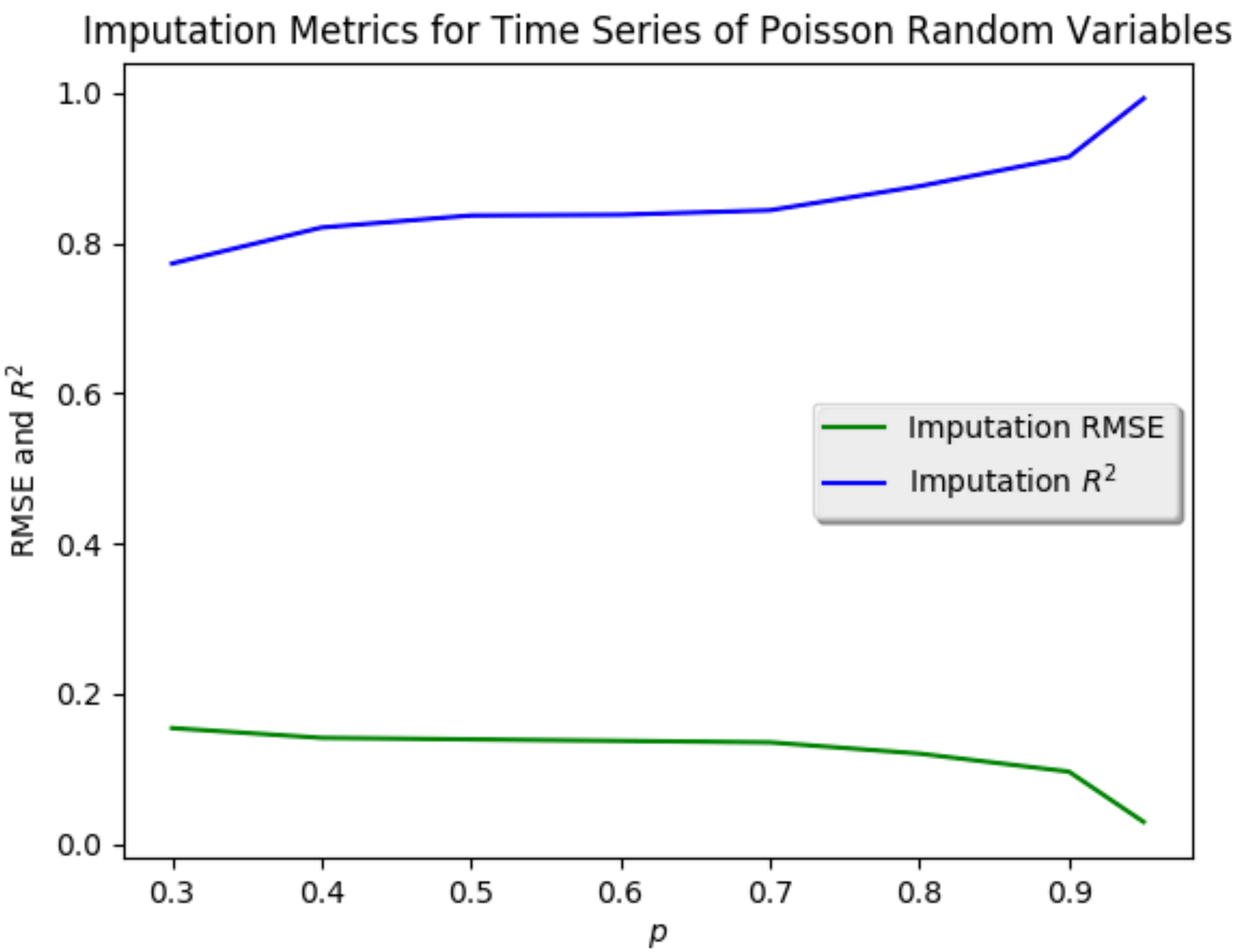
Answer 1: Time Series Imputation



mixture of periodic, trend and auto-regressive

with Poisson "noise" and randomly missing values

Answer 1: Time Series Imputation



mixture of periodic, trend and auto-regressive

with Poisson "noise" and randomly missing values

Answer 2: Time Series Forecasting

"Ground Truth" of interest: $f(t),\ t \in \mathbb{R}$, for example

$$f(t) = \sum_{k=1}^{K} \alpha_k \sin(\omega_k t) + \beta_k \cos(\omega_k t)$$

$$f(t) = \sum_{k=1}^{K} \alpha_k t^{\beta_k} \quad \text{or} \quad f(t) = \sum_{k=1}^{K} \alpha_k f(t-k) \quad \text{or, their combination....}$$

Observation: $X(t),\ t \in \{0, 1, \ldots, T\}$ s. t. if observed (w.p. *p*)

$$\mathbb{E}[X(t)] = f(t) \text{ (+ independence, conditions on "noise")}$$

Goal: produce estimate $\hat{X}(T+1)$ so that

$$\mathbb{E}\left[\left(\hat{X}(T+1) - X(T+1)\right)^2\right] \text{ is small}$$

Answer 2: Time Series Forecasting

Algorithm:

Transform to Matrix, Do Matrix Estimation, Regression, Prediction

X(1) X(2) ... X(L)	X(L+1) ... X(2L)	... ...

↓

$$\mathcal{X}^1 \equiv \begin{bmatrix} X(1) & X(L+1) & & \\ X(2) & X(L+2) & & \\ \vdots & \vdots & \cdots\ \cdots\ \cdots & \\ X(L) & X(2L) & & \end{bmatrix}$$

Answer 2: Time Series Forecasting

Algorithm:

Transform to Matrix, Do Matrix Estimation, Regression, Prediction

| X(k) ... X(L+k-1) | X(L+k)...X(2L+k-1) | ... ... |

$$\mathcal{X}^k \equiv \begin{bmatrix} X(k) & X(L+k) & \\ X(k+1) & X(L+k+1) & \\ \vdots & \vdots & \cdots\ \cdots\ \cdots \\ X(L+k-1) & X(2L+k-1) & \end{bmatrix} \qquad 1 \le k \le L$$

Answer 2: Time Series Forecasting

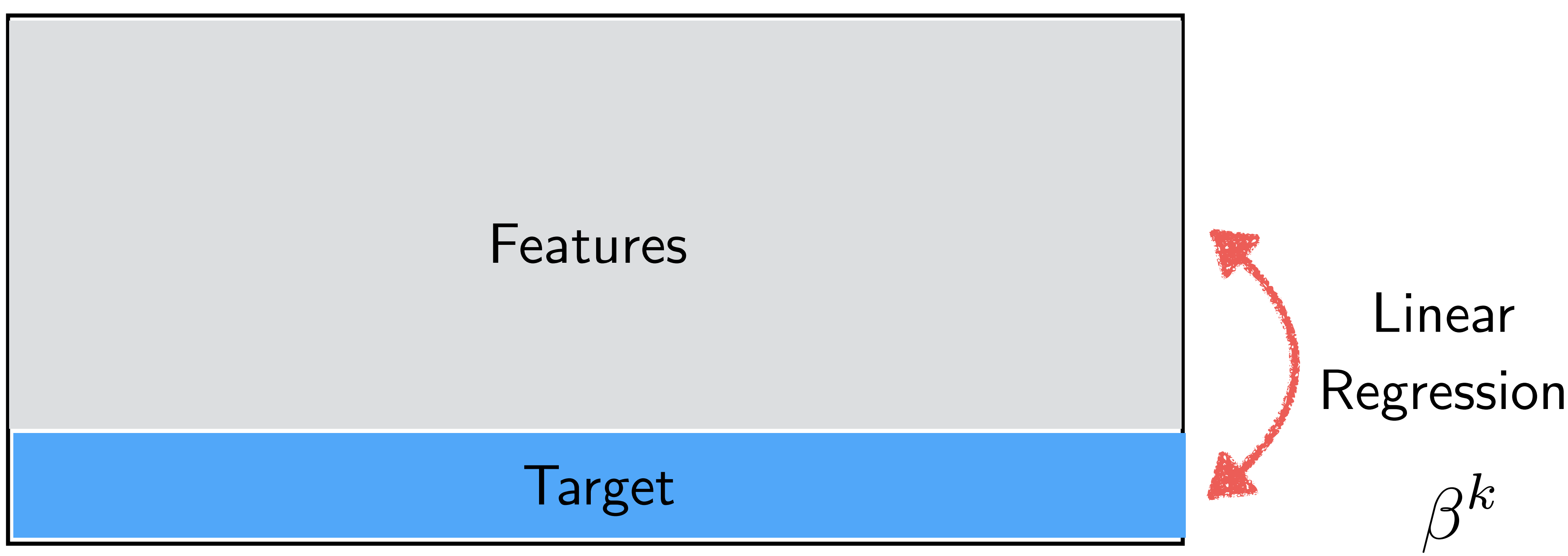
Algorithm:

Transform to Matrix, Do Matrix Estimation, Regression, Prediction

BB-ME

$$\mathcal{X}^k \rightarrow M^k \qquad 1 \leq k \leq L$$

$M^k \equiv$



Answer 2: Time Series Forecasting

Algorithm:

Transform to Matrix, Do Matrix Estimation, Regression, **Prediction**

[X(T-L+2) ... X(T)] → $V(T+1)$

Project on Coln Space of M^k

Inner product with β^k

↓

$\hat{X}(T+1)$

where $k = \big(T + 1 \mod L\big) + 1$

Answer 2: Time Series Forecasting

Theorem (Informal):

For L *large enough* (depending upon model params) with $L^2 \ll T$

the expectation of the transformed matrix obeys approx "linear regression"

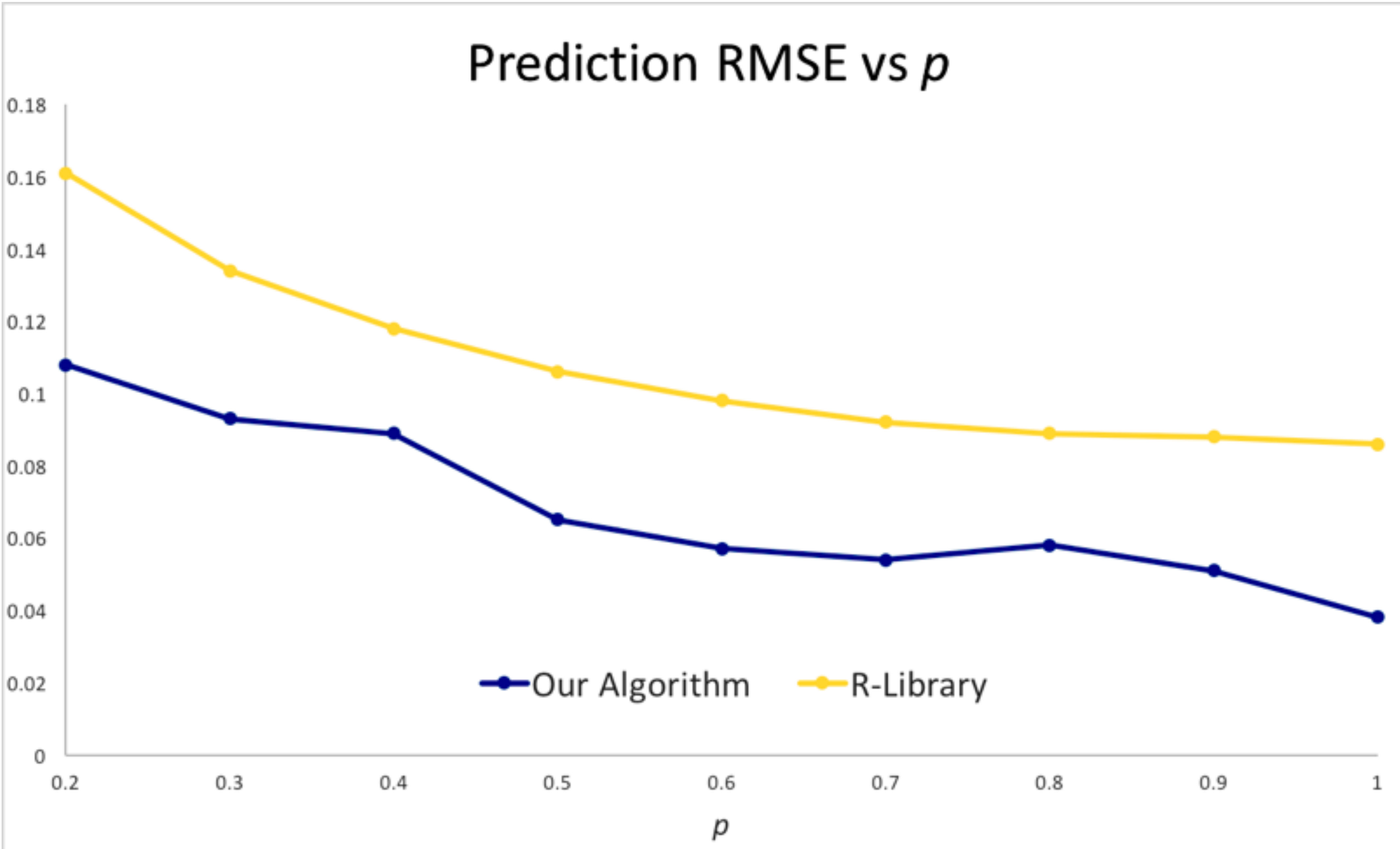
structure in addition to Latent Variable Model. For additive symmetric

noise model with good Matrix Estimation Black-Box

$$\mathbb{E}\left[(\hat{X}(T+1) - X(T+1))^2\right] \leq C\sigma^2/p$$

where σ^2 is noise variance, *C* is a universal constant

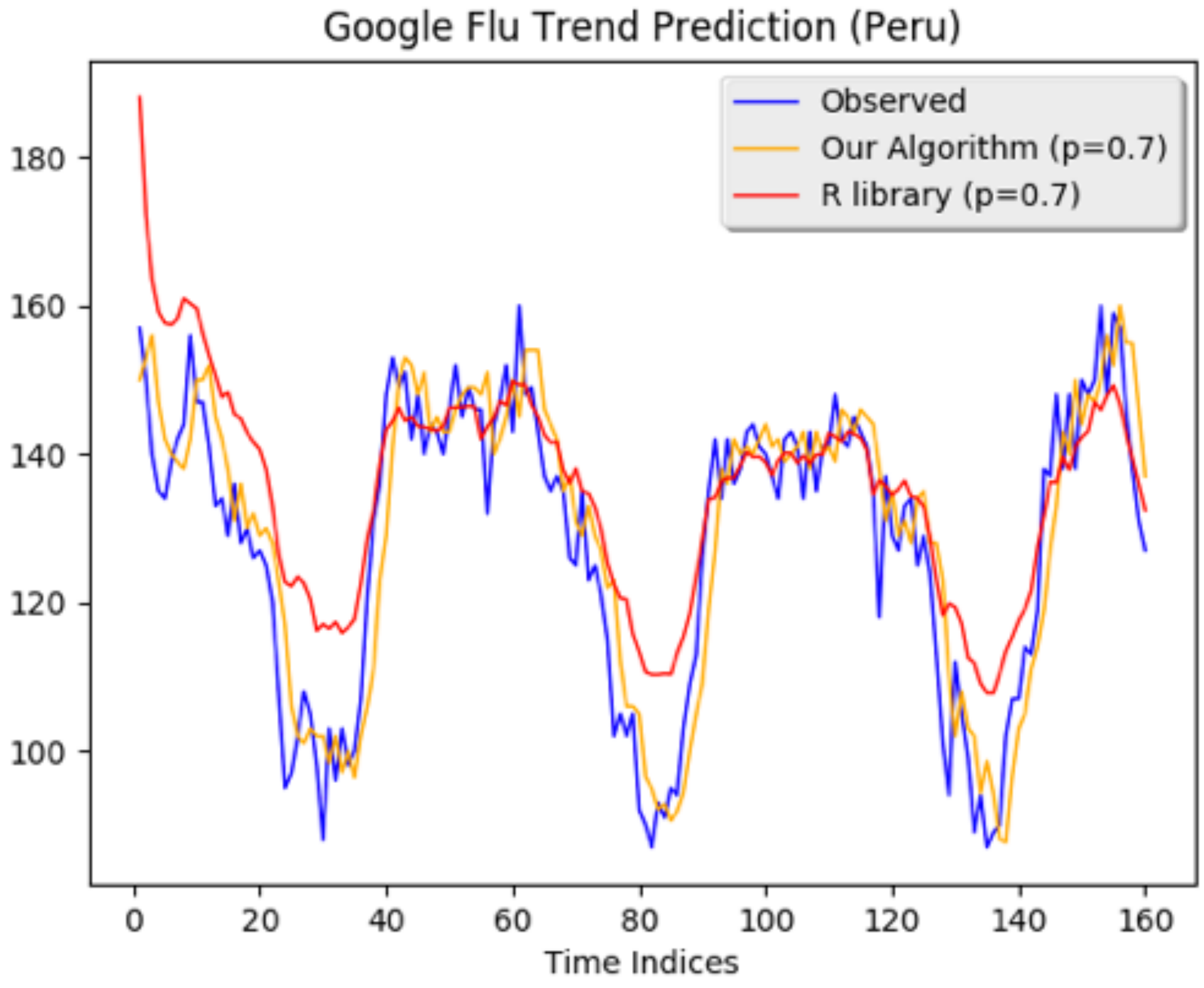
Answer 2: Time Series Forecasting

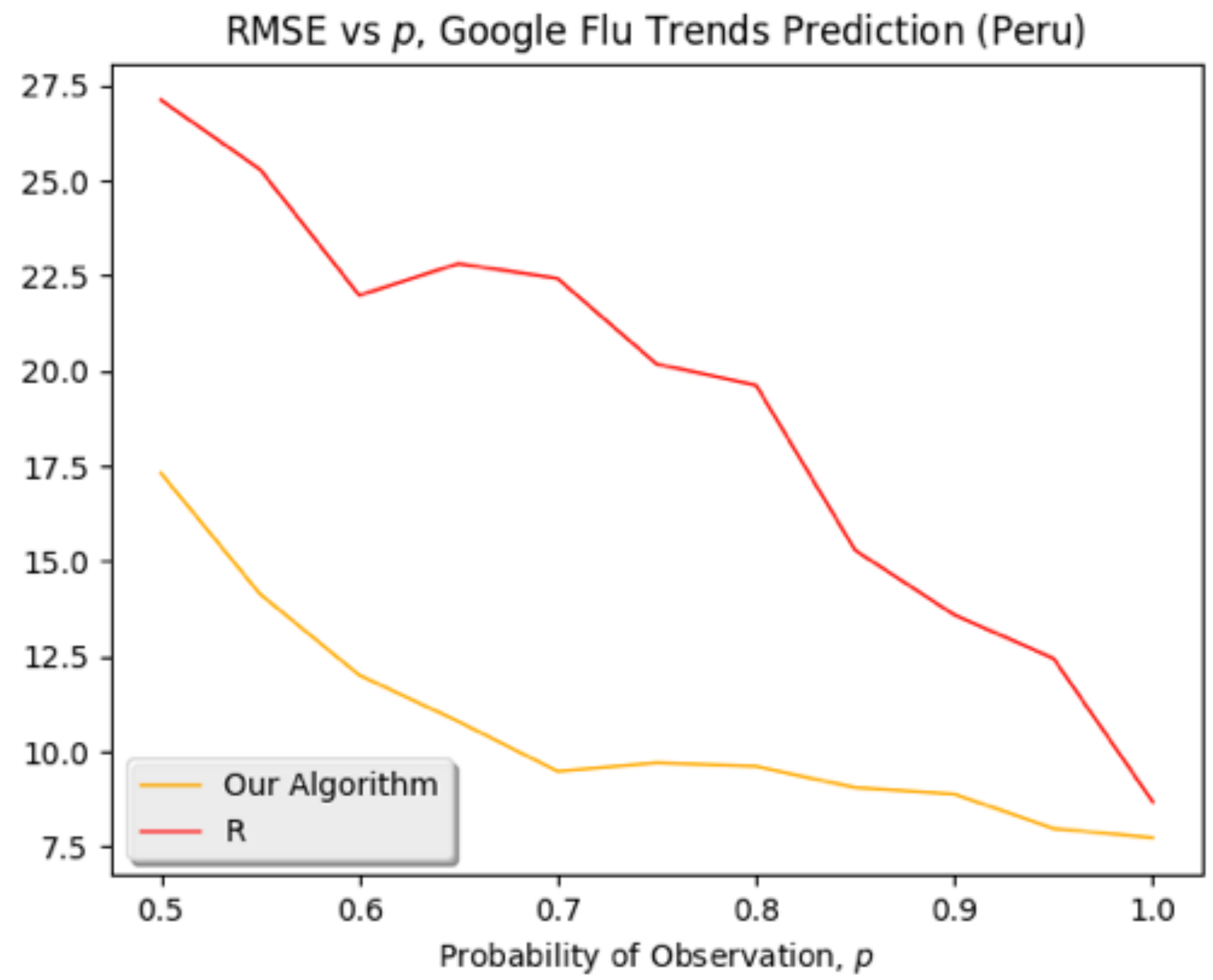


mixture of periodic, trend and auto-regressive

with additive zero-mean noise and randomly missing values

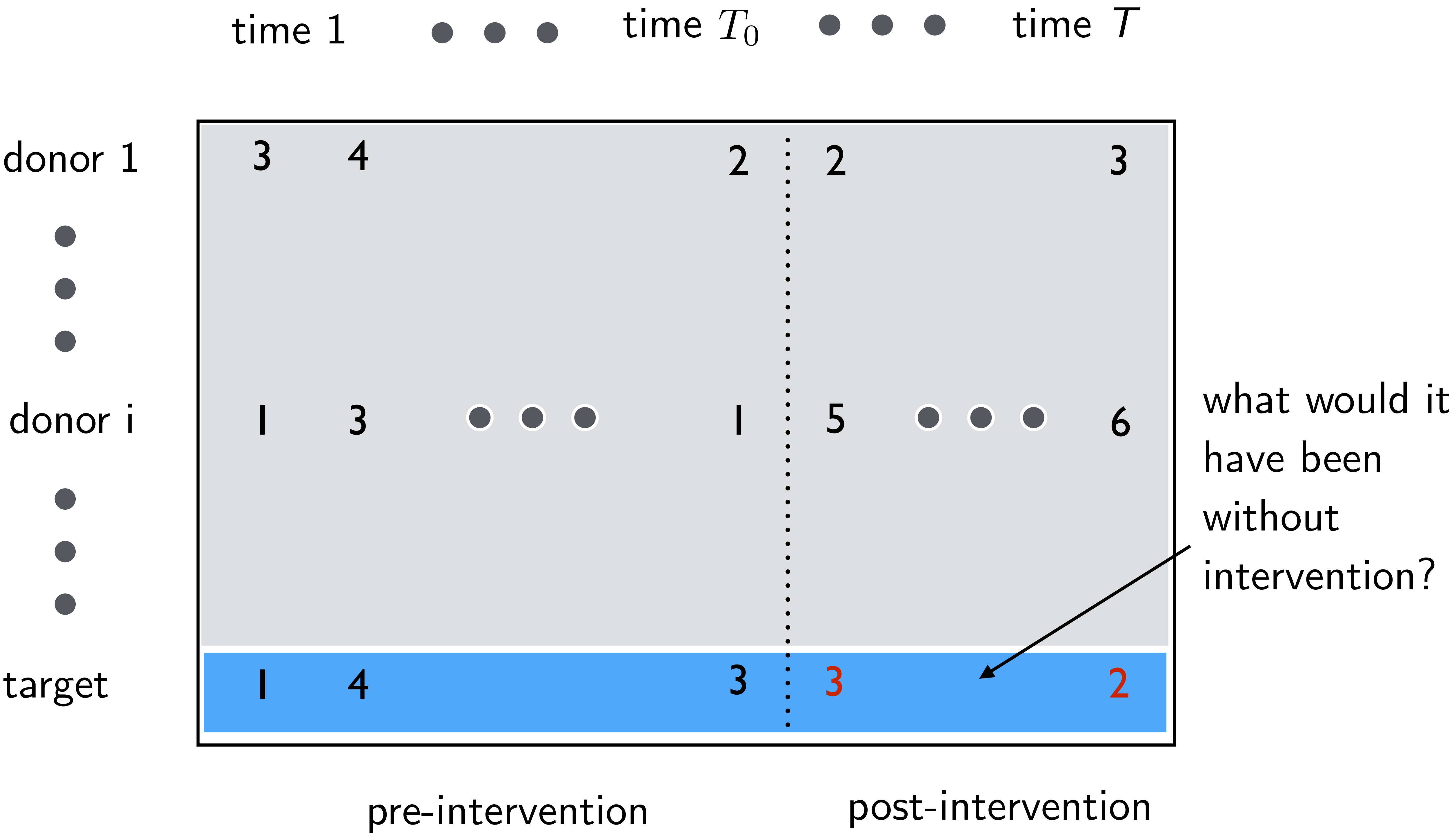
Answer 2: Time Series Forecasting





Google Flu Trend in Peru

Answer 3: Synthetic Control



Answer 3: Synthetic Control

Algorithm:

Grey Matrix through BB-ME

Regression

Target: Blue

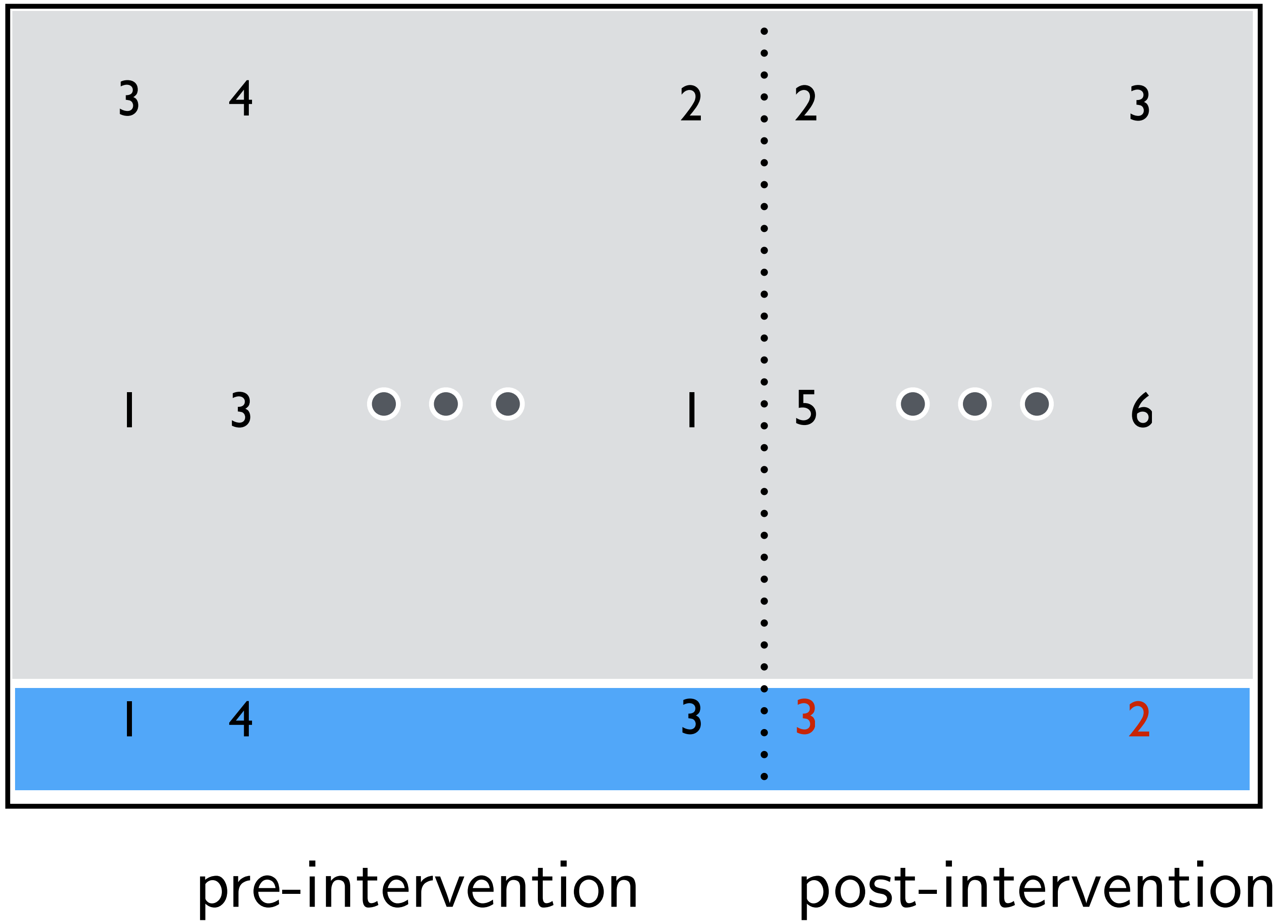
Features: Denoised Grey

Restriction: pre-intervention

Prediction (Synthetic control)

Predict Blue post-intervention

using post-intervention denoised Grey



Answer 3: Synthetic Control

Theorem (Informal):

If the matrix satisfies Latent Variable Model

then the synthetic control estimates true non-intervention outcome

such that the mean-squared error decays as $\widetilde{O}\big(T_0^{-1/2}/p\big)$

where p is the fraction of observed data

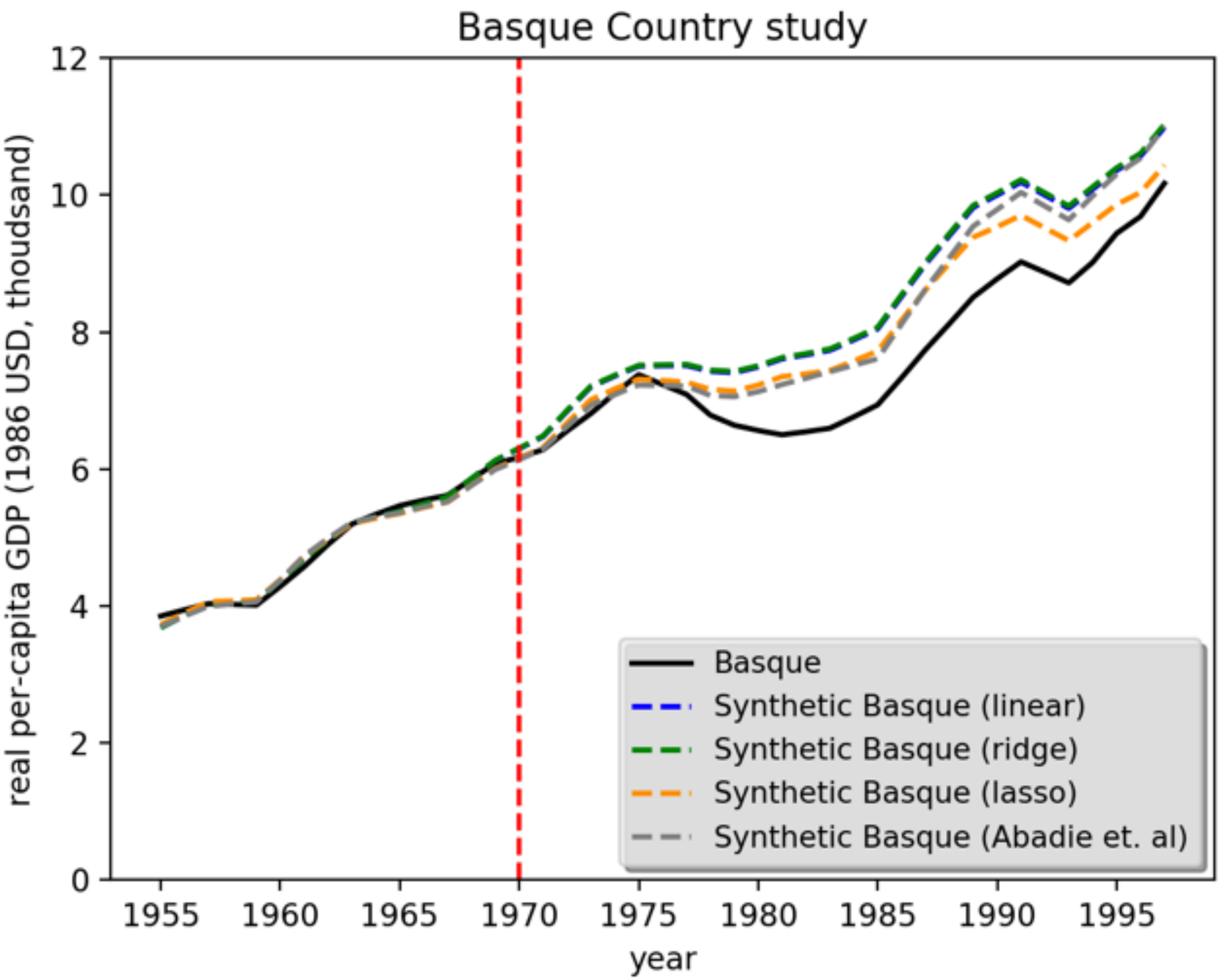
Popular factor model from Econometrics literature (cf. Abadie et al):

satisfy low-rank (rank 2) structure and a (very) special instance of

Latent Variable Model.

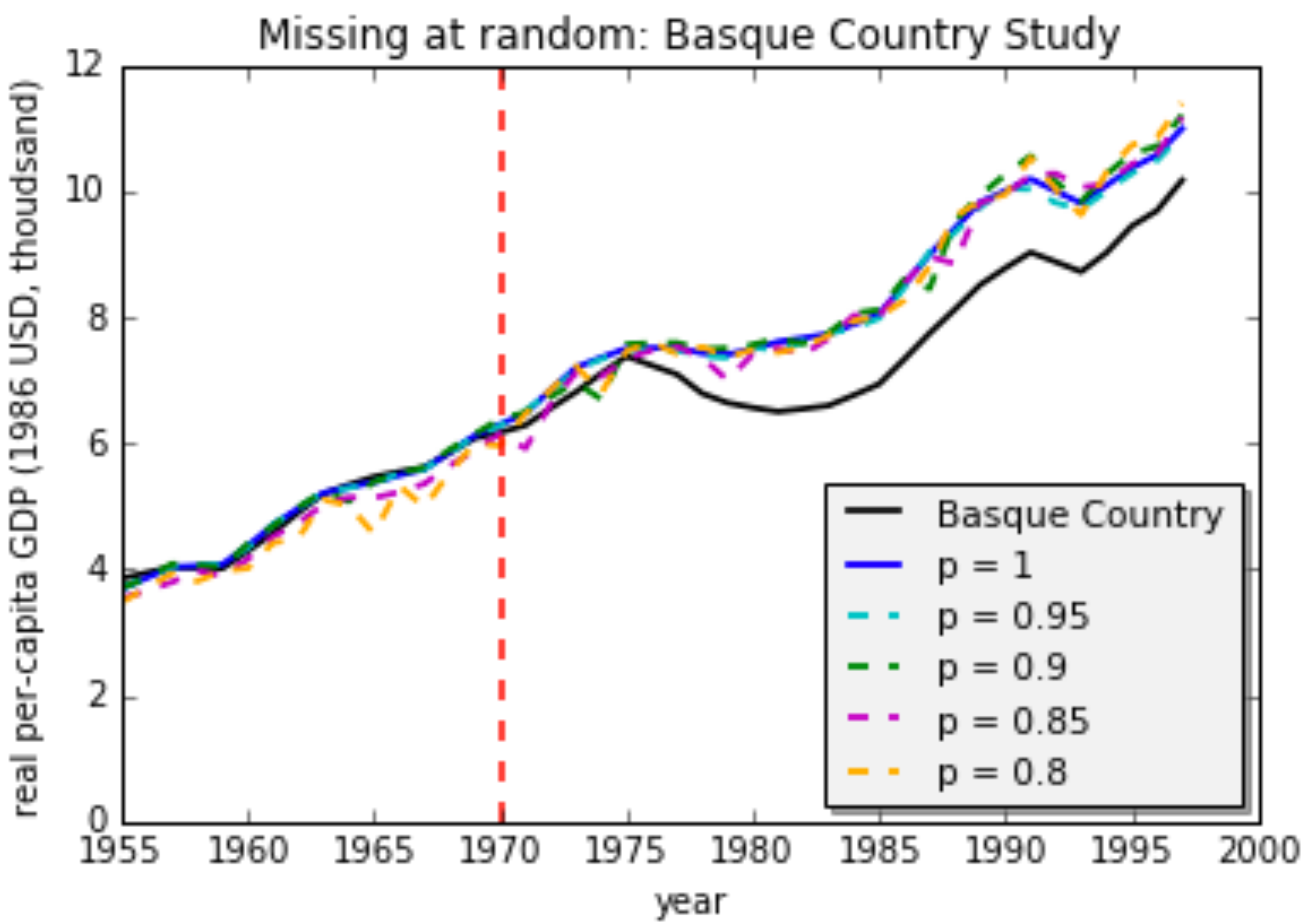
Answer 3: Synthetic Control

Did Terrorism have impact on Economy (of Basque Country)?



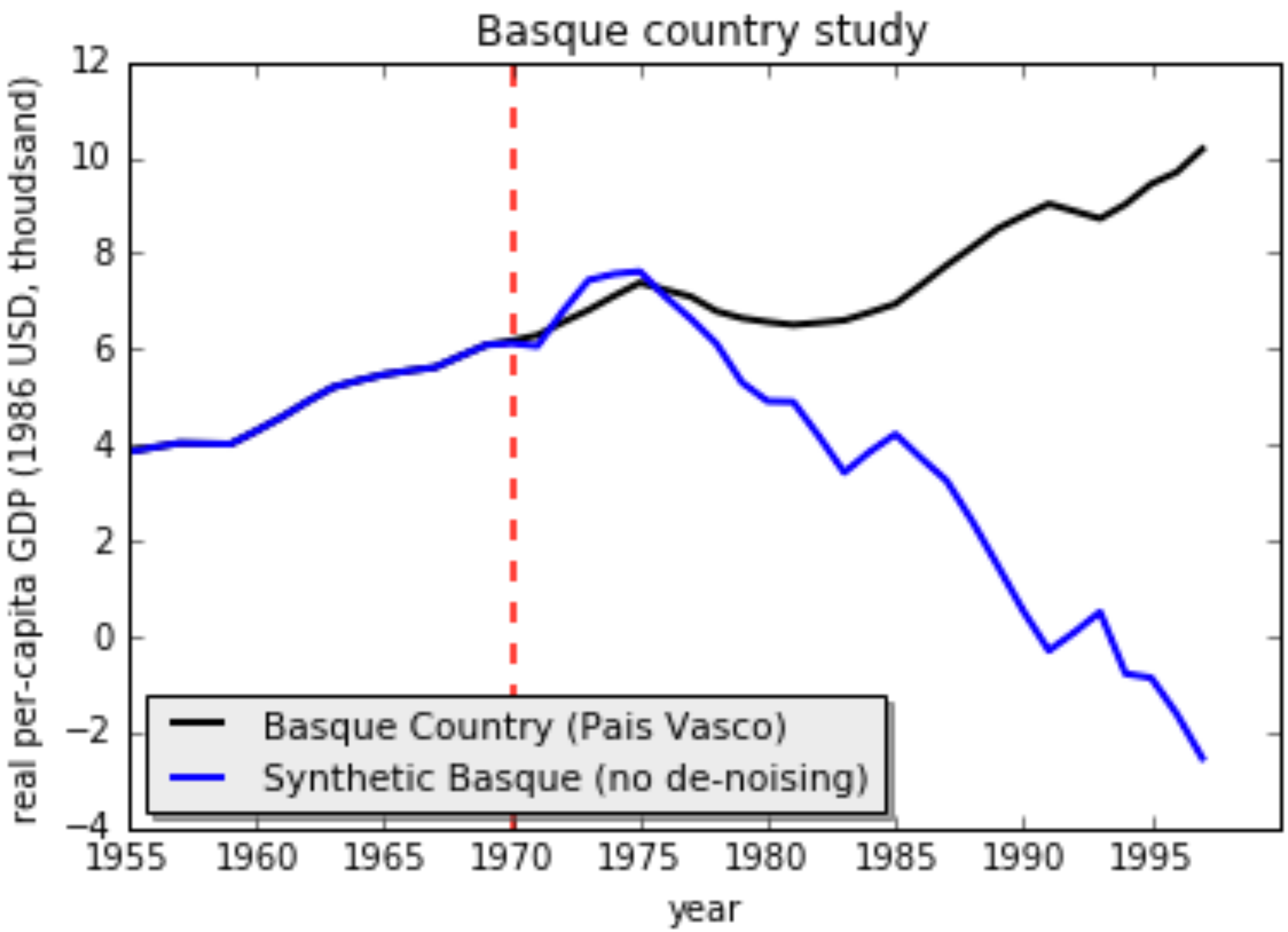
Answer 3: Synthetic Control

Did Terrorism have impact on Economy (of Basque Country)?



Answer 3: Synthetic Control

Did Terrorism have impact on Economy (of Basque Country)?



Summary

Matrix Estimation

A remarkable method with applications beyond *obvious*

Time Series Analysis

Matrix Estimation provides “universal” solution

Going forward

“Time Series Prediction DataBase” using such algorithm