Target-PCA: Transfer Learning Large Dimensional Panel Data

Junting Duan, Markus Pelger and Ruoxuan Xiong

Stanford University and Emory University

Motivation

Goal: Estimate latent factor model and impute missing values for large dimensional panel data with missing data

- Large dimensional panel data with missing entries is prevalent: e.g. macroeconomic data, financial data, recommendation system
- Latent factor model is commonly used to summarize large panels and impute missing values

Problem: Panel is not informative enough to estimate the factor structure:

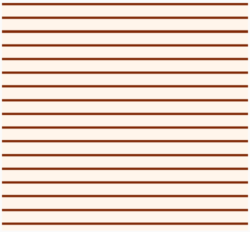
- Observed data insufficient to estimate full factor model:
  - Some times are never observed in panel, e.g. low-frequency data
  - Missingness depends on factor structure, e.g. selection bias
- Weak factor signals: Factors affect only small subset of units
  - Weak factors not identified by principal component analysis (PCA)

Our Solution: Target-PCA optimally combines information from multiple auxiliary panels to estimate factor structure in target panel

A Motivating Example: Low-Frequency Macroeconomic Data

Question: How to obtain high frequency macroeconomic time series that are only observed at lower frequency?

- Naive imputation by lagged values cannot capture fluctuation between observations
- Latent factors for time periods without observations cannot be estimated

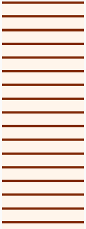


Low-frequency observation pattern: Rows denote time periods, columns denote macroeconomic variables (dark color: observed, light color: missing)

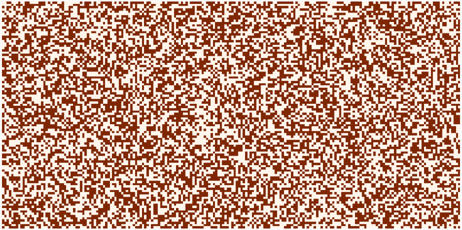
This Paper: Use Auxiliary Data

Use auxiliary data that share some common factors with target data, e.g.,

- Target data: Quarterly observed macroeconomic time-series
- Auxiliary data: Daily observed price-based information (stock returns)



Target data



Auxiliary data

Learn the latent factor structure for target data by optimally weighting auxiliary and target data

- Identify weak signals in target data
- Increase estimation efficiency of common latent factors

Challenge: How to optimally use auxiliary data?

- Factor model difference
  Auxiliary data may not contain all factors for the target
  $\Rightarrow$ Auxiliary data not sufficient to learn factor model
- Dimension difference
  Auxiliary data may have much more units than target data
  $\Rightarrow$ Too low weight for target panel when simply concatenating panels
- Missing pattern issues
  Missing pattern can depend on factor model
  Missing pattern affects the effective sample size

This paper: Novel method Target-PCA:
Optimally combines auxiliary and target data to estimate latent factor model and impute missing entries for the target

Contribution

Methodology:

- New setup to estimate latent factors for target data using auxiliary data
- Identify two effects in combining auxiliary with target data:
  (1) detection of weak signals, (2) efficient estimation
- Target-PCA: A novel estimator for latent factor model using the idea of transfer learning, and simultaneously achieves the two effects
- Inferential theory for target-PCA under very general assumptions on the approximate factor model and missing pattern

$\Rightarrow$ Easy-to-use and widely applicable estimator under general assumptions!

$\Rightarrow$ Importance: Imputation, factor estimation, causal inference

Empirics:

- Demonstrate superior performance of target-PCA, compared to benchmarks, to impute unbalanced macroeconomic panel

Related Literature (Incomplete and Partial List)

Factor modeling

- Full observations with inferential theory: Bai and Ng 2002, Bai 2003, Fan, Liao and Mincheva 2013, Pelger and Xiong 2021a+b
- Partial observations: Stock and Watson 2002, Jin, Miao and Su 2021, Bai and Ng 2021a, Cahan, Bai and Ng 2022, Xiong and Pelger 2022
- Weak factor detection: Lettau and Pelger 2020a+b, Bai and Ng 2021b, Giglio, Xiu, and Zhang 2021, Onatski 2022, Huang, Jiang, Li, Tong, and Zhou 2022

Matrix completion

- Independent sampling: Candes and Recht 2009, Mazumder, Hastie and Tibshirani 2010, Negahban and Wainright 2012
- Independent sampling with inferential theory: Chen, Fan, Ma and Yan 2019

Model and Estimation

Model Setup: Approximate Factor Model for Target and Auxiliary Data

Approximate factor models with k common factors (union of all factors)

Target data: N_y units over T time periods

$$Y_{ti} = \underbrace{F_t^\top}_{1\times k} \underbrace{(\Lambda_y)_i}_{k\times 1} + (e_y)_{ti} \quad \text{and} \quad \underbrace{Y}_{T\times N_y} = \underbrace{F}_{T\times k} \underbrace{\Lambda_y^\top}_{k\times N_y} + \underbrace{e_y}_{T\times N_y}$$

Auxiliary data: N_x units over T time periods

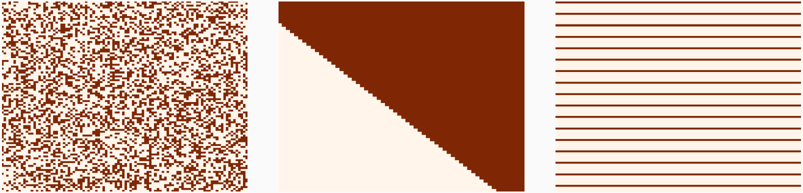
$$X_{ti} = \underbrace{F_t^\top}_{1\times k} \underbrace{(\Lambda_x)i}_{k\times 1} + (e_x)_{ti} \quad \text{and} \quad \underbrace{X}_{T\times N_x} = \underbrace{F}_{T\times k} \underbrace{\Lambda_x^\top}_{k\times N_x} + \underbrace{e_x}_{T\times N_x}$$

- Common factors F in Y and X: without loss of generality, loadings can be 0 for some factors, and $\Lambda_y^\top \Lambda_y$ and $\Lambda_x^\top \Lambda_x$ not full rank
- Dimension: N_y, N_x, and T are large, N_x can be much larger than N_y
- Strength of factor j in Y:
  - $\Rightarrow$ Strong on Y: $\sum_j (\Lambda_y)_{ij}^2 / N_y = O_p(1)$
  - $\Rightarrow$ Weak on Y: $\sum_j (\Lambda_y)_{ij}^2 / N_y = o_p(1)$, e.g., $(\Lambda_y)_{ij} \neq 0$ for small subset
  - $\Rightarrow$ Not existant on Y: $(\Lambda_y)_{ij} = 0$ for all i
- Similar for X, but assume weak factors on Y are strong on X
- Common component $C_{ti} = F_t^\top (\Lambda_y)_i$, idiosyncratic errors $(e_x)_{ti}$ and $(e_y)_{ti}$

Model Setup: Observation Pattern of Target Y

Observation matrix $W^Y = [W_{ti}] : W_{ti} = \begin{cases} 1 & \text{observed} \\ 0 & \text{missing} \end{cases}$

W^Y can be quite general



(a) Randomly missing (b) Staggered adoption (c) Low frequency

Assumption: Observation Pattern of Y

1. W^Y is independent of F and e_y (but can depend on Λ_y)
2. Sufficiently many time-series observations: $\frac{|Q_{ij}^Y|}{T} \geq q > 0$, where Q_{ij}^Y denotes the set of time periods when both units i and j of Y are observed

$\Rightarrow$ For exposition, assume X is fully observed

Target-PCA Estimator

Motivation: Combine PCA objective functions for auxiliary data X and target data Y with a positive target weight γ: For fully observed Y,

$$\min_{F,\Lambda_x,\Lambda_y} \underbrace{\sum_{i=1}^{N_x}\sum_{t=1}^{T}(X_{ti} - F_t^\top(\Lambda_x)_i)^2}_{\text{auxiliary error}} + \gamma \cdot \underbrace{\sum_{i=1}^{N_y}\sum_{t=1}^{T}(Y_{ti} - F_t^\top(\Lambda_y)_i)^2}_{\text{target error}}$$

Equivalent to (with normalization assumption $\frac{1}{T}F^\top F = I_k$)

$$\max_F \operatorname{trace}\Big(F^\top(XX^\top + \gamma \cdot YY^\top)F\Big) = \max_F \operatorname{trace}\Big(F^\top Z^{(\gamma)}(Z^{(\gamma)})^\top F\Big)$$

where $Z^{(\gamma)} = [X, \sqrt{\gamma}Y] \in \mathbb{R}^{T\times(N_x+N_y)}$

Target-PCA:

1. Estimate sample covariance matrix $\tilde{\Sigma}^{(\gamma)} \in \mathbb{R}^{(N_x+N_y)\times(N_x+N_y)}$ of $Z^{(\gamma)}$ using only observed entries
2. Estimate loadings $\tilde{\Lambda}_x$ and $\tilde{\Lambda}_y$ by applying PCA to $\tilde{\Sigma}^{(\gamma)}$
3. Estimate factors $\tilde{F}$ by regressing observed $Z^{(\gamma)}$ on $\tilde{\Lambda}_x$ and $\tilde{\Lambda}_y$
4. Estimate common components/impute missing entries $\tilde{C}_{ti} = \tilde{F}_t^\top(\tilde{\Lambda}_y)_i$

$\Rightarrow$ Xiong and Pelger (2022) applied to $Z^{(\gamma)}$

Key Element of Target-PCA Estimator: Target Weight γ

$$\min_{F,\Lambda_x,\Lambda_y} \underbrace{\sum_{i=1}^{N_x}\sum_{t=1}^{T}(X_{ti} - F_t^\top(\Lambda_x)_i)^2}_{\text{auxiliary error}} + \gamma \cdot \underbrace{\sum_{i=1}^{N_y}\sum_{t=1}^{T}(Y_{ti} - F_t^\top(\Lambda_y)_i)^2}_{\text{target error}}$$

Three special cases:

- $\gamma = 0$: PCA on X
- $\gamma = \infty$: PCA on observed Y
- $\gamma = 1$: PCA on concatenated data $Z = [X, Y]$

Two fundamental effects of target weight γ:

- Consistency effect in factor identification
  (We need to select γ at the right rate)
- Efficiency effect in the estimation of factors and loadings
  (We need to select γ at the right scale)

Optimal target weight γ achieves both effects in one-step
$\Rightarrow$ Optimal combination of multiple data sets in one step

Two Important Effects of Target Weight

Effect 1: Consistency Effect of Target Weight γ

Consistency: Select weight γ to identify strong factors from both panels

- Allows to estimate weak factors on Y with target-PCA
- Intuition: Top eigenvalues of $XX^\top$ and $\gamma YY^\top$ should be of the same scale. Top eigenvalues $XX^\top$ and $YY^\top$ proportional to N_x and N_y

$\Rightarrow$ Select $\gamma = O(N_x/N_y)$

Illustrative example: A two-factor model

- Panel Y: Factor 1 strong, but factor 2 weak $\Rightarrow$ Y only identifies factor 1
- Panel X: Factor 2 strong, but factor 1 missing $\Rightarrow$ X only identifies factor 2

Specifically

- Loadings of Y: factor 1 is strong: $(\Lambda_y)_{i1} \overset{i.i.d.}{\sim} (0, \sigma^2_{\Lambda_y})$; factor 2 is weak: $(\Lambda_y)_{i2} \overset{i.i.d.}{\sim} (0, \sigma^2_{\Lambda_y})$ if $i < N_y^{1/2}$, otherwise, $(\Lambda_y)_{i2} = 0$
- Loadings of X: only second factor exists $(\Lambda_x)_{i2} \overset{i.i.d.}{\sim} (0, \sigma^2_{\Lambda_x})$ and $(\Lambda_x)_{i1} = 0$
- $F_{t1} \overset{i.i.d.}{\sim} (0, \sigma^2_F)$, $F_{t2} \overset{i.i.d.}{\sim} (0, \sigma^2_F)$, $(e_x)_{ti} \overset{i.i.d.}{\sim} (0, \sigma^2_{e_x})$, $(e_y)_{ti} \overset{i.i.d.}{\sim} (0, \sigma^2_{e_y})$

Effect 1: Consistency Effect of Target Weight γ

Without errors and missing observations, target-PCA estimates factors from

$$\frac{1}{N_x+N_y} Z^{(\gamma)} Z^{(\gamma)\top} = \frac{1}{N_x+N_y}\left[XX^\top + \gamma YY^\top\right]$$
$$= \begin{bmatrix} F^{(1)} & F^{(2)} \end{bmatrix} \left(\hat{\Sigma}_{\Lambda,t}^{(\gamma)} + o_p(1)\right) \begin{bmatrix} F^{(1)\top} \\ F^{(2)\top} \end{bmatrix},$$

where $F^{(1)}, F^{(2)} \in \mathbb{R}^T$ denote the vector of the first and second factors and

$$\hat{\Sigma}_{\Lambda,t}^{(\gamma)} = \frac{N_x}{N_x+N_y}\left(\underbrace{\begin{bmatrix} 0 & 0 \\ 0 & \sigma_{\Lambda_x}^2 \end{bmatrix}}_{\Sigma_{\Lambda_x}} + \gamma \frac{N_y}{N_x} \cdot \underbrace{\begin{bmatrix} \sigma_{\Lambda_y}^2 & 0 \\ 0 & \frac{\sigma_{\Lambda_y}^2}{N_y^{1/2}} \end{bmatrix}}_{\Sigma_{\Lambda_y},t} \right)$$

- Key idea: $\gamma = O\left(N_x/N_y\right)$ gives asymptotically full rank of $\hat{\Sigma}_{\Lambda,t}^{(\gamma)}$
- Both eigenvalues in the limit of $\hat{\Sigma}_{\Lambda,t}^{(\gamma)}$ are of the same order

$\Rightarrow$ Both factors can be identified from $\frac{1}{N_x+N_y} Z^{(\gamma)} Z^{(\gamma)\top}$

$\Rightarrow$ Naive concatenating ($\gamma = 1$) can have rank deficiency for $\hat{\Sigma}_{\Lambda,t}^{(1)}$

Effect 2: Efficiency Effect of Target Weight γ

Efficient weighting of panels:

- First ensure identification of all factors in Y, then fine-tune the scaling of γ
- For a target weight γ with the right rate $O(N_x/N_y)$, which scaling constant improves the efficiency?

$\Rightarrow$ Balance noise level between observed target data and auxiliary data

Illustrative Example: A one-factor model

- Key ingredient: Noise level is different in X and Y
- $F_{t1} \overset{i.i.d.}{\sim} (0, \sigma_F^2)$, $(e_x)_{ti} \overset{i.i.d.}{\sim} (0, \sigma_{e_x}^2)$, $(e_y)_{ti} \overset{i.i.d.}{\sim} (0, \sigma_{e_y}^2)$
- Loadings of Y: $(\Lambda_y)_{i1} \overset{i.i.d.}{\sim} (0, \sigma_{\Lambda_y}^2)$
- Loadings of X: $(\Lambda_x)_{i1} \overset{i.i.d.}{\sim} (0, \sigma_{\Lambda_x}^2)$
- Observation pattern:
  Entries in Y are missing uniformly at random with $\mathbb{P}(W_{ti}^Y = 1) = p$

Effect 2: Efficiency Effect of Target Weight γ

Proposition

Let $\delta_{N_y,T} = \min(N_y, T)$ and assume $N_y/N_x \to c \in (0, \infty)$. For the one-factor model example, as T, N_x, $N_y \to \infty$, for any i and t, we have

$$\sqrt{\delta_{N_y,T}}(\Sigma_{C,ti}^{(\gamma)})^{-1/2}\left(\tilde{C}_{ti} - C_{ti}\right) \xrightarrow{d} \mathcal{N}(0,1),$$

where

$$\begin{aligned}\Sigma_{C,ti}^{(\gamma)} =& \frac{\delta_{N_y T}}{T}\frac{\sigma_{e_y}^2}{p\sigma_F^2}F_t^2 + 2\frac{\delta_{N_y T}}{T}\left(\frac{1}{p} - 1\right)(\Lambda_y)_i^2 F_t^2 \\ &+ \frac{\delta_{N_y T}}{N_y}(\Lambda_y)_i^2\left(\sigma_{\Lambda_x}^2 + \gamma\frac{N_y}{N_x}p\sigma_{\Lambda_y}^2\right)^{-2}\left(\frac{N_y}{N_x}\sigma_{\Lambda_x}^2\sigma_{e_x}^2 + \gamma^2\frac{N_y^2}{N_x^2}p\sigma_{\Lambda_y}^2\sigma_{e_y}^2\right)\end{aligned}$$

- The optimal γ that minimizes $\Sigma_{C,ti}^{(\gamma)}$ is $\gamma^* = \sigma_{e_x}^2/\sigma_{e_y}^2$ for any i and t
- Interpretation: Up-weight the panel with smaller idiosyncratic error to improve estimation efficiency (similar to GLS)

Asymptotic Results

Theorem 1: Consistency

Let $\delta_{N_y,T} = \min(N_y, T)$ and suppose that $N_y/N_x \to c \in [0, \infty)$. Under general observation pattern and approximate factor model ▸ assumptions, for $T, N_x, N_y \to \infty$:

1. If $\gamma = r \cdot N_x/N_y$ with some constant r, then $\Sigma_{\Lambda,t}^{(\gamma)} := \lim_{N_x,N_y\to\infty} \frac{N_x}{N_x+N_y}\left(\Sigma_{\Lambda_x} + \gamma \frac{N_y}{N_x}\Sigma_{\Lambda_y,t}\right)$ is positive definite, and

$$\delta_{N_y,T}\left(\frac{1}{N_x+N_y}\sum_{i=1}^{N_x+N_y}\left\|\tilde{\Lambda}_i^{(\gamma)} - H^{(\gamma)}\Lambda_i^{(\gamma)}\right\|^2\right) = O_p(1)$$

$$\delta_{N_y,T}\left(\frac{1}{T}\sum_{t=1}^{T}\left\|\tilde{F}_t - (H^{(\gamma)\top})^{-1}F_t\right\|^2\right) = O_p(1)$$

   This implies that $\tilde{C}_{ti}$ of Y is consistent.

2. If $\gamma \neq r \cdot N_x/N_y$ for any constant r, then $\Sigma_{\Lambda,t}^{(\gamma)}$ may not be positive definite. If $\Sigma_{\Lambda,t}^{(\gamma)}$ is not positive definite, then $\tilde{F}_t$ is inconsistent for some t

Theorem 2: Asymptotic Normality

Let $\delta_{N_y,T} = \min(N_y, T)$ and suppose that $N_y/N_x \to c \in [0,\infty)$ and $\gamma = r \cdot N_x/N_y$ for some constant r. Under general observation pattern and approximate factor model ▸ assumptions, as $T, N_x, N_y \to \infty$:

- Loadings of Y: for $\sqrt{T}/N_y \to 0$,

$$\sqrt{T}(\Sigma^{(\gamma)}_{\Lambda_y,i})^{-1/2}\left((H^{(\gamma)})^{-1}(\tilde{\Lambda}_y)_i - (\Lambda_y)_i\right) \xrightarrow{d} \mathcal{N}(0, I_k),$$

where $\Sigma^{(\gamma)}_{\Lambda_y,i} = \Sigma_F^{-1}(\Sigma^{(\gamma)}_{\Lambda})^{-1}(\Gamma^{(\gamma),\text{obs}}_{\Lambda_y,i} + \Gamma^{(\gamma),\text{miss}}_{\Lambda_y,i})(\Sigma^{(\gamma)}_{\Lambda})^{-1}\Sigma_F^{-1}$

- Factors: for $\sqrt{T}/N_y \to 0$ and $\sqrt{N_y}/T \to 0$,

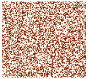
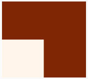
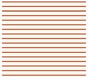
$$\sqrt{\delta_{N_y,T}}(\Sigma^{(\gamma)}_{F,t})^{-1/2}\left(H^{(\gamma)\top}\tilde{F}_t - F_t\right) \xrightarrow{d} \mathcal{N}(0, I_k),$$

where $\Sigma^{(\gamma)}_{F,t} = (\Sigma^{(\gamma)}_{\Lambda,t})^{-1}\left[\frac{\delta_{N_y,T}}{N_y}\Gamma^{(\gamma),\text{obs}}_{F,t} + \frac{\delta_{N_y,T}}{T}\Gamma^{(\gamma),\text{miss}}_{F,t}\right](\Sigma^{(\gamma)}_{\Lambda,t})^{-1}$

- Common components of Y: for $\sqrt{T}/N_y \to 0$ and $\sqrt{N_y}/T \to 0$,

$$\sqrt{\delta_{N_y,T}}(\Sigma^{(\gamma)}_{C,ti})^{-1/2}(\tilde{C}_{ti} - C_{ti}) \xrightarrow{d} \mathcal{N}(0,1)$$

Optimal γ^* for Different Missing Patterns and Noise Ratios

	p	$N_x/N_y = 1$			$N_x/N_y = 4$		
		NR=0.25	NR=1	NR=4	NR=0.25	NR=1	NR=4
	60%	0.25	1.00	4.00	0.25	1.00	4.00
	75%	0.25	1.00	4.00	0.25	1.00	4.00
	90%	0.25	1.00	4.00	0.25	1.00	4.00
	60%	0.61	1.75	4.25	1.95	5.09	7.00
	75%	0.42	1.53	4.35	1.06	3.62	6.12
	90%	0.28	1.15	4.18	0.40	1.62	4.61
	60%	0.55	1.96	4.66	1.69	5.52	7.84
	75%	0.39	1.46	4.34	0.92	3.23	5.84
	90%	0.28	1.12	4.13	0.40	1.56	4.50
	60%	0.70	2.16	4.96	2.24	6.20	8.63
	75%	0.47	1.48	4.30	1.30	3.56	5.90
	90%	0.32	1.12	4.04	0.61	1.91	4.50

$\Rightarrow$ Missing at random: Optimal γ^* only depends on NR (noise ratio), but not on N_x/N_y and fraction of observed entries p

$\Rightarrow$ Other observation patterns: Optimal γ^* depends on NR, N_x/N_y, p and other quantities related to correlations in observation pattern

Empirical Results

Goal: Compare imputation accuracy of target-PCA with benchmark methods

- XP_Y: PCA on Y only (Xiong and Pelger 2022)
- XP_Z: PCA on $Z = [X, Y]$ (Xiong and Pelger 2022)
- SE-PCA: Combining factors extracted from separate PCAs on X and Y

Data: 120 fully observed monthly U.S. macroeconomic indicators from FRED-MD from 01/1960 to 12/2020

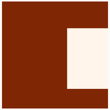
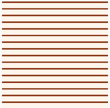
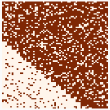
- Target Y: 19 series in interest and exchange rates category
- Auxiliary X: 101 series from the other 7 categories

Mask Y according to four types of missing patterns

- Missing at random
- Block missing
- Low-frequency observation
- Censoring

Compare the relative MSE $\sum_{i,t}(\tilde{C}_{it} - Y_{it})^2 / \sum_{i,t} Y_{it}^2$ on the masked entries

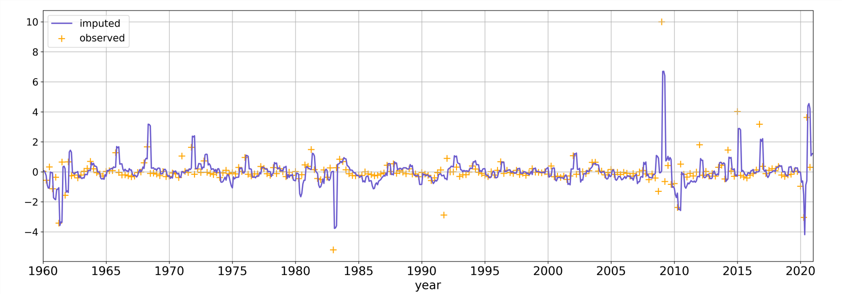
Empirical Study 1 – Relative MSE of Different Methods

	Observation Pattern (Missing Ratio)	factor number	T-PCA	XP_Y	XP_Z	SE-PCA
	missing at random (40%)	$k=1$	**0.787**	0.796	0.986	0.806
		$k=2$	**0.486**	0.503	0.969	0.499
		$k=3$	**0.483**	0.635	0.927	0.627
		$k=4$	**0.491**	0.813	0.793	0.795
		$k=5$	**0.479**	1.363	0.613	1.355
	block missing (19%)	$k=1$	**0.958**	1.018	0.971	1.003
		$k=2$	**0.700**	0.805	0.961	0.852
		$k=3$	**0.713**	0.796	0.974	0.803
		$k=4$	**0.741**	0.783	0.974	0.781
		$k=5$	**0.786**	2.601	0.935	2.584
	low-frequency (92%)	$k=1$	**0.942**	0.949	1.019	1.009
		$k=2$	**0.927**	1.140	0.931	1.149
		$k=3$	**0.926**	1.213	0.936	1.223
		$k=4$	**0.910**	1.212	1.095	1.234
		$k=5$	**1.017**	1.251	1.092	1.280
	censoring (40%)	$k=1$	**0.927**	-	0.996	0.995
		$k=2$	**0.881**	-	0.996	0.994
		$k=3$	**0.892**	-	0.993	0.992
		$k=4$	**0.882**	-	0.990	0.987
		$k=5$	**0.869**	-	0.984	0.981

$\Rightarrow$ Target-PCA provides the most precise imputation for all cases

Empirical Study 2 – Imputation of Low-Frequency Macro Time Series

Quarterly observed GDP vs. monthly imputed GDP by target-PCA



⇒ Target-PCA captures the (unknown and unobserved) variation in between two quarterly GDP observations using monthly observed auxiliary data

⇒ Target-PCA can be used for nowcasting low-frequency macro time-series

Conclusion

Target-PCA:

- Novel method to estimate a latent factor model for a target panel with missing observations using supplementary panel data
- Transfer learning perspective: Optimally extracts information from supplementary data that is useful for the target
- Easy-to-adopt method to estimate factor structure and impute missing observations that is broadly applicable
- Benefits of target-PCA:
  1. Estimation of weak factors in target panel
  2. Efficient combination of multiple panels
  3. Estimation of factor structure under challenging missing patterns
- Asymptotic inferential theory under very general assumptions on the approximate factor model and missing patterns:
  $\Rightarrow$ Provides guidance for the optimal selection of γ^*

Appendix

We present the assumptions of a simplified factor model which captures the main insight of the general approximate factor model

Assumption S1: Simplified Factor Model

There exists a constant $C < \infty$ such that

1. Factors: $F_t \overset{\text{i.i.d.}}{\sim} (0, \Sigma_F)$ and $\mathbb{E}\|F_t\|^4 \leq C$ for any t.
2. Loadings: $(\Lambda_x)_i \overset{\text{i.i.d.}}{\sim} (0, \Sigma_{\Lambda_x})$, where Σ_{Λ_x} is positive semidefinite. $(\Lambda_y^{\text{full}})_i \overset{\text{i.i.d.}}{\sim} (0, \Sigma_{\Lambda_y}^{\text{full}})$ and the loading of the j-th factor $(\Lambda_y)_{ij} = (\Lambda_y^{\text{full}})_{ij} \cdot (U_y)_{ij}$ where $\Sigma_{\Lambda_y}^{\text{full}}$ is positive definite, and Bernoulli random variable $(U_y)_{ij} \in \{0, 1\}$ is independent in i with $\mathbb{P}((U_y)_{ij} = 1) = p_j$ for some $p_j \in [0, 1]$. Furthermore, $\mathbb{E}\|(\Lambda_x)_i\|^4 \leq C$, $\mathbb{E}\|(\Lambda_y)_i\|^4 \leq C$ and $\Sigma_{\Lambda_x} + \Sigma_{\Lambda_y}$ is positive definite, where $\Sigma_{\Lambda_y} = \mathbb{E}[(\Lambda_y)_i(\Lambda_y)_i^\top]$.
3. Idiosyncratic errors: $(e_x)_{ti} \overset{\text{i.i.d.}}{\sim} (0, \sigma_{e_x}^2), (e_y)_{ti} \overset{\text{i.i.d.}}{\sim} (0, \sigma_{e_y}^2)$, $\mathbb{E}(e_x)_{ti}^8 \leq C, \mathbb{E}(e_y)_{ti}^8 \leq C$, and the ratio $\sigma_{e_x}/\sigma_{e_y}$ is bounded away from 0.
4. Independence: $F, \Lambda_x, \Lambda_y, e_x$ and e_y are independent.

Assumption S2: Moments of Observation Pattern and Simplified Factor Model

1. Missing pattern: for any i, there exist constants $\omega_i^{(1)}$, $\omega_i^{(2,1)}$, $\omega_i^{(2,2)}$, $\omega_i^{(2,3)}$ and $\omega_i^{(3)}$, such that $\frac{1}{N_y}\sum_{j=1}^{N_y}\frac{q_{ij}}{q_{ii}q_{jj}}\xrightarrow{p}\omega_i^{(1)}$, $\frac{1}{N_y^2}\sum_{j,l=1}^{N_y}\frac{q_{ii,jl}}{q_{ii}q_{jl}}\xrightarrow{p}\omega_i^{(2,1)}$, $\frac{1}{N_y^2}\sum_{j,l=1}^{N_y}\frac{q_{jj,il}}{q_{jj}q_{il}}\xrightarrow{p}\omega_i^{(2,2)}$, $\frac{1}{N_y^2}\sum_{j,l=1}^{N_y}\frac{q_{ij,il}}{q_{ij}q_{il}}\xrightarrow{p}\omega_i^{(2,3)}$, and $\frac{1}{N_y^3}\sum_{j,l,h=1}^{N_y}\frac{q_{il,jh}}{q_{il}q_{jh}}\xrightarrow{p}\omega_i^{(3)}$. Furthermore, there exist constants $\omega^{(1)}$, $\omega^{(2)}$ and $\omega^{(3)}$, such that $\frac{1}{N_y}\sum_{i=1}^{N_y}\omega_i^{(1)}\xrightarrow{p}\omega^{(1)}$, $\frac{1}{N_y}\sum_{i=1}^{N_y}\omega_i^{(3)}\xrightarrow{p}\omega^{(3)}$, and $\frac{1}{N_y}\sum_{i=1}^{N_y}\omega_i^{(2,1)}=\frac{1}{N_y}\sum_{i=1}^{N_y}\omega_i^{(2,2)}\xrightarrow{p}\omega^{(2)}$.
2. Systematic loadings for observed data: For any t, $\frac{1}{N_y}\sum_{i=1}^{N_y}W_{ti}^{Y}(\Lambda_y)_i(\Lambda_y)_i^\top\xrightarrow{p}\Sigma_{\Lambda_y,t}$ and $\Sigma_{\Lambda_x}+\Sigma_{\Lambda_y,t}$ is positive definite.

Effect 1: Consistency Effect of Target Weight γ

Proposition

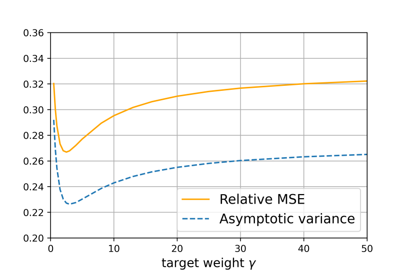
Under the data generating process and observation pattern described for the two-factor model, let $\delta_{N_y,T} = \min(N_y, T)$ and assume that $N_y/N_x \to 0$. Target-PCA with $\gamma = r \cdot N_x/N_y$ for some constant $r \in (0, \infty)$ can consistently estimate the latent factors. As $T, N_x, N_y \to \infty$, there exists some rotation matrix H such that

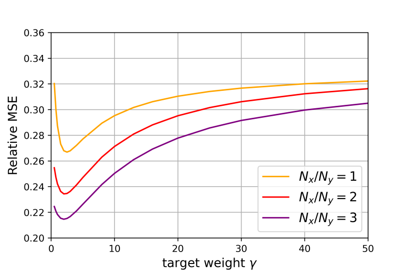
$$\delta_{N_y,T}\left(\frac{1}{T}\sum_{t=1}^{T}\left\|\tilde{F}_t - HF_t\right\|^2\right) = O_p(1).$$

If $\gamma = O(1)$, then $\tilde{F}_t$ is inconsistent.

Effect 2: Simulation Results of Efficiency Effect

Relative MSE of $\tilde{C}_{ti}$ for all i and t





- The optimal γ^* only varies with $\sigma_{e_x}/\sigma_{e_y}$ but not N_x/N_y in this missing at random example
- The optimal γ^* that minimizes the relative MSE coincides with the optimal γ^* that minimizes the asymptotic variance

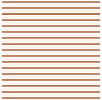
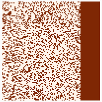
$\Rightarrow$ Use inferential theory to select the optimal γ^*

Simulation

- Comparison between target-PCA method and three benchmark methods
  - T-PCA: Our target-PCA method
  - XP_Y: PCA method (Xiong and Pelger (2020)) using only target Y
  - XP_Z: PCA method (Xiong and Pelger (2020)) using directly concatenated panel $Z = [X, Y]$
  - SE-PCA: Separate PCAs method combining factors separately extracted from X and Y as the factor estimators
- Two-factor model with three missing mechanisms:
  - Missing at random
  - Low-frequency observation
  - Missing depends on loadings
    Entries in Y are missing conditional on $S_i = \mathbb{1}(|(\Lambda_y)_{i2}| > \text{threshold})$
- We compare the relative mean square error (relative MSE) for the observed, missing and all entries of the common component of Y:

$$\text{relative MSE}_{\mathcal{M}} = \frac{\sum_{(t,i)\in\mathcal{M}} (\tilde{C}_{ti} - C_{ti})^2}{\sum_{(t,i)\in\mathcal{M}} (C_{ti})^2}$$

Simulation: Relative MSE for Different Estimators

Observation Pattern	$\mathcal{M}$	T-PCA	XP_Y	XP_Z	SE-PCA
	obs	**0.182**	0.407	0.224	0.531
	miss	**0.179**	0.411	0.222	0.563
	all	**0.181**	0.409	0.223	0.547
	obs	**0.279**	-	0.844	1.052
	miss	**1.011**	-	1.124	1.104
	all	**0.645**	-	0.980	1.077
	obs	**0.213**	0.234	0.256	0.276
	miss	**0.247**	0.290	0.281	0.352
	all	**0.239**	0.276	0.275	0.335

- Target-PCA estimator is robust in different settings
- Target-PCA estimator is efficient and achieves the smallest relative MSE compared with other three methods in most cases

Empirical Study 1 – Comparison with Benchmark Methods

Illustration of the performance of different methods:

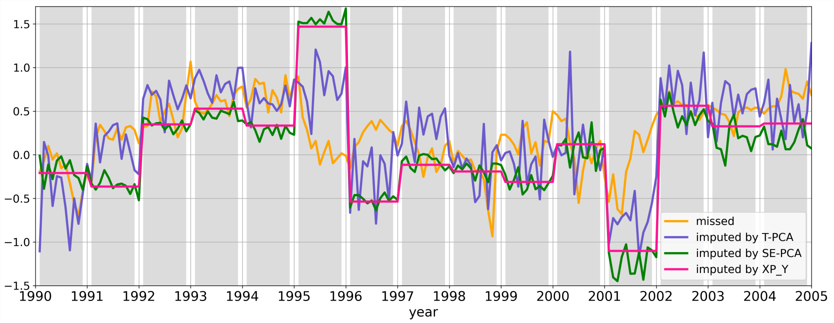


Figure 1: Real time series v.s. imputed time series of the spread between 3-month treasury and Fed Funds rate