# Large Dimensional Latent Factor Modeling with Missing Observations and Applications to Causal Inference

Ruoxuan Xiong and Markus Pelger

Stanford University

# Motivation: Publication Effect on Investment Strategies

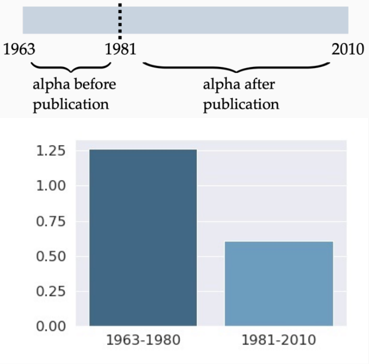
Question: Does academic publication of a strategy affect this strategy's return?

- Intuition: After publication traders exploit strategy and drive down profits
- Illustrative example (Banz 1981): Size strategy (small-minus-big portfolio)
  Smaller companies have higher average returns (published in 1981)
- Investment performance measure: Mean return in excess of a market index
  (alpha= outperformance relative to market)

# Motivation: Publication Effect on Investment Strategies

Question: Does academic publication of a strategy affect this strategy's return?

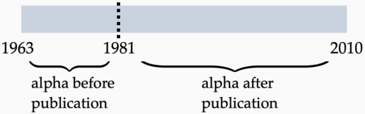
- Intuition: After publication traders exploit strategy and drive down profits
- Illustrative example (Banz 1981): Size strategy (small-minus-big portfolio)
  Smaller companies have higher average returns (published in 1981)
- Investment performance measure: Mean return in excess of a market index (alpha= outperformance relative to market)

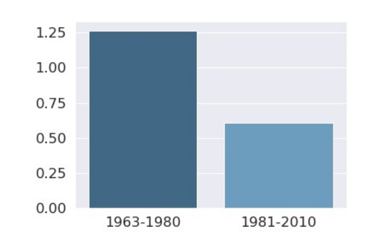


# Motivation: Publication Effect on Investment Strategies

Question: Does academic publication of a strategy affect this strategy's return?

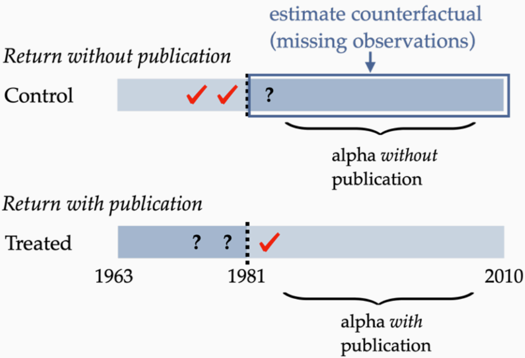
- Intuition: After publication traders exploit strategy and drive down profits
- Illustrative example (Banz 1981): Size strategy (small-minus-big portfolio)
  Smaller companies have higher average returns (published in 1981)
- Investment performance measure: Mean return in excess of a market index (alpha= outperformance relative to market)




Simple before-after analysis not appropriate!
It does not control for time-varying features.

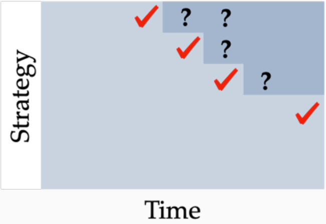
# This Paper: A Causal Inference Approach




- Experiments have identical control and treatment groups
- Fundamental problem here: Only observe treated or control outcomes
- Our approach: Model counterfactual as missing observations and impute missing values
- Counterfactual = mimicking average of untreated observations

## This Paper: New Methodology

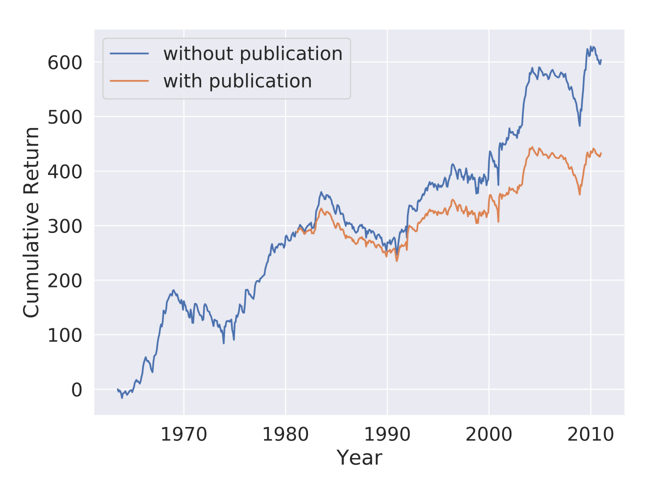
- Large-dimensional panel data: Many strategies' returns over many periods.
- Complex treatment pattern: Strategies are published at different times with different probabilities




Observational pattern for the control panel

- No pre-specified model: Use general statistical factors to impute counterfactual returns without a prior what makes strategies similar
- A general causal inference approach: Model counterfactual outcomes as missing observations to obtain entry-wise control and test individual and weighted effects

# Motivating Example: Cumulative Return of Size Strategy




- Our approach: Obtain the full time-series of counterfactual outcomes
- Advantage: Analyze more than simple mean effects

# Contribution

## Methodology

1. Easy-to-adopt method to impute missing observations on panel data with general observation patterns
2. Inferential theory for estimated factor model and each imputed value under various observation patterns
3. Tests for entry-wise and weighted treatment effects
4. Generalization of Principal Component Analysis (PCA) to incomplete panels

## Empirical results

1. Study the publication effect on strategies' returns and alphas
2. 15% of strategies exhibit significant reduction, different from and fewer than those from the naive before-after analysis

**Causal inference on panel data:**

Example: Telehealth, pricing algorithms on demand
Problem: When and where is the intervention effective?
Our solution: Tests for entry-wise and weighted treatment effects
Importance: Goes beyond mean effects without assuming prespecified covariates

## Broader Applications

**Causal inference on panel data:**

Example: Telehealth, pricing algorithms on demand
Problem: When and where is the intervention effective?
Our solution: Tests for entry-wise and weighted treatment effects
Importance: Goes beyond mean effects without assuming prespecified covariates

**Large-dimensional factor modeling**

Example: Macroeconomic data, stock returns
Problem: How to estimate a factor model from incomplete data?
Our solution: Estimator for the factor model with confidence interval
Importance: Input for other applications, for example risk factors

## Broader Applications

**Causal inference on panel data:**

Example: Telehealth, pricing algorithms on demand
Problem: When and where is the intervention effective?
Our solution: Tests for entry-wise and weighted treatment effects
Importance: Goes beyond mean effects without assuming prespecified covariates

**Large-dimensional factor modeling**

Example: Macroeconomic data, stock returns
Problem: How to estimate a factor model from incomplete data?
Our solution: Estimator for the factor model with confidence interval
Importance: Input for other applications, for example risk factors

**Missing data imputation**

Example: Financial data, users' ratings at Amazon, mixed frequency data
Problem: Whether to use imputed value?
Our solution: Estimator for each entry with confidence interval
Importance: Include observations with incomplete data instead of leaving them out for analysis which can lead to bias and efficiency loss

## Related Literature (Incomplete)

**Causal inference on panel data**

- Difference in differences: Card 90, Bertrand et al. 04, Athey and Imbens 18, Arkhangelsky 18
- Synthetic control methods:
  Abadie et al. 10, Haiao et al. 12, Abadie et al. 15, Doudchenko and Imbens 16, Li and Bell 17, Li 17, Masini and Medeiros 18, Arkhangelsky 18

**Matrix completion**

- Independently sampling: Candes and Recht 09, Candes and Plan 10, Mazumder et al 10, Negahban and Wainright 12, Klopp 14
- Dependently sampling: Athey et al. 18
- Independently sampling with inferential theory: Chen et al. 19

**Factor modeling**

- Full observations with inferential theory: Bai and Ng 02, Bai 03, Fan et al. 16, Kelly et al. 18, Pelger and Xiong 20a+b, Lettau and Pelger 20a+b
- Partial observations without inferential theory: Stock and Watson 02, Banbura and Modugno 14
- Partial observations with inferential theory: Jin et al 20, Bai and Ng 19

# Theory: Model and Estimation

## Model Setup: Approximate Latent Factor Model

Approximate factor model: Observe $Y_{it}$ for $N$ units over $T$ time periods

$$Y_{it} = \underbrace{\Lambda_i^\top}_{1\times k} \underbrace{F_t}_{k\times 1} + e_{it}$$

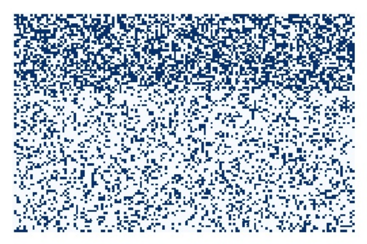
In matrix notation:

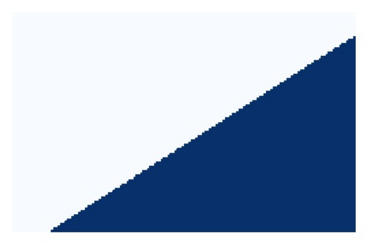
$$\underbrace{Y}_{N\times T} = \underbrace{\Lambda}_{N\times k} \underbrace{F^\top}_{k\times T} + \underbrace{e}_{N\times T}$$

- $N$ and $T$ large
- Factors $F_t$ explain common time-series movements
- Loadings $\Lambda_i$ capture correlation between units
- Factors and loadings are latent and estimated from the data
- Common component $C_{it} = \Lambda_i^\top F_t$
- Idiosyncratic errors $\mathbb{E}[e_{it}] = 0$
- Number of factors $k$ fixed

## General Observational Pattern

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





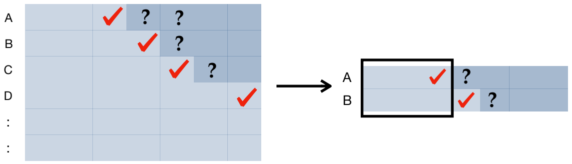
- Missing uniformly at random
  $P(W_{it} = 1) = p$
- Cross-section missing at random
  $P(W_{it} = 1) = p_t$
- Time-series missing at random
  $P(W_{it} = 1) = p_i$

- Staggered treatment adoption
  $P(W_{it} = 1) = p_{it}$
  Once missing stays missing:
  $W_{is} = 0$ for $s \geq t$
- Mixed-frequency observations
  $P(W_{it} = 1) = p_{it}$
  Equivalent to staggered design after reshuffling

# Estimation of the Factor Model

**Step 1** Estimate sample covariance matrix $\tilde{\Sigma}$ of $Y$ using only observed entries

$Q_{ij} = \{t : W_{it} = 1 \text{ and } W_{jt} = 1\}$ are times where both units are observed




$$\tilde{\Sigma}_{ij} = \frac{1}{|Q_{ij}|} \sum_{t \in Q_{ij}} Y_{it} Y_{jt}$$

**Step 2** Estimate loadings $\tilde{\Lambda}$ (standard):

Apply principal component analysis (PCA) to $\tilde{\Sigma} = \frac{1}{N} \tilde{\Lambda} \tilde{D} \tilde{\Lambda}^\top$

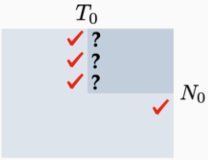
**Step 3** Estimate factors $\tilde{F}$ with regression on loadings for observed entries:

$$\tilde{F}_t = \left( \sum_{i=1}^{N} W_{it} \tilde{\Lambda}_i \tilde{\Lambda}_i^\top \right)^{-1} \left( \sum_{i=1}^{N} W_{it} \tilde{\Lambda}_i Y_{it} \right)$$

**Step 4** Estimate common components/missing entries $\tilde{C}_{it} = \tilde{\Lambda}_i^\top \tilde{F}_t$

# Illustration of Distribution Theory: A Toy Example

One factor model $X_{it} = \lambda_i F_t + e_{it}$ with




$$F_t \overset{\text{i.i.d.}}{\sim} \mathcal{N}(0,1)$$

$$\Lambda_i \overset{\text{i.i.d.}}{\sim} \mathcal{N}(0,1)$$

$$e_{it} \overset{\text{i.i.d.}}{\sim} \mathcal{N}(0,\sigma_e^2)$$

Inferential theory for $\tilde{\Lambda}_i$ for $i = 1, \cdots, N_0$,

$$\sqrt{T}\left(\tilde{\Lambda}_i - \Lambda_i\right) = \underbrace{\sqrt{\frac{T}{T_0}}\frac{1}{\sqrt{T_0}}\sum_{t=1}^{T_0} F_t e_{it}}_{\substack{\xrightarrow{d} \mathcal{N}\left(0, \frac{T}{T_0}\sigma_e^2\right) \\ \text{conventional term}}} + \underbrace{\sqrt{T}\left(\frac{1}{T_0}\sum_{t=1}^{T_0} F_t F_t^\top - \frac{1}{T}\sum_{t=1}^{T} F_t F_t^\top\right)\Lambda_i}_{\substack{\xrightarrow{d} \mathcal{N}\left(0, 2\frac{T-T_0}{T_0}\Lambda_i^2\right) \\ \text{variance correction term}}} + o_p(1),$$

- Conventional term:
  Asymptotic distribution of standard regression coefficients
- Variance correction term:
  Estimation uncertainty from using different number of observations

# Assumptions: Approximate Factor Model

## Assumption 1: Approximate Factor Model

1. Systematic factor structure: $\Sigma_F$ and $\Sigma_\Lambda$ full rank

$$\frac{1}{T}\sum_{t=1}^{T} F_t F_t^\top \xrightarrow{p} \Sigma_F \qquad \frac{1}{N}\sum_{i=1}^{N} \Lambda_i \Lambda_i^\top \xrightarrow{p} \Sigma_\Lambda$$

2. Weak dependence of errors: bounded eigenvalues of correlation and autocorrelation matrix for errors
   Simplification for presentation: $e_{it} \overset{iid}{\sim} (0, \sigma_e^2)$, $\mathbb{E}[e_{it}^8] < \infty$
3. Factors $F_t$ and errors $e_{it}$ independent
4. Uniqueness of factor rotation: Eigenvalues of $\Sigma_\Lambda \Sigma_F$ distinct
5. Bounded moments: $\mathbb{E}[\|F_t\|^4] < \infty$, $\mathbb{E}[\|\Lambda_i\|^4] < \infty$
   Simplification for presentation: $F_t \overset{\text{i.i.d.}}{\sim} (0, \Sigma_F)$, $\Lambda \overset{\text{i.i.d.}}{\sim} (0, \Sigma_\Lambda)$

- Standard assumptions on large dimensional approximate factor model
- Largest singular values estimate loadings and factors consistently up to rotation (in the case of no missing values)

## Assumptions: Observation Pattern

### Assumption 2: Observational Pattern

1. $W$ independent of $F$ and $e$ (but can depend on $\Lambda$)
2. "Sufficiently many" cross-sectional observed entries

$$\frac{1}{N}\sum_{i=1}^{N}\Lambda_i\Lambda_i^\top W_{it} \xrightarrow{p} \Sigma_{\Lambda,t} \qquad \text{full rank for all } t$$

3. "Sufficiently many" time-series observed entries

$$\frac{1}{N}\sum_{i=1}^{N}\Lambda_i\Lambda_i^\top \frac{1}{|Q_{ij}|}\sum_{t\in Q_{ij}} F_tF_t^\top \xrightarrow{p} \text{full rank matrix for all } j$$

4. "Not too many" missing entries: $q_{ij} = \lim_{T\to\infty} |Q_{ij}|/T \geq \underline{q} > 0$ and $q_{ij,kl} = \lim_{T\to\infty} \frac{|\mathcal{Q}_{ij}\cap\mathcal{Q}_{kl}|}{T}$; limits of $\frac{1}{N^2}\sum_{i=1}^N\sum_{l=1}^N \frac{q_{ij,lj}}{q_{ij}q_{lj}}$, $\frac{1}{N^3}\sum_{i=1}^N\sum_{l=1}^N\sum_{k=1}^N \frac{q_{li,kj}}{q_{li}q_{kj}}$ and $\frac{1}{N^4}\sum_{i=1}^N\sum_{l=1}^N\sum_{j=1}^N\sum_{k=1}^N \frac{q_{li,kj}}{q_{li}q_{kj}}$ exist.

# Asymptotic Results

**Theorem 1: Loadings**

Under Assumptions 1 and 2, it holds for $N, T \to \infty$ and $\sqrt{T}/N \to 0$:

$$\boxed{\sqrt{T}(H^{-1}\tilde{\Lambda}_j - \Lambda_j) \xrightarrow{d} \mathcal{N}\left(0, \omega_{jj} \cdot \Sigma_{\Lambda}^{\text{obs}} + (\omega_{jj} - 1)\Sigma_{\Lambda,j}^{\text{miss}}\right)}$$

- $H$ is a rotation matrix
- $\omega_{jj} = \lim_{N\to\infty} \frac{1}{N^2} \sum_{i=1}^{N} \sum_{l=1}^{N} \frac{q_{ij,lj}}{q_{ij}q_{lj}}$, $\omega_{jj} \geq 1$ (full observations: $\omega_{jj} = 1$)
- $\Gamma_{\Lambda}^{\text{obs}} = \Sigma_F^{-1}\sigma_e^2$ conventional covariance matrix
- $\Sigma_{\Lambda,j}^{\text{miss}} = \Sigma_F^{-1}\Sigma_{\Lambda}^{-1}(\Lambda_j^\top \otimes \Sigma_{\Lambda})\Xi_F(\Lambda_j \otimes \Sigma_{\Lambda})\Sigma_{\Lambda}^{-1}\Sigma_F^{-1}$ variance correction term
  $\Xi_F = \mathbb{E}[\mathrm{vec}(F_t F_t^\top)\mathrm{vec}(F_t F_t^\top)^\top]$

Fundamental expansion:

$$\sqrt{T}(H^{-1}\tilde{\Lambda}_j - \Lambda_j) = \left(\frac{1}{T}F^\top F\right)^{-1}\left(\frac{1}{N}\Lambda^\top\Lambda\right)^{-1}\left[\left(\frac{1}{N}\sum_{i=1}^{N}\Lambda_i\Lambda_i^\top\sqrt{\frac{T}{|Q_{ij}|}}\frac{1}{\sqrt{|Q_{ij}|}}\sum_{t\in Q_{ij}}F_t e_{jt}\right)\right.$$
$$\left. + \left(\frac{1}{N}\sum_{i=1}^{N}\Lambda_i\Lambda_i^\top\sqrt{T}\left(\frac{1}{|Q_{ij}|}\sum_{t\in Q_{ij}}F_t F_t^\top - \frac{1}{T}F^\top F\right)\right)\Lambda_j\right] + o_p(1)$$

$\Rightarrow$ Convergence rate is $\sqrt{T}$.

# Inferential Theory

## Theorem 2: Factors

Under Assumptions 1 and 2, it holds for $N, T \to \infty$ and $\sqrt{N}/T \to 0$:

$$\boxed{\sqrt{\delta}(H^\top \tilde{F}_t - F_t) \xrightarrow{d} \mathcal{N}\left(0, \frac{\delta}{N}\Sigma_{F,t}^{\text{obs}} + \frac{\delta}{T}(\omega - 1)\Sigma_{F,t}^{\text{miss}}\right)}$$

- $\delta = \min(N, T)$
- $\omega = \lim_{N\to\infty} \frac{1}{N^4} \sum_{i=1}^N \sum_{l=1}^N \sum_{j=1}^N \sum_{k=1}^N \frac{q_{li,kj}}{q_{li} q_{kj}}$ (full observations/missing uniformly at random: $\omega = 1$)
- $\Sigma_{F,t}^{\text{obs}} = \Sigma_{\Lambda,t}^{-1}\sigma_e^2$ conventional covariance matrix
- $\Sigma_{F,t}^{\text{miss}} = \Sigma_{\Lambda,t}^{-1}(I_r \otimes (F_t^\top \Sigma_F^{-1}\Sigma_\Lambda^{-1}))(\Sigma_{\Lambda,t}\otimes\Sigma_\Lambda)\Xi_F(\Sigma_{\Lambda,t}\otimes\Sigma_\Lambda)(I_r\otimes(\Sigma_\Lambda^{-1}\Sigma_F^{-1}F_t))\Sigma_{\Lambda,t}^{-1}$ variance correction term

Fundamental expansion:

$$\sqrt{\delta}(H^\top \tilde{F}_t - F_t) = \left(\frac{1}{N}\sum_{i=1}^N W_{it}\Lambda_i\Lambda_i^\top\right)^{-1}\Bigg[\left(\sqrt{\frac{\delta}{N}}\frac{1}{\sqrt{N}}\sum_{i=1}^N W_{it}\Lambda_i e_{it}\right)$$
$$+ \left(\frac{\sqrt{\delta}}{N}\sum_{i=1}^N W_{it}\left(H^{-1}\tilde{\Lambda}_i - \Lambda_i\right)\Lambda_i^\top F_t\right)\Bigg] + o_p(1)$$

$\Rightarrow$ Convergence rate is $\min\left(\sqrt{T}, \sqrt{N}\right)$.

## Inferential Theory

### Theorem 3: Common Components

Under Assumptions 1 and 2, it holds for $N, T \to \infty$:

$$\sqrt{\delta}(\tilde{C}_{jt} - C_{jt}) \xrightarrow{d} \mathcal{N}\left(0, \frac{\delta}{T}\left[F_t^\top\left(\omega_{jj} \cdot \Sigma_\Lambda^{\text{obs}} + (\omega_{jj} - 1) \cdot \Sigma_{\Lambda,j}^{\text{miss}}\right) F_t + (\omega - 1)\Lambda_j^\top \Sigma_{F,t}^{\text{miss}} \Lambda_j \right.\right.$$
$$\left.\left. - 2(\omega_j - 1) F_t^\top \Sigma_{\Lambda,F,j,t}^{\text{miss, cov}} \Lambda_j\right] + \frac{\delta}{N} \Lambda_j^\top \Sigma_{F,t}^{\text{obs}} \Lambda_j\right)$$

- $\omega_j = \lim_{N\to\infty} \frac{1}{N^3} \sum_{i=1}^N \sum_{l=1}^N \sum_{k=1}^N \frac{q_{li,kj}}{q_{li} q_{kj}}$ (full observations/missing uniformly at random: $\omega = 1$)
- $\Sigma_{\Lambda,F,j,t}^{\text{miss, cov}} = \Sigma_F^{-1}\Sigma_\Lambda^{-1}(\Lambda_j^\top \otimes \Sigma_\Lambda)\Xi_F(\Sigma_{\Lambda,t} \otimes \Sigma_\Lambda)\left(I_r \otimes (\Sigma_\Lambda^{-1}\Sigma_F^{-1} F_t)\right)\Sigma_{\Lambda,t}^{-1}$ covariance between variance correction terms of $\tilde{F}$ and $\tilde{\Lambda}$

Fundamental expansion:

$$\sqrt{\delta}\left(\tilde{C}_{it} - C_{it}\right) = \sqrt{\delta}\left(H^{-1}\tilde{\Lambda}_i - \Lambda_i\right)^\top F_t + \sqrt{\delta}\Lambda_i^\top\left(H^\top \tilde{F}_t - F_t\right) + o_p(1)$$

$\Rightarrow$ Convergence rate is $\min\left(\sqrt{T}, \sqrt{N}\right)$.

- Factor covariance matrix $\Sigma_{F,t}^{\text{obs}}$ neglected for $T/N \to 0$.
- Plug-in estimator for covariance matrices.

## Propensity-Weighted Estimator

### Assumption 3: Conditional Observational Pattern

Assume observations depend on observed, time-invariant covariates $S \in \mathbb{R}^{N \times K}$:

1. The probability of $W_{it} = 1$ depends on $S_i$ and $P(W_{it} = 1|S_i) > 0$.
2. Conditional cross-sectional independence: $W$ independent of $\Lambda$ conditional on $S$.
3. $W_{it}$ is independent of $W_{js}$ conditional on $S_i, S_j$.

- Conditional on $S_i$, $W_{it}$ is i.i.d.
- $S_i$ can actually include $\Lambda_i$
- Motivates "propensity score" estimator: Re-weight entries to obtain "missing-at-random"

Alternative estimator for loadings and common components:

$$\tilde{F}_t^S = \left( \sum_{i=1}^{N} \frac{W_{it}}{P(W_{it} = 1|S_i)} \tilde{\Lambda}_i \tilde{\Lambda}_i^\top \right)^{-1} \left( \sum_{i=1}^{N} \frac{W_{it}}{P(W_{it} = 1|S_i)} Y_{it} \tilde{\Lambda}_i \right)$$

- $\tilde{F}^S = \tilde{F}$ for cross-section missing at random: $P(W_{it} = 1|S_i)$ is the same for all $i$

## Inferential Theory for Propensity-Weighted Estimator

### Theorem 4: Propensity-Weighted Factors

Under Assumptions 1 and 3 it holds for $N, T \to \infty$:

$$\boxed{\sqrt{\delta}(H^\top \tilde{F}_t^S - F_t) \xrightarrow{d} \mathcal{N}\left(0, \frac{\delta}{N}\Sigma_F^{\text{obs},S} + \frac{\delta}{T}(\omega - 1)\Sigma_{F,t}^{\text{miss},S}\right)}$$

- $\Sigma_{F,t}^{\text{obs},S} = \Sigma_\Lambda^{-1}\Sigma_{\Lambda,S,t}\Sigma_\Lambda^{-1}\sigma_e^2$, where $\Sigma_{\Lambda,S,t} = \lim_{N\to\infty} \frac{1}{N}\sum_{i=1}^N \frac{1}{P(W_{it}=1|S_i)}\mathbb{E}[\Lambda_i\Lambda_i^\top|S_i]$ (conditional weighted second moment)
- $\Sigma_{F,t}^{\text{miss},S} =$ $\Sigma_\Lambda^{-1}((F_t^\top\Sigma_F^{-1}\Sigma_\Lambda^{-1}) \otimes I_r)(\Sigma_\Lambda \otimes \Sigma_\Lambda)\Xi_F(\Sigma_\Lambda \otimes \Sigma_\Lambda)((\Sigma_\Lambda^{-1}\Sigma_F^{-1}F_t) \otimes I_r)\Sigma_\Lambda^{-1}$ variance correction term (independent of $S$)

- Propensity-weighted estimator uses more information.
- If $S$ is independent of $\Lambda$, $\tilde{F}^S$ is less efficient than $\tilde{F}$ (concavity of $\frac{1}{P(W_{it}=1|S_i)}$).
- In one-factor case, for any $S$, $\tilde{F}^S$ is less efficient than $\tilde{F}$ ($\Sigma_{\Lambda,S,t}/(\Sigma_\Lambda)^2 \geq 1/\Sigma_{\Lambda,t}$ always holds from Cauchy-Schwartz Inequality).
- In the presence of omitted factors, $\tilde{F}^S$ can be more efficient.

# Inferential Theory for Propensity-Weighted Estimator

**Theorem 5: Propensity-Weighted Common Components**

Under Assumptions 1 and 3 it holds for $N, T \to \infty$:

$$\sqrt{\delta}(\tilde{C}^S_{jt} - C_{jt}) \xrightarrow{d} \mathcal{N}\Bigg(0, \frac{\delta}{T}\Big[F_t^\top \left(\omega_{jj} \cdot \Sigma_\Lambda^{\text{obs}} + (\omega_{jj} - 1) \cdot \Sigma_{\Lambda,j}^{\text{miss}}\right) F_t + (\omega - 1) \cdot \Lambda_j^\top \Sigma_{F,t}^{\text{miss},S} \Lambda_j$$
$$- 2(\omega_j - 1) F_t^\top \Sigma_{\Lambda,F,j,t}^{\text{miss, } S, \text{ cov}} \Lambda_j\Big] + \frac{\delta}{N} \Lambda_j^\top \Sigma_{F,t}^{\text{obs},S} \Lambda_j\Bigg)$$

- $\Sigma_{\Lambda,F,j,t}^{\text{miss, } S, \text{ cov}} = \Sigma_F^{-1} \Sigma_\Lambda^{-1} (\Lambda_j \otimes I_r)(\Sigma_\Lambda \otimes \Sigma_\Lambda) \Xi_F((\Sigma_\Lambda^{-1} \Sigma_F^{-1} F_t) \otimes \Sigma_\Lambda) \Sigma_\Lambda^{-1}$ covariance between variance correction terms of $\tilde{F}^S$ and $\tilde{\Lambda}$ (independent of $S$)

- The estimated probability weight can lead to additional correction terms in the asymptotic variance.

# Examples of Feasible Estimators of the Probability Weight

| Description | $P(W_{it}=1\|S_i)$ | Estimator | Asymptotic distribution of $\hat{p}_{it}-p_{it}$ | Variance correction |
|---|---|---|---|---|
| Missing at random | $p$ | $\hat{p}=\frac{1}{NT}\sum_{i,t}W_{it}$ | $O_p\left(\frac{1}{\sqrt{NT}}\right)$ | no |
| Cross-section missing at random | $p_t$ | $\hat{p}_t=\frac{\|O_t\|}{N}$ | $\frac{1}{\sqrt{N}}\mathcal{N}\left(0,p_t(1-p_t)\right)$ | no |
| Time-series missing at random (parametric) | $p(S_i)$ | logit on full panel $W$ | $O_p\left(\frac{1}{\sqrt{NT}}\right)$ | no |
| Time-series missing at random (non-parametric) | $p(S_i)$ | kernel on full panel $W$ | $O_p\left(\frac{1}{\sqrt{NTh}}\right)$ | yes/no |
| Cross-section and time-series dependency (parametric) | $p_t(S_i)$ | logit on $W_t$ | $O_p\left(\frac{1}{\sqrt{N}}\right)$ | yes |
| Cross-section and time-series dependency (discrete $S$) | $p_t(s)$ | $\hat{p}_t(s)=\frac{\|\mathcal{O}_{s,t}\|}{N_s}$ | $\frac{1}{\sqrt{N_s}}\mathcal{N}\left(0,p_t(s)(1-p_t(s))\right)$ | yes |
| Staggered treatment adoption without $S$ | $p_t$ | $\hat{p}_t=\frac{\|O_t\|}{N}$ | $\frac{1}{\sqrt{N}}\mathcal{N}\left(0,p_t(1-p_t)\right)$ | no |
| Staggered treatment adoption with $S$ (parametric) | $p_t(S_i)$ | hazard rate model | $O_p\left(\frac{1}{\sqrt{N}}\right)$ | yes |
| Mixed frequencies | $p_t$ | $\hat{p}_t=\frac{\|O_t\|}{N}$ | $\frac{1}{\sqrt{N}}\mathcal{N}\left(0,p_t(1-p_t)\right)$ | no |

# Tests for Causal Effects

Treatment effect for staggered design with $T_{0,i}$ control and $T_{1,i}$ treated

$$Y_{it}^{(\omega)} = \underbrace{\Lambda_i^{(\theta)\top} F_t^{(\theta)}}_{C_{it}^{(\theta)}} + e_{it}^{(\theta)}, \quad \theta = \begin{cases} 1 & \text{treated (missing)} \\ 0 & \text{control (observed)} \end{cases}$$

We consider two different effects:

1. Individual treatment effect: $\tau_{it} = C_{it}^{(1)} - C_{it}^{(0)}$
2. Weighted average treatment effect: $\tau_{\beta,i} = (Z^\top Z)^{-1} Z^\top \tau_{i,(T_{0,i}+1):T}$

Inferential theory of $\tilde{C}_{it}$ provides the test statistics.
Two sets of results: treatment changes only loadings or changes factor and loadings.

In empirical applications, we test $\mathcal{H}_0 : \tau_{\beta,i} \geq 0$ vs $\mathcal{H}_1 : \tau_{\beta,i} < 0$.

# Simulation

## Simulation Design

Comparison between the four methods that provide inferential theory

1. $\mathrm{XP}_{\mathrm{SIM}}$: Our simple method $\tilde{C}$
2. $\mathrm{XP}_{\mathrm{COND}}$ Our propensity-weighted method $\tilde{C}^S$
3. JMS (Jin, Miao and Su (2020)): Assuming missing at random
4. BN (Bai and Ng (2020)): Combined block PCA

We compare the relative MSE $\sum_{i,t}(\tilde{C}_{it} - C_{it})^2 / \sum_{i,t} C_{it}^2$

- The data generating process is $X_{it} = \Lambda_i^\top F_t + e_{it}$
- 2 factors
- $\Lambda_i \overset{\text{i.i.d.}}{\sim} \mathcal{N}(0, I_2)$, $F_t \overset{\text{i.i.d.}}{\sim} \mathcal{N}(0, I_2)$ and $e_{it} \overset{\text{i.i.d.}}{\sim} \mathcal{N}(0, 1)$

$\Rightarrow$ Our method allows for the most general observation pattern

$\Rightarrow$ Out method provides the most efficient estimation

# Simulation $N = 100$, $T = 150$

| Observation Pattern | $W_{it}$ | $\text{XP}_{\text{SIM}}$ | $\text{XP}_{\text{COND}}$ | JMS | BN |
|---|---|---|---|---|---|
| Random | obs | 0.03 | 0.03 | 0.05 | 215.31 |
| | miss | 0.03 | 0.03 | 0.04 | 224.14 |
| | all | 0.03 | 0.03 | 0.05 | 217.46 |
| Simultaneous | obs | 0.03 | 0.03 | 0.14 | 0.03 |
| | miss | 0.04 | 0.04 | 0.20 | 0.04 |
| | all | 0.03 | 0.03 | 0.16 | 0.03 |
| Staggered | obs | 0.03 | 0.03 | 0.39 | 0.15 |
| | miss | 0.08 | 0.08 | 0.33 | 0.19 |
| | all | 0.05 | 0.05 | 0.37 | 0.17 |
| Random | obs | 0.04 | 0.04 | 0.09 | 1714.10 |
| $W$ depends on $S$ | miss | 0.04 | 0.04 | 0.10 | 1538.62 |
| | all | 0.04 | 0.04 | 0.09 | 1650.06 |
| Simultaneous | obs | 0.04 | 0.04 | 0.63 | 0.09 |
| $W$ depends on $S$ | miss | 0.11 | 0.14 | 0.41 | 0.13 |
| | all | 0.07 | 0.09 | 0.53 | 0.11 |

$\Rightarrow$ XP is the most precise

# Simulation Omitted Factor and $N = 1000$, $T = 50$

| Observation Pattern | $W_{it}$ | $\text{XP}_{\text{SIM}}$ | $\text{XP}_{\text{COND}}$ | JMS | BN |
|---|---|---|---|---|---|
| Random | obs | 0.028 | 0.028 | 0.046 | 215.305 |
| | miss | 0.031 | 0.031 | 0.044 | 224.141 |
| | all | 0.029 | 0.029 | 0.045 | 217.456 |
| Simultaneous | obs | 0.025 | 0.025 | 0.143 | 0.026 |
| | miss | 0.043 | 0.043 | 0.200 | 0.038 |
| | all | 0.030 | 0.030 | 0.158 | 0.029 |
| Staggered | obs | 0.033 | 0.033 | 0.388 | 0.152 |
| | miss | 0.084 | 0.084 | 0.333 | 0.188 |
| | all | 0.053 | 0.053 | 0.367 | 0.166 |
| Random | obs | 0.035 | 0.035 | 0.089 | 1714.098 |
| $W$ depends on $S$ | miss | 0.043 | 0.044 | 0.095 | 1538.617 |
| | all | 0.038 | 0.038 | 0.091 | 1650.059 |
| Simultaneous | obs | 0.038 | 0.045 | 0.629 | 0.090 |
| $W$ depends on $S$ | miss | 0.111 | 0.144 | 0.413 | 0.129 |
| | all | 0.071 | 0.090 | 0.532 | 0.108 |
| Simultaneous | obs | 0.259 | 0.288 | 0.777 | 0.476 |
| $W$ depends on $S$ | miss | 0.558 | 0.422 | 0.673 | 0.606 |
| (Omitted one factor) | all | 0.390 | 0.346 | 0.730 | 0.532 |
| Random | obs | 0.035 | 0.035 | 0.068 | 0.987 |
| $W$ depends on $S$ | miss | 0.050 | 0.049 | 0.072 | 1.000 |
| $N \gg T$ | all | 0.040 | 0.039 | 0.069 | 0.991 |

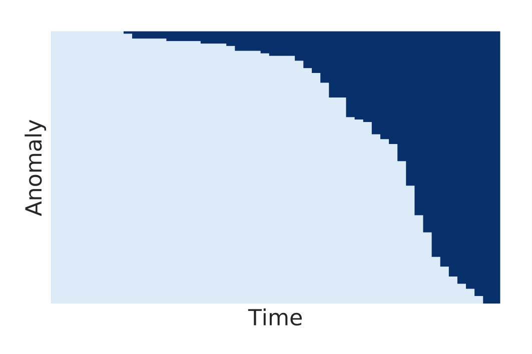
$\Rightarrow \text{XP}_{\text{COND}}$ can be the most precise for omitted factor, $N \gg T$ and $W$ depending on $S$

# Empirical Results

# Data Set and Observation Pattern

Data (Chen and Zimmermann, 2018): monthly returns of investment strategies from July 1963 to December 2013

Test publication effect on 100 strategies




Light: before publication; dark: after publication

Use 10-factor model as benchmark: Result robust to number of factors

## Testing the Publication Effect

Assumptions:

1. Underlying (latent) risk factors are not affected by publication.
2. Exposure to (latent) risk factors can be affected by publication.

$$C_{it}^{(\theta)} = (\lambda_i^{(\theta)})^\top F_t$$

Questions:

1. Does publication reduce average returns of anomalies?

$$\mathcal{H}_0 : \bar{C}_i^{(1)} - \bar{C}_i^{(0)} \geq 0,$$

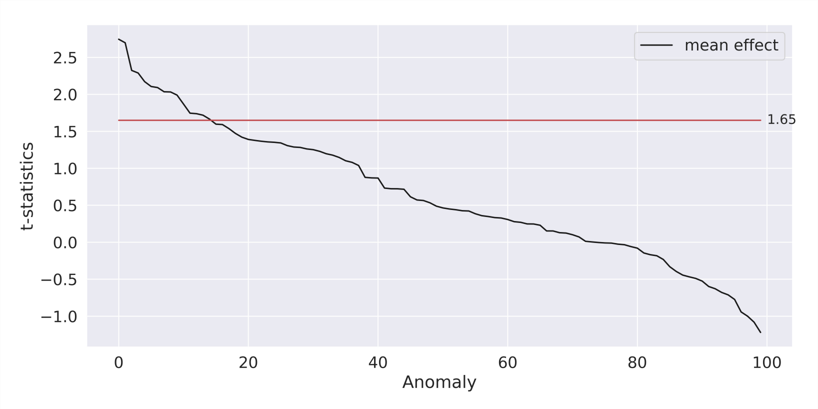
   where $\bar{C}_i^{(\theta)}$ is the mean of $C_{it}^{(\theta)}$ after publication
2. Does publication decrease the market pricing errors (alphas)?

$$\mathcal{H}_0 : \alpha_i^{(1)} - \alpha_i^{(0)} \geq 0,$$

   where $C_{it}^{(\theta)} = \alpha_i^{(\theta)} + \beta_i^{(\theta)}(R_{mt} - R_f) + \epsilon_{it}^{(\theta)}$, $R_{mt}$ is the market return and $R_f$ is the risk-free rate

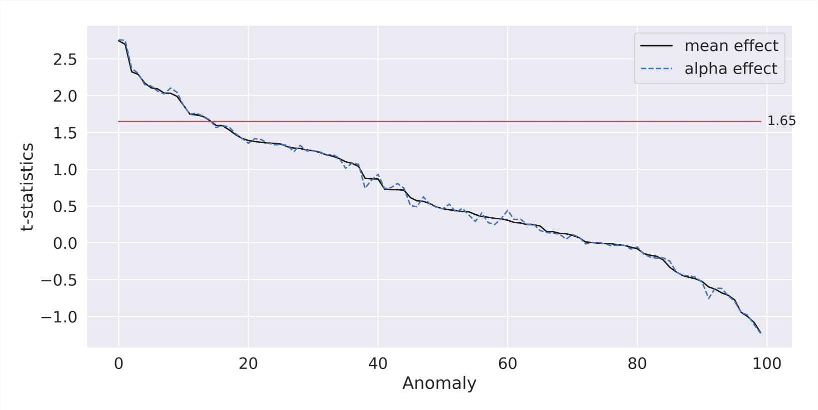
$\Rightarrow$ Test statistics for both questions use CLT for weighted treatment effect

# t-value of Publication Effect on Mean Returns



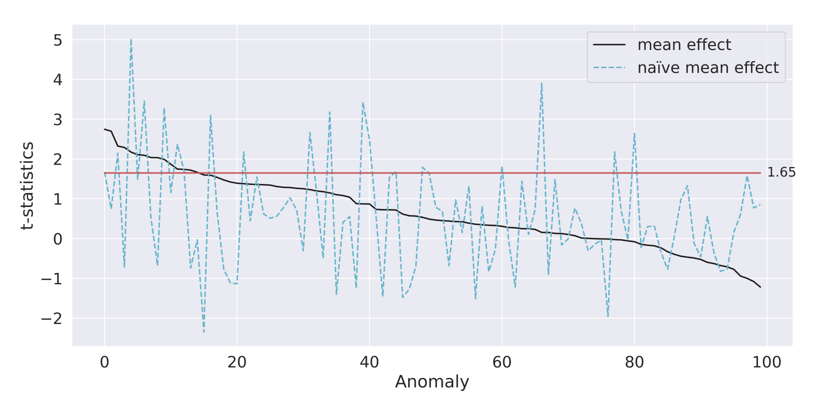

- 15% of strategies exhibit significant reduction at 95% confidence level
- Multiple hypothesis testing issue: 15% is an upper bound.
  Bonferroni correction is too conservative!

# t-value of Publication Effect on Alphas (Outperformance of Market)




- Almost identical results as with mean returns
- Long-short anomaly portfolios are constructed to be "market neutral." Most of their mean returns should not be explained by a market portfolio

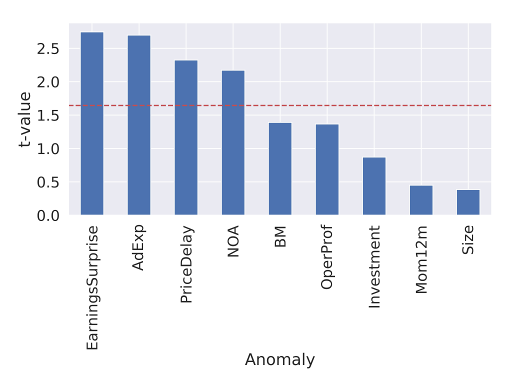
## Comparison with Before-After Analysis




Different results with simple before-after analysis

⇒ Time effects and correct estimation uncertainties are important!

# t-values of Selected Anomalies




*Significant*:

Earning Surprise

Advertising Expense (AdExp)

Price Delay

Net Operating Assets (NOA)

*Insignificant*:

Book to Market (BM)

Operating Profits (OperProf)

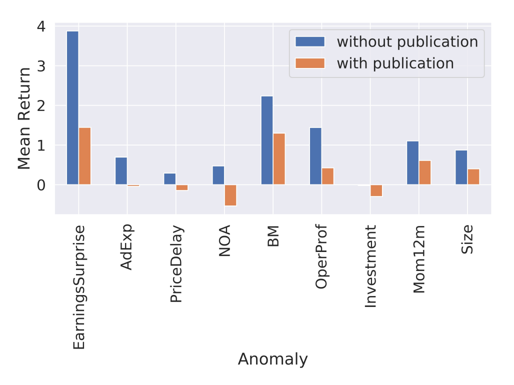
Investment

Momentum (Mon12m)

Size

- Most well-known strategies have statistical insignificant publication effect
- These anomalies represent systematic risk with constant exposure to the risk factors

# Economic Magnitude of Effect for Selected Anomalies




*Significant*:

Earning Surprise

Advertising Expense (AdExp)

Price Delay

Net Operating Assets (NOA)

*Insignificant*:

Book to Market (BM)

Operating Profits (OperProf)

Investment

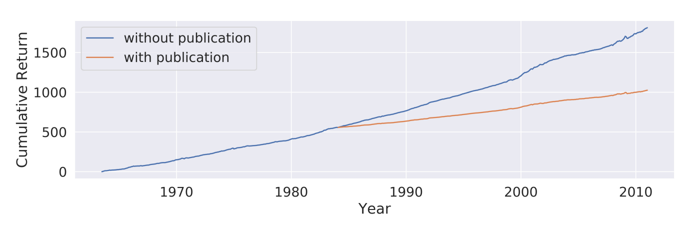
Momentum (Mon12m)

Size

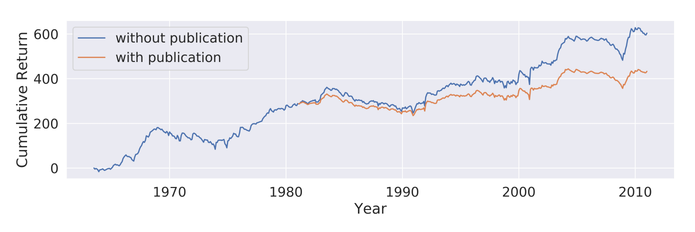
- An economically large difference does not imply statistical significance
- Estimation uncertainty matters!

# Cumulative Returns

Significance also depends on variance of returns




Earning Surprise: *Significant*




Size: *Insignificant*

# Conclusion

# Conclusion

Methodology

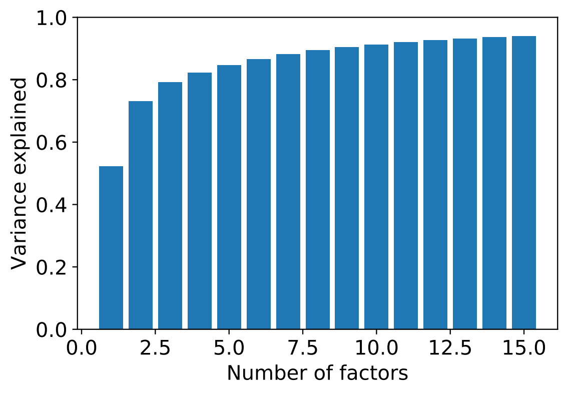
- Easy-to-adopt method to impute missing observations that is broadly applicable
- Confidence interval for each estimated entry under general and nonuniform observation patterns
- General tests for entry-wise and weighted treatment effects
- Generalizes conventional causal inference techniques to large panels and without assumptions on covariates

Empirical results

- 15% of strategies exhibit statistical significant reduction in average returns and outperformance of market
- Weaker publication effect than naive before-after analysis
- Well-known strategies have no significant publication effect $\Rightarrow$ consistent with compensation for systematic risk

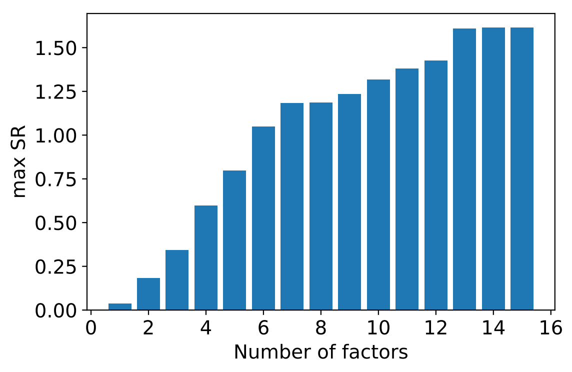
# Appendix

# How Many Latent Factors? Variance Explained



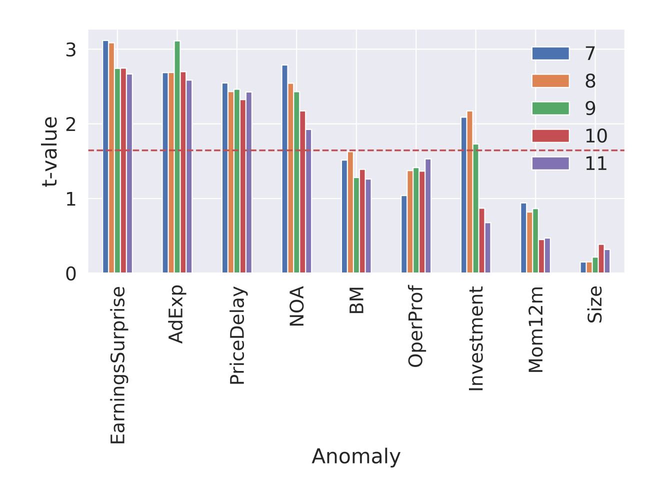

- Variance explained for different number of latent factors
- Benchmark: 10-factor model. Results are robust to using 5 to 10 factors

# Max Sharpe Ratio




Marginal increase in Sharpe ratio is small from 7 factors.

## Some Strategies' t-value (Different Number of Factors)




Results are generally robust to the choice of number of factors