時間序列分析中的因果推斷：
概述、學界和業界概況、發展策略（個人觀點）

Causal Inference in Time Series Analysis:
Overview, Academic and Industry Landscape, Development Strategies (Personal Perspective)

presented by Steven Ho (何思賢)
2024/05/27

VIZURO
Turn AI into Actions

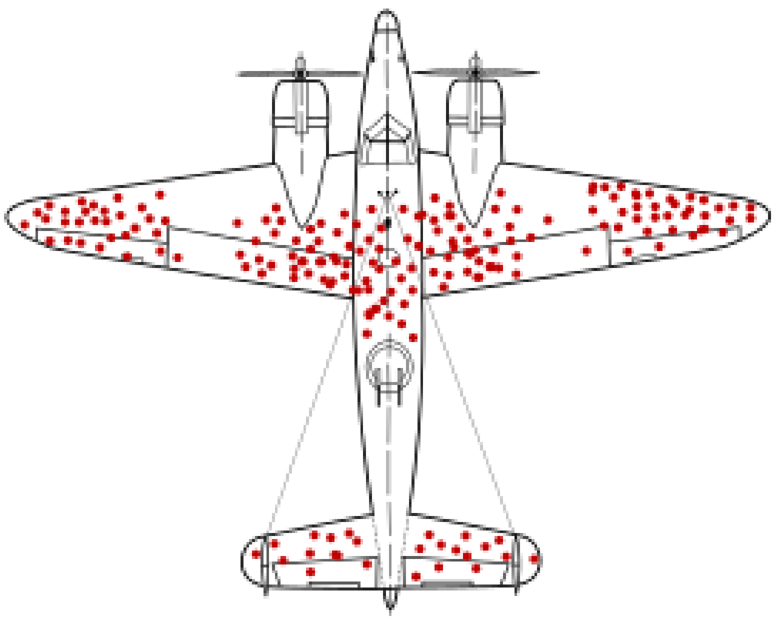
Outline

- Why Causal Inference
- Causal Inference in Academia
- Causal Inference in Industry
- Case Study: causaLens
- Vizuro’s Strategy for Developing Causal AI
- Q&A

VIZURO
Turn AI into Actions

Why Causal Inference

- Everyone knows, “Correlation does not imply causation.”
- Without a **model**, it's easy to make erroneous judgments. Example: survivorship bias.



VIZURO
Turn AI into Actions

Why Causal Inference (Cont'd)

In traditional econometrics:

- central theme:
  understanding the causes of bias in estimated coefficients
- various methods to obtain (asymptotically) unbiased-estimators for coefficients

In traditional machine learning:

- We essentially do not see these discussions how to obtain unbiased-estimators.

Why?

VIZURO
Turn AI into Actions

Why Causal Inference (Cont'd)

My personal view:

- If the data is *sufficient* and *representative*, and no policy decisions are involved, but only *predictions* based on the data, causal analysis may not be crucial.

- predictions vs explanation (+ policy recommendations): When it comes to *explanation* and *policy recommendations*, causal analysis is essential.
- reduce reliance on data representativeness and lower costs: Additionally, when data is *insufficient* or *unrepresentative*, a correct causal model can save substantial costs from erroneous inferences.

VIZURO
Turn AI into Actions

Why Causal Inference: An Example

A numerical example:

- True Model:

$$y = \beta_1 x_1 + \beta_2 x_2 + \epsilon, \quad (\beta_1, \beta_2) = (0.9, 0.1).$$

- If multicollinearity occurs ($\mathrm{Corr}(x_1, x_2) \simeq 1$), the estimated coefficients by OLS could be highly inaccurate.

e.g. (R-code)

```
set.seed(121)
mu <- c(0, 0)
Sigma <- matrix(c(1, 0.98, 0.98, 1), ncol=2)
data <- mvrnorm(n, mu, Sigma)
x1 <- data[, 1]
x2 <- data[, 2]
epsilon <- rnorm(n, mean = 0, sd = 1)
y <- 0.9 * x1 + 0.1 * x2 + epsilon
model <- lm(y ~ x1 + x2)
summary(model)
```

VIZURO
Turn AI into Actions

Why Causal Inference: An Example (Cont'd)

A numerical example (cont'd):

- `Result:`

```
 Coefficients:
            Estimate Std. Error t value Pr(>|t|)
(Intercept)  0.03597    0.03239   1.111   0.2670
x1           1.38972    0.16651   8.346 2.33e-16 ***
x2          -0.35426    0.16692  -2.122   0.0341 *
```

- The estimators are biased, leading to errors in *explanation*.
- From a *prediction* standpoint, these biased estimators are acceptable because $\hat{\beta}_1 + \hat{\beta}_2 \simeq 1$. Given the high correlation (0.98) between x_1 and x_2, $\hat{y}$ closely matches actual y.

VIZURO
Turn AI into Actions

Causal Inference in Academia: Overview

1980s: Foundations and Early Developments

- Key Figures:
- Donald Rubin, Paul Rosenbaum
- Methods:
  Potential Outcomes Framework, Propensity Score Matching
- Characteristics:
  - Emphasis on **randomized controlled trials** (RCTs)
  - Introduction of propensity score matching to handle confounding in observational studies (ar5iv) (Harvard Scholars)

VIZURO
Turn AI into Actions

Causal Inference in Academia: Overview (Cont'd)

1980-2000s: Expansion and Formalization

- Key Figures: **Judea Pearl**
- Methods: **Causal Diagrams, Structural Causal Models (SCM)**
- Characteristics:
  - Use of **graphical models** to represent causal relationships
  - Development of **do-calculus** for formal causal inference
  - Increased focus on **counterfactual reasoning** and **interventions** (ar5iv)

VIZURO
Turn AI into Actions

Causal Inference in Academia: Overview (Cont'd)

1990-2000s: Applied Methods and Policy Impact

- Key Figures: **James Heckman, Guido Imbens**
- Methods: **Instrumental Variables, Natural Experiments, Difference-in-Differences** (mainly in Economics)
- Characteristics:
  - Addressing **endogeneity** and **selection bias**
  - Emphasis on **quasi-experimental designs** for policy evaluation and impact assessment
  - Expansion of econometric techniques for causal inference (ar5iv) (Harvard Scholars)

VIZURO
Turn AI into Actions

Causal Inference in Academia: Overview (Cont'd)

2010s: **Integration with Big Data and Machine Learning**

- Key Figures: **Susan Athey, Stefan Wager**
- Methods: **Causal Trees, Double Machine Learning, Synthetic Control Methods**
- Characteristics:
  - Integration of ML techniques to **enhance causal inference accuracy**
  - Development of scalable methods for high-dimensional data, such as **causal forests** and **ensemble learning**
  - Emphasis on **robustness** and **generalizability** of causal estimates (Oxford University Press) (ar5iv)

VIZURO
Turn AI into Actions

Causal Inference in Academia: Overview (Cont'd)

2020s: **Current Trends and Future Directions**

- Methods: **Deep Learning for Causal Inference, Causal Discovery Algorithms**
- Characteristics:
  - Cross-disciplinary approaches integrating AI and causal inference
  - Advanced techniques for causal discovery in complex systems
  - Focus on improving **transparency** and **interpretability** of causal models

VIZURO
Turn AI into Actions

Causal Inference in Academia: GitHub Projects

Institution	Project Name	Description
PyWhy (Carnegie Mellon University)	causal-learn	A Python package for causal discovery implementing various causal discovery algorithms. **(Discovery)**
University of Hamburg	DoubleML	An R and Python package for double machine learning for causal inference. **(Inference)**
Stanford University	GRF	A package for generalized random forests for estimating heterogeneous treatment effects. **(Inference)**

VIZURO
Turn AI into Actions

Causal Inference in Industry: Overview (2010s to now)

- **Marketing and Customer Analytics**:
  Analysis of the impact of marketing strategies on customer behavior and sales
- **Healthcare**:
  Analysis of treatment effects and healthcare policies
- **Finance**:
  Risk assessment and investment strategy optimization
- **Real-Time Personalization:**
  Personalized recommendations and advertisements in real-time
- **Supply Chain Optimization:**
  Improving supply chain efficiency by understanding the causal impact of various factors on logistics and inventory management

VIZURO
Turn AI into Actions

Causal Inference in Industry: GitHub Projects

Company	Project Name	Description
Google	CausalImpact	An R package for causal inference using Bayesian structural time-series models. **(Inference)**
	TFP CausalImpact	Python version of above **(Inference)**
Uber	CausalML	A Python package for uplift modeling and causal inference with ML. **(Inference)**
Microsoft	DoWhy	A Python library supports explicit modeling and validation of causal assumptions, combining causal graphical models and the potential outcomes framework. **(Discovery)**
	EconML	A Python package for interpretable machine learning methods to estimate heterogeneous treatment effects. **(Inference)**

VIZURO
Turn AI into Actions

Case Study: causaLens
(leading company specializing in causal AI)

- **Causal AI Platform:**
  designed to help enterprises optimize decision-making by using causal inference techniques
- **Key Features**:
  - **Comprehensive Causal Inference Tools**:
    providing a wide range of tools for identifying and quantifying causal relationships
  - **Automated Causal Discovery**:
    offering automated tools that help users uncover causal structures from their data without manual intervention
  - **Real-Time Causal Analysis**:
    supporting real-time causal analysis, allowing for dynamic decision-making and immediate feedback
  - **Industry Applications**:
    versatile and applicable across various industries, including finance, healthcare, manufacturing, and retail

VIZURO
Turn AI into Actions

Vizuro's Strategy for Developing Causal AI

(i) Implement the functions outlined in the causaLens product roadmap (2024).

(ii) Our technology is temporarily behind, so we need to establish an advantage through our customers.

- This might be related to the characteristics of the customer datasets.
- Continue with (i) and integrate customers into our platform, offering after-sales services to enhance user retention.

Vizuro's Strategy for Developing Causal AI (Cont'd)

(iii) Market Forecast:

- I believe the causal AI field is similar to the markets for mathematical or statistical softwares, where multiple companies can coexist.
- While R can perform functions of multiple statistical software packages, SAS, SPSS, and STATA all have their markets.
- Causal AI is far more complex than these statistical software packages, leading to greater reliance on bundled services.
- To build loyal customers, a robust platform is essential.

VIZURO
Turn AI into Actions

Vizuro’s Strategy for Developing Causal AI (Cont’d)

(iv) Generative AI: Threats, Opportunities, and Challenges

- Vizuro must focus on the latest developments and choose promising technologies to invest in.
- Companies like H2O and causaLens are working to integrate LLMs into their products.
- There is no need to worry about the sudden emergence of large-scale causal AI, as that would signify a level of general artificial intelligence (AGI) that would bring societal changes beyond our concerns.

VIZURO
Turn AI into Actions

Q&A Session

- Questions welcome
- Thank you for listening