ADIA ¦ Lab

Why Has Factor Investing Failed?

Marcos López de Prado
Vincent Zoonekynd

Performance of Factor Investing

- Bloomberg – Goldman Sachs Asset Management US Equity Multi-Factor Index (BBG code: BGSUSEMF <Index>)
  - It tracks the long/short performance of the momentum, value, quality, and low-risk factors in U.S. stocks
  - Annualized Sharpe ratio: 0.22 (t-stat=0.90, *p*-value=0.19)
  - Average annualized excess return has been 0.82%
- This performance does not include:
  - transaction costs
  - market impact of order execution
  - cost of borrowing stocks for shorting positions
  - management and incentive fees

After 17 years of out-of-sample performance, factor investing's Sharpe ratio is statistically insignificant at any reasonable confidence level.

Seminar's Objective

- *p*-hacking is a well-understood source of false positives in investing
- A far less understood source of false positives is model specification
- The econometric canon favors models that over-control for colliders, because these models
  - always exhibit higher R-squared; and
  - exhibit, under general conditions, higher adjusted R-squared and lower *p*-values
- Factor strategies that over-control for colliders can yield systematic losses, even if all correlations remain constant and the risk premia are estimated with the correct sign
- **Specification errors explain the erratic performance of factor investing strategies**
- To read the full manuscript, visit: http://ssrn.com/abstract_id= 4697929

My Background & Experience

Science	Industry	Innovations	Publications	Govt Policy

Science

BERKELEY LAB

ADIA Lab

CORNELL UNIVERSITY FOUNDED A.D. 1865

RCC.HARVARD.EDU Real Colegio Complutense RCC

Industry

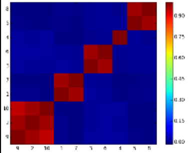
ADIA
جهاز أبوظبي للاستثمار
Abu Dhabi Investment Authority

GUGGENHEIM

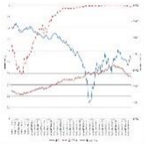
CITADEL

UBS

Innovations

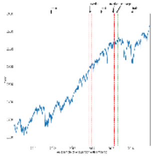


HRP|NCO

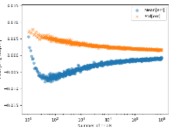


VPIN|OEH

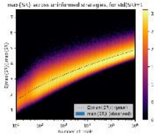
Corrective AI



CPCV

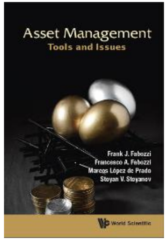


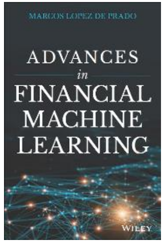
False Strat Theorem

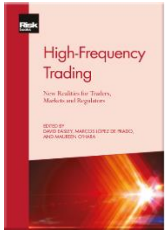


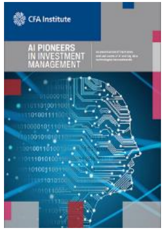
PSR|DSR

Publications









Govt Policy

UNITED STATES CONGRESS

COMMODITY FUTURES TRADING COMMISSION ★ 1975 ★

CONFIDENTIAL

Sovereign Wealth Funds (SWFs)

- SWFs are public organizations that manage
  - funds derived from a country's excess revenue;
  - set aside for global investment opportunities;
  - for the benefit of the country's citizens
- Objectives:
  - **Stabilization**: Shield public finances (buffer)
  - **Revenue**: Reduce or eliminate the need for taxes
  - **Savings**: Save for future generations
  - **Growth**: Maximize capital growth
- SWFs are influential global actors
  - Promote global financial stability and liquidity
  - Help develop economies through equity investments and the purchase of debt (without crowding-out)
  - Are a major instrument of economic diplomacy

Distribution of SWFs based on region of home country

SWF	Number of SWFs	AUM of SWFs* ($ Billion)
MiddleEast	13	2669.20 (41.46%)
Asia	16	2294.14 (35.63%)
Europe	10	1161.53 (18.04%)
North America	10	159.60 (2.48%)
Australia and New Zealand	3	111.39 (1.73%)
Africa	12	25.98 (0.40%)
Latin America and Caribbean	5	16.64 (0.26%)
Grand Total	**69**	**6438.48** (100.0%)**

*AUM as reported on SWFI website in Jan-2021

The IMF, the World Bank and SWFs play distinct but complementary roles that promote economic development and international stability. Thanks to their active participation in global markets, SWFs can act faster than other public organizations (e.g., during the GFC and COVID crises).

Factor Investing vs. Forecasting Strategies

Factor Investing

- Factor strategies decide whether a security i is bought or sold based on the value of $\hat{\beta}X_{i,t}$ at time t, where
  - $X_{i,t}$ measures i's exposure at time t to a risk characteristic
  - $\hat{\beta}$ is the estimated linear premium, as determined by a cross-sectional factor model with $|\hat{\beta}| \gg 0$
- The goal of the analysis is to establish whether the market rewards on average holding securities with exposure X (a common non-diversifiable rewarded risk)
- Estimating the risk premium β without bias requires that the factor model is correctly specified

ELSEVIER

Journal of Financial Economics
Volume 33, Issue 1, February 1993, Pages 3-56

Common risk factors in the returns on stocks and bonds ☆

Article preview

Eugene F. Fama , Kenneth R. French

Abstract

References (32)

Show more

Cited by (15586)

Add to Mendeley   Share   Cite

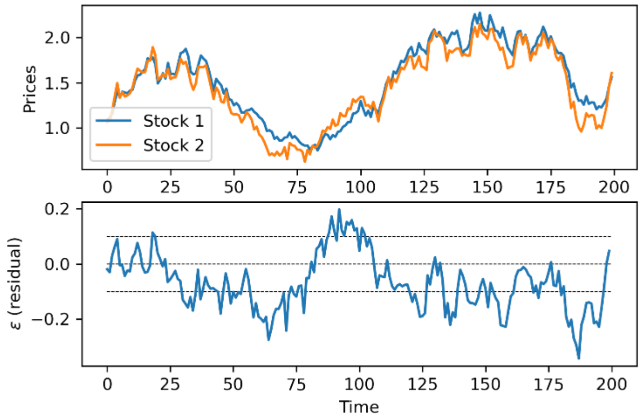
https://doi.org/10.1016/0304-405X(93)90023-5

Get rights and content

The Fama-French 3-factor (1993) and 5-factor (2015) models are among the most influential in finance, and the foundation for the USD 3T+ factor investing industry.

Forecasting Strategies

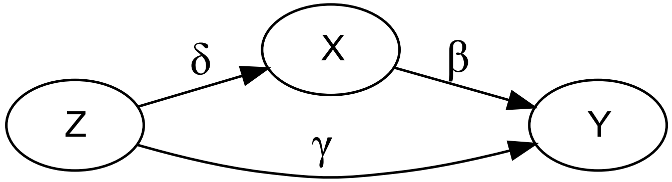
- Forecasting strategies decide whether a security i is bought or sold based on the forecasted return of i at time t
- For instance, a pairs trading strategy (a type of statistical arbitrage)
  - may buy or sell the spread i between two cointegrated securities, where $\hat{\varepsilon}_{i,t}$ measures the value of the residual of an error-correction model, a type of time series analysis
  - profits from the forecast that $\hat{\varepsilon}_{i,t}$ will converge to zero over time
- The purpose of the model is <u>not</u> to attribute a reward to a risk characteristic
  - A forecasting model can be misspecified, and still be profitable

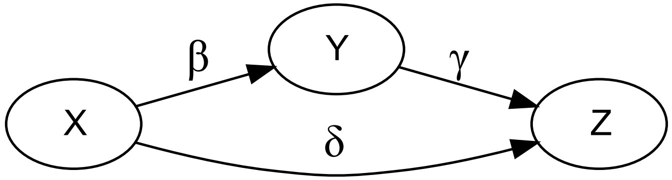


In a pairs trading model, the investor profits from the mean-reversion of ε, even if the model is misspecified.

The Assumption of Correct Model Specification

- The earlier explanation highlights that the assumption of correct model specification sets factor investing strategies apart
  - Other investment styles do not make decisions based on premia attributed to risk characteristics
- The violation of this assumption has practical consequences in terms of profitability
  - Under-controlling for confounders (see here)
  - Over-controlling for colliders (the focus of today's session)
- Factor model misspecification has two primary consequences for investors: Type-1 errors (false strategies) and type-2 errors (missed strategies)
  - Out of the type-1 errors, some of them result in systematic losses (a factor strategy with an expected negative return)





In the top graph, Z is a confounder to $X \rightarrow Y$. In the bottom graph, Z is a collider to $X \rightarrow Y$.

Mistaking one causal graph for the other has adverse consequences in factor investing strategies.

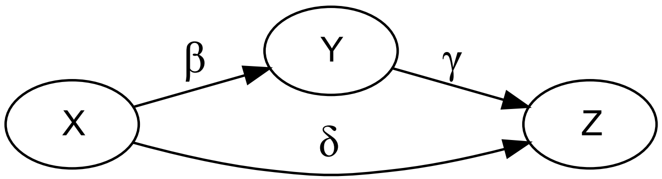
Introduction to Colliders

Data-Generating Process (DGP)

- Consider a DGP with equations:

$$Y := X\beta + u$$

$$Z := Y\gamma + X\delta + v$$



where $\gamma \neq 0$, $\delta \neq 0$, and variables (u, v, X) are independent and identically distributed as a standard Normal, $(u, v, X) \sim N[0, I]$

- Only variable X is a factor exposure, because only X is a cause of Y

In the language of causal inference, Z is a collider, because it is influenced by both the cause (X) and the effect (Y).

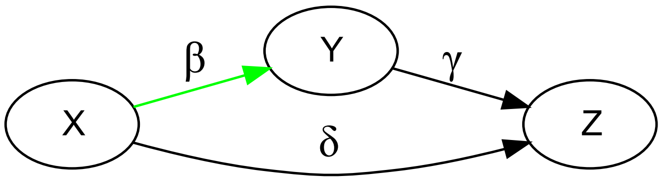
Correct Model Specification

- With proper knowledge of the causal graph, observers would estimate the effect of X on Y (green arrow) by fitting the equation

$$Y = X\beta + \varepsilon$$

- With that model specification, parameter β is estimated without bias,

$$E\left[\hat{\beta}|X\right] = \beta$$
$$E\left[\hat{Y}|X\right] = XE\left[\hat{\beta}|X\right] = X\beta = E[Y|X]$$



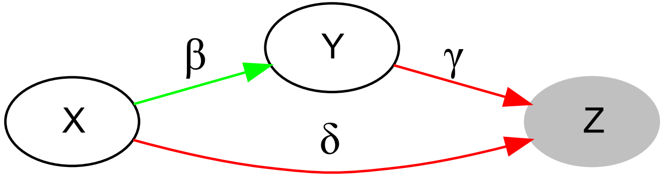
Without conditioning on Z, the collider blocks the non-causal path from X to Y (i.e., $X \rightarrow Z \leftarrow Y$).

Only β is a risk premium, and (δ, γ) are not risk premia.

Over-Controlling for a Collider

- Suppose that observers incorrectly attempt to estimate the causal effect of X on Y by fitting

$$Y = X\beta + Z\theta + \varepsilon$$



- Controlling for collider Z (shaded node) unblocks the non-causal path (red arrows), thus β and θ are estimated with a bias

$$E[\hat{\beta}|X, Z] = \frac{\beta - \delta\gamma}{1 + \gamma^2} \qquad E[\hat{\theta}|X, Z] = \frac{\gamma}{1 + \gamma^2}$$

$$E[\hat{Y}|X, Z] \quad = X\frac{\beta - \delta\gamma}{1 + \gamma^2} + Z\frac{\gamma}{1 + \gamma^2}; \; E[Y|X] = X\beta$$

Controlling for the collider unblocks the **non-causal path** from X to Y through Z (i.e., $X \to Z \leftarrow Y$).

Collider parameters (δ, γ) bias the estimate of risk premium β.

Notably, the model that over-controls for a collider estimates Y with lower variance of the error than the correctly specified model, i.e. $Var[\varepsilon] < Var[u]$.

Factor & Forecasting Performance Analysis

Factor Strategy Under Constant Correlations

- With the correctly specified factor model, the factor strategy that sizes positions by $E\left[\hat{Y}|X\right]$ produces expected returns

$$E\left[YE\left[\hat{Y}|X\right]|X\right] = (X\beta)^2 \geq 0$$

- However, with the incorrect model specification $Y = X\beta + Z\theta + \varepsilon$, the investor controls for a variable $\tilde{Z}$ whose value is set before Y, hence

$$E[YE[\hat{Y}|X,\tilde{Z}]|X,\tilde{Z}] = X\beta \frac{X\beta + \gamma(\tilde{Z} - \delta X)}{1 + \gamma^2}$$

By the time Z is known and it is possible to condition on its value, the value of Y has already been set. In other words, it is not possible to estimate $E[\hat{Y}|X,Z]$ *before* the value of Y has been set.

Attempting to condition on a value of Z before the value of Y is set is tantamount to conditioning on a variable $\tilde{Z}$ that adds no information beyond what is known when the value of X is set (a Point in Time issue).

For appropriate combinations of real values of $(X,\tilde{Z},\beta,\gamma,\delta)$, the strategy yields systematic losses, i.e. $E[YE[\hat{Y}|X,\tilde{Z}]|X,\tilde{Z}] < 0$. For example, the strategy yields systematic losses under $(X,\tilde{Z},\beta,\gamma,\delta) = (1,1,1,1,3)$.

Forecast Strategy Under Constant Correlations

- With the correctly specified factor model, a forecasting strategy that sizes positions by $E[\hat{Y}|X]$ has expected returns

$$E\left[YE[\hat{Y}|X]\right] = \beta^2 \geq 0$$

- However, with the incorrect model specification $Y = X\beta + Z\theta + \varepsilon$,

$$E[YE[\hat{Y}|X,\tilde{Z}]] = \beta\frac{\beta - \gamma\delta}{1 + \gamma^2}$$

For appropriate combinations of real values of (β, γ, δ), the strategy yields systematic losses, i.e. $E[YE[\hat{Y}|X,\tilde{Z}]] < 0$. For example, the strategy yields systematic losses under $(\beta, \gamma, \delta) = (1,1,2)$.

The Econometric Canon

Explanatory Power & *p*-Values

- The R-squared of a factor model that over-controls for a collider always exceeds the R-squared of the correct model
  - Given enough observations, the adjusted R-squared of a factor model that over-controls for a collider exceeds the adjusted R-squared of the correct model
- When the collider effects are strong and of similar magnitude, or the causal effect is weak, the *p*-values of estimated coefficients in a factor model that over-controls for a collider are smaller than those of the correct model
  - The presence of colliders facilitates *p*-hacking under rather general conditions

$$R^2_{\text{correct}} = 1 - \frac{Var[u]}{Var[Y]}$$

$$= 1 - \frac{1}{1+\beta^2}$$

$$R^2_{\text{over-controlled}}$$

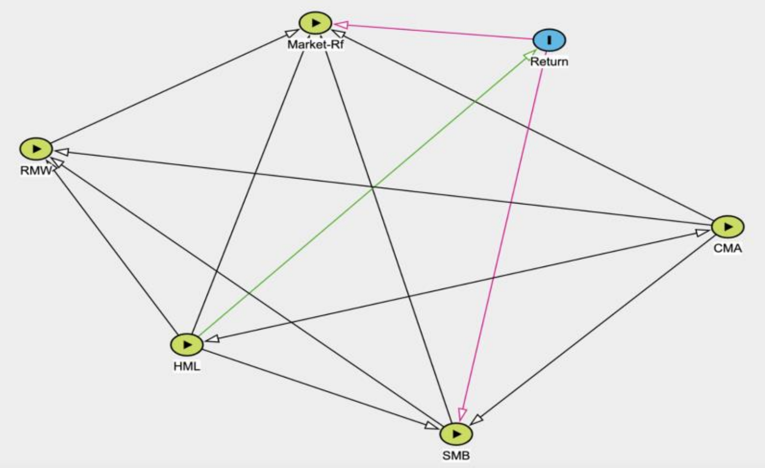
$$= 1 - \frac{Var[\varepsilon]}{Var[Y]}$$

$$= 1 - \frac{1}{(1+\gamma^2)(1+\beta^2)}$$

Controlling for collider Z introduces non-causal association, with the effect of inflating the R-squared and, given enough observations, also the adjusted R-squared.

Canonical Model Selection

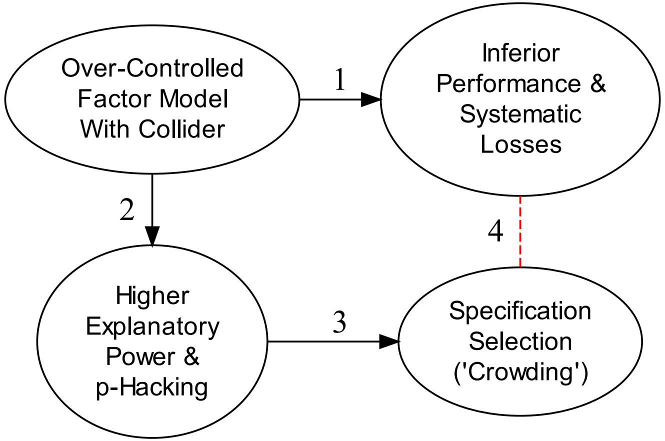
- The econometric canon is seemingly ignorant of the problem of collider bias
  - Greene [2012, p. 136] indicates that the downside of including irrelevant variables is “the reduced precision of the estimates,” and that “[o]mitting variables from the equation seems generally to be the worse of the two errors”
  - Greene [2012, p. 137] recommends a “general-to-simple” approach, where a researcher starts with a “kitchen sink regression, which contains every variable that might conceivably be relevant”
  - Fama and French [2015] argument for adding two factors to their initial three-factor model specification was that “the five-factor model performs better than the three-factor model when used to explain average returns”



Sadeghi et al. [2024] apply the PC algorithm for causal discovery to the variables used in Fama and French [1993], and find that the HML factor may be a collider.
Gu et al. [2023] find that Fama and French [2015] may also be misspecified, with the presence of at least one collider (market returns net of the risk-free rate).

The Causal Mechanism for Adverse Outcomes

1. Strategies based on misspecified factor models can yield systematic losses in the collider case
2. Misspecified factor models that over-control for a collider exhibit a higher R-squared than correctly specified models
   – Under certain conditions, colliders make it easier to *p*-hack a factor
3. Canonical approaches for specification-selection in econometrics favor over-controlled models, crowding out correctly specified models



The combined effect of (1,2,3) is an increased association between selected (either published or deployed) factor investing strategies and underperformance, including systematic losses.

This causal mechanism provides a channel for "few-shot *p*-hacking."

The Solution

Causal Discovery Algorithms

- Over the past three decades, statisticians have developed numerous computational methods and algorithms for the discovery of causal relations, represented as directed acyclic graphs (DAGs)
- These methods can be divided into three types:
  - **Score-based algorithms** optimize a defined score function, and can be used in the absence latent confounders (e.g., GES)
  - **Constraint-based methods** exploit conditional independence relationships in the data (e.g., PC, FCI)
    - FCI can give asymptotically correct results even if there are latent confounders
  - **Functional causal models** (FCMs) distinguish between DAGs in the same equivalence class

Social Science Computer Review

Impact Factor: 4.1 / 5-Year Impact Factor: 4.7

Restricted access | Research article | First published April 1991

An Algorithm for Fast Recovery of Sparse Causal Graphs

Peter Spirtes and Clark Glymour View all authors and affiliations

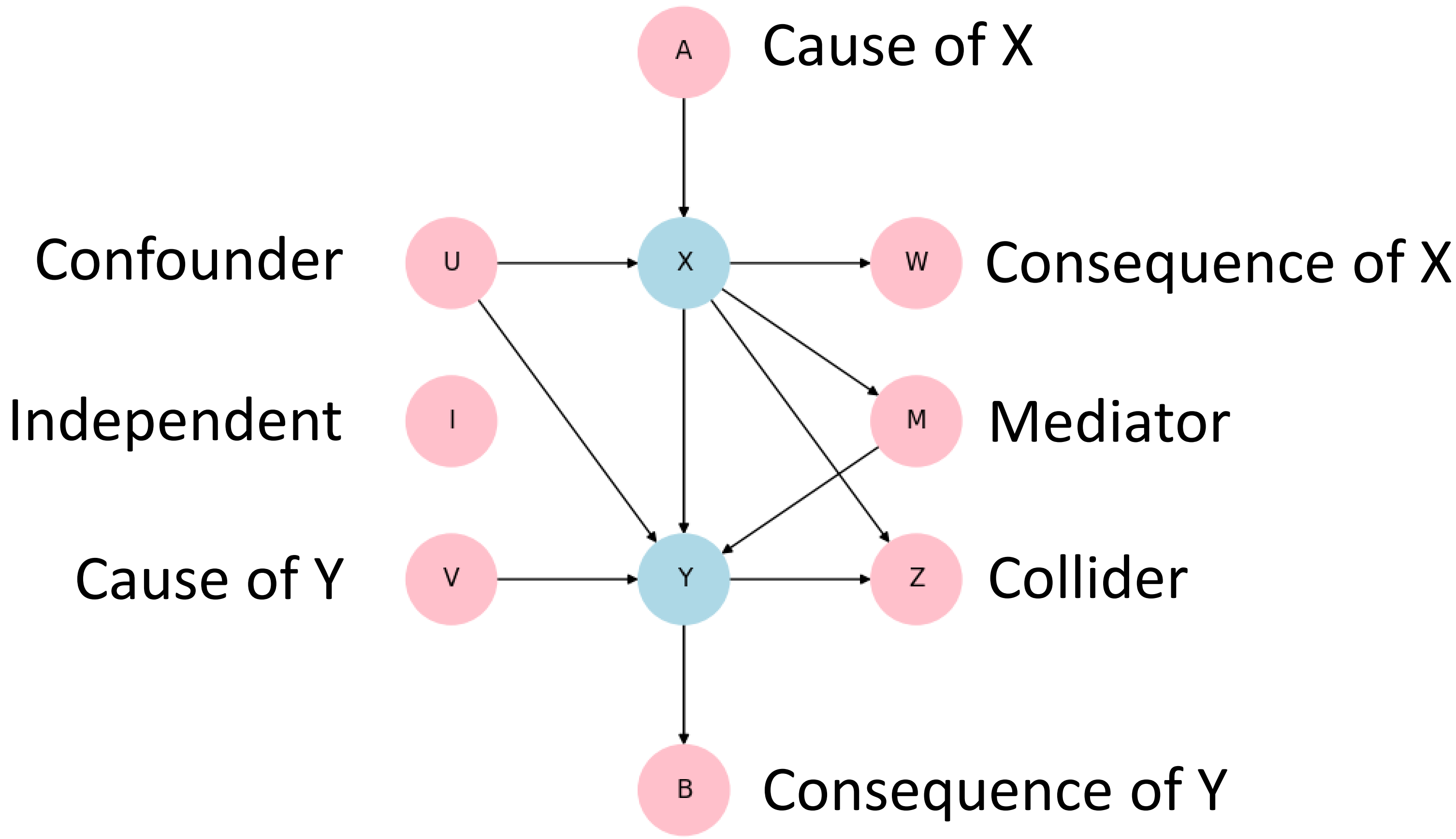
Volume 9, Issue 1 | https://doi.org/10.1177/089443939100900106

Contents | Get access | Cite article | Share options | Information, rights and permissions

Abstract

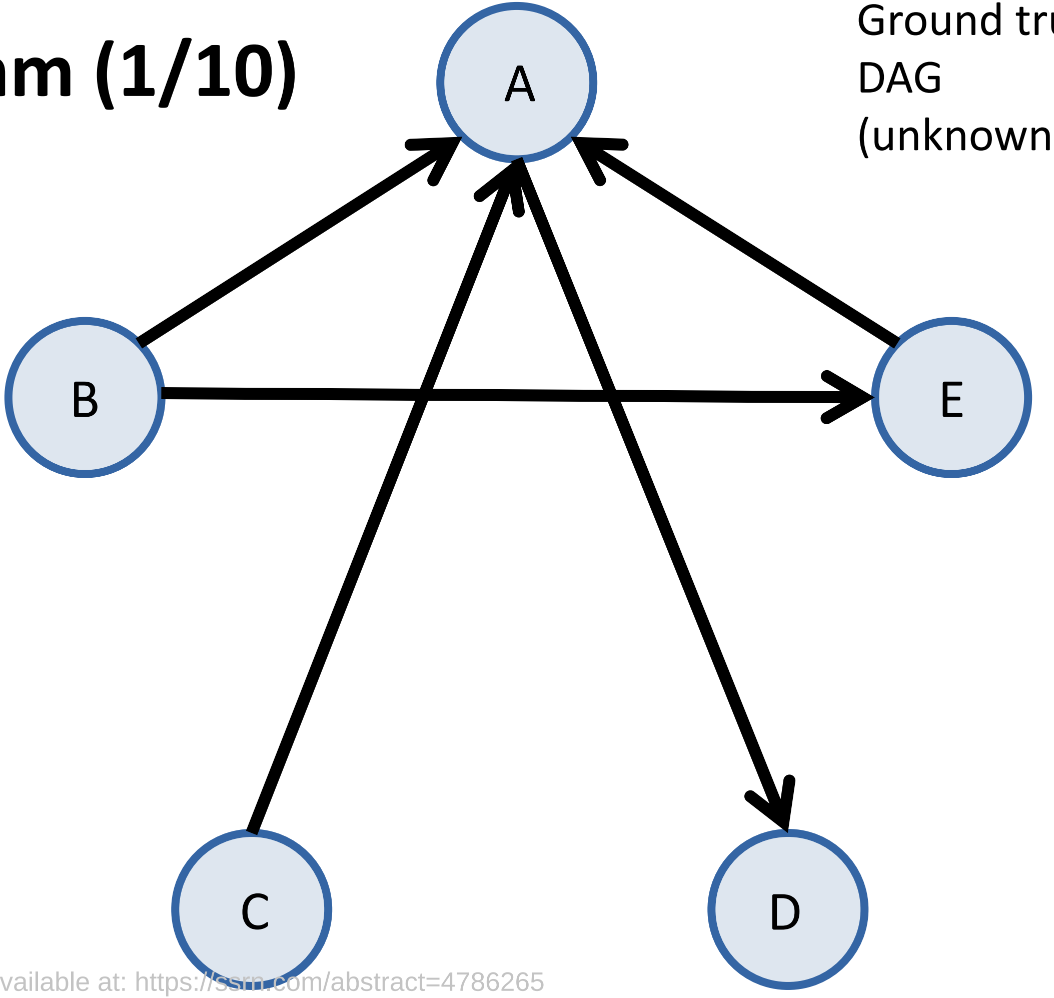
Previous asymptotically correct algorithms for recovering causal structure from sample probabilities have been limited even in sparse causal graphs to a few variables. We describe an asymptotically correct algorithm whose complexity for fixed graph connectivity increases polynomially in the number of vertices, and may in practice recover sparse graphs with several hundred variables. From sample data with n = 20,000, an implementation of the algorithm on a DECStation 3100 recovers the edges in a linear version of the ALARM network with 37 vertices and 46 edges. Fewer than 8% of the undirected edges are incorrectly identified in the output. Without prior ordering information, the program also determines the direction of edges for the ALARM graph with an error rate of 14%. Processing time is less than 10 seconds. *Keywords* DAGS, Causal Modelling.

Discovering a causal graph typically requires extra-statistical (beyond observational) information, in the form of an experiment or expert knowledge. However, under some assumptions, it is possible to derive causal graphs that are *consistent with* the observed sample.

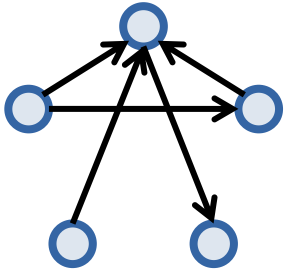

A
Cause of X
Confounder
U
X
W
Consequence of X
Independent
I
M
Mediator
Cause of Y
V
Y
Z
Collider
B
Consequence of Y

PC Algorithm (1/10)

- Start with a complete (undirected) graph
- Remove the edge $X — Y$ if $X \perp Y | Z$ for some (possibly empty) set of nodes Z
- For all $X — Z — Y$ (with no edge between X and Y), if $X \perp Y$ and $\neg(X \perp Y)|Z$, we have a collider
- Propagate the orientation, assuming we have found all the colliders

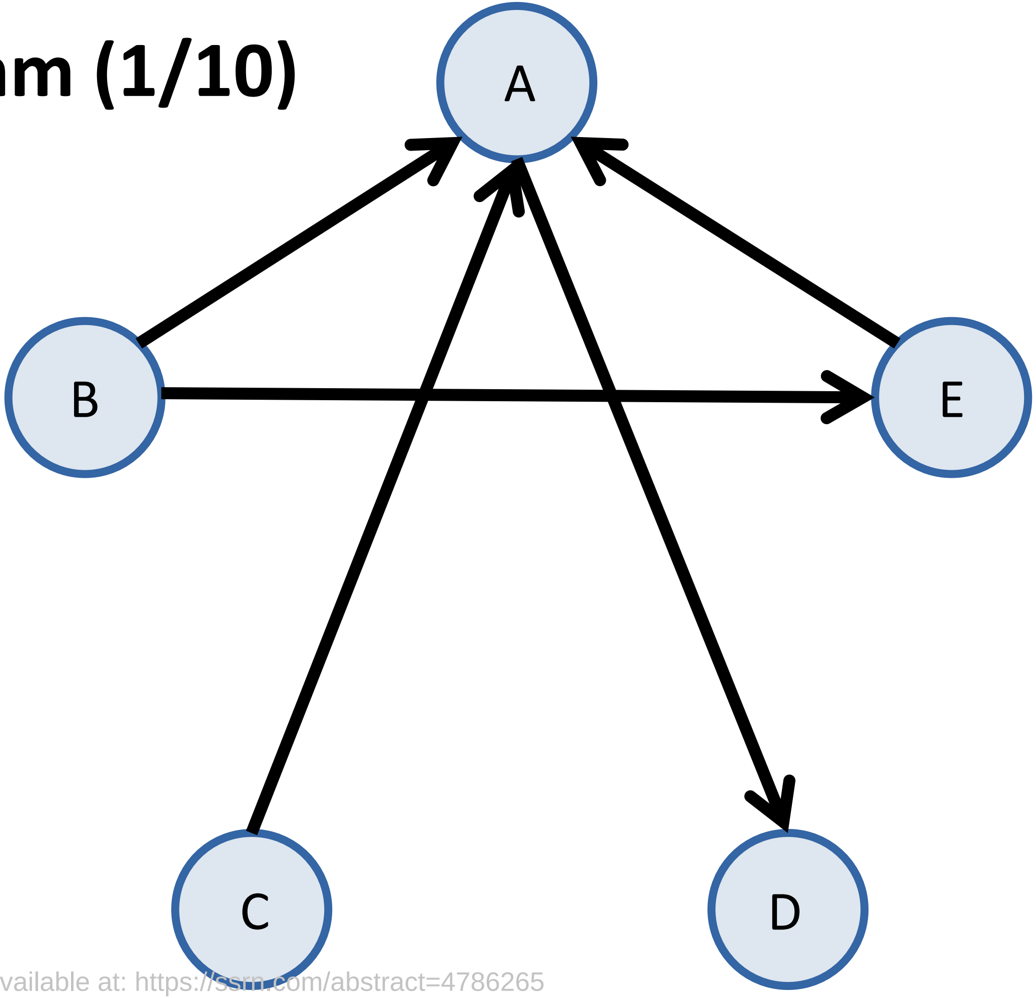


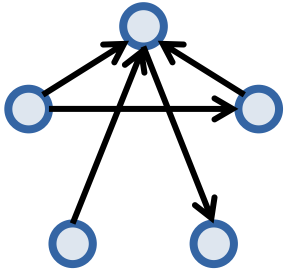
Ground truth DAG (unknown)



PC Algorithm (1/10)

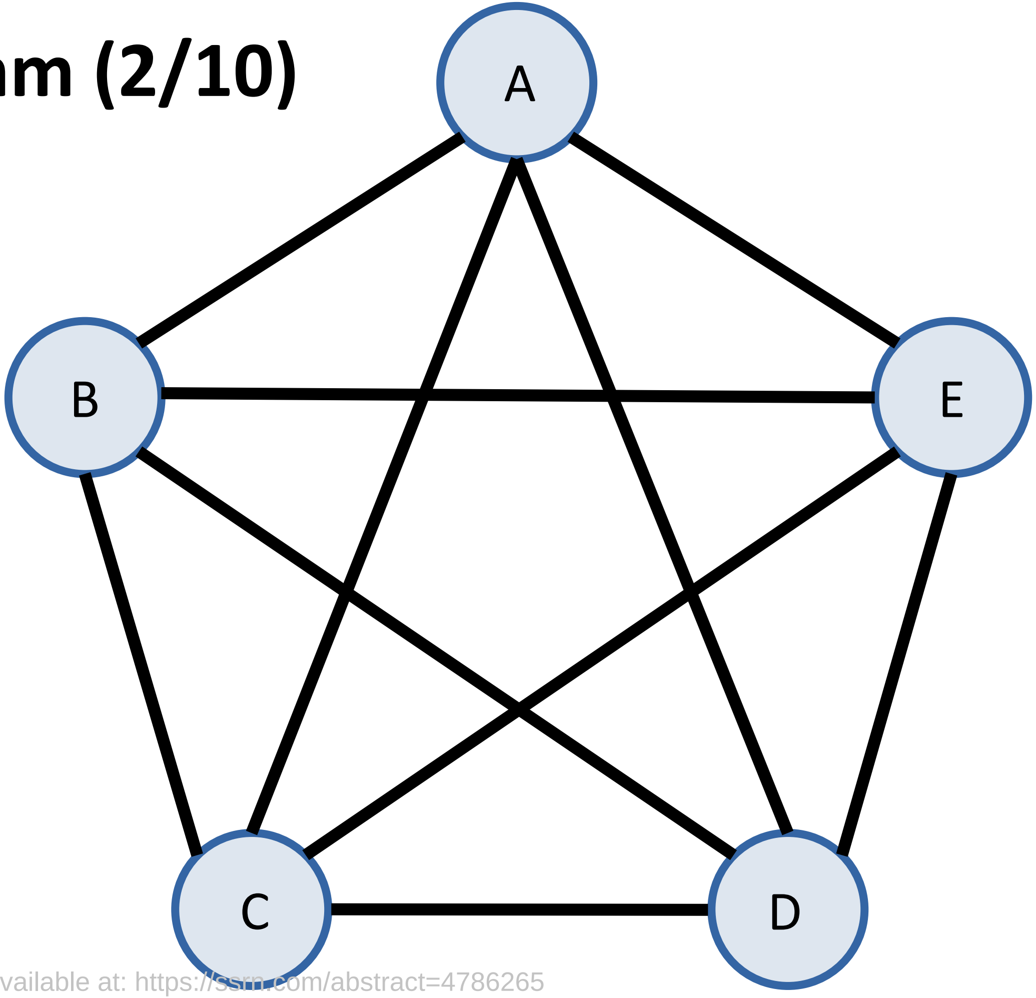
- Start with a complete (undirected) graph
- Remove the edge $X — Y$ if $X \perp Y | Z$ for some (possibly empty) set of nodes Z
- For all $X — Z — Y$ (with no edge between X and Y), if $X \perp Y$ and $\neg(X \perp Y) | Z$, we have a collider
- Propagate the orientation, assuming we have found all the colliders

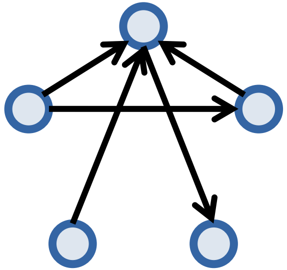




PC Algorithm (2/10)

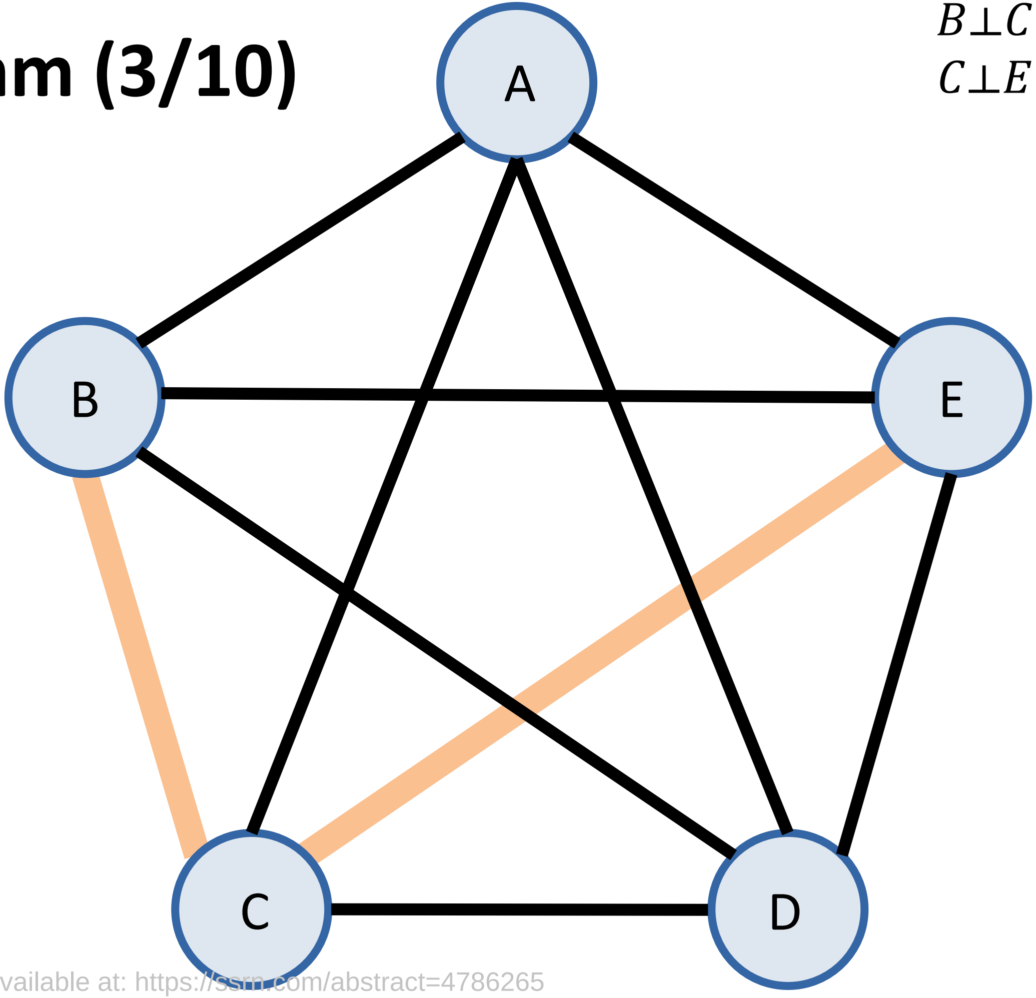
- Start with a complete (undirected) graph
- Remove the edge $X — Y$ if $X \perp Y | Z$ for some (possibly empty) set of nodes Z
- For all $X — Z — Y$ (with no edge between X and Y), if $X \perp Y$ and $\neg(X \perp Y)|Z$, we have a collider
- Propagate the orientation, assuming we have found all the colliders



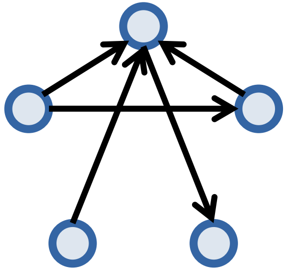


PC Algorithm (3/10)

- Start with a complete (undirected) graph
- Remove the edge $X — Y$ if $X \perp Y | Z$ for some (possibly empty) set of nodes Z
- For all $X — Z — Y$ (with no edge between X and Y), if $X \perp Y$ and $\neg(X \perp Y)|Z$, we have a collider
- Propagate the orientation, assuming we have found all the colliders

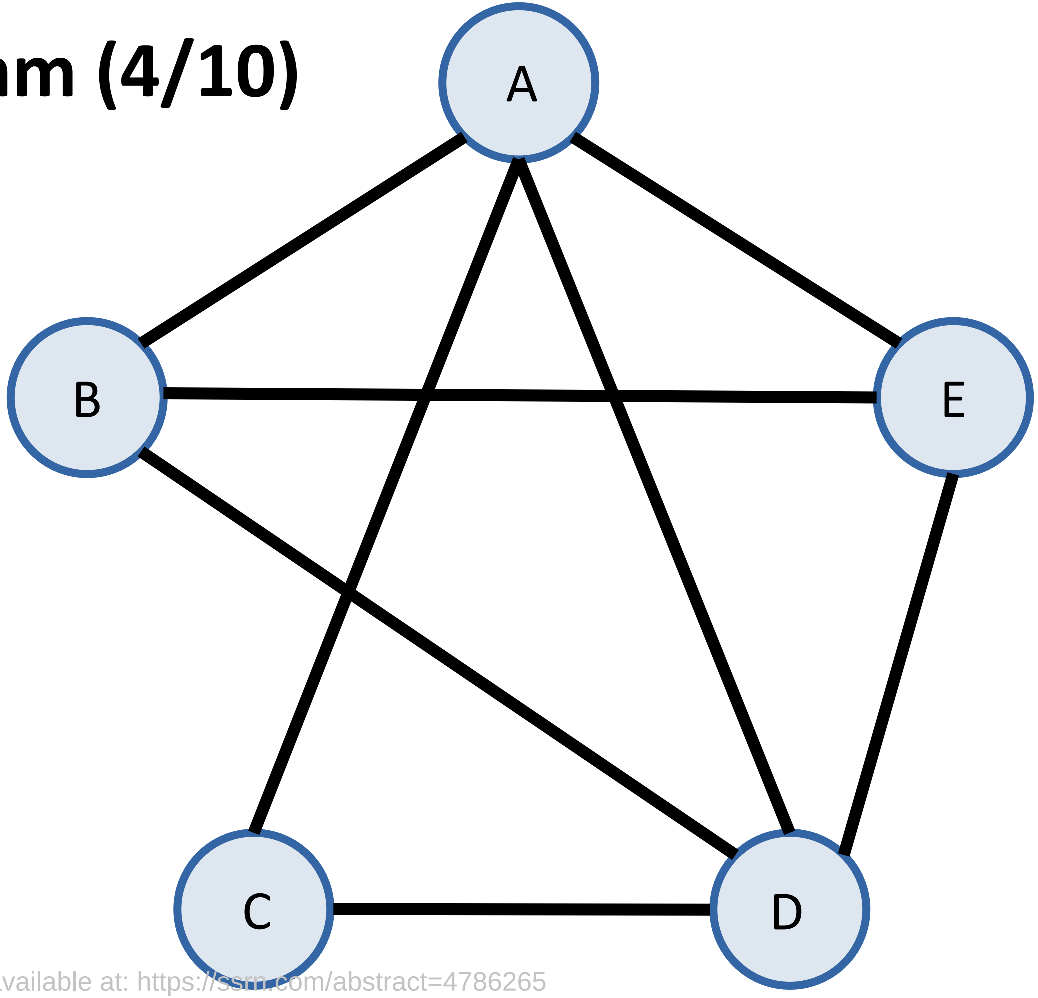


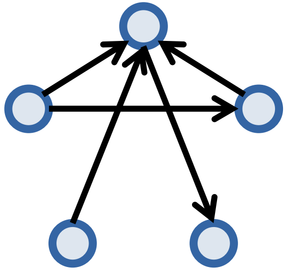
Classification: Public



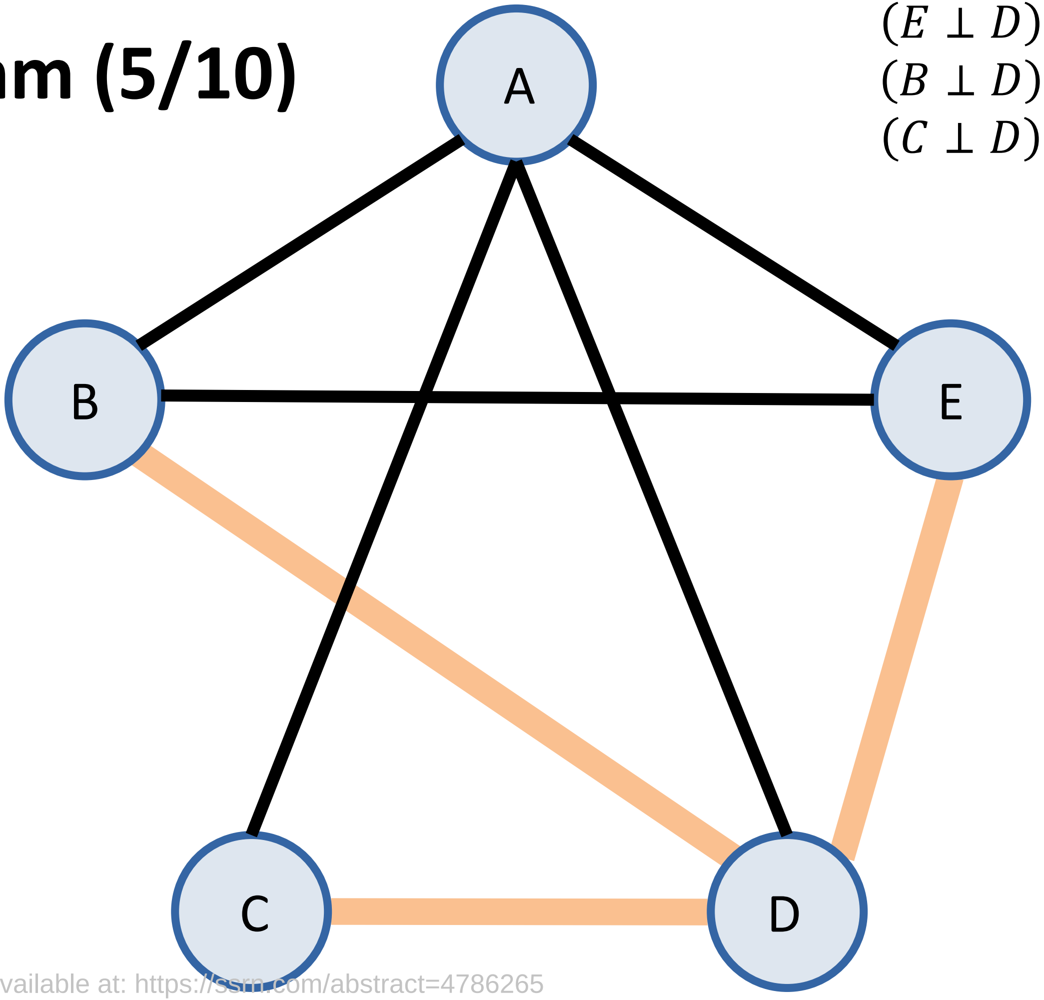
PC Algorithm (4/10)

- Start with a complete (undirected) graph
- Remove the edge $X — Y$ if $X \perp Y | Z$ for some (possibly empty) set of nodes Z
- For all $X — Z — Y$ (with no edge between X and Y), if $X \perp Y$ and $\neg(X \perp Y)|Z$, we have a collider
- Propagate the orientation, assuming we have found all the colliders



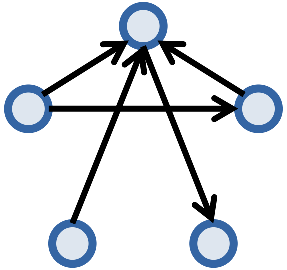


PC Algorithm (5/10)

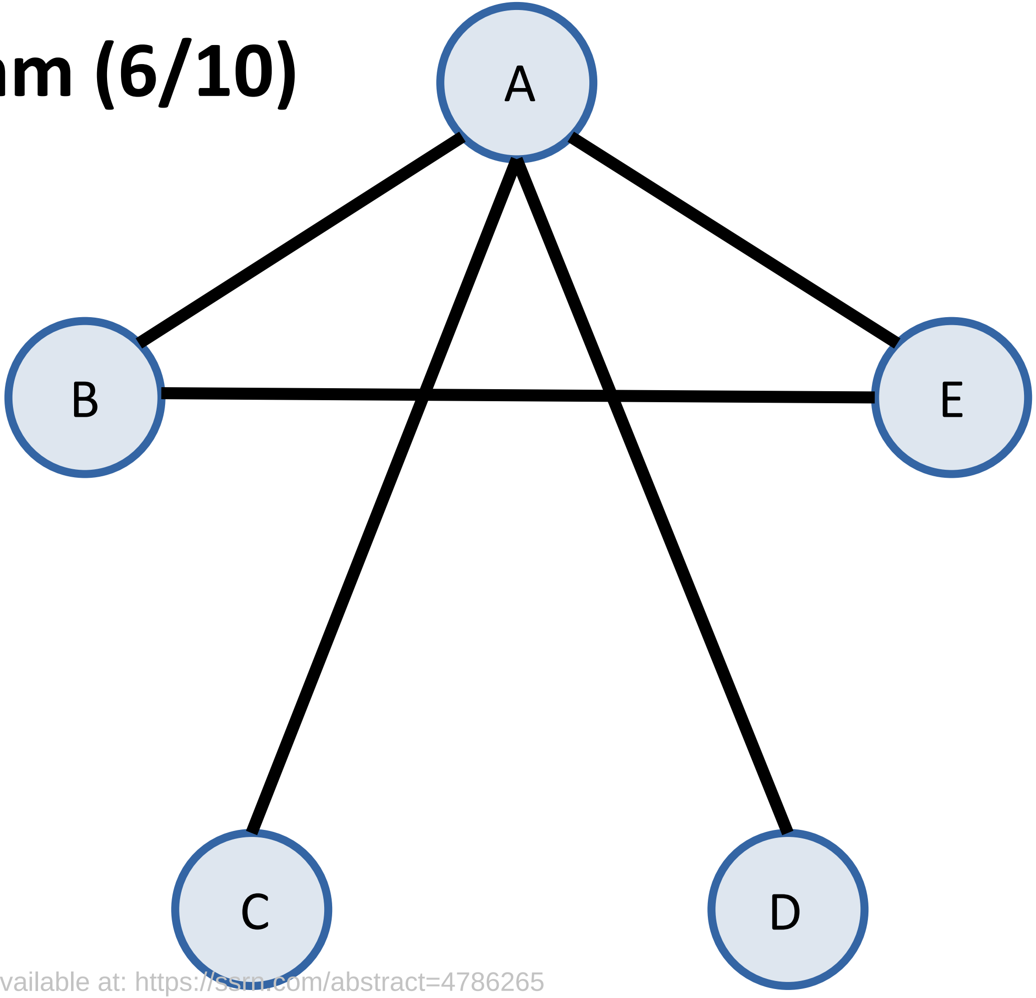


$(E \perp D)|A$
$(B \perp D)|A$
$(C \perp D)|A$

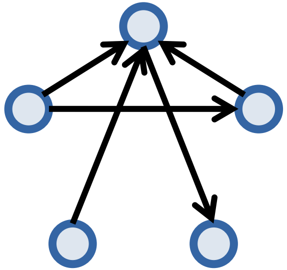
- Start with a complete (undirected) graph
- Remove the edge $X — Y$ if $X \perp Y|Z$ for some (possibly empty) set of nodes Z
- For all $X — Z — Y$ (with no edge between X and Y), if $X \perp Y$ and $\neg(X \perp Y)|Z$, we have a collider
- Propagate the orientation, assuming we have found all the colliders



PC Algorithm (6/10)



- Start with a complete (undirected) graph
- Remove the edge $X — Y$ if $X \perp Y | Z$ for some (possibly empty) set of nodes Z
- For all $X — Z — Y$ (with no edge between X and Y), if $X \perp Y$ and $\neg(X \perp Y) | Z$, we have a collider
- Propagate the orientation, assuming we have found all the colliders

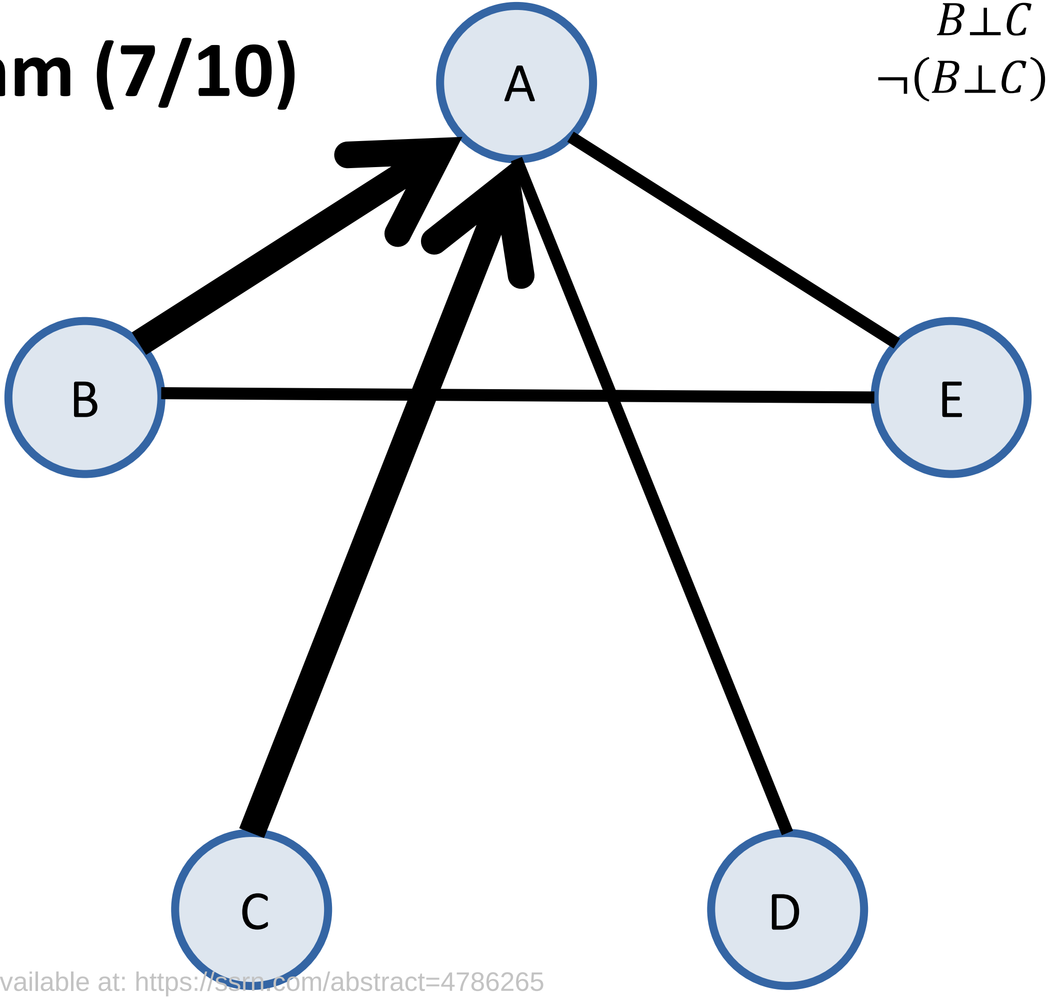


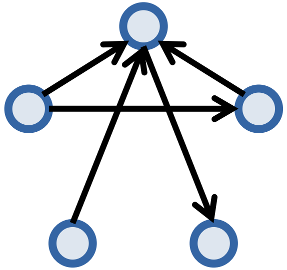
PC Algorithm (7/10)

- Start with a complete (undirected) graph
- Remove the edge $X — Y$ if $X \perp Y | Z$ for some (possibly empty) set of nodes Z
- For all $X — Z — Y$ (with no edge between X and Y), if $X \perp Y$ and $\neg(X \perp Y)|Z$, we have a collider
- Propagate the orientation, assuming we have found all the colliders

$B \perp C$

$\neg(B \perp C)|A$

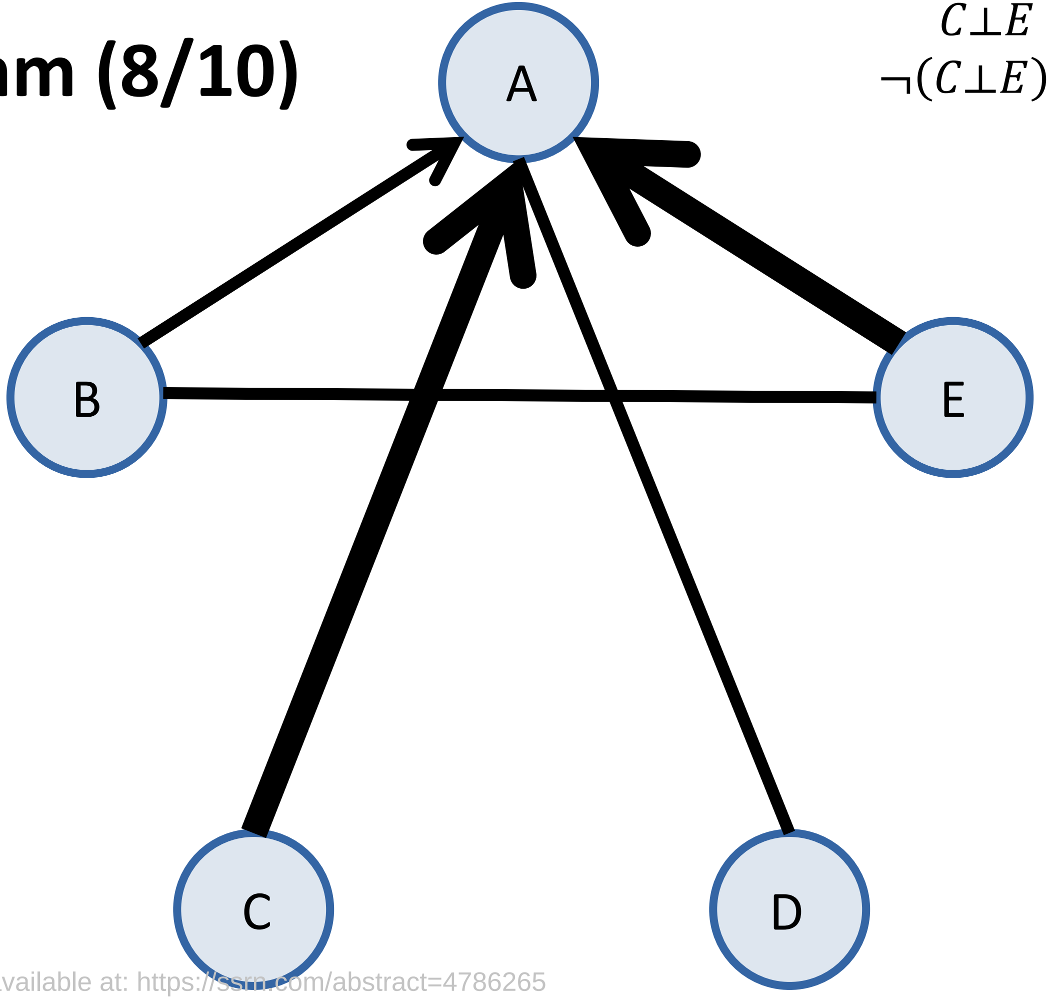




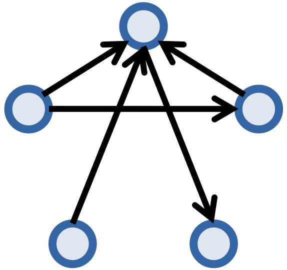
PC Algorithm (8/10)

$C \perp E$

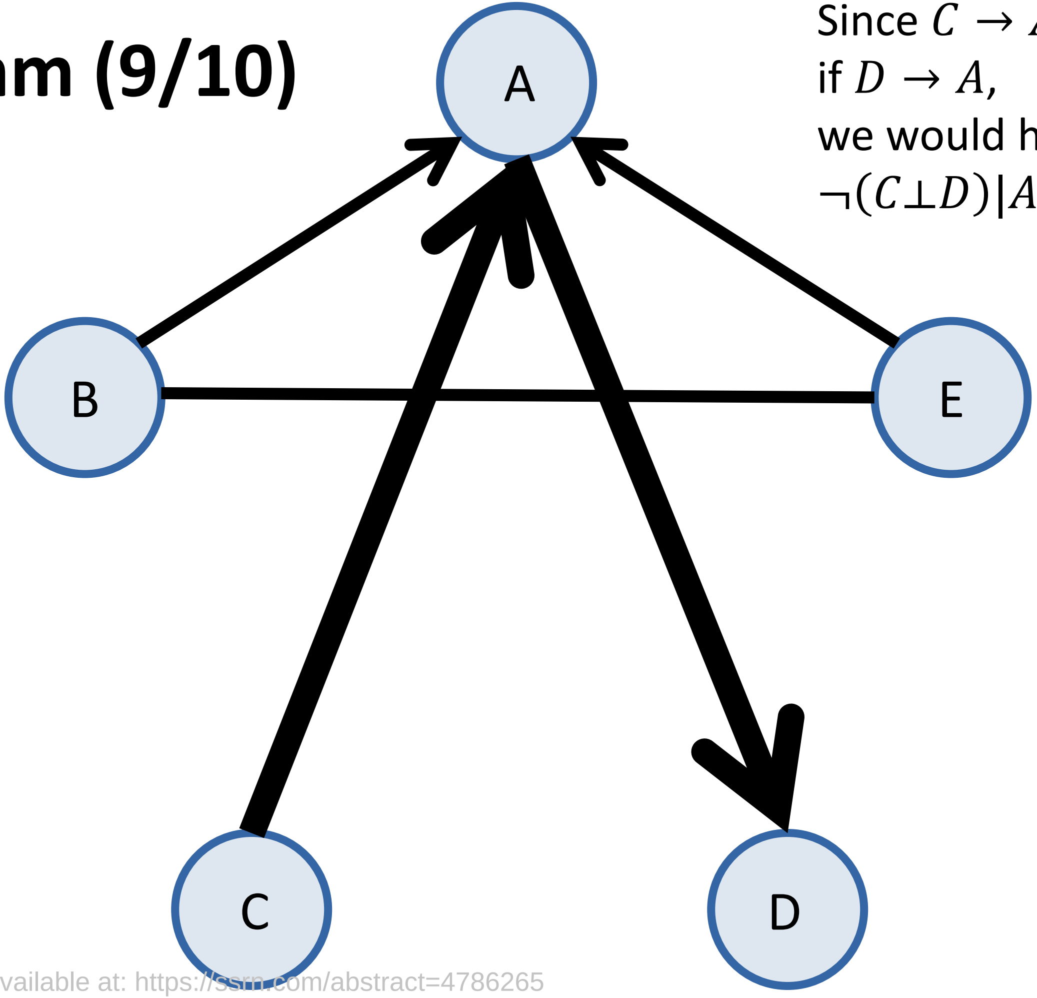
$\neg(C \perp E)|A$



- Start with a complete (undirected) graph
- Remove the edge X—Y if $X \perp Y|Z$ for some (possibly empty) set of nodes Z
- For all X—Z—Y (with no edge between X and Y), if $X \perp Y$ and $\neg(X \perp Y)|Z$, we have a collider
- Propagate the orientation, assuming we have found all the colliders

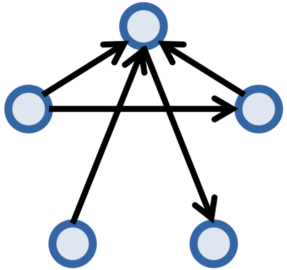


PC Algorithm (9/10)

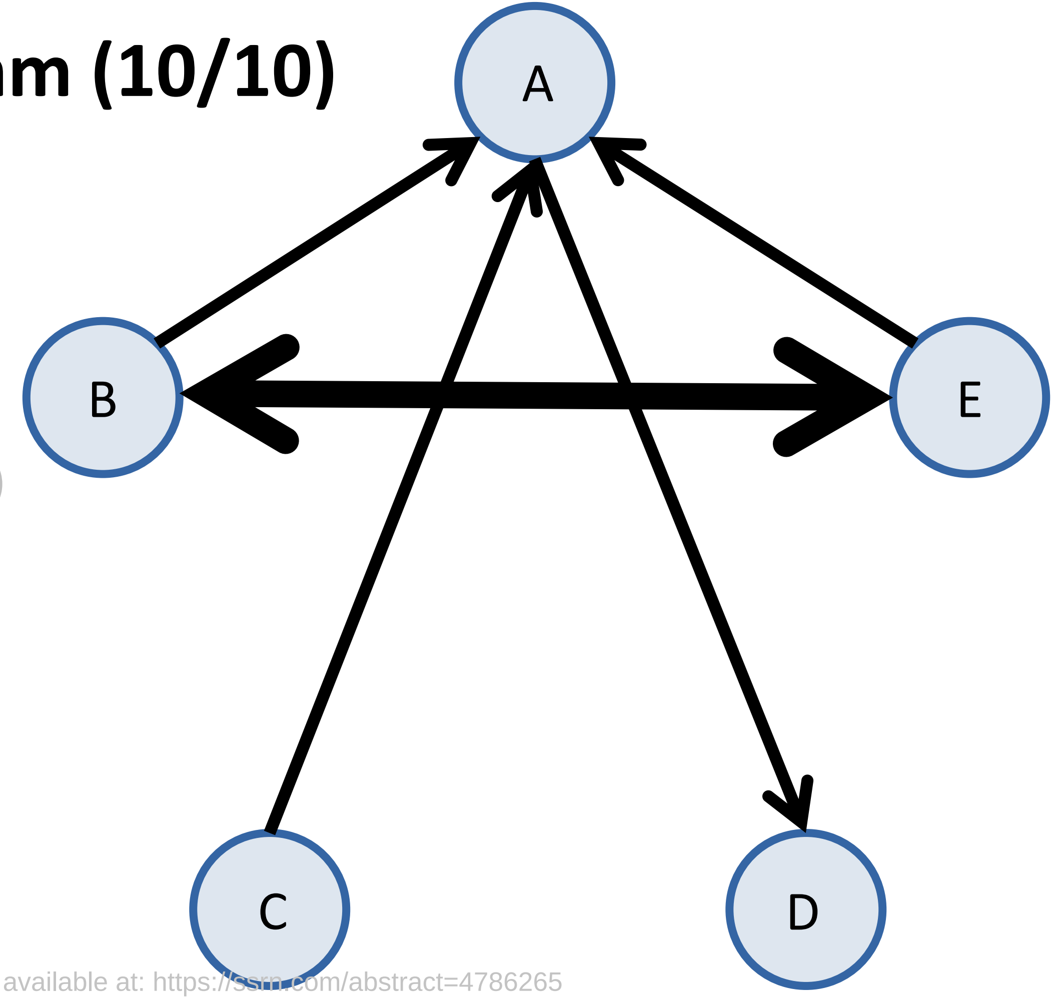


Since $C \to A$,
if $D \to A$,
we would have
$\neg(C \perp D)|A$

- Start with a complete (undirected) graph
- Remove the edge $X — Y$ if $X \perp Y|Z$ for some (possibly empty) set of nodes Z
- For all $X — Z — Y$ (with no edge between X and Y), if $X \perp Y$ and $\neg(X \perp Y)|Z$, we have a collider
- Propagate the orientation, assuming we have found all the colliders

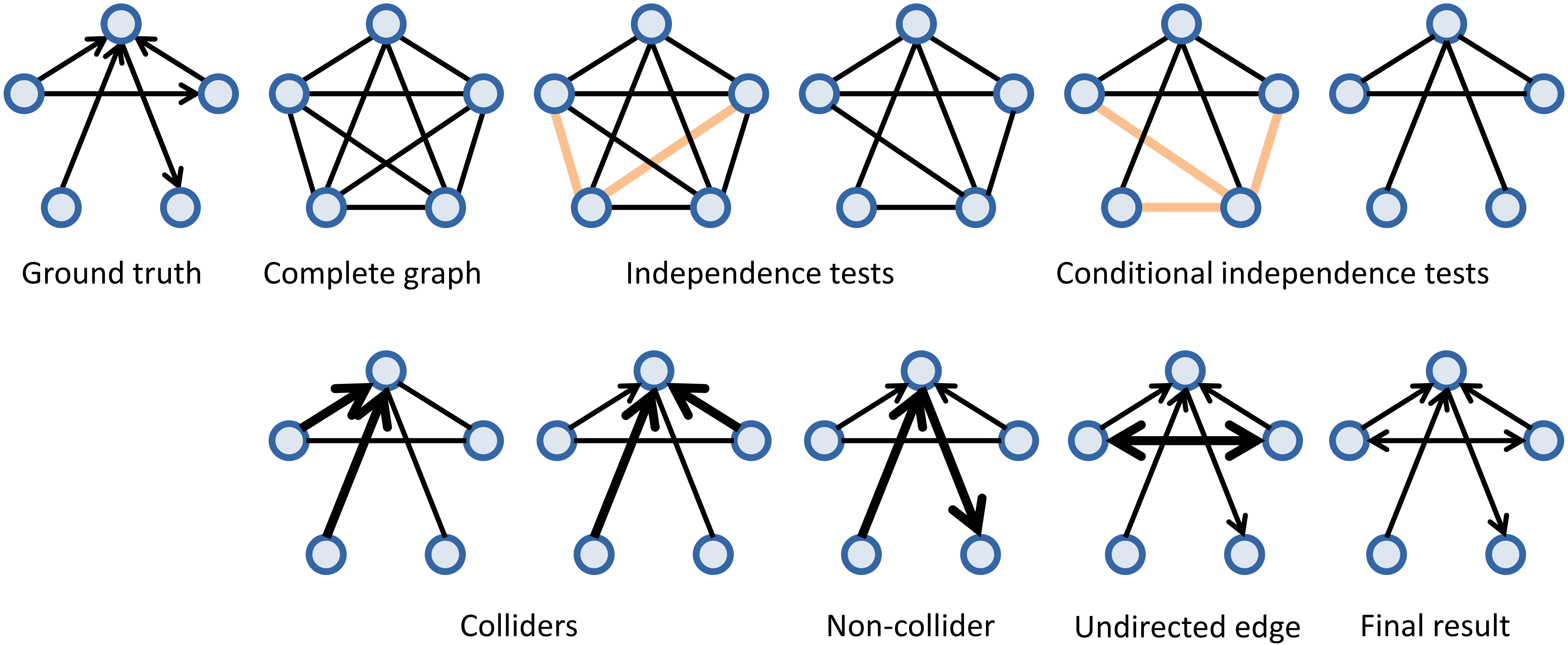


PC Algorithm (10/10)



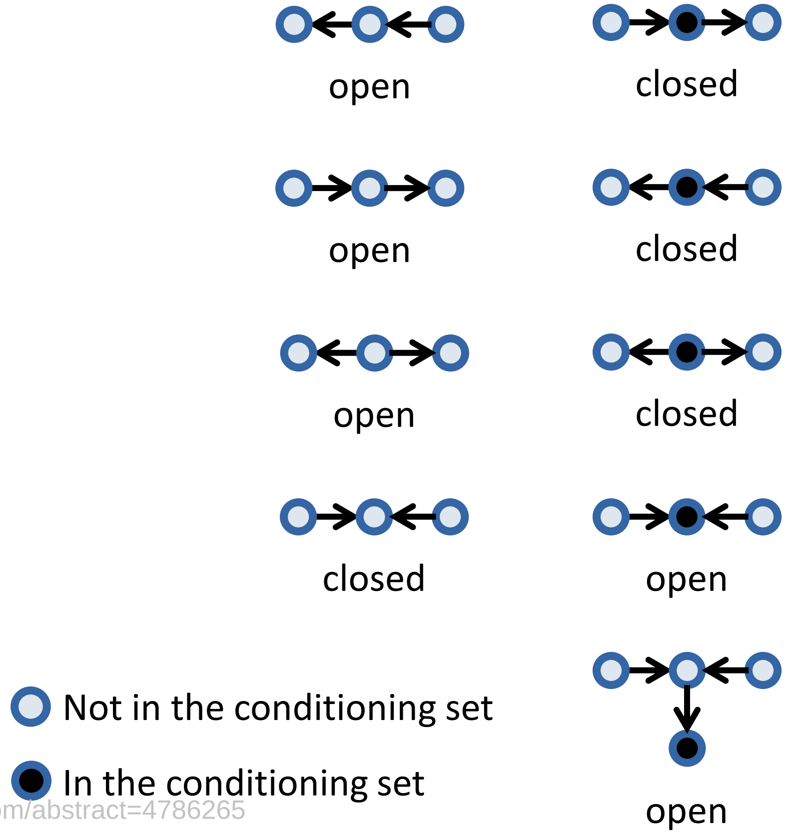
- Start with a complete (undirected) graph
- Remove the edge $X — Y$ if $X \perp Y | Z$ for some (possibly empty) set of nodes Z
- For all $X — Z — Y$ (with no edge between X and Y), if $X \perp Y$ and $\neg(X \perp Y) | Z$, we have a collider
- Propagate the orientation, assuming we have found all the colliders

PC Algorithm At A Glance



Specification Selection (1/2)

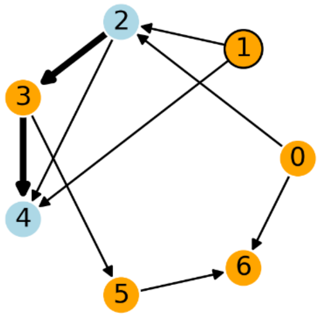
- **The sufficient conditioning set is not necessarily unique**
  - For instance, the parents of X may form a sufficient conditioning set for $X \to Y$, but it may be needlessly large
- More generally, a sufficient conditioning set is a set of nodes blocking all the non-causal paths from X to Y, while leaving all the causal paths open
- If some variables are not observed, things get more complicated, but do-calculus can tell us if the effect can be estimated from observational data alone, and how



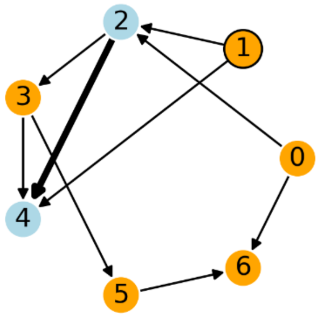
Specification Selection (2/2)

To assess the strength of the causal relation $X \rightarrow Y$:

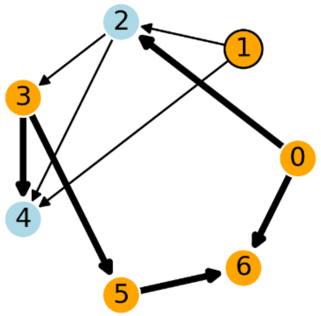
- List all the (undirected) paths from X to Y
- All the non-causal paths should be blocked
  - if not, condition on one or more nodes to block them
- All the causal paths should be open
  - if not, adjust the conditioning set to unblock them



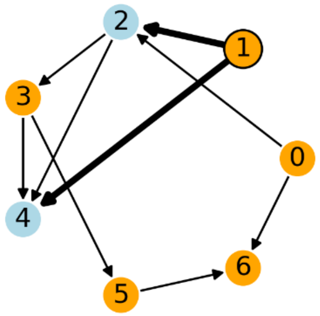
Causal path



Causal path



Blocked by the collider 6



Blocked if we condition on 1

The Dawn of Causal Factor Investing

- A scientific theory is a falsifiable statement of the form "X causes Y through mechanism M"
- Scientific theories matter to investors because
  - causality is a necessary condition for investment efficiency
    - associational models misattribute risks and performance, thus preventing investors from building efficient portfolios
  - causal models enable counterfactual reasoning, hence the stress-testing of investment portfolios in a coherent and forward-looking manner
    - associational models cannot answer counterfactual questions, such as what would be the effect of Y on a not-yet-observed scenario X, thus exposing those relying on associations to black-swan events
- **Financial economists' adoption of causal inference has the potential to transform investing into a truly scientific discipline**

Type	Rigor	Example
Randomized controlled trials	Very high	Algo-wheel experiments
Natural experiments	High	Market-maker reaction to random spikes in order imbalance
Simulated interventions	Medium	Estimate effect of HML using a causal graph
Econometric (observational) studies	Low	Factor investing literature; backtested investment strategies
Case studies	Very low	Broker report / analysis
Expert opinion	Anecdotal	Investment guru's prediction

Hierarchy of evidence: Pre-scientific vs. scientific evidence in financial research.

Call For Papers

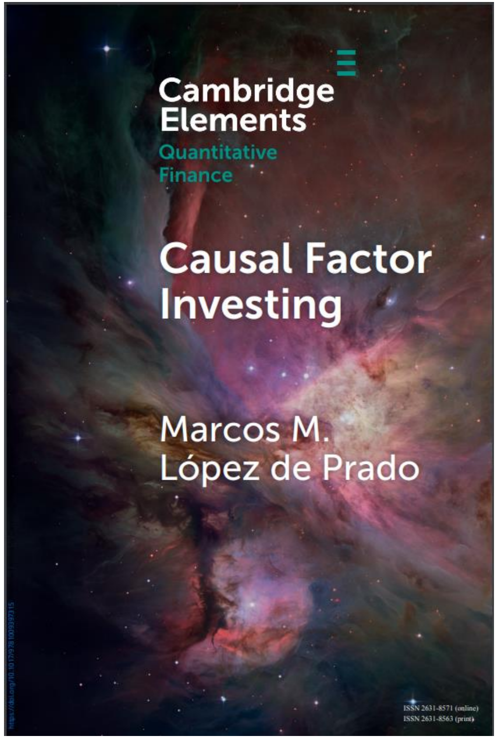
ADIA Lab encourages researchers to move factor investing beyond its current pre-scientific stage.

To help dawn the discipline of Causal Factor Investing, ADIA Lab has called for papers that promote the use of the formal language of causal inference in investing.

To learn more, visit:
https://www.adialab.ae/call-for-papers

We look forward to your papers!

For More Information



Download for free

Causal Factor Investing

(Cambridge University Press, 2023)

Available at:

https://www.cambridge.org/core/elements/causal-factor-investing/9AFE270D7099B787B8FD4F4CBADE0C6E

Disclaimer

- The views expressed in this presentation are my own, and do not necessarily reflect the views of Cornell University, the Abu Dhabi Investment Authority, or ADIA Lab
- No investment decision or particular course of action is recommended by this presentation
-

www.adialab.ae