Price Impact Models and Applications

Introduction to Algorithmic Trading

Kevin Webster

Spring 2023

Columbia University

Plan

Last Week
The Mathematics of Causal Inference (1/2)

For this Week
The Mathematics of Causal Inference (2/2)

(a) The three rules of do-calculus
(b) The back-door criterion
(c) The front-door criterion
(d) Application to prediction bias

Next Week
Transaction Cost Analysis

Last Week's Summary

(a) A causal model $(\Omega, \mathcal{F}, \mathbb{P}, \mathcal{G})$ is an extension of a standard probability space for a compatible DAG $\mathcal{G}$.

(b) Conversely, a causal structure $\mathcal{G}$ can be seen as a set of constraints on the probability measure $\mathbb{P}$.

(c) d-separation graphically describes conditional independence assumptions.

(d) do()-action is a new operator that generates a new probability measure $\tilde{\mathbb{P}}$ for interventions of the form $\mathrm{do}(X = x)$.

Do-Calculus

Refresher: Do-Action (1/4)

"Counterfactuals"
The action $\text{do}(X)$ mathematically formalizes the following "counterfactual":

What if I had done X?

Definition (The do() operator)

Given two variables X and Y on a causal model $\mathcal{M}$, define the $\text{do}(X)$ action as follows

$$\mathbb{P}(Y|\,\text{do}(X = x)) = \tilde{\mathbb{P}}(Y)$$

where $\tilde{\mathbb{P}}$ is obtained by replacing the function f_i defining variable X with the constant function $X = x$.

Refresher: Do-Action (2/4)

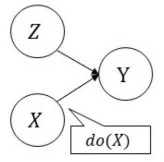

Figure 1: Visual representation of the action do(X) on the causal graph $\mathcal{S}$.

Example of do($X = x$) on the linear causal model

$$
\begin{aligned}
X &= \cancel{\alpha Z + \epsilon_1} \\
&= x \\
Y &= \beta X + \gamma Z + \epsilon_2 \\
Z &= \epsilon_3
\end{aligned}
$$

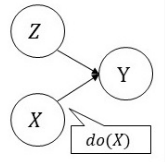

Figure 2: Truncated causal graph $\mathcal{G}_{\overline{X}}$.

Proposition (Do action)
Let $\mathcal{G}$ be a causal structure and $\mathbb{P}$ a probability measure consistent with $\mathcal{G}$. Then the probability measure $\tilde{\mathbb{P}}$ obtained from the do-action $do(X)$ is consistent with the causal structure $\mathcal{G}_{\overline{X}}$.

Refresher: Do-Action (4/4)

Relationship with Bayesian conditioning
The choice to use a similar notation for Bayesian conditioning and do actions is not by accident. Under certain conditions, one has the so called *naive* estimation formula

$$\mathbb{P}(Y|\mathrm{do}(X)) = \mathbb{P}(Y|X).$$

More generally, one goal of causal inference is to establish *purely Bayesian* formulas for *action* estimates.

Example on the linear causal model

$$\begin{aligned} \mathbb{E}[Y|\mathrm{do}(X)] &= \beta X \\ \mathbb{E}[Y|X] &= \beta X + \gamma \mathbb{E}[Z|X]. \end{aligned}$$

Mixing Bayesian with do-calculus

Proposition (Mixing do() with Bayesian conditioning)
Given three disjoint sets of variables X, Y, and Z on a causal model $\mathcal{M}$, one has

$$\mathbb{P}(Y \mid do(X), Z) = \frac{\mathbb{P}(Y, Z \mid do(X))}{\mathbb{P}(Z \mid do(X))}.$$

Three Rules of Do-Calculus (1/3)

Rule 1: Insertion/deletion of observations

$$\mathbb{P}(Y | \mathrm{do}(X), Z, W) = \mathbb{P}(Y | \mathrm{do}(X), W)$$

if Y and Z are conditionally independent given X and W on the graph $\mathcal{G}_{\overline{X}}$.

Example on blackboard.

Three Rules of Do-Calculus (2/3)

Rule 2: Action/observation exchange

$$\mathbb{P}(Y \mid \mathrm{do}(X), \mathrm{do}(Z), W) = \mathbb{P}(Y \mid \mathrm{do}(X), Z, W)$$

if Y and Z are conditionally independent given X and W on the graph $\mathcal{G}_{\overline{X}\underline{Z}}$.

Example on blackboard.

Three Rules of Do-Calculus (3/3)

Rule 3: Insertion/deletion of actions

$$\mathbb{P}(Y|\,\mathrm{do}(X), \mathrm{do}(Z), W) = \mathbb{P}(Y|\,\mathrm{do}(X), W)$$

if Y and Z are conditionally independent given X and W on the graph $\mathcal{G}_{\overline{X}\,\overline{Z(W)}}$, where $Z(W)$ is the subset of Z that are not ancestors of nodes in W in $\mathcal{G}_{\overline{X}}$.

The three rules make do-calculus complete.
A causal expression is identifiable iff it can be broken down using a combination of the three rules. This is a very axiomatic approach: not recommended in practice.

A Straightforward Identification Criterion*

Corollary (A sufficient naive identifiability condition)
Let $\mathcal{M}$ be a causal model with causal structure $\mathcal{G}$. Let Z and Y be two variables such that Z has no parents. Then

$$\mathbb{P}(Y \mid do(Z)) = \mathbb{P}(Y \mid Z).$$

Proof.
Z is an isolated node on $\mathcal{G}_{\underline{Z}}$. Rule 2 applies: $Y \perp Z$ on $\mathcal{G}_{\underline{Z}}$. Hence,

$$\mathbb{P}(Y \mid \mathrm{do}(Z)) = \mathbb{P}(Y \mid Z).$$

$\square$

Intuition: with no parents for X, only direct paths from X to Y can cause correlation between X and Y. The notion of back-door generalizes this insight.

The Back-Door Criterion

Back-Doors*

Definition (Back-door)
Let X and Y be two variables in a causal structure $\mathcal{G}$. Define the set of variables Z to satisfy the *back-door* criterion *from* X *to* Y if:

(a) Z *d*-separates every path between X and Y with a link *pointing into* X.

(b) No descendant of X is in Z.

The *back-door* criterion extends to sets of variables X and Y by requiring Z to satisfy the back-door criterion for every *pair* formed from X and Y.

This definition generalizes the naive scenario: conditioning on the correct variables leaves node X with only direct paths to Y.

The Back-Door Criterion

Theorem (Back-door criterion)
Let $\mathcal{M}$ be a causal model with causal structure $\mathcal{G}$. Assume the set Z satisfies the back-door criterion from X to Y. Then controlling for Z identifies the effect of action $do(X)$ on Y.

For discrete variables, this leads to

$$\mathbb{P}(Y|\,do(X)) = \sum_{z} \mathbb{P}(Y|\,X, Z = z)\,\mathbb{P}(Z = z).$$

For continuous variables, this leads to

$$p(Y|\,do(X)) = \int p(Y|\,X, Z = z)\,p(Z = z)dz.$$

Implications of the Formula (1/3)

Implications for expectations

$$\begin{aligned}\mathbb{E}\left[\, Y \middle| \operatorname{do}(X=x)\right] &= \int y p\left(Y=y \middle| \operatorname{do}(X=x)\right) dy \\ &= \int y \int p\left(Y=y \middle| X=x, Z=z\right) p(Z=z) dz dy \\ &= \int \int y p\left(Y=y \middle| X=x, Z=z\right) dy \; p(Z=z) dz \\ &= \int E\left[\, Y \middle| X=x, Z=z\right] p(Z=z) dz.\end{aligned}$$

This differs from the tower property if $X \not\perp Z$.

$$\begin{aligned}\mathbb{E}\left[\, Y \middle| X=x\right] &= \mathbb{E}\left[\mathbb{E}\left[\, Y \middle| X=x, Z\right] \middle| X=x\right] \\ &= \int E\left[\, Y \middle| X=x, Z=z\right] p\left(Z=z \middle| X=x\right) dz.\end{aligned}$$

Implications of the Formula (2/3)

Implications for linear regression
Let $Y = \beta X + \gamma Z + \epsilon$ and $X = \eta Z + \epsilon'$ be a linear structural model. Then,

$$\begin{aligned}\mathbb{E}\left[\, Y \middle|\, \mathrm{do}(X = x)\right] &= \int E\left[\, Y \middle|\, X = x, Z = z\right] p(Z = z) dz \\ &= \int (\beta x + \gamma z)\, p(Z = z) dz \\ &= \beta x + \gamma \mathbb{E}[Z]\end{aligned}$$

In particular, one has the *differences-in-differences* formula for regression:

$$\mathbb{E}\left[\, Y \middle|\, \mathrm{do}(X = x_1)\right] - \mathbb{E}\left[\, Y \middle|\, \mathrm{do}(X = x_0)\right] = \beta(x_1 - x_0).$$

Implications of the Formula (3/3)

Do-calculus generalizes linear econometrics

(a) $\mathbb{E}[\,Y \mid \mathrm{do}(X = x)]$ generalizes the role of β in assigning causal meaning to regression.

(b) *Given a causal graph*, do-calculus removes philosophical discussions on which variables to control for.

Compatibility with machine learning
Do-actions define causal expressions even for non-parametric models!

This abstraction aligns with machine learning, which focus on predictions and actions over model parameters.

Application of the Back-Door Criterion (1/2)

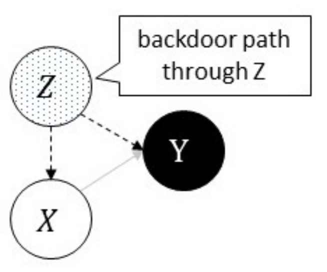

Figure 3: Back-door for estimating $X \rightarrow Y$.

s

Estimating $X \rightarrow Y$ requires controlling for Z.

Application of the Back-Door Criterion (2/2)

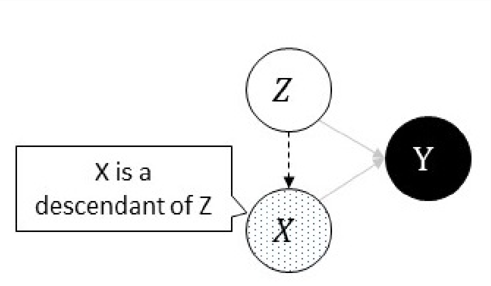

Figure 4: Negative result for estimating $Z \rightarrow Y$ conditioning on X.

However, one can have too many controls.
For instance, controlling for X hurts, rather than helps, when estimating $Z \rightarrow X$.

The Front-Door Criterion

Back-doors are sometimes impractical.

The back-door criterion requires controlling for every indirect path.

(a) What if there are many distinct backdoors? Sizable controls requires sizable data.

(b) What if the confounder (backdoor) is known but non-observable?

(c) What if the confounder is unknown?

The front-door criterion is an alternative trick to circumvent those problems.
Key idea: if we can *funnel* all the direct paths $X \to ... \to Y$ through one variable Z, then Z can proxy for X.

Front-Doors*

Definition (Front-door)
Let X and Y be two variables in a causal structure $\mathcal{G}$. The set of variables Z satisfy the *front-door* criterion *from X to Y* if:

(a) Z blocks every *directed* path *from X to Y*,

(b) there is no unblocked back-door path from X to Z, and

(c) X blocks all back-door paths from Z to Y.

A Minimal Example

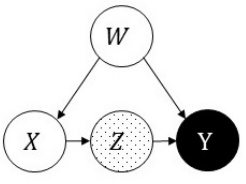

Figure 5: Example front-door through Z for identifying $X \rightarrow Y$.

W **confounds** $X \rightarrow Y$**.**
Assume W is unobserved, e.g., another trader's alpha. However:

(a) Z is observable,

(b) Z intercepts all $X \rightarrow ... \rightarrow Y$, and

(c) W does not *directly* affect Z.

Therefore, $Z \rightarrow Y$ proxies for $X \rightarrow ... \rightarrow Y$.

The Front-Door Criterion

Theorem (Front-door criterion)
Let $\mathcal{M}$ be a causal model with causal structure $\mathcal{G}$. Assume the set Z satisfies the front-door criterion from X to Y. Then one has the following identifiability equation for discrete variables

$$\mathbb{P}(Y = y | do(X = x)) = \sum_{z} \mathbb{P}(Z = z | X = x) \sum_{x'} \mathbb{P}(Y = y | X = x', Z = z) \mathbb{P}(X = x')$$

and the equivalent formula for continuous variables.

Application of the Front-Door Criterion (1/3)

Instrumental variable regression
Consider a linear causal model

$$\begin{aligned} X &= \eta W + \epsilon \\ Z &= \gamma X + \epsilon' \\ Y &= \theta Z + \kappa W + \epsilon''. \end{aligned}$$

An econometrician will (correctly) assign $\beta = \gamma \cdot \theta$ as the causal effect of X on Y. Note that

$$\begin{aligned} \mathbb{E}\left[Y \middle| X \right] &= \gamma \cdot \theta X + \kappa \mathbb{E}\left[W \middle| X \right] \\ &\neq \beta X + \kappa \mathbb{E}[W]. \end{aligned}$$

Direct regression is biased! Backdoor through W is unobservable.

Front-door

$$
\begin{aligned}
&\mathbb{E}\left[Y \middle| \mathrm{do}(X)\right] \\
&\quad = \sum_z \mathbb{P}\left(Z = z \middle| X = x\right) \sum_{x'} \mathbb{E}\left[Y \middle| X = x', Z = z\right] \mathbb{P}\left(X = x'\right) \\
&\quad = \sum_z \mathbb{P}\left(Z = z \middle| X = x\right) \sum_{x'} \left(\theta z + \kappa \mathbb{E}\left[W \middle| X = x'\right]\right) \mathbb{P}\left(X = x'\right) \\
&\quad = \sum_z \mathbb{P}\left(Z = z \middle| X = x\right) \theta z + \sum_z \mathbb{P}\left(Z = z \middle| X = x\right) \kappa \mathbb{E}[W] \\
&\quad = \theta E\left[Z \middle| X = x\right] + \kappa \mathbb{E}[W] \\
&\quad = \theta \cdot \gamma X + \kappa \mathbb{E}[W].
\end{aligned}
$$

Application of the Front-Door Criterion (3/3)

Instrumental variable regression
Therefore, in the linear case, the front-door criterion is equivalent to two regressions:

$$Z = \gamma X + \epsilon'$$
$$Y = \theta Z + \lambda X + \epsilon'''.$$

Then, discard λ, as it channels the unknown confounding through W. Instead, use $\beta = \gamma \cdot \theta$ as it channels direct paths $X \to ... \to Y$.

Machine learning application
Like the back-door criterion, the front-door criterion extends beyond linear econometrics to allow non-parametric models.

Disentangling Alpha from Impact

Prediction bias

Bouchaud et al. (2018)

“**Prediction bias 1:** *The larger the volume Q of a metaorder, the more likely it is to originate from a stronger prediction signal.” (p. 238)*

“**Prediction bias 2:** *Traders with strong short-term price-prediction signals may choose to execute their metaorders particularly quickly” (p. 238).*

Interpretation of the causal model

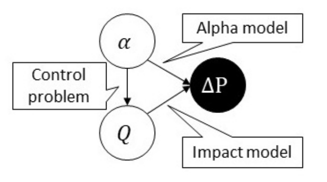

Figure 6: Causal graph $\mathcal{S}$.

(a) $\alpha \to \Delta P$ corresponds to the *alpha model*.
It states that the features captured by the alpha model cause price moves.

(b) $\alpha \to Q$ corresponds to the *control problem*.
It states that changes in alpha cause changes in trading behavior.

(c) $Q \to \Delta P$ corresponds to *price impact*.
It states that trades cause price moves.

Prediction bias

Proposition (Prediction bias)
Under the causal structure $\mathcal{S}$, one has

$$\mathbb{E}\left[\Delta P \mid do(Q)\right] \neq \mathbb{E}\left[\Delta P \mid Q\right].$$

The action associated with the impact counterfactual is identified conditioning on alpha

$$p\left(\Delta P \mid do(Q)\right) = \int p\left(\Delta P \mid Q, \alpha\right) p\left(\alpha\right) d\alpha.$$

Conclusion
"Simply" cofit impact with alpha.

A practical consideration (1/2)

Why conditioning on alpha is hard
The only way a researcher can estimate impact using observational data is if they condition on α. Furthermore, from a practical perspective, α should be the alpha *used point in time* by the live trading process that generated Q, *not the latest version of the alpha model*.

$$\underbrace{\Delta P \sim \beta \Delta I(Q)}_{\text{prediction bias}}; \quad \underbrace{\Delta P \sim \beta \Delta I(Q) + \gamma \alpha_{\text{latest}}}_{\text{prediction bias}}; \quad \underbrace{\Delta P \sim \beta \Delta I(Q) + \gamma \alpha_{\text{historical}}}_{\text{correct}}$$

Of the above three statistical regressions, only the third one will converge to the correct β as the size of the observational dataset increases.

A practical consideration (2/2)

Example (Impact-adjusted alpha research)
Consider a trader who can reproduce historical alphas but fits a novel model going forward. Let $\alpha_{\text{historical}}$ and α_{new} be the historical and new alpha signals. Then, a two-step regression resolves the ambiguity in a bias-free way.

(a) The historical regression

$$\Delta P = \Delta I(\lambda, Q) + \beta^h \alpha_{\text{historical}} + \epsilon$$

yields the correct λ for impact.

(b) Given the impact model $I(\lambda, Q)$, one computes *unperturbed* returns ΔS via $\Delta S = \Delta P - \Delta I(\lambda, Q)$. Finally, the regression

$$\Delta S = \beta^n \alpha_{\text{new}} + \epsilon'$$

provides the correct β^n for the novel alpha model.

Implementation Bias

Two trading strategies in a hierarchy

Implementation bias graph

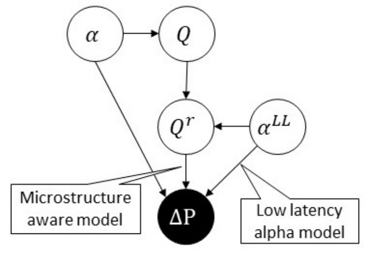

Figure 7: Strategic trading Q and tactical trading Q^r.

Two new options to measure price impact

Using micro-trades when alpha is unknown
Q^r satisfies the front-door criterion: one can use

$$Q = \gamma Q^r + \epsilon$$
$$\Delta P = \Delta I(Q, \lambda) + \Delta I(Q^r, \lambda^r) + \epsilon'$$

to estimate the impact $Q \to ... \to \Delta P$ without controlling for α!

Using a market making algo to estimate micro-impact
The action do($Q = 0$) corresponds to turning off the upstream algo and *only do market making*.

(a) This experiment estimates $Q^r \to \Delta P$ by controlling for α^{LL}.
(b) This experiment is free if you already have a market-making business.
(c) This experiment is *not* possible if you are not allowed to make markets, e.g. some brokers.

Example of experiment design (1/3)

Question
To estimate $\mathbb{E}\left[\Delta \mid \mathrm{do}(Q)\right]$, one must control for α. Is it ok to also control for α^{LL}?

Why this matters
Stakeholders may ask for excessive control variables. It's important to be able to tell when these hurt rather than help, and make systems robust to such choices.

Example of experiment design (2/3)

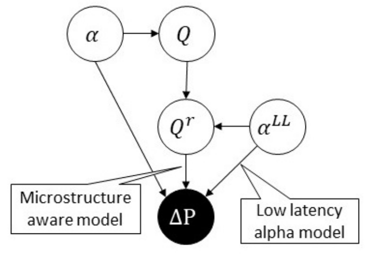

Figure 8: Clean hierarchy.

No causal bias
One can control for both α, α^{LL}:

$$E\left[\Delta P \middle| \mathrm{do}(Q)\right] = \int E\left[\Delta P \middle| Q, \alpha, \alpha^{LL}\right] p\left(\alpha, \alpha^{LL}\right) d\alpha d\alpha^{LL}.$$

Example of experiment design (3/3)

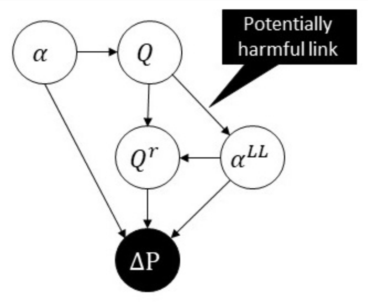

Figure 9: Alternative hierarchy.

No causal bias
One cannot control for both α, α^{LL}:

$$E\left[\Delta P \middle| \mathrm{do}(Q)\right] \neq \int E\left[\Delta P \middle| Q, \alpha, \alpha^{LL}\right] p\left(\alpha, \alpha^{LL}\right) d\alpha d\alpha^{LL}.$$

Weekly Summary

Methods to identify causal expressions

(a) The back-door criterion identifies control variables to block confounding paths.

(b) The front-door criterion funnels non-confounding paths through a proxy.

(c) For complex scenarios, use the three rules of do-calculus.

Questions?

Next week
No class!