Price Impact Models and Applications

Introduction to Algorithmic Trading

Kevin Webster

Spring 2023

Columbia University

Plan

Last Week
Bouchaud's list of four trading biases.

For this Week
The Mathematics of Causal Inference (1/2)

(a) Causal graphs
(b) *d*-separation criterion
(c) do() operator

Next Week
The Mathematics of Causal Inference (2/2)

Last Week's Summary

(a) **Prediction bias:**
Trades triggered by or considering an alpha signal exhibit bias when estimating their price impact.

(b) **Synchronization bias:**
"The impact of a metaorder can change according to whether or not other traders are seeking to execute similar metaorders at the same time."

(c) **Implementation bias:**
Tactical deviations from the strategic trading trajectory introduce biases.

(d) **Issuer bias:**
"Another bias may occur if a trader submits several dependent metaorders successively."

The Mathematics of Causal Inference

Mathematical setup

Simplified notation
Given a probability space $(\Omega, \mathcal{F}, \mathbb{P})$ with random variables X, Y and Z, one is interested in expressing statements about conditional distributions.

Equalities such as

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

should be understood as

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

holds for all valid outcomes x, y and z for the random variables X, Y and Z.

Refresher: Bayesian Statistics (1/4)

Observations as a change of probability measure
Let $(\Omega, \mathcal{F}, \mathbb{P})$ with random variables X, Y. Define the conditional probability of $Y = y$ given the observation $X = x$ as a new probability measure $\tilde{\mathbb{P}}$ over values of Y such that

$$\tilde{\mathbb{P}}(Y = y) = \frac{\mathbb{P}(Y = y, X = x)}{\mathbb{P}(X = x)}$$

Interpretation
The new probability measure is obtained by taking the subset of all events where $\{X = x\}$ happens, and re-normalizing them to a new probability measure.

Refresher: Bayesian Statistics (2/4)

Bayesian notation
For simplicity, one introduces the Bayesian notation

$$\mathbb{P}\left(Y = y \middle| X = x \right) = \tilde{\mathbb{P}}\left(Y = y \right).$$

Hence, in our compressed notation,

$$\mathbb{P}\left(Y \middle| X \right) = \frac{\mathbb{P}\left(Y, X \right)}{\mathbb{P}\left(X \right)}.$$

Refresher: Bayesian Statistics (3/4)

Bayes's Formula
Let $(\Omega, \mathcal{F}, \mathbb{P})$ with random variables X, Y. Then, the equation

$$\mathbb{P}(Y|X) = \frac{\mathbb{P}(X|Y)\mathbb{P}(Y)}{\mathbb{P}(X)}$$

holds. The corresponding change of probability measure on Y considers the information gained from observing X, also called the *Bayesian update rule.*

Refresher: Bayesian Statistics (4/4)

In the context where Y is a model parameter to be estimated,

(a) $\mathbb{P}(X|Y)$ is the likelihood of observing the data X given the model parameters Y.

(b) $\mathbb{P}(Y)$ is the prior on the model parameters Y. This is the distribution of Y prior to observing X.

(c) $\mathbb{P}(X)$ is the evidence.

(d) $\mathbb{P}(Y|X)$ is the posterior on the model parameters Y. This is the updated distribution of Y *after* observing X.

Causal Structure*

Definition (Causal structure)
A *causal structure* of a set of variables V is a directed acyclical graph (DAG) in which each node corresponds to a distinct element of V.

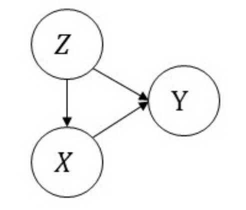


Figure 1: Example of a causal structure $\mathcal{G}$ for three variables X, Y and Z.

Causal Model

Definition (Causal model)
A *causal model* consists of

(a) A causal structure $\mathcal{G}$.

(b) A set of functions f_i compatible with $\mathcal{G}$,

$$f_i : (\text{parents}(x_i), \epsilon_i) \mapsto f_i(\text{parents}(x_i), \epsilon_i)$$

where parents(x_i) are outcomes of the *parent variables* of X_i, and ϵ_i a noise term idiosyncratic to X_i.

(c) A probability space $(\Omega, \mathcal{F}, \mathbb{P})$ that assigns probabilities to all the ϵ_i, with each ϵ_i being independent.

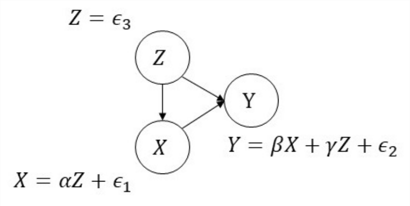
Example of a causal model (1/2)

A simple *linear model* can be attached to the causal structure $\mathcal{G}$ to promote it to a *causal model* $\mathcal{M}$.

$$X = \alpha Z + \epsilon_1$$
$$Y = \beta X + \gamma Z + \epsilon_2$$
$$Z = \epsilon_3$$

for parameters α, β and γ. The noise terms ϵ_i are independent Gaussians with mean 0 and variance σ_i respectively.

Example of a causal model (2/2)



Constructivist point of view

(a) A causal model begins by assigning probabilities to the ϵ_i.

(b) These propagate along the graph to assign a joint probability distribution over the variables X_i.

(c) The causal DAG creates a natural template for a computational DAG, improving reproducibility for numerical simulations in High Performance Computing (HPC).

HPC implementation a Monte-Carlo Simulation using a Graph

```
sampleZ:{[seed]
        ... \Swap the implementation of this node based on
    your model
        :Z;}

sampleX:{[Z; seed]
        ... \Swap the implementation of this node based on
    your model
        :X;}

sampleY:{[Z; X; seed]
        ... \Swap the implementation of this node based on
    your model
        :Y;}

\The propagation pattern is independent of the nodes'
    implementation and only relies on the causal graph
tbl: update Z: sampleZ[seed] by seed from tbl;
tbl: update X: sampleX[Z; seed] by seed from tbl;
tbl: update Y: sampleY[Z; X; seed] by seed from tbl;
```

Causal structures as constraints on probability measures

Definition (Consistency)
A probability distribution $\mathbb{Q}$ over the set of variables V is said to be consistent with a causal structure $\mathcal{G}$ if there exists a causal model $\mathcal{M}$ based on the causal structure $\mathcal{G}$ that generates $\mathbb{Q}$ for the variables V.

Probabilistic point of view
Causal structures can be seen as *constraints* placed on a probability measure.

A causal statement under $\mathcal{G}$
is a statement that holds for *any* probability measure consistent with the causal structure $\mathcal{G}$.

d-separation

Three Special Paths*

Definition (DAG paths)
Let $\mathcal{G}$ be a DAG. Define a *path* connecting nodes X and Y as a sequence of consecutive edges, *regardless of their directionality*, leading from X to Y.

Furthermore, define the following three-node paths for three nodes i, m, and j.

(a) *A chain*

$$i \rightarrow m \rightarrow j$$

(b) *A fork*

$$i \leftarrow m \rightarrow j$$

(c) *A collider*

$$i \rightarrow m \leftarrow j$$

Causal independence (d-separation)*

Definition (d-separation)
Let $\mathcal{G}$ be a DAG. Define a path p to be d-separated (or blocked) by a set of nodes Z if and only if

(a) the path p contains a chain $i \rightarrow m \rightarrow j$ or a fork $i \leftarrow m \rightarrow j$ such that m is in Z.

(b) the path p contains a collider $i \rightarrow m \leftarrow j$ such that neither m nor its descendants are in Z.

Define two sets of nodes X and Y to be d-separated by the set Z if Z d-separates every path from a node in X to a node in Y.

Causal independence implies probabilistic independence

Theorem (Probabilistic implications of d-separation)
Let $\mathcal{G}$ be a causal structure and X, Y, and Z be three sets of variables on $\mathcal{G}$.

(a) If Z d-separates X and Y, for every probability measure $\mathbb{P}$ consistent with $\mathcal{G}$, X and Y are conditionally independent given Z.

(b) Conversely, if Z does not d-separate X and Y, then there exists at least one probability measure $\mathbb{P}$ consistent with $\mathcal{G}$ for which X and Y are not conditionally independent given Z.

Examples of d-separation (1/2)*

Crowding bias graph $\mathcal{C}$

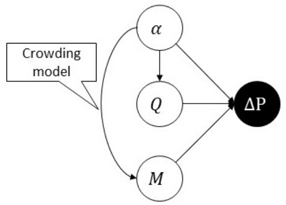


Figure 2: Two traders Q, M trade with a common alpha α. Note that each trader may consider additional, independent alphas.

Under $\mathcal{C}$, α d-separates Q and M.

(a) $Q \not\perp M$ due to the fork $Q \leftarrow \alpha \rightarrow M$.

(b) For any probability measure consistent with $\mathcal{C}$, Q and M are independent conditional on α.

Examples of d-separation (2/2)

Survivorship bias in a crosser

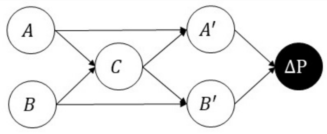


Figure 3: Two independent algorithms A, B cross internally in a crosser C to save the bid-ask spread. The resulting market trades are A', B'.

For any implementation of the above graph,

(a) $A \perp B$ and $A' \not\perp B'$.

(b) Conditional on C, $A \not\perp B$ and $A' \perp B'$.

Document your trading algorithms' dependency graph to uncover self-inflicted trading biases!

Causal independence is equivalent to probabilistic independence

General result (out of scope for class)
There is a more general result for *stable* sets of probability measures.

(a) A set of probability measures is *stable* with respect to a causal structure if
   (*) they are consistent with the graph and
   (*) there is no probability measure for which a causal link is trivial (e.g., there doesn't exist a parameter choice for which the correlation is zero).

(b) For stable sets of probability measures, *d*-separation and probabilistic independence are equivalent.

(c) Conversely, this equivalence characterizes stable sets of probability measures. This is the probabilistic point of view on stability.

Revisiting the Monty Hall Problem (1/3)

The Monty Hall problem, named after the original host of the American television game show "Let's make a deal", involves a game with four rules:

(*) The host presents three doors to a game participant. Behind one of the three doors, chosen by the host before the start of the game, stands a valuable prize.
Mathematically, let H be the variable denoting the door choice made by the host. H takes value in $\{1, 2, 3\}$.

(*) The participant chooses one of the three doors. The host does not reveal the chosen door at this stage.
Denote by P this choice made by the participant. P takes value in $\{1, 2, 3\}$ and is independent of H.

Revisiting the Monty Hall Problem (2/3)

(*) The host then *reveals* one of the remaining worthless doors. Denote by R the revealed door. R takes value in the set difference $\{1, 2, 3\} \setminus \{P, H\}$ given knowledge of P and H.

(*) Given the revealed worthless door R, the participant has the option to reveal the door P chosen initially or to pick the third door. Their final door choice determines their prize.

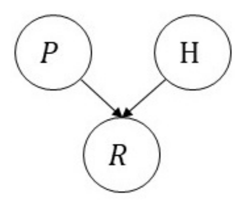


Figure 4: Causal graph $\mathcal{H}$ for the Monty Hall problem. P and H represent the participant and host's independent choices. R is the revealed door, which depends on both P and H.

Revisiting the Monty Hall Problem (3/3)

From d-separation, under the causal graph $\mathcal{H}$, one has

$$\mathbb{P}(H | P) = \mathbb{P}(H),$$

as $P \perp H$, but

$$\mathbb{P}(H | P, R) \neq \mathbb{P}(H | R)$$

due to the collider path $P \rightarrow R \leftarrow H$.

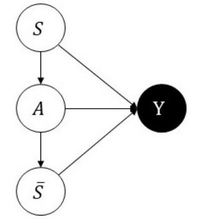
Pearl (2009)

> *"Observations on a common consequence of two independent causes tend to render those causes dependent."*

Conditional on observing the revealed door, the decisions made by the host and the participant become (statistically) dependent.

Simpson's Paradox in Transaction Cost Analysis (TCA)

Example causal structure $\mathcal{A}$ for evaluating trading algorithms.



Interpretation

(a) An order of size S is submitted upstream.

(b) A trading algorithm A from a set $\{A_1, ..., A_k\}$ of algorithms is chosen based on the order size S

(c) The algorithm A affects the trading speed $\bar{S}$ of the order.

(d) Y measures the *arrival slippage* of the order.

Proposition (An example of a causal statement)
S and $\bar{S}$ are independent conditional on A under any probability measure consistent with $\mathcal{A}$.

The mathematics of causal inference

Summary so far

(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) Causal structures $\mathcal{G}$ encapsulate a (potentially large) number of conditional independence assumptions in a (hopefully) intuitive format.

Next step
Introduce *do-calculus*, an extension to Bayes' rule under a given causal structures.

A motivational example: Simpson's paradox

A	S	Sample size	$E[Y\|A,S]$
a	s	40k	-10bps
a	l	10k	-40bps
p	s	10k	-5bps
p	l	40k	-25bps

Figure 5: Expected arrival slippages across order sets.

Discrete case

Assume that A and S take binary values in $\{a, p\}$ and $\{s, l\}$ respectively. a stands for aggressive, p for passive, s for small and l for large.

The aggressive algorithm was allocated smaller orders.

Simpson's paradox (1/4)

A	S	Sample size	$E[Y\|A,S]$
a	s	40k	-10bps
a	l	10k	-40bps
p	s	10k	-5bps
p	l	40k	-25bps

Two performance estimators
Note the unconditional estimators

$$\mathbb{E}[Y|A=a] = -16; \quad \mathbb{E}[Y|A=p] = -21$$

lead to the opposite conclusion from the conditional estimators $\mathbb{E}[Y|A, S=s]$ and $\mathbb{E}[Y|A, S=l]$.

Simpson's paradox (2/4)

Two formulas, two conclusions
The apparent paradox is resolved in a Bayesian world by noting the negative result

$$\mathbb{E}[Y|A] \neq \sum_{x \in \{s,l\}} \mathbb{E}[Y|A, S=x]\,\mathbb{P}(S=x).$$

Indeed, by the tower property,

$$\mathbb{E}[Y|A] = \sum_{x \in \{s,l\}} \mathbb{E}[Y|A, S=x]\,\mathbb{P}(S=x|A).$$

Hence, whenever $S \not\perp A$, the two formulas disagree!

Simpson's paradox (3/3)

But which formula is correct?
Bayes does not provide an answer. Intuitively, traders will tell you that one should condition by S to "compare apples to apples"

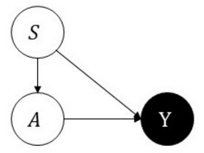
But which formula is correct?
Those same traders will flip their answer if S is replaced with $\bar{S}$!

This is because the trade size S was "forced" onto the algorithm A, but the trade speed $\bar{S}$ was a "decision" made by A.

Simpson's paradox (4/4)

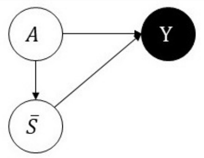
Summary
Goal is to mathematically formalize the trader intuition to establish the performance of an algorithm under a given causal structure.

Conditioning on S



$$\sum_{x \in \{s,l\}} \mathbb{E}\left[Y | A, S = x \right] \mathbb{P}(S = x).$$

Not conditioning on $\bar{S}$



$$\mathbb{E}\left[Y | A \right].$$

do()-actions and Interventions

Definition (Causal model)
A *causal model* consists of

(a) A causal structure.

(b) A set of functions f_i compatible with the causal structure,

$$f_i : (\text{parents}(x_i), \epsilon_i) \mapsto f_i(\text{parents}(x_i), \epsilon_i)$$

where parents(x_i) are outcomes of the *parent variables* of X_i, and ϵ_i a noise term idiosyncratic to X_i.

(c) A probability space $(\Omega, \mathcal{F}, \mathbb{P})$ that assigns probabilities to all the ϵ_i, with each ϵ_i being independent.

Introducing the do()-operator

"Counterfactuals"
The action 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 do(X) action as follows

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

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

Refresher

Observations as a change of probability measure
Let $(\Omega, \mathcal{F}, \mathbb{P})$ with random variables X, Y. Define the conditional probability of $Y = y$ given the observation $X = x$ as a new probability measure $\tilde{\mathbb{P}}$ over values of Y such that

$$\tilde{\mathbb{P}}(Y = y) = \frac{\mathbb{P}(Y = y, X = x)}{\mathbb{P}(X = x)}$$

Interpretation
The new probability measure is obtained by taking the subset of all events where $\{X = x\}$ happens, and re-normalizing them to a new probability measure.

Example

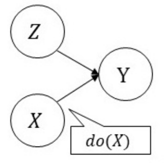


Figure 6: Visual representation of the action do(X) on the causal graph $\mathcal{S}$.

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

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

Example

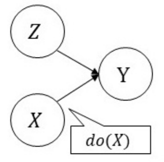


Figure 7: Visual representation of the action do(X) on the causal graph $\mathcal{S}$.

Constructivist point of view

```
tbl: update Z: sampleZ[seed] by seed from tbl;
/tbl: update X: sampleX[; seed] by seed from tbl; /
    commenting out this link in the graph!
tbl: update X: x by seed from tbl; /constant intervention
tbl: update Y: sampleY[Z; X; seed] by seed from tbl;
```

Probabilistic point of view on *do*() action (1/2)*

Definition (Trimming causal graphs)
Let X and Y be two sets of variables on a DAG $\mathcal{G}$. Define the following *link erasures*.

(a) Define *erasing the parents* of X as the graph $\mathcal{G}_{\overline{X}}$ where one erases all links pointing to nodes in X from $\mathcal{G}$.

(b) Define *erasing the children* of X as the graph $\mathcal{G}_{\underline{X}}$ where one erases all links emerging from nodes in X from $\mathcal{G}$.

(c) One can combine multiple erasure operations by appending graph indexes. For example, $\mathcal{G}_{\overline{X}\underline{Y}}$ erases the parents of X and the children of Y.

Probabilistic point of view on $do()$ action (2/2)

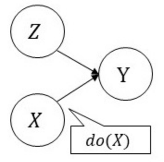


Figure 8: Truncated causal graph $\mathcal{G}_{\overline{X}}$.

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 $\mathcal{G}_{\overline{X}}$.

Example

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

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

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

Example on the linear causal model*

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

Relationship with AB testing

Slight extension
The do$(X = x)$ operator can be extended to an independent random variable x.

AB test example
Consider do$(A = \epsilon)$ where

$$\mathbb{P}\left(\epsilon = a\right) = \mathbb{P}\left(\epsilon = p\right) = \frac{1}{2}.$$

Interventional data
One refers to data generated under $\tilde{\mathbb{P}}$ as *interventional data*, in contrast to the *observational data* generated by $\mathbb{P}$.

$$\tilde{\mathbb{P}}\left(A = a\right) = \tilde{\mathbb{P}}\left(A = p\right) = \frac{1}{2}.$$

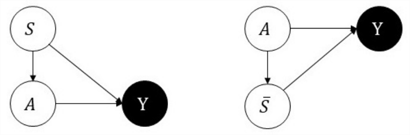
General goals of the do()-operator

Causal bias
If the naive Bayesian estimator for a causal expression doesn't hold, one says that a *causal bias* is present.

Fixing a causal bias
There are two methods.

(a) A non-naive identification formula can be proven using *do-calculus*.
(b) An *interventional dataset* can be designed such that a naive identification formula holds.

Example on Simpson's paradox



Proposition
For any probability measure consistent with the above causal structures:

$$\mathbb{E}[Y|do(A)] = \sum_{x\in\{s,l\}} \mathbb{E}[Y|A, S = x] \cdot \mathbb{P}(S = x)$$

and

$$\mathbb{E}[Y|do(A)] \neq \sum_{x\in\{s,l\}} \mathbb{E}\left[Y|A, \bar{S} = x\right] \cdot \mathbb{P}\left(\bar{S} = x\right),$$

in line with trader intuition.

Proof.
Next week! □

Weekly 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 $\mathrm{do}(X = x)$.

Questions?

Next 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