Association-based vs Causal Research:
the Hype, the Contrasts, and the Stronger-than-expected Complementary Overlaps

QuantStrats, Roundtable2, NYC, March 12, 2024

JD Opdyke, Chief Analytics Officer, Partner, Sachs Capital Group Asset Management

Background:
"The clarion call for causal reduction in the study of capital markets is intensifying. However, in self-referencing and open systems such as capital markets, the idea of unidirectional causation (if applicable) may be limiting at best, and unstable or fallacious at worst." Polakow et al., 2023

"…the distinction between prediction and causation, taken to its limit, melts away." Daoud & Dubhashi, 2023

"… a DAG structure recovering algorithm, which is **based on the Cholesky factorization of the covariance matrix of the observed data** (emphasis added). …achieves the state-of-the-art performance." Cai et al., 2023

"Most of the literature on causality considers the structural framework of Pearl and the potential-outcomes framework of Neyman and Rubin to be formally equivalent, and therefore interchangeably uses the do-notation and the potential-outcome subscript notation to write counterfactual outcomes. In this paper, we … prove that structural counterfactual outcomes and potential outcomes do not coincide in general – not even in law." De Lara, 2023.

"… potential outcomes (PO) and structural causal models (SCMs) stand as the predominant frameworks. However, these frameworks face notable challenges in practically modeling counterfactuals … we identify an inherent model capacity limitation, termed as the "degenerative counterfactual problem", emerging from the consistency rule that is the cornerstone of both frameworks. ..we introduce a novel *distribution-consistency* assumption … We hope it opens new avenues for future research of counterfactual modeling, ultimately enhancing our understanding of causality and its real-world applications." Gong et al, 2024

"Try to come up with some form of efficient frontier for the 500 constituents of the SP500 with a causal model. Or ask yourself if that even make sense on the first place.", Alejandro Rodriguez Dominguez, 2024, Causal Researcher, Head of Quantitative Analysis at Miraltabank

"The explicit study of causality in AI fields has officially hit the 'hype cycle', at least according to Gartner" Grimbly, 2022

"We show that in the presence of cycles, many of the convenient properties of acyclic SCMs do not hold in general: they do not always have a solution; they do not always induce unique observational, interventional and counterfactual distributions; a marginalization does not always exist, and if it exists the marginal model does not always respect the latent projection; they do not always satisfy a Markov property; and their graphs are not always consistent with their causal semantics.", Bongers et al., 2021

"we describe the recent debate on whether the hazard function should be used for causal inference in time-to-event studies and consider three different potential outcomes frameworks (by Rubin, Robins, and Pearl, respectively) as well as use the single-world intervention graph to show mathematically that the hazard function has causal interpretations under all three frameworks" Ying and Xu, 2023

Talking Points:
The advent of causal modeling (CM) was a seminal paradigm shift, providing original methodologies to better answer what many claim were previously unanswerable, or at least only partially answerable research questions under association-based modeling (ABM). However, CM is not new. The original contributions of Pearl (1996, 2000: structural causal models / DAGs / do calculus) and Angrist, Imbens, and Rubin (1996) and Imbens and Rubin (2006) (potential outcomes) over a quarter century ago sat under-appreciated and under-utilized for far too long, making the recent explosion of CM in applied settings long overdue.

Yet perhaps in part because of this delay, and also perhaps inevitably, CM arguably is now part of a 'hype cycle' (Grimbly, 2022) wherein its ranges of appropriate application are sometimes stretched beyond what is scientifically defensible in many settings. This happens with many groundbreaking paradigms, and in an imperfect world this may be the only way to discover exactly where those appropriate boundaries and limits of application reside in specific settings (a flawed but not unreasonable analogy is to Efron's discovery of the bootstrap (1979) and its subsequent, ubiquitous implementation). But as the CM research dust begins to settle, we must remain cognizant of, and vigilant regarding, research(ers) caught up in the hype cycle, overstating the originality and/or breadth and/or distinct uniqueness of setting-specific solutions proffered by CM, especially when presented strictly and solely as a superior alternative to ABM. Reasons for the need for such vigilance are manyfold:

1. there appear to be unforeseen, fundamental, and perhaps irreconcilable contradictions between the basic causal paradigms of structural causal models and potential outcomes (De Lara, 2023)
2. there remain many settings in which a CM framework doesn't (yet?) make much sense (Rodriguez Dominguez, 2024);
3. in contrast to some early claims, and the persistent marketing of many, the use cases for quasi-causal modeling based on observational data and traditionally ABM methods appear to be growing rapidly (see Liu et al., 2021; Ying and Xu, 2023; and Cai et al., 2023)
4. the definitions of what causality means have been rigorously re-examined and sometimes called into question, and time-tested ABM frameworks have been shown to have straightforward, causal interpretations (Ying and Xu, 2023)
5. much of what is viewed as entirely and separately original in CM is actually incremental advancement on, and/or convergence with, previously established ABM results (Daoud & Dubhashi, 2023)

The main point here obviously is not to denigrate CM, as the pace of new, promising causal model development has never been more impressive (one such example for causal portfolio analysis uses information geometry and focuses on causal manifolds, which can more readily incorporate time as a dynamic variable compared to other causal models (see Rodriguez Dominguez, 2024)). Rather, the point is to avoid a hype cycle and encourage CM's scientifically responsible application via recognition of the modest thesis proposed herein: simply, that while the differences between CM and ABM are, in fact, sometimes sharp and distinct and clearly defined, the areas of overlap, continuation and continuity, and blurred lines appear to be much greater than the differences. What's more, CM researchers will unnecessarily limit the breadth and utility of their work if they succumb to the siren's songs that portray all of CM as a sharp break from the ABM past, if not an unambiguously superior and self-contained mode of inquiry entirely separate from ABM. This is a fairly widespread but misleading strawman, and in some cases a disingenuous one.

In contrast and in reality, some of the most cutting edge CM work is built squarely and transparently on ABM. One recent and compelling example (Cai et al., 2023) is based on the unarguable "king of ABM" – the covariance matrix – and uses its Cholesky factor to not only perform significantly faster, but also to achieve state-of-the-art performance in recovering the ground truth directed acyclic graphs. Can, and should, this still be classified strictly as "CM"? What good (or harm!) do such bright-line classifications do? Who benefits, and who is harmed, from such rigid classifications?

Bringing this home, one of today's QuantStrats presentations (Opdyke, "Beating the Correlation Breakdown") uses this exact construct – the Cholesky factor of the correlation/concordance matrix – to derive the finite sample distributions of ALL positive definite, matrix-based measures of dependence, under challenging, real-world data conditions. Can this, too, be directly and readily useful in CM contexts? Given the above, should we still be defining this rigidly and strictly as an ABM method/framework at all? These lines are becoming increasingly blurred, and rightly so. The old chants of "correlation is not causation" are not expanding our knowledge frontier, and can even be misleading. Correlation may not be causation, but neither is it not NOT causation, as the above research shows. Simplistic, tired mantras like those remain very limiting for responsible causal research that seeks broad and effective application, and that can benefit tremendously by avoiding its hype cycle, which would wrongly classify it as superior to and entirely separate from association-based modeling.

In summary, even though causal modeling is not new, its widespread application is, and we are still learning new, surprising, and critically important things in this field (see De Lara, 2023). Researchers of all persuasions will benefit by identifying and avoiding its hype cycle, eschewing artificial, strawman divisions between it and association-based modeling, and agnostically and scientifically embracing the most effective methodological combinations of both.