Joint Discussion Paper Series in Economics

by the Universities of
Aachen • Gießen • Göttingen
Kassel • Marburg • Siegen

ISSN 1867-3678

No. 01-2026

Bernd Hayo

Economic Causality in Light of Philosophical Concepts: Distinctly Idiosyncratic

Coordination: Bernd Hayo
Philipps-University Marburg
School of Business and Economics
Universitätsstraße 24, D-35032 Marburg
Tel: +49-6421-2823091, Fax: +49-6421-2823088, e-mail: hayo@wiwi.uni-marburg.de

Economic Causality in Light of Philosophical Concepts:

Distinctly Idiosyncratic

Bernd Hayo

Philipps-Universität Marburg

This version:

16 May 2026

Bernd Hayo
Macroeconomic Research Group
School of Business and Economics
Philipps-Universität Marburg
D-35032 Marburg
Germany
Phone: +49-6421-2823091
Email: hayo@wiwi.uni-marburg.de

Acknowledgements

I thank Andreas Landmann, Claudia Schwirplies and Johannes Zahner for helpful comments. All remaining errors are my own.

Economic Causality in the Design-Based Tradition:

Distinctly Idiosyncratic

Abstract

Design-based econometrics, centred on average treatment effects in the potential-outcomes framework, treats causal effects as population-level expectations under hypothetical interventions and thereby offers an expectation-based, mechanism-agnostic conception of causality. This conception is counterfactual in form yet local, context-dependent and largely 'black-box' in content, making it distinctly idiosyncratic by comparison with counterfactual, regularity, dispositional and interventionist traditions. The paper situates design-based causality within this taxonomy, highlighting its reliance on properties of an underlying data-generating process and the resulting importance of external validity. It then places core identification strategies – difference-in-differences, instrumental variables and regression discontinuity designs – within a genealogy extending from medieval causal logic through Mill and Fisher to the Neyman–Rubin framework. This perspective shows that robustness and transportability are as epistemologically central as identification.

Keywords: Economic causality; economic methodology; design-based econometrics; causal inference; counterfactual, dispositional and regularity causation; policy evaluation

JEL codes: B40; C10; C90

1. Introduction

Mostly Harmless Econometrics by Angrist and Pischke (2009), together with related work, has had a marked influence on empirical economics. It has contributed decisively to establishing what is now widely regarded as best practice in applied microeconometrics: empirical analysis should rely on experimental or quasi-experimental designs that allow for a clear causal interpretation. Methods commonly associated with this 'design-based' perspective include difference-in-differences (DiD), regression discontinuity designs (RDD), instrumental variable estimation (IV) and randomised controlled trials. These tools are typically formulated within the Neyman–Rubin potential-outcomes framework, in which causal effects are differences between hypothetical outcomes under treatment and control.

This design-based strand is only one part of the broader econometric landscape. Other approaches, such as traditional linear regression and vector autoregressive (VAR) models, can be interpreted as instantiating forms of probabilistic or regularity-based causality (e.g. Moneta, 2003; Maziarz and Mróz, 2020), while structural econometric and macroeconomic models articulate causal structure through systems of equations and behavioural assumptions. The present paper deliberately focuses on the design-based potential-outcomes tradition, rather than on econometrics in general. Its concern is how causality is conceptualised and operationalised in this specific and now highly influential part of empirical economics.

Within the potential-outcomes framework, each observational unit i is associated with at least two potential outcomes: $Y_i(1)$ under treatment and $Y_i(0)$ under control. The causal effect for unit i is often written as $\tau_i = Y_i(1) - Y_i(0)$, while empirical analysis typically focuses on population-level parameters such as the average treatment effect $\tau = E[\tau_i]$. Formally, one should think of $Y_i(0)$ and $Y_i(1)$ as random variables $Y_i(0, \omega)$ and $Y_i(1, \omega)$ defined on an underlying probability space Ω, where ω indexes states of the world. Design-based econometrics is primarily concerned with identifying and estimating expectation-valued causal parameters such as $E[Y_i(1, \omega) - Y_i(0, \omega)]$, rather than with attributing causal responsibility for particular realised outcomes $Y_i(D_i, \omega)$. This expectation-centred conception of causality sits uneasily with classical philosophical traditions that locate causality in concrete events, stable powers or regular successions in the world.

The existing literature on economic methodology and econometric causality has addressed many aspects of this development, but not the particular combination considered here. Classic methodological works such as Blaug (1992), Hausman (1992) and Boumans and Davis (2010) discuss the role of econometrics in economics without focusing on the design-based causal turn. Earlier econometric methodology, for example Pagan (1987), Darnell and Evans (1990) and Hayo (1998), predates the widespread adoption of the potential-outcomes framework. The new approach to econometrics is discussed from a variety of perspectives in a symposium published in the *Journal of Economic Perspectives* in 2010. After a summary of the main arguments by Angrist and Pischke (2010), prominent empirical researchers provide their own views. Leamer (2010) is sympathetic – after all, the claim of Angrist and Pischke is to 'Take the Con out of Econometrics' – but remains critical. Even harsher is the response by Sims (2010), who flatly states: 'The fact is, economics is not an experimental science and cannot be' (p. 59). However, none of these contributions relates this new concept of economic causality to those discussed in philosophy. Heckman (2000, 2008) and Heckman and Pinto (2024) trace the history of ‘econometric causality’ with particular emphasis on structural models, policy parameters and the evolution of causal thinking within economics. Hoover (2001) examines causality in macroeconomics, again with a strong structural flavour. More recently, Maziarz (2020) connects modern econometric approaches to counterfactual, dispositional and regularity theories of causation, while Crespo (2020) and Henschen (2025) offer broader philosophical perspectives on causality in economics and macroeconomic policy.

Against this background, the purpose of the present paper is fourfold. First, it situates design-based economic causality within a philosophical taxonomy that encompasses counterfactual, regularity and interventionist accounts, whilst also considering dispositional views. I argue that the potential-outcomes framework is counterfactual in form, but that its emphasis on population expectations and local causal parameters (such as LATE) produces a distinctive hybrid: causal claims are about expectations of stochastic processes rather than about individual events or stable powers.

Second, the paper highlights the expectation-based nature of design-based causality. Causal parameters such as the average treatment effect are properties of the data-generating process, not of any single realised history; they may be well defined and identifiable even if no individual outcome ever coincides with the estimated causal effect. This feature, while familiar from a technical point of view, has received limited explicit philosophical discussion.

Third, the paper argues that modern identification strategies such as DiD, IV and RDD can be illuminated by placing them within a longer intellectual genealogy extending back to medieval causal logic. William of Ockham’s formulation of a Method of Difference and Duns Scotus’s Concurrence Method anticipated two complementary epistemic strategies for causal analysis: isolating differences under controlled comparison and testing whether putative causes and effects reliably occur together across varying contexts. Mill’s nineteenth-century restatement of the Method of Difference, Fisher’s development of experimental design and the Neyman–Rubin potential-outcomes framework can be read as successive formal elaborations of these ideas. The design-based ‘credibility revolution’ in econometrics thus stands in a longer tradition of thinking about how to use data for causal inference.

Finally, the paper examines the epistemological consequences of the design-based emphasis on identification and expectation-valued parameters. In current practice, Ockham-style concerns – ensuring that treatment status is as good as random, controlling for confounders, exploiting quasi-experimental variation – have clear priority. Scotus-style concerns with the stability and transportability of causal relations across contexts are often treated as secondary robustness checks. The analysis here suggests that, if causality is understood as a property of the data-generating process, then questions of robustness and external validity are not optional extras but integral to causal assessment. Treating robustness and transportability as a modern analogue of the Concurrence Method helps to clarify why design-based results may fail to travel across institutional settings and why mechanism-based analysis, while not required for identification in a given context, becomes crucial for generalisation.

The remainder of the paper proceeds as follows. Section 2 outlines the design-based framework and relates it to major philosophical traditions. Section 3 examines the expectation-based nature of these causal parameters and the tension between local identification and external validity. Section 4 contrasts the Ockhamite Method of Difference with the Scotist Concurrence Method to illuminate modern robustness practice. Section 5 concludes by summarising the philosophical distinctiveness of design-based economic causality and its implications for empirical work.

2. Design-based causality in economics and philosophical traditions

Design-based econometric causality traces its formal origins to two foundational contributions. Neyman (1923) articulated the principle that causal inference requires the comparison of potential outcomes under different treatments, a framework later generalised by Rubin (1974). Its adoption in economics gained momentum in the mid-1990s and became mainstream following the work of Angrist, Imbens, and Rubin (1996) and the methodological codification by Angrist and Pischke (2009). Within this framework, each unit i is associated with (at least) two potential outcomes: $Y_i(1)$ under treatment and $Y_i(0)$ under control. The individual causal effect is $\tau_i = Y_i(1) - Y_i(0)$, and the average treatment effect (ATE) is $\tau = E[\tau_i]$, defined as an expectation with respect to the relevant distribution over units and states of the world (Imbens and Rubin, 2015).

Several characteristics of this conception of causality are noteworthy. First, causality is explicitly counterfactual: it is defined over hypothetical outcomes, not observed regularities. Second, causal effects are defined at the level of individual units, even though empirical analysis focuses on population-level parameters such as the ATE or local average treatment effects (LATE). Thirdly, the design-based framework is largely mechanism-agnostic: it specifies what changes under intervention, not why. Finally, causality is tied directly to hypothetical interventions, rather than to temporal precedence or mere statistical association.

It is important to stress that this design-based potential-outcomes framework is one strand within econometrics, not a characterisation of econometrics as a whole. Other econometric traditions, such as linear regression and vector autoregressive (VAR) models, can be read as instantiating probabilistic or regularity-based notions of causality (Moneta, 2003; Maziarz and Mróz, 2020), while structural equation models articulate causal structure via systems of behavioural and technological relations. The present section therefore considers how design-based economic causality relates to four prominent philosophical families: counterfactual, dispositional, regularity, and interventionist accounts.

Relation to the counterfactual concept of causality

Among philosophical theories of causation, counterfactual accounts most obviously resemble design-based practice. Modern counterfactual theories, notably those of Lewis (1973, 1986), analyse causation in terms of counterfactual dependence: event A causes event B if, had A not occurred, B would not have occurred. The potential-outcomes framework mirrors this logic. Each potential outcome corresponds to a counterfactual scenario, and causal effects are defined as differences across such scenarios.

Design-based methods operationalise counterfactual reasoning through concrete identification strategies. In a standard difference-in-differences (DiD) setting, the estimand is interpreted as the average treatment effect on the treated, recovered by comparing the treated group's observed trajectory with its counterfactual trajectory inferred from the control group (Angrist and Pischke, 2009). Let D_{it} denote treatment status and $Y_{it}(d)$ the potential outcome for unit i in period t under treatment state $d \in \{0,1\}$. Under the parallel-trends assumption, the average change in untreated potential outcomes is the same in both groups.

In instrumental variables (IV) designs, an instrument Z_i isolates exogenous variation in treatment D_i. Under the exclusion and monotonicity assumptions, the LATE identifies the causal effect for those units whose treatment status is shifted by the instrument (Imbens and Angrist, 1994). Regression discontinuity designs (RDD), in turn, exploit discontinuities in treatment assignment at a threshold in a running variable X_i. Under continuity of potential outcomes in X_i at the cut-off, the jump in the conditional expectation of the outcome at the threshold identifies a local treatment effect for units at that boundary. (Lee and Lemieux, 2010).

In each case, the key causal quantities are expectations defined over potential outcomes in counterfactual scenarios. Yet there are also important differences between design-based counterfactuals and classical philosophical counterfactuals. Philosophical discussions typically concern singular historical or metaphysical claims – for example, whether the assassination of Archduke Franz Ferdinand caused the First World War. Design-based econometrics, by contrast, is oriented towards population-level questions such as what would happen, on average, to earnings or health under an alternative policy regime. The question is not whether a particular individual's outcome would have differed under a different treatment, but what the expected outcome would be across a relevant group. However, design-based estimands are often local rather than universal. The IV LATE pertains only to 'compliers' and may differ from the population ATE (Imbens and Angrist, 1994). Causal parameters are thus defined relative to specific populations, instruments and settings, complicating any analogy with the universal counterfactuals that often preoccupy philosophers. Design-based causality is thus a pragmatic, expectation-based variant of counterfactual causation: it uses counterfactual structure but orients it firmly towards policy-relevant averages.

Relation to the dispositional concept of causality

Dispositional accounts emphasise that systems have causal powers or capacities which, under suitable conditions, bring about particular effects (Harré and Madden, 1975; Mumford and Anjum, 2011). A fragile glass has the capacity to shatter when struck; an expansionary monetary policy may have the capacity to raise output, given appropriate institutional conditions. Under such accounts, causation involves the manifestation of these powers, often in conjunction with supporting or interfering conditions. In economics, this way of thinking appears most naturally in structural and macroeconomic modelling. Cowles-style structural equations and modern DSGE models attribute behaviour to parameters representing preferences, technologies and policy rules, which can be read, in Cartwright's terms, as capacities that generate outcomes under appropriate 'shielding' conditions (Cartwright, 1988, 1989; Hoover, 2001). Cartwright explicitly argues that at least some econometric and structural models are committed to such capacities.

By contrast, mainstream design-based econometric practice, especially reduced-form applications of DiD, IV and RDD, largely eschews explicit dispositional commitments. These methods do not typically model causal capacities or mechanisms, nor do they assert that a treatment possesses an intrinsic power to produce a given outcome in all contexts. Instead, they estimate probabilistic, context-dependent effects which may vary across populations and settings (Cartwright, 1989, 2007). A policy that raises employment in one jurisdiction may have a weaker or even negligible effect in another; such variation is routinely accommodated rather than treated as a violation of causal principles. The relationship is therefore asymmetric: structural traditions incorporate weak dispositional commitments by tying causal claims to explicit mechanisms, whereas reduced-form design-based econometrics deliberately avoids embedding strong metaphysical views about powers or capacities in its core inferential machinery. It prioritises answers to questions of the form 'What happens to the expected outcome if we implement this intervention in this context?' over questions of the form 'Why does this treatment have the power to produce this outcome across contexts?'. This makes structural macroeconomics a much closer fit to dispositional philosophies of causation than the design-based microeconometric tradition.

Relation to the regularity concept of causality

Humean accounts (Hume, 1739) treat causal relations as regular patterns of succession in experience, marked by constant conjunction and temporal precedence, and reject appeals to deeper causal mechanisms. Modern regularity theories, such as Mackie's INUS-condition analysis (Mackie, 1974), elaborate this into more sophisticated accounts of causal structure. In econometrics, however, Humean ideas have largely been reinterpreted as diagnostic tools rather than as definitions of causality.

Granger causality in time-series econometrics is a classic example (Granger, 1969). Variable X is said to 'Granger-cause' variable Y if past values of X improve forecasts of Y, conditional on the past of Y and other information. This notion formalises temporal precedence and predictive content, both central to Humean intuitions, but does not itself guarantee a causal relationship in the stronger sense required for interventionist policy analysis. Similarly, pre-trend tests in DiD designs apply Humean reasoning: if treated and control units exhibit parallel outcome trends before treatment, this supports – though does not prove – the assumption that the groups would have followed similar paths in the absence of treatment.

These Humean-inspired diagnostics play an important role in design-based practice. They function as empirical checks on the credibility of identifying assumptions rather than as constitutive definitions of causality. If treated and control units were already diverging prior to treatment, this signals a threat to DiD identification; if lagged values of a proposed cause do not help predict the outcome, this raises questions about the putative mechanism. At the same time, design-based causal claims are not defined by such regularities. The core definition remains counterfactual and expectation-based: causal effects are differences between potential outcomes under alternative treatments. Regularities in observed data serve to make those definitions empirically credible, not to constitute them.

Furthermore, once expectations and forward-looking behaviour are taken seriously, simple temporal precedence becomes a fragile indicator of causality. Agents may respond to anticipated future events based on past regularities, so that an apparent 'cause' can follow an effect in calendar time. This makes purely Humean notions particularly ill-suited as foundations for economic causality, even though their diagnostic descendants remain embedded in econometric practice.

Relation to the interventionist concept of causality

Interventionist accounts, notably Woodward (2003), characterise causal relations in terms of what would happen under suitably defined interventions: X causes Y if there is a possible intervention on X that would change Y, holding other relevant factors fixed. On a natural reading, the design-based potential-outcomes framework appears close to this perspective, since it explicitly conceptualises causal effects as differences between outcomes under alternative interventions. Maziarz (2020) has therefore suggested that design-based econometrics should be associated primarily with interventionist philosophy.

There are, however, reasons to be cautious about a straightforward identification. Woodward's account places considerable emphasis on invariance: causal claims are tied to stable relations that would persist under a range of hypothetical interventions. By contrast, standard design-based estimands such as the ATE or LATE are often local and context-dependent. They capture how expected outcomes respond to a particular intervention in a specific institutional environment, but may not be invariant across different policy regimes or populations.

Moreover, Woodward's notion of an ideal intervention differs from the interventions typically studied in applied economics. Real-world economic interventions may generate spillovers, induce general-equilibrium adjustments or alter institutional rules in ways that change the underlying causal structure. While Woodward does not require interventions to be practically feasible or ethically acceptable, empirical economics necessarily focuses on actually implemented or realistically implementable policies. As a result, what counts as an 'intervention' in econometric practice often falls short of the idealised manipulations envisaged in philosophical discussions.

The design-based framework can therefore be understood as interventionist in spirit but expectation-based and local in execution. It shares with interventionism the focus on 'what would happen if we were to do X?', and it uses potential outcomes to formalise such questions. Yet it typically refrains from making strong claims about invariant causal laws under all admissible interventions. Instead, it produces estimates of how expected outcomes change under particular interventions in specific settings, leaving broader questions of invariance and transportability to supplementary analysis. This tension between local identification and broader causal claims will be examined further in the subsequent section.

3. The expectation-based nature of design-based causality

It is important to distinguish identification, estimation and statistical inference within the design-based framework. Identification concerns the conditions under which a causal parameter is recoverable from observable data; estimation concerns how sample data are used to approximate that parameter; and inference concerns the uncertainty surrounding the estimate.

The present discussion focuses on what the framework delivers as its core causal claim, namely an expectation, rather than on the internal distinction between these operations.

A defining feature of design-based causality is its reliance on expectations rather than realised outcomes. Within the potential-outcomes framework, causal effects are defined in terms of potential outcomes, not realised individual experiences, and are aggregated into population-level parameters through the data-generating process (DGP). The ATE represents a population-level parameter, capturing the expected difference between potential outcomes under treatment and control. Individual outcomes may contradict the average effect without undermining the causal claim itself. Formally, the causal effect is defined as $\tau_i = Y_i(1) – Y_i(0)$, while population-level causal parameters such as the ATE are defined as expectations over the probability distribution implied by the DGP. It is not a property of any individual's realised experience, nor of particular events observed in the world.

That individual causal effects can be heterogeneous is well understood within this framework (see, for example, Angrist, 2004; Angrist and Pischke, 2009; Athey and Imbens, 2017).[1] The point here is not that heterogeneity is conceptually problematic. Rather, it marks a departure from philosophical traditions that ground causal claims in particular events or stable causal powers. In these traditions, a 'causal effect' is not an expectation over a probability distribution, but a feature of what actually happens. Moreover, if one focuses purely on identification, it is theoretically possible that the expected effect is such that it will never occur in practice with certainty. The expectation may exist as a property of the probability distribution even if no realised outcome ever matches it exactly. A simple lottery example illustrates this: suppose payouts are £0, £5 and £10, with probabilities 0.7, 0.2 and 0.1, respectively. The expected value of this lottery is £2, but this value lies outside the support of the distribution of realised outcomes.

This creates the curious situation in which a causal effect may be well defined and identifiable by design, yet may never be realised in concrete events. The tension is not internal to the econometric framework, which is explicit about working with expectations, but arises when one compares it with more traditional philosophical conceptions of causality. In those conceptions, causal relations attach to actual sequences of events, to powers or dispositions or to stable regularities in the world. By contrast, design-based causal effects attach to expectations defined over potential-outcome distributions. Economic causality in the design-based sense is best treated as a statement about the DGP rather than about particular histories. This expectation-based conception allows for 'black-box' identification: causal effects can be estimated without modelling the mechanisms through which they arise. While pragmatically

[1] Richardson and Troost (2009) provide an instructive case in which a DiD design operates at the level of individual units: two adjacent Federal Reserve districts in Mississippi during the Great Depression. Even in this setting, however, the causal claim rests on a counterfactual comparison (how would the Atlanta Fed district have fared under St Louis Fed policy?) that remains fundamentally unobservable, illustrating that the epistemic structure discussed here applies irrespective of the level of aggregation.

powerful, this abstraction raises philosophical questions about the ontological status of causal effects and complicates claims about external validity.[2]

This tension is also visible when one contrasts design-based causal inference with the aspirations of much of economic theory. Traditional economic theory often seeks deterministic or quasi-deterministic regularities – for instance, that increases in price reduce quantity demanded, ceteris paribus. Modern causal inference, by contrast, conceptualises causal effects probabilistically: interventions are understood as shifting the expected distribution of outcomes across units, with causal parameters defined as expectations over potential outcomes. Realised outcomes may therefore deviate from the estimated causal effect without threatening the causal claim itself – a participant in a randomised controlled trial may experience an outcome opposite to the ATE, yet this is interpreted as statistical variation around the expectation rather than as evidence against causality.

A clarification is warranted. The claim that design-based econometrics treats causal effects as parameters of a DGP is primarily epistemological rather than ontological. For the purposes of identification and estimation, causal effects are represented and recovered as features of a stochastic process. This does not necessarily entail the metaphysical assertion that causality 'resides' in the DGP. Indeed, the DGP itself is not directly observable but inferred from realised data. Thus, treating causality as ontologically grounded in the DGP would reverse the usual inferential direction from data to the DGP, amounting to what might be described as an explanatory inversion. The more defensible reading is that the DGP serves as an analytical construct through which causal effects are defined and recovered, representing an epistemological commitment rather than a metaphysical one. This conception bears some resemblance to philosophical accounts of probabilistic causation (Suppes, 1970; Hitchcock, 2018), where causes raise the probability of their effects. Yet it departs from such accounts by defining those effects as population-level expectations rather than properties of particular occurrences. Such departures do not map neatly onto classical philosophical theories of causation, which typically locate causality in events, facts, powers or regularities, rather than in mathematical expectations over probability distributions.

The idiosyncrasy deepens further once one recognises that expected causal effects may be both well-defined and identified whilst remaining empirically elusive. If a policy intervention reshapes the distribution of outcomes but produces highly dispersed or multimodal responses, the average effect may rarely, if ever, be realised in practice. The causal effect then exists primarily as a population parameter. Within economics this is generally regarded as acceptable, since policy evaluation is assumed to concern average consequences: a policymaker asking whether a reform improves expected welfare is primarily interested in mean effects rather than in individual realisations. Yet political decision-making often hinges on how particular groups or individuals fare, not merely on expected outcomes. While these issues point towards normative questions

[2] In a more elaborate taxonomy – used, for instance, in psychology – this is called 'external reliability'. In this taxonomy, 'validity' refers to the question of whether an indicator actually measures what it is supposed to measure, something which economists often treat as given. See, for example, Cook and Campbell (1979) on reliability and external validity in quasi-experimental research.

beyond the scope of the present analysis, they highlight the philosophical peculiarity of grounding causality in expectations alone.

Closely connected to this expectation-based framework is an emphasis on identification relative to mechanism. A causal effect can be identified and robustly estimated without specifying how the treatment produces its outcome. This does not imply that mechanisms are inaccessible in principle. A growing literature on mediation and causal pathways seeks to decompose total effects into constituent processes (see, e.g., Ludwig et al., 2011; VanderWeele, 2015; Bullock and Green, 2021). Nonetheless, in much of applied microeconomics and policy evaluation, mechanism-based analysis remains secondary to 'identification-first' strategies. The prevailing approach treats treatments as exogenous shifters rather than as components within structured causal systems. At the same time, recent empirical work increasingly encourages, and in some cases effectively requires, some account of 'mechanism', for example via mediation analysis, heterogeneous-effects exploration or informal narrative pathways. Such accounts are often taken to enhance the credibility of a research design, especially where questions of robustness, persistence and welfare relevance are at stake (Grüne-Yanoff, 2016; Marchionni and Reijula, 2019). This creates a tension: at the level of the Neyman–Rubin potential-outcomes framework, identification of an ATE or LATE does not require a mechanism, whereas at the level of research practice and publication standards, mechanism-centred narratives are often treated as if they were integral to causal assessment. The framework thus permits black-box identification, but contemporary empirical norms often encourage opening the box, especially when results are used for policy advice or transported to new settings.

Instrumental variable estimation exemplifies this stance: the exclusion restriction requires that the instrument affects the outcome only through the treatment, yet the researcher need not specify the channels through which the treatment itself operates. The estimator recovers a causal effect without decomposing it into sub-mechanisms. In this sense, causal inference proceeds as a 'black box': what matters is that variation in treatment can be isolated, not how the treatment brings about its effect.

This pragmatism is intentional and defensible for the narrow purpose of identification. Policy evaluation ultimately seeks actionable answers: will the implementation of intervention X raise employment, earnings or health on average? The ATE is designed to address precisely this question within the maintained identification assumptions. Understanding why an intervention works is undoubtedly valuable for improving policy design and for assessing whether effects will generalise to new settings, but it is not logically required for establishing causal effects at the level of identification. An economist can assert that a programme increases expected employment without knowing whether this occurs through improved skills, employer responses, behavioural incentives or institutional frictions.

However, the limitations of black-box reasoning become evident when causal claims are extended beyond their original empirical environment. The central issue then becomes external validity. In practice, design-based identification often operates in a partial-equilibrium framework. The estimated causal effect captures the impact of the treatment conditional on the prevailing economic environment. Yet when interventions are scaled up, general-equilibrium adjustments may alter or even reverse the original effect. Price responses, displacement of untreated individuals and institutional adaptations are channels through which scaling can undermine the external validity of a partial-equilibrium estimate. The experience with microcredit programmes is instructive: randomised evaluations in individual settings produced positive local effects, but a series of large-scale replications found near-zero average effects once scaled beyond their initial experimental settings (Banerjee et al., 2015). The identified causal effect was not 'wrong' in its original context, but its policy relevance depended on equilibrium conditions that the black-box research design could not capture.

Formal results on causal transportability reinforce this point. Pearl (2009) and Pearl and Bareinboim (2014), for example, show that, without assumptions about the stability of underlying causal mechanisms, effects identified in one context cannot in general be transferred to another. Black-box identification is sufficient for establishing that a treatment changes the expected outcome locally, but validating whether that relationship persists across environments requires insight into the structure of the causal process itself. It is precisely this kind of generalisation that classical economic theory often presupposes when formulating laws such as the law of demand.

Viewed through the lens of dispositional philosophy, this tension takes a different form. On dispositional accounts, causation consists in latent powers or tendencies possessed by objects or systems, which may or may not manifest depending on surrounding conditions. A disposition represents the capacity to bring about an effect when appropriate triggers and enabling conditions are present. In empirical causal analysis, this implies not only isolating causal factors, but also recognising the possibility of 'interference conditions' that prevent causal powers from manifesting. In the presence of such interference conditions, genuine causal relationships may remain empirically invisible in the data.

Design-based econometrics, however, cannot simply adopt dispositionalism, since it consciously eschews appeals to unobservable powers. This resistance echoes Hume's scepticism towards such metaphysical entities, with causation instead reduced to relations among observable or at least measurable magnitudes. Yet, as argued in the previous section, the contemporary design-based conception of causality does not align neatly with Humean regularity either. It neither requires constant conjunction in realised outcomes nor grounds causality in mechanisms. Instead, it locates causation in expectations defined over stochastic processes, producing a distinctive hybrid that departs from both dispositional and regularity-based traditions. The resulting framework is philosophically coherent, but it is also distinctly idiosyncratic when set against classical accounts of causation.

4. Medieval causal logic and design-based econometric practice

So far, we have focused on the theoretical properties of various concepts of causality. The discussion now turns to the intellectual genealogy of analysing causality empirically. The question of how to conduct causal analysis empirically can be traced back to the Middle Ages (Knuuttila, 1993; Marenbon, 2007; Pasnau, 2011). Medieval philosophers developed systematic methods for reasoning about causes and effects. Two methods are particularly instructive in the present context: William of Ockham's 'Method of Difference' and Duns Scotus's 'Concurrence Method' (Cross, 2005; Lagerlund, 2015). Comparable epistemic strategies were later reformulated in Mill's methods of experimental enquiry (Mill, 1843), refined in Fisher's theory of experimental design (Fisher, 1935) and formalised statistically in Neyman's and Rubin's potential-outcomes framework (Neyman, 1923; Rubin, 1974). Design-based econometrics can, in this interpretive sense, be read as a further development of these epistemic strategies, translating abstract logical principles into concrete identification and validation procedures.

The Method of Difference and design-based econometric identification

Ockham's Method of Difference can be summarised as follows: if two situations are alike in all relevant respects except one, and an effect occurs in one situation but not the other, then the differing circumstance is the cause (or part of the cause) of the effect. Ockham illustrates this with examples in which one observes recovery from illness after the administration of a particular herb, while all other potential causes have been held fixed or excluded (Marenbon, 2007; Lagerlund, 2015). The underlying idea is that causal inference proceeds by isolating a single difference against a background of similarity.

Mill (1843) explicitly restated this as his 'Method of Difference': 'If an instance in which the phenomenon under investigation occurs, and an instance in which it does not occur, have every circumstance save one in common, that one circumstance, in which alone the two instances differ, is the effect, or the cause, or an indispensable part of the cause, of the phenomenon.' This became a cornerstone of nineteenth-century thought on causal enquiry. Fisher's work on agricultural experiments can be interpreted as a statistical formalisation of this principle: randomisation creates groups that are similar in expectation, so that any systematic difference in outcomes can be attributed to the treatment (Fisher, 1935; Neyman, 1923). Rubin (1974) subsequently generalised this logic into the potential-outcomes framework, in which causal effects are defined as differences between potential outcomes under alternative treatments.

Modern design-based identification strategies apply this principle in different guises. In difference-in-differences designs, treated and control units are compared before and after an intervention. Under the parallel-trends assumption, the difference in outcomes between the two groups after the intervention, net of pre-existing differences, is attributed to treatment. The econometric implementation using two-way fixed effects can be interpreted as enforcing Ockham's principle at the level of expectations: time-invariant differences between units and common shocks are differenced out, leaving the treatment indicator as the salient difference (Angrist and Pischke, 2009).

In an instrumental variable design, one isolates the exogenous component of variation in treatment – the part driven by the instrument – and attributes differences in outcomes associated with that component to the treatment itself. The exclusion restriction and relevance condition formalise Ockham's demand that only one causal channel should vary. Regression discontinuity designs similarly approximate the Method of Difference at a threshold: units just above and just below a cut-off are assumed to be locally similar in all relevant respects other than treatment status, so any discontinuity in the outcome at the threshold is attributed to treatment (Lee and Lemieux, 2010).

In this way, the Method of Difference is not merely of historical interest. It provides a unifying conceptual template for understanding the logic of design-based identification. Much like the principle of ‘frugal inference’, which favours parsimonious causal models (Forster et al., 2018), the aim is to hold as much as possible constant, isolate a single difference and attribute differences in expected outcomes to that variation.

The Concurrence Method and robustness

Duns Scotus articulated a complementary method for causal analysis, sometimes referred to as the ‘Concurrence Method’ (Cross, 2005). Where Ockham focuses on situations in which an effect occurs and does not occur, Scotus is concerned with cases in which the effect is observed. His question is: across a range of circumstances in which the effect occurs, which candidate causes are consistently present? The idea is to vary the background conditions as widely as possible and observe whether the association between a putative cause A and an effect B persists. If B is always found in conjunction with A across varying contexts, this supports the claim that A has a genuine, stable causal connection to B. If the association fails to ‘concur’ under some conditions, this suggests that the apparent causal link may be context-specific or spurious.

Where Ockham’s approach compares cases in which the effect occurs with cases in which it does not, Scotus’s approach focuses on cases in which the effect is observed. Econometrically, this implies that Ockham’s dataset will be larger than Scotus’, thereby increasing the information available for estimation. From a philosophy of science perspective, however, Ockham’s approach faces a counterintuitive problem: observing cases where the effect does not occur (e.g. treatment without outcome) may seem uninformative about causality, yet it can still logically support causal claims, since the absence of the effect implies the absence of the cause under the maintained causal structure (contraposition).[3]

Scotus’s method anticipates later concerns with robustness and external validity. Rather than relying on a single carefully controlled comparison, it asks whether a proposed causal relation holds across a variety of situations. In modern terms, one might say that Ockham’s method is primarily about identification – isolating a difference that can be attributed to treatment – whereas Scotus’s method is primarily about validation – testing whether the identified effect reflects a stable causal connection rather than an artefact of particular circumstances.

In contemporary econometrics, this logic is reflected in robustness checks that examine whether an estimated treatment effect survives variation in sample, specification and context. Researchers routinely re-estimate models across subsamples (for example, by gender, age, region or period), employ alternative sets of controls, use different functional forms and draw on multiple datasets. If the estimated effect remains similar in magnitude and sign across these perturbations, this is taken as evidence that the effect is not a fragile artefact of a particular specification. The practice of ‘test, test, test’ advocated by Hendry (1993) and the emphasis on robustness in Leamer’s (1983) ‘extreme bounds analysis’ can be interpreted as modern

[3] This resembles Hempel’s ‘Paradox of the Raven’ (Hempel, 1945). The paradox arises from the logical equivalence between ‘All ravens are black’ and its contrapositive ‘All non-black things are non-ravens’. Observing a black raven supports the first claim, but so does observing a non-black non-raven object (for example, a green apple), which intuitively appears irrelevant. Analogously, observing treatment without outcome can, through contraposition under the maintained causal structure, still provide support for a causal claim even though this may appear epistemically unintuitive.

counterparts to Scotus's insistence that genuine causal relations should manifest across varying conditions.

To be sure, there are important differences. Scotus did not have access to statistical concepts such as sampling variation or confidence intervals and his notion of 'always' is qualitative rather than probabilistic. Moreover, in many economic applications it is unrealistic to expect that a treatment effect will have the same magnitude in all contexts. Nonetheless, the epistemic concern is similar: stability of the effect across changes in background conditions provides support for its causal status, whereas sensitivity to small contextual changes raises doubts.

The complementarity of identification and validation

Viewed together, the Method of Difference and the Concurrence Method highlight an important complementarity in causal inference. Identification alone – in the sense of being able to express a causal parameter as a functional of the observable distribution under maintained assumptions – does not guarantee that the estimated effect captures a stable, policy-relevant relationship. A cleverly designed DiD, IV or RDD study may satisfy the Ockham-style requirement of isolating a difference attributable to treatment, but the resulting estimate may be highly context-specific. Conversely, seeking Scotus-style concurrence without a clear identification strategy risks mistaking stable associations for causal relations.

Medieval authors were already aware that these methods address different epistemic questions. The Method of Difference asks 'What causes what, under controlled comparison?'; the Concurrence Method asks 'Does this causal relation hold across a range of contexts?'. Both are needed for a well-founded causal claim. Contemporary design-based econometrics strongly emphasises the former – often under the heading of 'credibility' – but tends to treat the latter under the heading of 'robustness' as secondary or optional.[4] The medieval perspective suggests that this hierarchy may be epistemologically unbalanced.

In practical terms, one can interpret Scotus's insight as a call to integrate robustness and external-validity analysis more tightly into the design-based framework. Rather than viewing robustness checks as mere appendices to the main identification exercise, they can be seen as attempts to approximate the Concurrence Method under the constraints of modern data and institutional variation. In this spirit, work on causal transportability (Pearl and Bareinboim, 2014) and the limits of piecemeal causal inference from multiple studies (Mayo-Wilson, 2014) can be interpreted as an attempt to formalise when and how a causal effect identified in one context can be expected to hold in another. The theoretical message of this literature is closely aligned with Scotus's concern: without assumptions about the stability of underlying mechanisms across contexts, one cannot infer that a local causal effect generalises.

External validity and natural kinds

Ockham also recognised a deeper problem: how can one move from observing particular causal instances to asserting general principles? His answer appealed to something like 'natural kinds':

[4] In other approaches to econometrics, this type of robustness check is more widely used. A particularly strong emphasis on robustness is found in Leamer's extreme bounds analysis (1983).

if all causes of a certain kind produce the same effect on entities of another kind, then observation of singular instances can be generalised (Lagerlund, 2015). This maps neatly onto modern concerns with external validity and treatment-effect transportability. A randomised controlled trial that estimates the ATE of a programme in one population provides evidence about that population, but its relevance for another depends on whether the two populations are of the 'same kind' in the relevant causal respects.

In contemporary terms, this raises the question of whether the mechanisms linking treatment and outcome, and the distribution of relevant background factors, are sufficiently similar across contexts. The Concurrence Method provides a way of operationalising this question: does the effect A → B persist when we vary the surrounding conditions? If yes, then the causal mechanism is structurally robust; it is indeed a property of the 'kind' under investigation rather than an artefact of a particular setting. If no, then the effect is context-dependent and generalisations must be made with caution.

From this perspective, the modern design-based emphasis on local, expectation-based causal parameters has a clear strength – it facilitates credible identification within specific contexts – but also a corresponding vulnerability. Without systematic attention to Scotus-style concurrence across contexts, it risks producing a proliferation of highly local causal claims whose broader applicability remains unclear. Situating design-based econometrics within this medieval–modern genealogy clarifies why issues of robustness and external validity are not merely methodological details, but central to the epistemology of causal inference in economics.

5. Conclusion

Design-based econometrics defines causal effects through potential outcomes, with average treatment effects and related parameters formulated as expectations. Economic causality, in this sense, is counterfactual in form but population-level and expectation-based in content: it attributes causal significance to differences in expected outcomes under alternative interventions, for specified populations and contexts.

When examined through philosophical lenses, this produces a distinctive hybrid. The design-based framework clearly draws on counterfactual logic and, in spirit, shares the interventionist focus on 'what would happen if we did X?'. At the same time, its standard estimands are local and context-dependent: they describe how expected outcomes change under particular interventions in particular settings, rather than articulating invariant relations that hold across a wide range of interventions and environments. Regularity-based tools such as pre-trend tests and Granger-causality checks survive as diagnostics, while dispositional and mechanistic notions are largely bracketed. There is, therefore, no clean alignment with any single established philosophical tradition.

From the standpoint of policy evaluation, this framework is coherent and defensible because it provides estimates of average causal effects under interventions. In this respect, design-based causality is well suited to the practical demands of applied microeconomics and programme evaluation.

However, this principle itself is contestable. One might argue that policy evaluation concerns realised and individual effects rather than expected outcomes or that scientific epistemology should be unified across disciplines rather than allowing each field to define causality in its own terms. An obvious cost of this distinctiveness is that economic causality is thin on metaphysical detail about how causation actually works. Invoking the idea of a unified philosophy of science across disciplines, as pursued, for example, in logical positivism and critical rationalism (Ayer, 1936; Popper, 1959), suggests that field-specific definitions of causality may not be helpful. On this view, one might seek to define economic causality in a way that goes beyond the relatively narrow framework of policy intervention and expectation-based parameters. This idea has itself faced sharp criticism, most prominently in Feyerabend's (1975) 'anything goes!', but adopting a notion specific to economics nonetheless contributes to what Hausman (1992) termed the 'separate science of economics'.

This paper does not settle those broader methodological debates, but it does clarify what is distinctive about design-based causality and what follows from taking that distinctiveness seriously. First, if causal claims are understood as statements about expectation-valued parameters of a data-generating process, then issues of robustness and external validity are epistemologically co-equal with identification. Ockham-style methods of difference, as instantiated in DiD, IV and RDD designs, are indispensable for isolating local causal effects, but Scotus-style concerns with concurrence across contexts – realised today in robustness checks, subsample analyses and formal transportability work – are indispensable for assessing whether such effects are stable and policy-relevant beyond their original settings.

Second, interpreting modern practice through a longer genealogy – running from medieval causal methods through Mill and Fisher to the Neyman–Rubin framework – highlights that identification and validation were often treated as complementary rather than hierarchical. In practical terms, this suggests that design-based econometrics would benefit from treating robustness testing across contexts, subsamples and specifications as constitutive of causal inference, not merely supplementary. Integrating insights from robustness-oriented traditions, such as Leamer's (1983) 'extreme bounds analysis' and Hendry's (1993) 'test, test, test' principle, with the design-based approach would move econometrics closer to the medieval ideal in which both isolation and concurrence are required for justified causal claims.

Against this background, the conclusion is that economic causality in the design-based tradition is indeed 'distinctly idiosyncratic': it is a coherent, practice-guided synthesis of counterfactual reasoning, interventionist pragmatism and selective Humean diagnostics, deliberately stripped of strong metaphysical commitments. Whether one regards this as an acceptable endpoint or as a starting point for a more unified philosophical account of causation in economics depends on one's broader methodological commitments. What is clear, however, is that the expectation-based and context-dependent character of design-based causality reinforces the importance of robustness, external validity and the limits of generalisation in contemporary empirical work.

References

Angrist, J.D. (2004). Treatment effect heterogeneity in theory and practice. *Economic Journal* 114(494), C52–C83

Angrist, J.D. and J.-S. Pischke (2009). *Mostly harmless econometrics: an empiricist's companion*. Princeton University Press.

Angrist, J.D. and J.-S. Pischke (2010). The credibility revolution in empirical economics. *Journal of Economic Perspectives* 24, 3–30.

Angrist, J.D., G.W. Imbens and D.B. Rubin (1996). Identification of causal effects using instrumental variables. *Journal of the American Statistical Association* 91, 444–455.

Athey, S. and Imbens, G.W. (2017). The State of Applied Econometrics: Causality and Policy Evaluation. *Journal of Economic Perspectives* 31, 3–32.

Ayer, A.J (1936). *Language, truth and logic*. London: Victor Gollancz.

Banerjee, A., D. Karlan and J. Zinman (2015). Six randomized evaluations of microcredit: introduction and further steps. *American Economic Journal: Applied Economics* 7, 1–21.

Blaug, M. (1992). *The methodology of economics*. Cambridge: Cambridge University Press.

Boumans, M. and J. Davis (2010). *Economic methodology: understanding economics as a science*. Basingstoke: Palgrave Macmillan.

Bullock, J.G. and D.P. Green (2021). The failings of conventional mediation analysis and a design-based alternative. *Advances in Methods and Practices in Psychological Science* 4(4), 25152459211047227.

Cartwright, N. (1988). Probability and causality: why Hume and Indeterminism don't mix. *Noûs* 22, 569–587.

Cartwright, N. (1989). *Nature's capacities and their measurement*. Oxford University Press.

Cartwright, N. (2007). *Hunting causes and using them: approaches in philosophy and economics*. Cambridge: Cambridge University Press.

Cook, T.D. and D.T. Campbell (1979). *Quasi-experimentation: design and analysis issues for field settings*. Boston: Houghton Mifflin.

Crespo, R.F. (2020). *The nature and method of economic sciences: evidence, causality, and ends*. London: Routledge.

Cross, R. (2005). *Duns Scotus*. Oxford: Oxford University Press.

Darnell, A.C. and J.L. Evans (1990). *The limits of econometrics*. Aldershot: Elgar.

Feyerabend, P. (1975). *Against method*. London: New Left Books.

Fisher, R.A. (1935). *The design of experiments*. Edinburgh: Oliver and Boyd

Forster, M., G. Raskutti, R. Stern and Naftali Weinberger (2018). The frugal inference of causal relations. *British Journal for the Philosophy of Science* 69, 821–848.

Granger, C.W.J. (1969). Investigating causal relations by econometric models and cross-spectral methods. *Econometrica*, 37(3), 424–438.

Grüne-Yanoff, T. (2016). Why behavioural policy needs mechanistic evidence. *Economics and Philosophy* 32, 463–483.

Harré, R. and E.H. Madden (1975). *Causal powers: a theory of natural necessity*. Oxford: Blackwell.

Hausman, D. (1992). *The inexact and separate science of economics*. Cambridge: Cambridge University Press.

Hayo, B. (1998). Simplicity in econometric modelling: some methodological considerations. *Journal of Economic Methodology* 5, 247–261.

Heckman, J.J. (2000). Causal parameters and policy analysis in economics: a twentieth century retrospective. *Quarterly Journal of Economics* 115, 45–97.

Heckman, J.J. (2008). Econometric causality. *International Statistical Review* 76, 1–27.

Heckman, J.J. and R. Pinto (2024). Econometric causality: the central role of thought experiments. *Journal of Econometrics* 243, 105719.

Hempel, C. (1945). Studies in the logic of confirmation. *Mind* 54, 1–26.

Hendry, D.F. (1993). *Econometrics: alchemy or science?* Oxford: Blackwell.

Henschen, T. (2025). *Causality and objectivity in macroeconomics*. London: Routledge.

Hitchcock, C. (2018). Probabilistic causation. Stanford Encyclopedia of Philosophy. Stanford: Metaphysics Research Lab, Stanford University.

Hoover, K.D. (2001). *Causality in macroeconomics*. Cambridge: Cambridge University Press.

Hume, D. (1739). *A treatise of human nature*. (Book I, Part III, 'Of Knowledge and Probability').

Imbens, G.W. and J.D. Angrist (1994). Identification and estimation of local average treatment effects. *Econometrica* 62, 467–475.

Imbens, G.W. and D.B. Rubin (2015). *Causal inference for statistics, social, and biomedical sciences*. Cambridge: Cambridge University Press.

Knuuttila, S. (1993). *Modalities in medieval philosophy*. London: Routledge.

Lagerlund, H. (ed.) (2015). Medieval theories of causation. Stanford Encyclopedia of Philosophy. Stanford: Metaphysics Research Lab, Stanford University.

Leamer, E.E. (1983). Let's take the con out of econometrics. *American Economic Review* 73, 31–43.

Leamer, E.E. (2010). Tantalus on the road to Asymptopia. *Journal of Economic Perspectives* 24, 31–46.

Lee, D.S. and T. Lemieux (2010). Regression discontinuity designs in economics. *Journal of Economic Literature* 48, 281–355.

Lewis, D.K. (1973). *Counterfactuals*. Cambridge: Harvard University Press.

Lewis, D.K. (1986). *On the plurality of worlds*. London: Basil Blackwell.

Ludwig, J., J.R. Kling and S. Mullainathan (2011). Mechanism experiments and policy evaluations. *Journal of Economic Perspectives* 25, 17–38.

Mackie, J.L. (1974). *The cement of the universe: a study of causation*. Oxford: Oxford University Press.

Marchionni, C. and S. Reijula (2019). What is mechanistic evidence, and why do we need it for evidence-based policy? *Studies in History and Philosophy of Science Part A* 73, 54–63.

Marenbon, J. (2007). *Medieval philosophy: an historical and philosophical introduction*. London: Routledge.

Mayo-Wilson, C. (2014). The limits of piecemeal causal inference. *British Journal for the Philosophy of Science* 65, 213–249.

Maziarz, M. (2020). *The philosophy of causality in economics: causal inferences and policy proposals*. London: Routledge.

Maziarz, M. and R. Mróz (2020). Response to Henschen: causal pluralism in macroeconomics. *Journal of Economic Methodology* 27, 164–178.

Mill, J.S. (1843). *A system of logic: ratiocinative and inductive*. Book III, Chapter VIII.

Moneta, A. (2003). Graphical Models for Structural Vector Autoregressions. *LEM Papers Series* 2003/07, Laboratory of Economics and Management (LEM), Pisa: Sant'Anna School of Advanced Studies.

Mumford, S. and Anjum, R.L. (2011). *Getting causes from powers*. Oxford: Oxford University Press.

Neyman, J. (1923). On the application of probability theory to agricultural experiments. Essay on principles. Section 9. *Roczniki Nauk Rolniczych* 10, 1–51. (English translation: *Statistical Science*, 1990, 5(4), 465–472.)

Pagan, A. (1987). Three econometric methodologies: a critical appraisal. *Journal of Economic Surveys* 1, 3–24.

Pasnau, R. (2011). *Metaphysical themes 1274–1671*. Oxford: Oxford University Press.

Pearl, J. (2009). *Causality: models, reasoning, and inference* (2nd ed.). Cambridge: Cambridge University Press.

Pearl, J. and E. Bareinboim (2014). External validity: From do-calculus to transportability across populations. *Statistical Science* 29, 579–595.

Popper, K.R. (1959). *The logic of scientific discovery*. London: Hutchinson.

Richardson, G. and W. Troost (2009). Monetary intervention mitigated banking panics during the Great Depression: quasi-experimental evidence from a Federal Reserve district border, 1929–1933. *Journal of Political Economy* 117, 1031–1073.

Rubin, D.B. (1974). Estimating causal effects of treatments in randomized and nonrandomized studies. *Journal of Educational Psychology* 66, 688–701.

Sims, C. (2010). But economics is not an experimental science. *Journal of Economic Perspectives* 24, 59–68.

Suppes, P. (1970). *A probabilistic theory of causality*. Amsterdam: North-Holland.

VanderWeele, T.J. (2015). *Explanation in causal inference: methods for mediation and interaction*. Oxford: Oxford University Press.

Woodward, J. (2003). *Making things happen: a theory of causal explanation*. Oxford: Oxford University Press.