Not all causal effects are created equal: selecting pathway cases for in-depth analysis using causal forests

Vlad Surdea-Hernea: Central European University, Wien

August 19, 2024

Abstract

Providing complex answers to causal questions requires both cross-case evidence for the existence of a causal effect and within-case evidence for the mechanisms through which this effect propagates. The former requires a robust causal identification strategy, while the latter is more amenable to qualitative investigation using methods narrowly tailored to mechanistic analysis, such as process tracing. However, once a cross-case causal effect has been estimated, there is little guidance on how to select a case in which to trace the causal mechanism. After reviewing the limitations of the algorithms currently available, I present a novel case selection method that uses causal forests to recover granular causal effects for each case in the sample, and show how this approach can be used to select pathway cases in a manner consistent with their stated goal of recovering causal mechanisms. I then briefly discuss how this framework can be extended to other types of cases, including typical and deviant cases.

Keywords: case selection, machine learning, mixed methods, pathway case, causal forest

1 Introduction

While the credibility revolution has improved the internal validity of quantitative analyses that address "effects of causes" type questions, it has not improved our understanding of how causal effects propagate and through what mechanisms (Mahoney and Goertz, 2006). This puts political science at odds with other hard sciences, where an explanation of *why* something happens is almost always complemented by an explanation of *how* it happens, which is often more important to practitioners (Darden, 2002; Illari and Williamson, 2012; Skillings, 2015).

Causal mediation analysis, the state of the art in the quantitative study of causal mechanisms, allows the researcher to non-parametrically estimate and then confirm whether a variable is a mediator, i.e. whether it is causally situated between the treatment and the outcome, only under sequential ignorability, a stricter requirement than even randomising treatment and intermediate variables (Imai et al., 2011; Imai and Yamamoto, 2013) [1]. Moreover, even if sequential ignorability holds, such a research design effectively reduces causal mechanisms from complex and (possibly) overdetermined processes involving multi-agent interactions to a linear causal chain. Qualitative research, on the other hand, has traditionally emphasised complex mechanistic explanations, developing and refining formal methods for eliciting causal mechanisms and testing their validity against alternative explanations, such as the various variants of process tracing (Beach and Pedersen, 2019; Bennett and Checkel, 2015; Humphreys and Jacobs, 2015).

In light of this (latent) complementarity, many methodologists have proposed exploiting the different types of causal knowledge produced by qualitative and quantitative methods, highlighting the unique strengths of each to address the limitations of the other (Blair et al., 2019; Brady et al., 2006; Collier et al., 2010; Coppock and Kaur, 2022; Glynn and Ichino, 2015)[2]. Ideally, one could use the appropriate quantitative methods to establish that a

[1] There are alternative approaches to causal mediation anchored in structural equation modelling, but these are often based on implicit, untestable assumptions (Cook et al., 2002; MacKinnon, 2012).

[2] There are, however, those who argue that there is a fundamental incompatibility between the two

causal effect of X on Y exists at the cross-case level, and then use qualitative methods for within-case analysis to explain in detail how the mechanism M by which the effect $X \rightarrow Y$ propagates (and possibly also to identify the variables M_i that comprise the causal chain underpinning the mechanism M).

However, one challenge for mixed-methods approaches is how to make the leap from the cross-case level, which recovers "dataset observations", to the within-case level, which privileges "causal process observations" (Collier et al., 2004). In other words, what is the appropriate procedure for selecting one case out of many through which to identify and characterise a causal mechanism conditional on the causal effect being identified? Without a proper answer, fundamental ontological differences would likely inhibit the adoption of mixed-methods research designs (Gerring, 2004, pp. 351-352).

This paper proposes a data-driven method for making this ontological leap, focusing on the setting where the presence of the theorised mechanism M is necessary for theory generalisation over the population of cases: *pathway cases* (Gerring, 2007).

The rest of the paper is structured as follows. In Section 2, I explain in more detail what pathway cases are and why they are suitable as crucial cases for testing the presence of a causal mechanism M. In Section 3, I review the current algorithms used to select pathway cases for in-depth analysis, and discuss their limitations in light of the goal of drawing causal inferences. In Section 4, I present how causal forests, an algorithm for recovering heterogeneous causal effects, can serve as a data-driven algorithm for pathway case selection. Finally, in Section 5, I briefly discuss how the conceptual framework introduced in this paper can be easily extended to other types of cases, including typical and deviant cases.

traditions, one that precludes pure mixed-methods (Beach, 2020)

2 Pathway cases

2.1 The rationale behind pathway cases

One of the earliest statements on why and how to select one crucial case to test the validity of a theoretical argument was made by Eckstein (1975). He argued that social scientists should conduct in-depth investigations in that case in the sample most likely to be explained by that theory, but not by competing theories (p.118). However, this strategy did not distinguish between evidence that a theory is valid (i.e. cross-case analysis of the existence of a causal effect $X \rightarrow Y$) and evidence about how the theory works to explain the outcome (i.e. within-case analysis of what the causal mechanisms are $X \rightarrow M(?) \rightarrow Y$). This conflation runs counter to recent methodological advances, which show that while mechanistic evidence is useful for explanation and for making a compelling argument in favour of a theory (Gerring, 2011, 215-217), it is not a necessary condition for causal inference (Gerring, 2010).

Recognising the limitations of the crucial case approach, Gerring (2007) introduced the concept of a pathway case, which he argues follows the original intent of Eckstein (1975) but is adapted to the empirical realities of contemporary political science. A pathway case is one for which the cross-case causal relationship $X \rightarrow Y$ has been shown not only to be present but also to be strong, but for which the mechanism M by which this causal relationship propagates has yet to be elicited (Gerring, 2007, pp.238-239).

In this way, pathway cases are similar in spirit to typical cases in set-theoretic multi-method research (SMMR), which are selected for within-case process tracing based on how well they fit the cross-case relationship found through Qualitative Comparative Analysis (Rohlfing and Schneider, 2018; Schneider and Rohlfing, 2013). However, because SMMR relies on a regularity theory of causation (Mahoney, 2008; Mahoney and Acosta, 2022; Rohlfing and Schneider, 2018), its selection algorithm cannot be easily integrated into the standard causal inference framework used in most quantitative research.

Since pathway cases only become apparent after the cross-case causal inference was per-

formed, they are by design diagnostic tools aimed at further exploring the depth of the causal relationship, not at expanding its breadth (Gerring and Cojocaru, 2016, p.405-406)[3]. Assuming unconfoundedness and that the magnitude of the causal effect induced by X is high, if the theoretical argument explaining that causal effect is correct, then the pathway case is uniquely insightful for understanding whether and how the argument appears to work, through which processes, and through the interactions of which agents, institutions, and enabling conditions (Gerring, 2016; Seawright and Gerring, 2008; Weller and Barnes, 2014). Deductively, the diagnostic power of pathway cases derives from a simple consideration: if the mechanistic argument underlying an identified causal effect is correct, then the case in which the causal effect is strongest should provide the clearest instance of how that causal mechanism operates.

In other words, the visibility of the causal mechanism needs to be monotonic in the treatment effect, which makes the presence of the theorised causal mechanism in the pathway case a necessary condition for claiming the more general presence of that mechanism in the population of cases, and its absence strong evidence for rejecting the original theory (Gerring, 2007). If we think of pathway analysis as an empirical test of our theorised mechanism, then failure to identify a mechanism in the pathway case would provide robust evidence against the main theory, and its presence would provide some degree of evidence in favour of it. Thus, depending on the specifics of each research context, pathway analysis provides at worst a hoop test for the proposed mechanism, and at best a doubly decisive test (Collier, 2011).

2.2 Selecting pathway cases for in-depth analysis

While it is clear that pathway cases can help researchers open the black box of causality once a cross-case effect has been identified, a major challenge is how to accurately determine the magnitude of this causal effect in each case. Most pathway cases in the existing literature have been selected using heuristics rather than formal algorithms, potentially compromising

[3]Note that this objective is distinct from that of causal mediation analysis, which aims to establish a causal chain at the cross-case level (Imai et al., 2011).

the validity of results derived from such within-case analyses. Some selection algorithms have been proposed (Gerring, 2016; Weller and Barnes, 2014), but as I will show, they ultimately rely either on informed guesses based on patterns in the data, or on simple correlational and model fit-based measures that are potentially relevant for prediction but largely meaningless for causal inference.

A representative example of current pathway selection algorithms comes from Gerring (2016, pp.108-110), who proposes to compare the residuals of each observation in a full model including the treatment X and the covariates necessary for unconfoundedness with those in a reduced model from which X has been removed. Specifically, the author argues that if the residual of an observation i in the full model shrinks significantly compared to the reduced model, this provides evidence that the treatment X is responsible for moving that observation closer to the regression line, indicating the presence of a strong causal effect. This type of algorithm is quite common for various case selection techniques in qualitative research (Seawright and Gerring, 2008), but at least for the pathway case it relies on ambiguity about what represents the causal effect of a treatment X.

Even if we assume that X is conditionally ignorable in the full model, the difference in the residuals between the two models would only give us some measure of the predictive power of X, since the incomplete model cannot be expected to capture any causal effect of the other variables apart from X, with the estimated coefficients of these variables being merely nuisance parameters that cannot be meaningfully interpreted (Keele, 2015). More generally, this kind of algorithmic confusion arises because many social scientists do not treat the proximity of an observation i to a regression line and the strength of the causal effect in that individual case as two separate target quantities, with the former not serving as an approximation of the latter (Athey and Imbens, 2017, 2019). Measuring the strength of a causal effect in each case involves counterfactual comparisons between outcomes under different intervention states for the same i, whereas prediction involves comparisons between

the characteristics of different cases (Mullainathan and Spiess, 2017)[4]. Comparisons between units may be relevant for causal inference in limited circumstances, but making such claims relies on strong assumptions about how the unobservable characteristics of cases in the sample are distributed (Sekhon, 2009).

Other selection techniques raise even more straightforward issues. For example, Dafoe and Kelsey (2014) exploit the distribution of treatment X and outcome Y together with that of a third factor S that potentially drives heterogeneous treatment effects. The proposed algorithm selects pathway cases based on the joint distribution of these variables in light of the proposed theory, which is relevant to a descriptive rather than causal inference problem. This approach is very similar to the one proposed by Weller and Barnes (2014), and followed by many empirical political scientists, for selecting pathway cases, but the latter is comparativist in nature rather than focusing identifying a crucial case. The comparative case study design could alleviate concerns over selecting based on descriptive measures but, as a rule, the values of treatment and outcome are not indicative of the strength of the causal effect (Imbens and Rubin, 2015, pp.16-18).

One reason why these examples of algorithms rely on descriptive or predictive rather than causal measures is that it is inherently impossible to measure the strength of causal effects in each case using the traditional tools of econometrics. To explain this, I rely on the causal framework introduced by Rubin (1974) following the pioneering work of Neyman (1923). In the baseline scenario, where we assume the existence of a binary treatment $W_i \in \{0, 1\}$, each observation is endowed with two potential outcomes that depend on the treatment status: $Y_i(W_i = 0)$ and $Y_i(W_i = 1)$ or, for simplicity, $Y_i(1)$ and $Y_i(0)$. In the real world, however, we can only observe the outcome associated with whether or not the treatments were actually administered to unit i, which can be expressed by the switching equation as $Y_i = Y_i(1)W_i + Y_i(0)(1 - W_i)$. If $W_i = 1$ then $Y_i(0)$ is not known and if $W_i = 0$ then $Y_i(1)$

[4]A separate but equally problematic issue with this algorithm is that, since model complexity is not penalised, a reduced model that includes a large number of covariates that are sufficiently correlated with X could result in some very small differences in residuals that are not even indicative of predictive power, but of inappropriate model selection and large collinearity (Kadane and Lazar, 2004; Tibshirani, 1996)

is not known. The fact that we cannot observe realisations of both potential outcomes $Y_i(0)$ and $Y_i(1)$ is commonly referred to as the fundamental problem of causal inference (Holland, 1986), and transforms causal identification into a missing data problem (Rubin, 1976).

Therefore, the individual treatment effect $ITE = Y_i(1) - Y_i(0)$ cannot be calculated from the available data, whether experimental or observational. However, the ITE would have been precisely the causal estimand that would have allowed us to quantify the strength of the causal effect in each case and thus optimally select a pathway case for within-case analysis. In fact, both experimental and quasi-experimental methods recognise our inability to estimate the ITE as the status quo and instead attempt to obtain an unbiased estimate of an average causal effect, such as the average treatment effect, $ATE = \mathbb{E}[Y_i(1) - Y_i(0)]$, or local and conditional versions of the ATE for techniques such as instrumental variables. However, the ATE provides insufficiently granular causal knowledge for the selection of the pathway case, as its estimated value is constant for each observation in the dataset. This is acknowledged by many scholars who discuss the inadequacy of average causal effects for pathway analysis (Weller and Barnes, 2014, p.38), but their solution is to work with descriptive measures for each case as a second-best option, whereas I offer a data-driven solution that remains true to the goal of causal inference.

Nevertheless, this discussion highlights the necessary conditions for an ideal algorithm suitable for selecting pathway cases. First, it should be based on estimated causal effects rather than on descriptive or predictive measures. Second, it should follow from, or at least not contradict, claims about the existence of the estimated ATE. Third, it should go beyond the ATE and attempt to measure more personalised treatment effects that come as close as possible to the unobserved ITE. Then, by comparing these proxies for ITE, one would be able to select a pathway case for which the *causal effect* is strongest.

3 Causal forests for pathway case selection

3.1 Heterogeneous treatment effects estimation

The limitations of working with average causal quantities are well recognised in the existing literature, ranging from medicine to epidemiology to marketing and political campaigning (Chernozhukov et al., 2018; Imai and Strauss, 2011; Imai and Ratkovic, 2013). Traditionally, this has been addressed by searching for moderators after credibly recovering the ATE, which would ideally allow us to learn which features of the observations increase the causal power of treatment X. However, actually performing causal moderation, rather than just some version of subgroup analysis, requires very strong assumptions that are unlikely to be met in most observational studies (Bansak, 2021).

In practice, therefore, the identification of heterogeneous effects has been reduced to estimating linear models with interaction terms, which are then given explicit or implicit causal interpretations, even though there is no evidence that the theorised moderator is not correlated with unobservable characteristics (Keele and Stevenson, 2021). Apart from endogeneity, another problem with such an approach is that it allows researchers to iteratively search for more extreme values of causal effects for particular subgroups of the population and then report only these, despite their potential spuriousness.

Advances in machine learning (ML) have explicitly addressed these problems by developing novel techniques to identify more granular causal quantities in an honest way that still yields valid asymptotic confidence intervals for the underlying treatment effect (Athey and Imbens, 2016; Athey et al., 2019; Nie and Wager, 2021; Wager and Athey, 2018). At the heart of these developments is a very simple problem of applied policy: how to learn for which observations in the sample the treatment would be most effective? For example, for people with which socio-demographic characteristics would a vaccine against a new virus be most effective? Who should we offer a training programme to, knowing that they are likely to benefit?

However, the same caveat discussed earlier must be taken into account: most ML techniques have been developed for, and are therefore useful for, prediction problems (Grimmer et al., 2021). Causal inference, as a very specific type of prediction (i.e. predicting what would happen after manipulating a treatment X, holding all other factors constant) done through an ML lens, requires special attention and can only be achieved if certain assumptions hold, similar to the case of standard econometric tools. If these assumptions do not hold, even the most complex ML models will fail to recover unbiased causal estimates, or even to detect the minimally sufficient set of covariates necessary to satisfy unconfoundedness (Hünermund et al., 2023).

For a long time, this was a conceptual distinction that limited the scope of ML and its usefulness for causal inference. In recent years, however, rapid progress has been made in integrating big data methods with the rules for causal inference outlined in the potential outcomes framework, giving rise to a novel research programme whose goal is to learn from the data how treatment X has different effect sizes based on the characteristics of the units. Formally, it's important to note that this research programme still does not assume that we can recover the counterfactual $ITE = Y_i(1) - Y_i(0)$ from the observable data, but it does claim that we can reconstruct very granular conditional average treatment effects (CATE) given by $\mathbb{E}[Y_i(1) - Y_i(0)|Z = z]$, where $Z = z$ denotes that an observation has the set of features z from feature space Z (Vegetabile, 2021). For example, instead of learning about the effect of a vaccine on the whole population, we can measure the difference in effect between very specific groups, such as white women and black men. Or we might learn that a training programme focused on digital literacy would make the biggest difference to people aged 18-49 in rural areas.

I argue that beyond policy evaluation, these ML-based tools for identifying heterogeneous treatment effects can be fruitful for isolating sufficiently granular causal effects that would allow us to select appropriate cases for pathway analysis. Particularly in a setting with continuous covariates, the CATE computed for each observation can be considered as an

approximation of the ITE, or at least as the best approximation that current algorithms, ML or otherwise, are able to provide. While there are several ML-based estimators that could potentially be used for this task (Knaus et al., 2021), I proceed to describe causal forests, one of the most popular causal ML algorithms (Wager and Athey, 2018; Athey et al., 2019).

3.2 Causal forests as estimators for CATE

Causal forests are generalised random forest algorithms that aim to accurately measure how the effect of a causal factor X varies across the sample, while correctly predicting the value of the causal effect induced (Athey and Imbens, 2016; Wager and Athey, 2018; Athey et al., 2019). While random forests are ensemble learning methods based on constructing a large number of decision trees in order to minimize the prediction error for an outcome Y between the leaves of each tree (Breiman, 2001), causal forests attempt to simultaneously maximize the difference in ATEs between the leaves while accurately estimating the ATE (Wager and Athey, 2018). This dual objective distinguishes causal forests from many other ML-based causal inference tools that explicitly address only the prediction problem.

The exact algorithm that causal forests use to recover CATEs is quite technical, but it's core can be explained in a straightforward way. To ensure that the causal forest does not identify random noise and treat it as the source of causal heterogeneity, Wager and Athey (2018) proposes to split the data used to train the model into two sub-samples, one for splitting (i.e., capturing heterogeneity, the C in CATE) and one for estimation (i.e., recovering the causal effect, the ATE in CATE). The splitting sub-sample is used to construct each causal tree under the two conditions mentioned above. In turn, the causal tree is used on the estimation sub-sample to actually measure the causal effect. The treatment effect within each leaf of a decision tree is calculated as the difference in means between the score of the outcome Y in the two treatment and control groups.

Statistically, a crucial aspect of causal forests is the asymptotic normality of the treatment effects recovered from the causal forest, which means that as the sample size grows

large, the distribution of the estimated treatment effects approaches a normal distribution. Consequently, we can accurately calculate the variance and thus generate confidence intervals for the estimates. Overall, this means that not only can we predict the CATE for a given set of features $Z_i = z$ for an observation i, but we can also measure the uncertainty around the CATE. This is particularly important for case selection issues, where we want our empirical tests that we apply to the data to be as stringent as possible in order to provide as much mechanistic evidence as possible, which boils down to comparisons not only between point estimates, but also between the lower bounds of the effects in each case.

3.3 Selecting pathway cases with causal forests

As stated, the main objective of the pathway case selection procedure remains to identify the case from the population in which the causal effect is strongest. In the language of the potential outcomes framework, we need to identify the observation i that maximises $Y_i(1) - Y_i(0)$. Due to the fundamental problem of causal inference, this is impossible, so the second best becomes the one that maximises $Y_i(1) - Y_i(0)|Z_i = z$ for a detailed set of observable characteristics Z. Thus, the problem of selecting pathways boils down to estimating heterogeneous treatment effects, for which I have shown that causal forests provide a robust solution.

With this in mind, how should a researcher proceed? First, credible evidence for the existence of a causal effect needs to be established at the cross-case level. This involves either an experimental or quasi-experimental research design, coupled with robustness checks, placebo tests and sensitivity analyses, and possibly setting extreme bounds on the effect (Coppock and Kaur, 2022; Eggers et al., 2023).

Second, assuming that such an analysis has been performed and an ATE has been identified, we proceed with the selection of a pathway case. This involves transitioning from ATE to $CATE$, which, as discussed earlier, can be done efficiently using a causal forest. In practice, this would entail a researcher implementing the causal forest algorithm and then

predicting, for each observation in the dataset, the CATE together with the 95% confidence interval. A pathway case would be one with the highest $CATE$ of the causal factor X on the outcome Y. Alternatively, one could compare the lower bounds of the CATEs.

Two points are worth noting, as they often come up in discussions of pathways cases. First, how to deal with outliers? Gerring (2007) implies that cases that appear to be extreme outliers in the proposed causal model should be excluded from candidacy for in-depth within-case analysis. The argument against extreme outliers is that they generally do not provide confirmatory evidence for a proposition (p.248). While, in general, extreme outliers should be carefully examined using available case knowledge to learn why they have non-standard values in Y, there is no reason to automatically exclude them from consideration. In fact, it could be crucial to learn specifically why the CATE is highest for such an outlier, and whether outlier patterns correspond to higher CATEs, which could potentially call the model specification into question. However, if the researcher is determined to remove outliers, causal forest estimation would not preclude this; in fact, learning the CATE might even improve the researcher's ability to distinguish whether an outlier is the result of causally meaningful processes or just an idiosyncratic observation, possibly due to poor data quality.

Second, the best case for pathway analysis often displays stable causal patterns over time, in the case of panel data. This is because usually, to test a theoretical argument, we want the strong cross-case causal effect not to be strong in one year and then suddenly dissipate, but to evolve over time so that the in-depth investigation can follow these processes and identify the actors responsible for their unfolding. A simple solution would be to include the time of each observation as one of the characteristics of each observation. Then the CATE of the observations would also take into account the time dimension, as we would have a causal effect for each period in which a case is observed. With this information in hand, one could pursue several strategies for selecting pathways cases. First, one could select cases for which the CATE is stable over time, which would imply a constant effect. Second, when studying path-dependent processes with increasing returns, researchers could select cases where the

causal effect increases over time. Similarly, when studying processes that dissipate, maybe due to some post-treatment intervention, one could select cases with a high initial CATE that decreases steadily and monotonically over time.

One could possibly argue that it is not just the period that changes over time, but also other characteristics of observation i, i.e. $Z_{i,t} \neq Z_{t,\neg i}$. While future research, particularly in ML, is likely to provide more appropriate answers for recovering unbiased period averages of different CATEs, one solution would be to always start with the cases that have the highest CATE for period t, theorise how the features are likely to evolve over time alongside the new CATE, if at all, and select cases that best fit such theoretical hunches. To some, deviating from a purely data-driven approach to case selection may seem like a failure of such an approach. This is not the case, as causal inference is always based on untestable assumptions for which theory can provide guiding answers (Coppock and Kaur, 2022). However, even if one remains sceptical about the fusion of theoretical and empirical knowledge, causal forests still remain an appropriate solution for a large class of static causal models, or for models in which features remain stable over time.

4 Potential extensions

Pathway cases are not the only setting of interest for within-case analysis; on the contrary, most qualitative work, particularly that which uses process tracing, focuses on typical and deviant cases. Examining the presence of a theorised causal mechanism in typical and deviant cases allows us to understand the causal chain of actors and processes through which an effect propagates, as well as to define the scope conditions under which the causal mechanism operates (Elman et al., 2016; Gerring, 2016; Schneider, 2024).

While a general theory of case selection is beyond the scope of this paper, I argue that, conceptually, causal forests and the CATEs they estimate for each observation in the dataset provide intuitive ways of thinking about typical and deviant cases. In addition, I develop

the first case selection protocol in which typicality and deviance are defined in terms of the causal mechanism rather than the presence of the outcome, since the latter is likely to be determined by multiple causal factors not necessarily captured by the theory being tested. To unpack this claim, I introduce two new concepts, knowledge of which I argue is both necessary and sufficient to determine which cases are best suited for intensive analysis as either typical or deviant.

First, we need to know how much the size of the causal effect in a particular case i deviates from the average, the ATE, which constitutes an indicator of the *relative causal relevance (RCR)* of that case. Formally, we have to compute the standardized difference between the CATE and the ATE for each case i, as shown in Equation 1.

$$RCR_i = \frac{\mathbb{E}[Y_i(1) - Y_i(0)|Z = z_i] - \mathbb{E}[Y_i(1) - Y_i(0)]}{\sqrt{\frac{1}{N-1}\sum_{j=1}^{N}\left(\mathbb{E}[Y_j(1) - Y_j(0)|Z = z_j] - \mathbb{E}[Y_j(1) - Y_j(0)]\right)^2}} \tag{1}$$

Second, we need to know the size of the causal condition X in each case i relative to its mean value. This measures how "strong" the causal shock of X is, a metric I call *trigger severity (TS)*. Trigger severity is a descriptive metric that can be calculated directly from the data, as in Equation 2. In other words, TS_i tells us how many standard deviations away from the mean each X_i is in the dataset.

$$TS_i = \frac{X_i - \left(\frac{1}{N}\sum_{j=1}^{N} X_j\right)}{\sqrt{\frac{1}{N-1}\sum_{j=1}^{N}\left(X_j - \left(\frac{1}{N}\sum_{k=1}^{N} X_k\right)\right)^2}} \tag{2}$$

By exploiting information about the distribution of TS and RCR in the dataset, we can describe a variety of case study situations. For example, the most suitable pathway case is given by the observation with the highest RCR, regardless of TS. However, in the unlikely event that several observations share the same RCR, the TS can be used to distinguish between them. One solution would be to take the case with the highest RCR and lowest TS, which would give you the case where the smallest causal factor triggered the largest causal

effect. The remainder of this section explores how to select typical and deviant cases using the TS and RCR score distributions.

4.1 Typical cases

Typical cases are defined as cases that are representative of the observable distribution (Seawright and Gerring, 2008; Gerring, 2016; Gerring and Cojocaru, 2016). Again, we face the similar problem that such definitions imply causally relevant quantities, such as the generalisability of the causal properties identified in a particular setting, while using only descriptive or predictive metrics. Usually, typical cases are defined as being "close to the regression line", a metric that has no bearing in causal terms.

In the framework proposed in this paper, a typical case i would be one in which the TS_o is as close as possible to zero, implying an average trigger of the causal mechanism, as well as one in which RCR_i is also close to zero, indicating that its CATE is approaching the ATE, exactly what typical cases intuitively look for[5]. However, using RCR and TS scores instead of descriptive and predictive measures has other advantages for the researcher. In particular, it enables one to assess whether the impact of the causal mechanism is monotonic in the size of its trigger.

Typically, quantitative researchers assume that causal factors X make an additive contribution to the outcome Y. In other words, adding more of X will produce more of Y. To explain this in a theory, one needs a causal mechanism that allows for such additivity. However, this is largely unproven in either quantitative or qualitative scholarship and is assumed by fiat. By relying on TS and RCR, it is possible to select a set of cases that allow us to validate the additivity assumption. Specifically, we need to look at two cases and unpack the causal mechanism previously identified in a typical/pathway case. First, the causal mechanisms should be less powerful than average when RCR is negative and TS is negative.

[5]While both conditions are relevant for selecting a pure typical case, if one had to choose between a case with null RCR and non-null TS and a case with null TS but RCR is non-null , I recommend the former, because ultimately the search for causal mechanisms implies prioritising causal effects and not second-order notions of how such effects were induced.

Second, the same mechanisms should be stronger than average when RCR is positive and TS is positive. This is because, if the causal mechanisms allows for additivity in X, it must be the case that (more) less X leads to a lower impact of the mechanism, which ultimately translates to a (higher) lower value of Y.

4.2 Deviant cases

Deviant cases are usually defined as cases that, on the basis of observable properties, should conform to a causal argument but do not, making them anomalies in the report of an existing theory (Gerring and Cojocaru, 2016, p.399). They are essential for learning the limits of a causal mechanism, the conditions necessary for its presence and non-anomalous unfolding, and the scope conditions that define the population to which the mechanism is applied. Following the conceptual language of the Rubin causal model, deviant cases i are those in which one would expect the CATE to be of normal magnitude, given the magnitude of the causal condition X. In the framework I propose, this implies a TS above zero and a RCR below one; or conversely, a TS below zero and a RCR above one. Since several cases could satisfy these properties, I propose the following formula for the degree of deviation DD_i:

$$DD_i = |TS_i| + |RCR_i| - |TS_i + RCR_i| \tag{3}$$

In cases where TS and RCR are aligned in their directionality - both have either positive or negative values - the formula essentially nullifies the degree of deviation, as $DD_i = 0$. This finding is consistent with the conventional understanding that cases with aligned causal effect size and causal condition intensity do not represent deviant cases. Conversely, when TS and RCR diverge in their signs, the formula amplifies the degree of deviation, so that $DD_i > 0$. Such cases, where the size of the causal effect (RCR) and the intensity of the causal condition (TS) move counter-intuitively in opposite directions, are emblematic of the anomalous or deviant cases within causal analysis. Deviant cases should be selected as those

with the maximum DD_i, i.e. those where the gap between what we would expect (the CATE) based on the observable properties (X) is the largest. When available, both types of deviant cases should be analyzed: negative TS-positve RCR, positive TS-negative RCR.

5 Conclusions

Intensive case study methods have recently seen a resurgence, drawing more on the formality of quantitative methods to ensure that mixed-methods approaches are robust and not sensitive to subjective researcher decisions (Coppock and Kaur, 2022; Fairfield and Charman, 2017; Humphreys and Jacobs, 2015). However, most of the developments are related to how to conduct the analysis to identify and unpack complex causal mechanisms, without taking into account that such tools are only meaningful if the cases of interest have been selected according to the same rules as those driving the quantitative causal identification.

To fill the current lacuna, I present a data-driven method for selecting pathway cases that provide the necessary conditions for identifying a general causal mechanism. This algorithm is rooted in the Rubin causal model, making it compatible with the standard toolkit of methods on which political science relies for design-based identification, ranging from regression discontinuity designs to instrumental variables. I demonstrate the superiority of this algorithm in direct contract to currently available algorithms, which I argue are predicated on either describing predictive rather than causal quantities. I then provide a brief guide to implementing this pathway case selection algorithm, including a discussion of how to overcome potential challenges.

Finally, I move beyond pathway cases and provide more general metrics for selecting different types of cases. Using new metrics that assess the relative size of the causal shock and causal mechanism in each case, I show how researchers can identify the best typical and deviant cases for intensive analysis. Taken together, the proposed case selection framework should allow for a smoother ontological leap from cross-case quantitative to within-case

qualitative analysis.

References

Susan Athey and Guido Imbens. Recursive partitioning for heterogeneous causal effects. *Proceedings of the National Academy of Sciences*, 113(27):7353–7360, 2016.

Susan Athey and Guido W Imbens. The state of applied econometrics: Causality and policy evaluation. *Journal of Economic perspectives*, 31(2):3–32, 2017.

Susan Athey and Guido W Imbens. Machine learning methods that economists should know about. *Annual Review of Economics*, 11:685–725, 2019.

Susan Athey, Julie Tibshirani, and Stefan Wager. Generalized random forests. *Ann. Statist.*, 47(2):1148–1178, Apr 2019. doi: 10.1214/18-AOS1709.

Kirk Bansak. Estimating causal moderation effects with randomized treatments and non-randomized moderators. *Journal of the Royal Statistical Society Series A: Statistics in Society*, 184(1):65–86, 2021.

Derek Beach. Multi-method research in the social sciences: A review of recent frameworks and a way forward. *Government and Opposition*, 55(1):163–182, 2020.

Derek Beach and Rasmus Brun Pedersen. *Process-tracing methods: Foundations and guidelines*. University of Michigan Press, 2019.

Andrew Bennett and Jeffrey T Checkel. *Process tracing*. Cambridge University Press, 2015.

Graeme Blair, Jasper Cooper, Alexander Coppock, and Macartan Humphreys. Declaring and diagnosing research designs. *American Political Science Review*, 113(3):838–859, 2019.

Henry E Brady, David Collier, and Jason Seawright. Toward a pluralistic vision of methodology. *Political Analysis*, 14(3):353–368, 2006.

Leo Breiman. Random forests. *Machine learning*, 45:5–32, 2001.

Victor Chernozhukov, Mert Demirer, Esther Duflo, and Ivan Fernandez-Val. Generic machine learning inference on heterogeneous treatment effects in randomized experiments, with an application to immunization in india. Technical report, National Bureau of Economic Research, 2018.

David Collier. Understanding process tracing. *PS: Political Science & Politics*, 44(4):823–830, 2011.

David Collier, Henry E Brady, and Jason Seawright. Sources of leverage in causal inference: Toward an alternative view of methodology. In *Rethinking social inquiry: Diverse tools, shared standards*, pages 229–266. Rowman and Littlefield, 2004.

David Collier, Henry E Brady, and Jason Seawright. Outdated views of qualitative methods: time to move on. *Political Analysis*, 18(4):506–513, 2010.

Thomas D Cook, Donald Thomas Campbell, and William Shadish. *Experimental and quasi-experimental designs for generalized causal inference*, volume 1195. Houghton Mifflin Boston, MA, 2002.

Alexander Coppock and Dipin Kaur. Qualitative imputation of missing potential outcomes. *American Journal of Political Science*, 66(3):681–695, 2022.

Allan Dafoe and Nina Kelsey. Observing the capitalist peace: Examining market-mediated signaling and other mechanisms. *Journal of Peace Research*, 51(5):619–633, 2014.

Lindley Darden. Strategies for discovering mechanisms: Schema instantiation, modular subassembly, forward/backward chaining. *Philosophy of Science*, 69(S3):S354–S365, 2002.

Harry Eckstein. Case studies and theory in political science. In Fred I. Greenstein and Nelson W. Polsby, editors, *Handbook of political science. Political science: Scope and theory*, volume 7, pages 94–137. Addison-Wesley, Reading, MA, 1975.

Andrew C. Eggers, Guadalupe Tuñón, and Allan Dafoe. Placebo tests for causal inference. *American Journal of Political Science*, 2023.

Colin Elman, John Gerring, and James Mahoney. Case study research: Putting the quant into the qual, 2016.

Tasha Fairfield and Andrew E Charman. Explicit bayesian analysis for process tracing: Guidelines, opportunities, and caveats. *Political Analysis*, 25(3):363–380, 2017.

John Gerring. What is a case study and what is it good for? *American political science review*, 98(2):341–354, 2004.

John Gerring. Is there a (viable) crucial-case method? *Comparative political studies*, 40(3): 231–253, 2007.

John Gerring. Causal mechanisms: Yes, but.... *Comparative political studies*, 43(11):1499–1526, 2010.

John Gerring. *Social science methodology: A unified framework*. Cambridge University Press, 2011.

John Gerring. *Case study research: Principles and practices*. Cambridge university press, 2016.

John Gerring and Lee Cojocaru. Selecting cases for intensive analysis: A diversity of goals and methods. *Sociological Methods & Research*, 45(3):392–423, 2016.

Adam N Glynn and Nahomi Ichino. Using qualitative information to improve causal inference. *American Journal of Political Science*, 59(4):1055–1071, 2015.

Justin Grimmer, Margaret E Roberts, and Brandon M Stewart. Machine learning for social science: An agnostic approach. *Annual Review of Political Science*, 24:395–419, 2021.

Paul W Holland. Statistics and causal inference. *Journal of the American statistical Association*, 81(396):945–960, 1986.

Macartan Humphreys and Alan M Jacobs. Mixing methods: A bayesian approach. *American Political Science Review*, 109(4):653–673, 2015.

Paul Hünermund, Beyers Louw, and Itamar Caspi. Double machine learning and automated confounder selection: A cautionary tale. *Journal of Causal Inference*, 11(1):20220078, 2023.

Phyllis McKay Illari and Jon Williamson. What is a mechanism? thinking about mechanisms across the sciences. *European Journal for Philosophy of Science*, 2:119–135, 2012.

Kosuke Imai and Marc Ratkovic. Estimating treatment effect heterogeneity in randomized program evaluation. *The Annals of Applied Statistics*, 7(1):443–470, 2013.

Kosuke Imai and Aaron Strauss. Estimation of heterogeneous treatment effects from randomized experiments, with application to the optimal planning of the get-out-the-vote campaign. *Political Analysis*, 19(1):1–19, 2011.

Kosuke Imai and Teppei Yamamoto. Identification and sensitivity analysis for multiple causal mechanisms: Revisiting evidence from framing experiments. *Political Analysis*, 21 (2):141–171, 2013.

Kosuke Imai, Luke Keele, Dustin Tingley, and Teppei Yamamoto. Unpacking the black box of causality: Learning about causal mechanisms from experimental and observational studies. *American Political Science Review*, 105(4):765–789, 2011.

Guido W Imbens and Donald B Rubin. *Causal inference in statistics, social, and biomedical sciences*. Cambridge University Press, 2015.

Joseph B Kadane and Nicole A Lazar. Methods and criteria for model selection. *Journal of the American statistical Association*, 99(465):279–290, 2004.

Luke Keele. The statistics of causal inference: A view from political methodology. *Political Analysis*, 23(3):313–335, 2015.

Luke Keele and Randolph T Stevenson. Causal interaction and effect modification: same model, different concepts. *Political Science Research and Methods*, 9(3):641–649, 2021.

Michael C Knaus, Michael Lechner, and Anthony Strittmatter. Machine learning estimation of heterogeneous causal effects: Empirical monte carlo evidence. *The Econometrics Journal*, 24(1):134–161, 2021.

David MacKinnon. *Introduction to statistical mediation analysis*. Routledge, 2012.

James Mahoney. Toward a unified theory of causality. *Comparative Political Studies*, 41 (4-5):412–436, 2008.

James Mahoney and Laura Acosta. A regularity theory of causality for the social sciences. *Quality & Quantity*, 56(4):1889–1911, 2022.

James Mahoney and Gary Goertz. A tale of two cultures: Contrasting quantitative and qualitative research. *Political analysis*, 14(3):227–249, 2006.

Sendhil Mullainathan and Jann Spiess. Machine learning: an applied econometric approach. *Journal of Economic Perspectives*, 31(2):87–106, 2017.

Jerzy Neyman. On the application of probability theory to agricultural experiments. essay on principles. *Ann. Agricultural Sciences*, pages 1–51, 1923.

Xinkun Nie and Stefan Wager. Quasi-oracle estimation of heterogeneous treatment effects. *Biometrika*, 108(2):299–319, 2021.

Ingo Rohlfing and Carsten Q Schneider. A unifying framework for causal analysis in set-theoretic multimethod research. *Sociological Methods & Research*, 47(1):37–63, 2018.

Donald B Rubin. Estimating causal effects of treatments in randomized and nonrandomized studies. *Journal of educational Psychology*, 66(5):688, 1974.

Donald B Rubin. Inference and missing data. *Biometrika*, 63(3):581–592, 1976.

Carsten Q Schneider. *Set-Theoretic Multi-Method Research: A Guide to Combining QCA and Case Studies*. Cambridge University Press, 2024.

Carsten Q Schneider and Ingo Rohlfing. Combining qca and process tracing in set-theoretic multi-method research. *Sociological Methods & Research*, 42(4):559–597, 2013.

Jason Seawright and John Gerring. Case selection techniques in case study research: A menu of qualitative and quantitative options. *Political research quarterly*, 61(2):294–308, 2008.

Jasjeet S Sekhon. Opiates for the matches: Matching methods for causal inference. *Annual Review of Political Science*, 12:487–508, 2009.

Derek John Skillings. Mechanistic explanation of biological processes. *Philosophy of Science*, 82(5):1139–1151, 2015.

Robert Tibshirani. Regression shrinkage and selection via the lasso. *Journal of the Royal Statistical Society Series B: Statistical Methodology*, 58(1):267–288, 1996.

Brian G Vegetabile. On the distinction between" conditional average treatment effects"(cate) and" individual treatment effects"(ite) under ignorability assumptions. *arXiv preprint arXiv:2108.04939*, 2021.

Stefan Wager and Susan Athey. Estimation and inference of heterogeneous treatment effects using random forests. *Journal of the American Statistical Association*, 113(523):1228–1242, 2018.

Nicholas Weller and Jeb Barnes. *Finding pathways: Mixed-method research for studying causal mechanisms*. Cambridge University Press, 2014.