Visualizing environmental econometrics using directed acyclic graphs

Pierce Donovan[†]

August 2023

Abstract

Environmental and resource economists have gravitated toward writing empirical papers with an emphasis on causal inference. Yet efforts to promote statistical rigor have not facilitated improvements in the transparency and accessibility of the assumptions underlying effect identification. In this paper, I demonstrate how Directed Acyclic Graphs augment and standardize the communication of prior causal knowledge in environmental economics. I then provide suggestions for integrating this innovation into future presentations, publications, peer reviews, and pedagogy.

Keywords: environmental econometrics, causal inference, identification, research design, communication, directed acyclic graphs (JEL: A20, C12, C51, C52, Q50)

[†]Assistant Professor, Department of Economics, University of Nevada, Reno
email: `pierce.donovan@unr.edu`
web: `piercedonovan.github.io`

A framework for communicating causal knowledge

Most empirical papers in environmental economics today are structured around a single statistical inference of interest, with an emphasis on how the authors' identification strategy supports a measurement with a justifiably-causal interpretation (Segerson, 2019). In many cases, the storytelling appears to take more effort than the statistics, perhaps because designing clever identification strategies for causal inference is difficult. It is the thoughtful communication of our research design that now defines "rigor" in empirical papers, rather than a barrage of specification tests.

Despite this focus, economists have not adopted a systematic framework for presenting causal knowledge—although we ostensibly understand the core elements of one. In order to claim that our identifying variation is "as good as random," we invoke economic theory and our familiarity with a specific setting. Theory explains the causal channels through which an intervention may impact an outcome of interest. Familiarity with the setting informs empirical research design. Together, they help us understand the policy implications of our results and applicability to other settings. Yet our writing often tackles the issue of communicating theory and empirical assumptions—prior subjective information—in an incomplete and ad hoc manner.

I want to build on the "credibility revolution" with increased clarity, transparency, and testability of different *causal* hypotheses. Most economists have become comfortable reporting on a statistical inference's sensitivity to varying empirical assumptions about functional form or different sets of control variables. But this haphazard and non-exhaustive sensitivity analysis is less likely to reveal some objective truth than it is to invite opportunities for selective reporting and other [unintentionally] misleading research practices (Ferraro and Shukla, 2020). Further—and more pertinent to the present article—while this exercise in showing consistency across model permutations is likely to give researchers confidence in their results, it fails to address whether or not their approach estimates the effect of interest—and this is what a "robustness check" should be doing.

The spirit of specification testing is to learn if a statement like "the effect of X on Y is likely positive" is permissible. A test of a result's sensitivity to an assumption about a data generating process is suitable to bolster this claim, because it is a test of whether or not you are measuring the effect of interest. This change to a causal model will naturally motivate a specific change in empirical strategy and highlight a critical prior belief, while a mechanical regression permutation fails to imply anything about the correctness of a causal model. But as of this article, many economists employ a vague and tacit writing style when explaining causal models and identification assumptions in their research, and this makes the testing of causal hypotheses difficult.

Directed Acyclic Graphs (DAGs) (Pearl, 1995) can systematize our thinking about how we create and test causal hypotheses. DAGs display causal knowledge, motivate identification strategies and indicate the conditions under which a regression yields a causally-interpretable measurement. They provide us with a way to choose conditioning variables, sources of data, and empirical methods in a manner that can be easily understood and validated by a wider audience of researchers, students, and stakeholders. In this feature, I provide a complete introduction to this visualization technique, explore how it can facilitate causal hypothesis testing over a wide swath of settings, and share some insights discovered while investigating this idea. With this article, I aim to lower the costs of adoption for an emerging method of broad applicability in environmental economics.

This article takes a different track relative to that of a few prominent bears (e.g. Imbens (2020); Heckman and Pinto (2022), and their derivatives). The bearish case ignores the following facts: (1) most economists do not currently use any explicit framework to model causal knowledge, (2) the potential outcomes language that is commonly taught in graduate school and embedded in our writing is derived from a framework no more general than the graphical approach, and (3) the vast majority of contemporary research involves settings where the limitations of DAGs would not apply. Throughout this paper, I will provide a fairer assessment of the benefits and costs of the graphical approach. My view is that DAGs are an augmentative tool to economists, and that they generate the greatest benefits when integrated into our existing research workflow.

In the next two sections, I provide a primer on DAGs and generate several toy applications to illustrate their utility. In the penultimate section, I review a recent research article with the aid of a DAG. In the final section, I discuss the merits of integrating this innovation into future presentations, publications, peer reviews, and pedagogy.

A primer on Directed Acyclic Graphs

A DAG represents our best understanding of the various causal and spurious links between variables of interest—within the particular setting in which data are collected. It can help us determine an identification strategy, validate claims of causal identification in regression output, and suggest other scenarios where our results are likely to apply. This section provides a minimalist introduction to their interpretation.

Figure 1 displays a representative DAG. Our initial task is to center our thinking around a particular relationship of interest, say $T \rightarrow Y$. Directed arrows like $T \rightarrow Y$ convey statements like "the outcome Y is in part determined by the status of some treatment T." Additional [indirect] causal relationships between T and Y may be mediated by other variables (e.g. M). As we expand our model, we want to consider any variables which could distort the observed relationship between T and Y away from the sum of the causal effects.

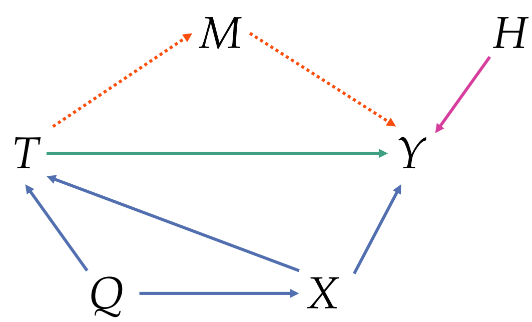

Figure 1: A representative DAG. Some variables—and the causal links between them—may not be observable, and we can emphasize these with dashed arrows. I use color in my figures to direct focus towards certain links of interest like $T \rightarrow Y$, although this is merely a stylistic preference.

Spurious relationships between T and Y are created by confounding variables—those which influence both T and Y—like X or Q. If we do not mitigate the effects of confounding variables, our estimate of the treatment effect will be biased. On the other hand, variables like H will not influence the causal interpretation of a measured correlation between T and Y, since they do not contribute to a spurious relationship. Thus the DAG has already given us a way to differentiate between malignant and benign sources of variation.

If data on X and Q are available, we can employ a simple matching strategy to control for these confounding variables. Regression will automatically isolate the causal relationship between T and Y from the variation induced by changes in X or Q. In the language of DAGs, we are closing these non-causal paths between T and Y.

What if a variable like Q wasn't observable? This is where the term "acyclic" deserves a moment of explanation. This feature disallows loops—no variable can be a descendant of itself (thus the endogenous variables of a model are able to be determined recursively). Importantly, this restriction is the main limitation to the complete applicability of DAGs to all of the potential research questions economists may want to ask. Questions involving simultaneous causality—e.g. in research using general equilibrium models—are not addressable in this framework, however, it is worth pointing out that the vast majority of contemporary articles in environmental economics explicitly or implicitly utilize causal models built upon recursive counterfactual logic. Therefore, DAGs apply to the vast majority of questions that environmental economists currently ask.

Nevertheless, the restriction allows us to use our DAG to make statements about conditional independence, e.g. "Y is independent of Q conditional on X." Since variation in Q can only impact Y through an intermediate effect on X, controlling for the variation in X also halts the pass-through of information from Q to Y. Thus the requirement for closing a spurious path between T and Y can be relaxed to either controlling for a confounding variable along that path, or one of its descendants. Here, once X is controlled for, the status of Q becomes immaterial with respect to the identification of the effect of T on Y.

When reporting regression output, the marginal "effect" on Y attributable to T reflects all remaining open paths from T to Y. It is our responsibility to determine whether or not this variation generates a measurement of a causal effect. In the present case, a viable control strategy for estimating $T \rightarrow Y$ would be to control for X, since X additionally blocks the path including Q. The decision to include or exclude H (or Q) in our regression will not effect the causal interpretation, although including these controls may improve the precision of our $T \rightarrow Y$ estimate. We cannot disentangle the direct effect from the indirect effect here, as the variable M is unobservable.

Applications of DAGs to environmental economics

Now that the groundwork has been laid, I will demonstrate how to implement DAGs through the use of several bespoke toy examples in environmental economics. Although these examples are fairly straightforward without the use of DAGs, my point is to show how discussions of research design can be made more clearly and succinctly with the aid of the graphical approach. In this way, DAGs are hardly redundant, but augmentative.

Experimental sources of variation sever spurious links

Does the built environment influence transportation modality? To many, induced demand is taken as a maxim. But this ignores the idea of residential sorting; individuals with preferences for transit or walkability will tend to move to places with those amenities. This confounding factor is not observable, and a naïve regression of transit decisions on built environment would pick up both the causal and spurious paths in Figure 2.

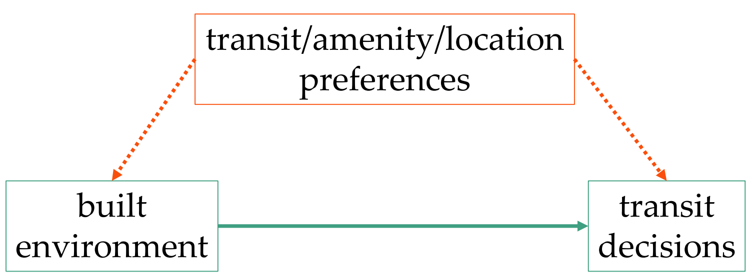

Figure 2: A naïve regression of transit decision on built environment without controlling for preferences results in a biased estimate. Randomization would trim the arrow from preferences to the built environment, thus removing the confounding factor.

One source of useful identifying variation comes from housing lotteries for below market-rate housing. While a traditional housing voucher program grants potential tenants some agency in locating a rental unit, a housing lottery provides a chance to win the option to live in a specific, randomly-assigned building. When market-rate housing is cost-prohibitive, the preferences of prospective tenants become irrelevant. Some properties may be closer to metro lines and others could have more readily-available parking, but this variation in the built environment is assigned to the tenant by the lottery. Using the DAG above, random assignment would sever the arrow between preferences and the built environment, thus eliminating the confounding issue.

Understanding mechanisms prevents over-conditioning

Matching strategies are more complicated than we think. I have witnessed researchers consider control variables simply because they are available for inclusion. Many of these control decisions are made without any causal logic to back them up. To illustrate, Figure 3 warns us of a potential pitfall involving over-conditioning when addressing the following question: how does an urban tree canopy improve housing values?

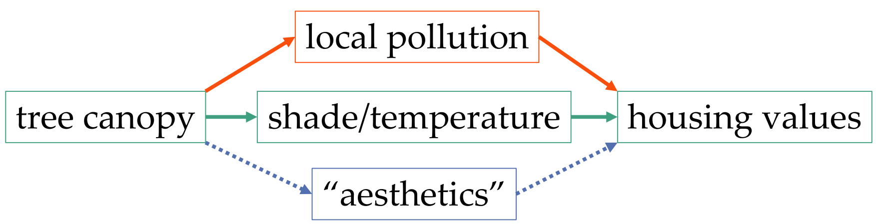

Figure 3: Controlling for a mediator will close a causal path, preventing the estimation of an indirect effect, or the inclusion of the indirect effect in a total effect estimate.

The above DAG implies that there are three mechanisms through which increasing tree cover conceivably impacts housing values, to wit, trees filter local air and groundwater pollutants, lower energy bills and provide passive cooling, and increase landscape aesthetic. In general, the extent to which a list like this is exhaustive is something that must be assumed by the researcher, as with any structural causal model.

If we were working on a hedonic analysis linking tree cover to housing values, I'd imagine a popular suggestion would be to control for local pollution levels—the logic being that tree cover and pollution are likely correlated, and both things impact housing values. But this would be a blunder, as we have not properly considered the role the pollution variable plays in our DAG. Presently, it looks more like a mediator than a confounder, because it facilitates a causal path from the tree canopy variable to housing values. By controlling for local pollution, we close a causal path that we meant to leave open.

DAGs can save us from eliminating some of the variation that we wanted to exploit or generating estimates of something other than what we desired. In the present setting, an unconditional regression is best. However, sometimes we may want to estimate only a portion of the full effect. For example, the isolation of the aesthetic effect would be possible if we controlled for the two measurable mediators, but the immeasurability of the aesthetic variable prevents us from isolating the other channels.

Some control variables are better left alone

Some research design strategies can inadvertently introduce bias to estimators. To illustrate, let's consider the estimation of the impact of intense, short-lived increases in wildfire smoke exposure on respiratory health, using readily-available data from hospital admissions. Figure 4 provides a model of a data-generating process that takes into account how our sample is being collected. Because smoke exposure also increases the likelihood of emergency room visits due to interactions with cardiovascular disease, I have drawn a second pathway from smoke exposure to hospital admission (while suppressing the mediator for simplicity) that is unrelated to respiratory health.

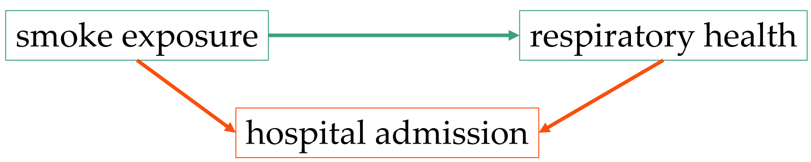

Figure 4: Collecting data using a selected sub-sample will result not only in a lack of external validity, but a lack of internal validity as well, due to collider bias.

Certain variables create spurious correlations between treatment and outcome once they are included as controls—or in this case, baked into the data collection strategy. In the present example, weakened cardiovascular health and weakened respiratory health are both sufficient conditions for being more likely to showing up in the hospital data. But the sufficiency of either condition implies that among hospital admits, those with poor cardiovascular health are more likely to have better respiratory health, conditional on smoke exposure. This is not a causal claim, but a spurious correlation generated by the data generating process. Since those impacted by wildfires are more likely to have worsened cardiovascular health—and we see them in the hospital data—the spurious link will contribute to an under-estimation of the negative respiratory health effect of interest.

DAGs will alert you to potential identification pitfalls—like the one above—that would be tricky to suss out in your head. Whenever you spot a variable that invites a collision of two arrows, the variable is called a collider. Stratifying on a collider variable—through inclusion in the control set or the data collection process—will open an otherwise closed path. The abstraction of the identification problem to a graphical one helps us understand if this is desirable.

Some people may mistakenly reduce this example to an external validity concern. However, the measured relationship between smoke exposure and respiratory health won't even be *internally*-valid here. The collider bias will act to reduce—or even flip the sign of—the measured smoke exposure "effect" within the hospitalized sub-sample. However, since we understand the second mechanism through which smoke exposure increases the likelihood of hospital admission, we could control for cardiovascular health and sever the non-causal path from smoke exposure to respiratory health, even when the collider is in play. This allows us to measure an unbiased respiratory health effect—but only for the hospitalized population.

A new flavor of sensitivity analysis

DAGs are agnostic to the functional forms you choose to model relationships between variables. The abstraction frees us up to think about alternative *causal* hypotheses, each represented by a different DAG. This gives us an opportunity to explore a new type of "robustness check" that involves re-validating (and modifying) an identification strategy based on different DAG configurations.

Let's consider another example. Does adopting an electric vehicle decrease household emissions? We would expect the direct effect to be positive due to increased mineral extraction and fossil fuel-derived electricity demand. However, the bulk of the adoption effect is likely indirect and negative, through the replacement of a gas-powered car. As a straw-man, consider the economist who controls for the number of gas-powered cars in the household. This would be a blunder, as they have closed a causal path. By controlling for the number of gas-powered cars, we remove the replacement channel.

But previous economist wouldn't even measure the direct effect with their strategy, given my understanding of car buying preferences. The inclusion of other relevant household characteristics in the DAG (Figure 5) will manifest a collider bias. For example, households with car enthusiasts are drawn both to gas-powered cars and to driving more often; this makes our control variable into a collider on the enthusiast path and introduces a spurious correlation.

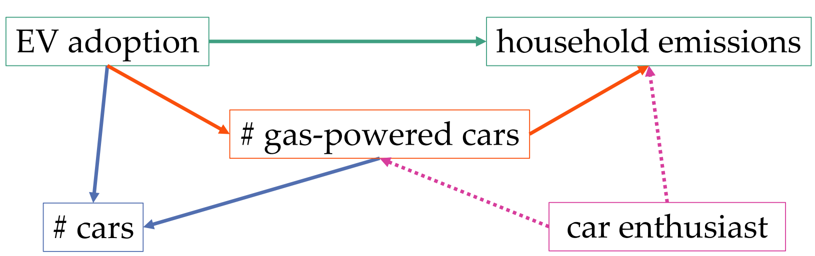

Figure 5: The true impact of electric vehicle adoption on household emissions is only revealed if we omit controls for the number of cars in the household, gas-powered or otherwise.

Many economists say that the replacement story is obvious when I start this example (although no one spots the collider). But I've also been told that we apparently want to control for the total number of cars instead. In my mind, this is also a blunder. Figure 5 shows why. In my mental model, the total number of cars acts as another collider. When we stratify on the number of cars in the household, we implicitly assume that there is a replacement of a gas-powered car for every electric vehicle adopted—which is unlikely. Thus we move from an underestimate to an overestimate of the decrease in emissions.

My model says that we should not control for the number of gas-powered vehicles or the number of cars in the household. Perhaps your data source or setting implies a different causal structure than mine. In this case, we can re-evaluate whether or not my no-control strategy holds by checking the new DAG for any threats to identification.

Efficient communication of identification assumptions

Fishing quota systems have revolutionized the organization of fisheries, ensuring the biological sustainability of the exploited species and maximizing the value of an authorized total allowable catch. In these systems, fishers explicitly own shares of this total allowable catch and are free to trade these fishing rights among themselves. The market for quota shares autonomously guides ownership into the hands of the most efficient fishers, therefore increasing the economic rents that accrue to each share.

How much of the variation in the price of quota is due to changes in fleet efficiency (i.e. catch per unit effort), as opposed to changes in the demand for fish? A potential problem with separating these two impacts is the fact that any market development that increases the value of the quota may additionally incentivize investment that leads to greater efficiency. Figure 6 shows how uncontrolled demand side factors create spurious relationships when interpreting the effect of changes in fleet efficiency on the price of quota.

Now imagine a potential regulation brought about due to concerns about the bycatch of an endangered species often found co-mingling with the targeted species. Perhaps a restriction on a certain type of gear or fishing location. This restriction reduces the efficiency of the fleet, impacts the value of fishing quota through this efficiency mechanism, and is plausibly exogenous with respect to shifts in demand.

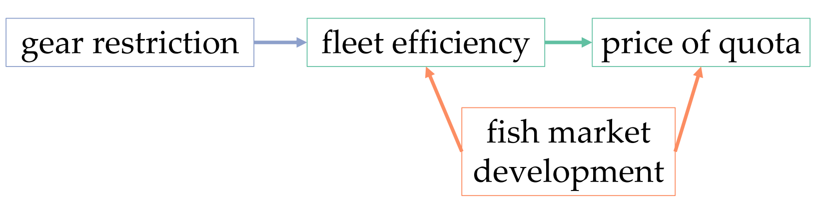

Figure 6: Given our mental model above, the corresponding DAG reflects the assumptions required for the gear restriction variable to be a useful instrument.

Figure 6 shows exactly what these features buy us. The regulation turns fleet efficiency into a collider variable along the spurious path from gear restriction to the quota price.

And it acts as a mediator on the causal path. Thus an unconditional regression of the price of quota on the gear restriction variable provides a causally-interpretable estimate. A regression of fleet efficiency on the gear restriction is similarly unbiased. Dividing the two relevant regression coefficients yields the instrumental variables estimator for the effect of changing fleet efficiency on the price of quota—another causally-interpretable estimate.

Our DAG reveals the familiar identification assumptions for the instrumental variables approach—"relevancy, independence, and validity"—just by checking for the presence or absence of three key arrows involving the gear restriction variable. Given our fishery story, we draw an arrow from gear restriction to fleet efficiency (relevance, though the strength must be checked), but not to or from the price of quota or fish market development variables (validity and independence, respectively). With the DAG, we're able to transparently communicate to others what we believe to be true about our instrument, and readers can more easily follow (or dispute) our assumptions.

The credibility revolution in economics can be extended

In the previous section, we used the DAG to find links between variables that we could estimate without bias. The resulting strategy happened to line up with the familiar instrumental variables approach. But each new DAG structure provides an opportunity to discover a novel identification strategy that would have otherwise remained hidden. With one more toy example, I'll illustrate an estimation method that exists outside of the contemporary applied econometrics literature. Here, I will show that we can still generate an unbiased estimate of a treatment effect even with selection into treatment, as long as an exogenous mediator variable exists between treatment and outcome.

How beneficial are climate resilience-motivated crop insurance programs in developing nations? We would expect the most savvy farmers to have the most interest in an insurance program if made available—as well as the most sophisticated farming operations. Any naïve regression strategy would clearly pick up this selection bias, but Figure 7 shows how to remedy this selection problem. Consider the mechanism through which insurance would lead to a benefit. If there is no adverse weather event (and thus no crop damage/insurance claim), then crop insurance will have no impact on income.

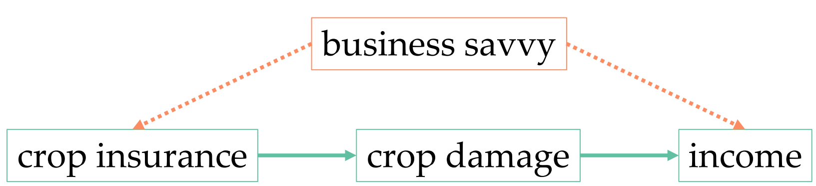

Figure 7: When an exogenous mediator exists, we can separately identify the effect of treatment on the mediator and the effect of the mediator on the outcome, then chain these effects together.

As drawn, the mediator variable provides a way out. We can identify the first link in the causal chain with a simple regression of crop damage on crop insurance. We do not have to worry about the spurious path from crop insurance to crop damage, since income assumes the role of a collider and omitting income closes this path. We can also identify the second link by regressing income on crop damage while controlling for crop insurance, which breaks the other non-causal path. By chaining these impacts together, we recover a causally-interpretable estimate of the effect of crop insurance on income.

Figure 7 implies three assumptions about this mediator. To start, our mediator must intercept all causal paths from crop insurance to income. If there was some sort of direct causal path, then we would not be able to avoid the initial bias we were trying to remedy. This is similar to the validity assumption in instrumental variables. The other two assumptions require the crop damage variable to have unconfounded relationships with crop insurance and income. This means the business savvy variable—or any other unobservable—may not create a non-causal path involving the mediator.

You might think that I've hidden an arrow here. Why couldn't the savviest of farmers mitigate their exposure to weather-related damages through siting or some other mechanism? In this case, I can propose a modification to the DAG. Conditional on the known risk of a disastrous weather event, the event itself is now plausibly-exogenous with respect to the business savvy of individual farmers. Thus adding a measurable control variable like "risk of damage" in between the business savvy and crop damage variables would signal that a control strategy is available to sever this spurious link. We should then run the two aforementioned regressions while controlling for exposure risk.

A DAG-augmented review of a recent JAERE article

When I attend a research presentation or read a journal article, I often sketch a DAG to highlight any key assumptions or barriers to identification presented by the setting and empirical strategy. This is part of an iterative exercise that requires me to regularly check my understanding of the research setting, second-guess my grasp of the identification assumptions, and revise the DAG until some level of confidence is attained. Most economists likely do something involving the first two activities already; I find that adding in the third generative step forces a deeper level of engagement with other people's research.

In this section, I provide an example review of the identification strategy from a recent paper by West, Fairlie, Pratt, and Rose (2021). They studied the impact of receiving an irrigation restriction violation notice on household water consumption in southern California, using a fuzzy regression discontinuity design. I then go on to share the limitations of my review if it was restricted just to the casual inference concepts addressed by DAGs.

Automated enforcement of irrigation regulations for water conservation

In early July, 2015, the water utility serving the city of Burbank sent warnings to households who would be violating the upcoming summer irrigation restrictions based on their behavior in late-June. The notices contained details about a new monitoring system that used real-time data to automatically inform the utility of irrigation outside of permitted times. This treatment resulted in a tenfold increase in the number of households ever notified of an irrigation violation—and signaled a change in the likelihood of enforcement and threat of financial penalties should overuse continue.

West, Fairlie, Pratt, and Rose report a substantial decrease in water consumption in response to receiving a notice—roughly 600 gallons per week, or 31% of mean household use—using a fuzzy regression discontinuity design. Treatment was predominantly determined by whether or not a household was found to be non-compliant with upcoming summer restrictions during a week in late-June; the noncompliance algorithm counted the number of days with peak hourly water usage exceeding an arbitrarily-chosen threshold (by which point irrigation was evident). Since the summer restrictions only allowed for irrigation two days per week, when the third-highest daily peak consumption hour exceeded 125 gallons, a household would be deemed non-compliant.

Thus the running variable of interest is household water use during the third-highest daily peak consumption hour. The utility allowed for some leniency for historically efficient customers, so crossing the above threshold only implied a jump in the *probability* of receiving a warning. The lower-volume customers were instead evaluated based on their fourth-highest daily peak consumption hour, but this "consumption tier" information was not available to the research team. The imperfect compliance in treatment assignment using the stricter rule supported a fuzzy regression discontinuity approach.

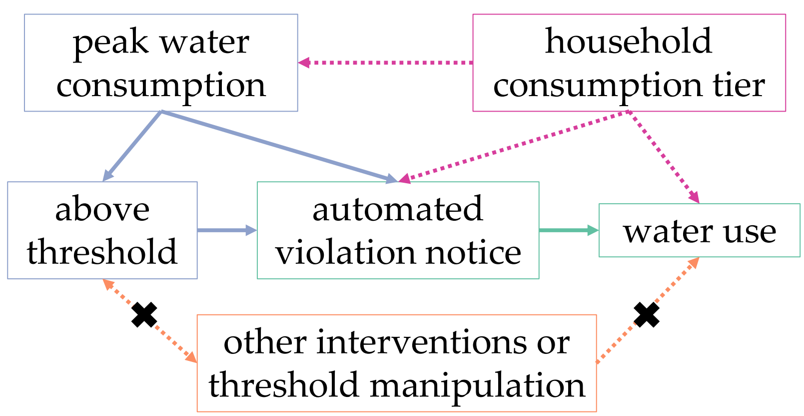

Figure 8: Using our summary of the paper's setting, we can generate a DAG that communicates the assumptions that must hold in order for the empirical strategy to be successful. If the summary admits an arrow inconsistent with the approach, we have reason to be critical. If an author supplies a DAG that doesn't seem to match their story, we can point this out as well.

We can validate this empirical strategy by drawing a DAG that reflects the authors' understanding of the data generating process. Figure 8 aids us in running through the requirements for a satisfactory fuzzy regression discontinuity design. An easy place to start is with the typical instrumental variables assumptions: relevance, independence, and validity. Clearly, the instrument (whether or not the stricter peak consumption threshold is exceeded) partly determines whether or not a household receives an irrigation restriction violation notice—it was a necessary precondition. And because the crossing of the peak consumption threshold is defined entirely by a household's peak water consumption, when we condition on the running variable, the instrument is independent of the unobserved confounder—the household consumption tier. Lastly, the instrument only impacts household water use through the delivery of a notice.

While the consumption tier confounds the relationship between receiving a notice and summer water use—those who use less water historically would likely continue to do so and not receive a notice, and the opposite would be true for the higher-volume households, leading to an under-estimate of the treatment effect due to unbalanced treatment and control groups—the instrumental variables approach blocks the confounding path.

We can now discuss the assumptions specific to the regression discontinuity approach. First, we hope to separate the drop in water use due to receiving a notice from the water use trend in the peak consumption running variable (there should be an upward trend since a household with higher peak usage will have higher total consumption). Upon controlling for the running variable, any paths that utilize continuous variation in water use across the threshold are severed, leaving only the potential discontinuity to be measured.

Second, we must avoid the possibility of there being an additional cause for a discrete change in water use when comparing households just over the threshold to those just below it. There are several ways to state the "local randomization" assumption, but I prefer to symbolize it with a non-path from the threshold to the outcome. Because we ultimately regress water use on the peak consumption threshold—and not the notice status itself—we merely *wish* to attribute the threshold-derived jump in outcomes to the notice, rather than some other factor. The authors make this assertion easy to believe, as no concurrent intervention utilizing the threshold existed (eliminating the arrow from the threshold to a potential intervention mediator), and no households were privy to the threshold or the automated detection algorithm (eliminating the arrow from a potential sorting confounder to the threshold).

From this discussion, we can be confident that the authors' fuzzy regression discontinuity approach generates an unbiased estimate of the effect of the irrigation restriction violation notices on summertime water use. But you may notice that the abstractions our DAG uses leave us with an incomplete empirical review. I will address this below.

DAGs are not the be-all and end-all

Because DAGs provide a non-parametric representation of the underlying causal structure of the data generating process, any interaction between two variables that each impact a third variable is implicitly allowed. For example, it is a combination of the consumption tier and peak water consumption variables that modify whether or not a notice is sent to a notice-eligible household. We'd of course control for the running variable, but as the consumption tiers are not measurable, any interaction between the two variables would've still had an opportunity to confound the notice-water use relationship. The DAG doesn't explicitly represent this interaction nuance, but it does suggest a consumption tier-related issue of some kind is still possible since controlling for the peak consumption variable only addresses two of the three non-causal paths involving the consumption-tier variable. The instrumental variables strategy tackles the remaining path. The only way to rule out an unwanted interaction effect is to make a model assumption separately from the DAG that it does not exist, or use an empirical design that controls for both factors.

It should be clear that DAGs cannot provide a one-to-one mapping to everything you may know about econometrics, and for this reason I believe their main role is augmentative. They are a tool for building confidence in the causal-interpretability of an estimate. The interaction example is a subset of a general concern: a DAG does not tell us *how* we should control for a particular variable, and this is where we must rely on more traditional econometric knowledge. The DAG imposes no restriction on model specification (e.g. local linear regression, bandwidth, etc.), and assumes that we can choose the appropriate regression model for our setting. There is a different set of assumptions related to functional form to be considered separately from the causal ones shown by the DAG.

The other shortcoming is that DAGs cannot describe the extent of the external validity of the estimate of interest. In the case of heterogeneous treatment effects, an instrumental variables strategy typically makes use of a monotonicity assumption (trivially met here) to show that our estimate is a local average treatment effect (although no treated households were below the threshold, so this is equivalently an average treatment effect on the treated). And our reliance on the extrapolation of counterfactuals in regression discontinuity means that this treatment effect is limited in scope to settings where the peak consumption threshold for notice eligibility is around 125 gallons. Our traditional counterfactual logic used here does not map to any feature in the graph.

Even as environmental economists become more comfortable with the graphical approach, it is evident that additional discussion outside of the DAG framework will still be needed in our research. And as we saw in the primer, certain questions involving nonrecursive model designs cannot be conveyed via DAGs. These two facts are often seen as evidence that DAG adoption is impractical, but this extreme view ignores their potential for improving communication of causal knowledge in the majority of settings and implies a false dichotomy that forces us to choose one framework over another. In reality, we don't give anything up by using a new tool in conjunction with the old ones.

Opportunities for implementation

When we lack the facilities to present causal knowledge, we approach causal inference blindfolded. Since identification strategies are nuanced and borne out from a place of deep familiarity with specific data and settings, our current approach to their explanation involves the additional challenge of speaking to other blindfolded individuals.

Economics is one of the rare disciplines that uses presentations to improve papers in progress—rather than report on finished products. Yet the status quo presentation moves quickly and inadvertently utilizes abundant hand-waving through claims of causal effect identification. I must often suspend disbelief in order to enjoy a talk, but this puts me in a weaker position of understanding and makes me less likely to be able to follow along with or provide feedback to the presenter. The inclusion of a DAG could increase the benefits of presenting our work. It is easy to evaluate whether you believe a displayed DAG represents the data generating process, turn any discrepancy into a question for the presenter, or test if a variation still admits the use of the proposed identification strategy.

We have an opportunity to make the research process more transparent and responsive to criticism. DAGs engender more cogent feedback from our peers. When reviewers ask us for some sort of sensitivity analysis, they now have a language for raising concerns about causal hypotheses, which are [arguably] of more value than semi-inspired modifications to the authors' preferred regression specification. A new causal hypothesis generates a new DAG, which then organically motivates a new regression. This is objectively better than suggesting different sets of fixed-effects.

DAGs create tremendous value for pedagogy (e.g. Cunningham (2021); Huntington-Klein (2022)). Graduate studies in economics today highlight causal effect identification, yet many of the insights that I share in this article are absent from applied econometrics instruction. Even more value is created at the undergraduate level; my introductory econometrics course attracts a diverse group of students to economics through its primary emphasis on the causal storytelling skills that we pick up as econometricians. The use of DAGs only increases access to my class and further facilitates the building of this desirable competency. Prioritizing this objective in economics education may similarly increase our ability (and that of our students) to communicate with policymakers and stakeholders.

As we saw in the last section, DAGs are not limited to toy models. A common "anti-arrow" complaint discards DAGs because they are technically redundant. But DAGs are better-suited to building confidence in the causal-interpretability of an estimate and sharing our logic with others. Environmental economists ultimately hope to generate research that informs environmental policy, and this tool may make that more likely.

References

Cunningham, S. (2021). *Causal Inference: The Mixtape*. Yale University Press.

Ferraro, P. J. and Shukla, P. (2020). Feature—Is a Replicability Crisis on the Horizon for Environmental and Resource Economics? *Review of Environmental Economics and Policy*, 14(2):339–351.

Heckman, J. J. and Pinto, R. (2022). The Econometric Model for Causal Policy Analysis. *Annual Review of Economics*, 14(1):893–923.

Huntington-Klein, N. (2022). *The effect: an introduction to research design and causality*. CRC Press, Taylor & Francis Group.

Imbens, G. W. (2020). Potential Outcome and Directed Acyclic Graph Approaches to Causality: Relevance for Empirical Practice in Economics. *Journal of Economic Literature*, 58(4):1129–1179.

Pearl, J. (1995). Causal diagrams for empirical research. *Biometrika*, 82(4):669–688.

Segerson, K. (2019). Reflections—On the Role of Theory in Contemporary Environmental and Natural Resource Economics. *Review of Environmental Economics and Policy*, 13(1):124–129.

West, J., Fairlie, R. W., Pratt, B., and Rose, L. (2021). Automated Enforcement of Irrigation Regulations and Social Pressure for Water Conservation. *Journal of the Association of Environmental and Resource Economists*, 8(6):1179–1207.