GfÖ
GfÖ Ecological Society of Germany, Austria and Switzerland
Basic and Applied Ecology 19 (2017) 1–10
Basic and Applied Ecology
www.elsevier.com/locate/baae

REVIEW

Quasi-experimental methods enable stronger inferences from observational data in ecology

Van Butsic[a,b], David J. Lewis[c], Volker C. Radeloff[d], Matthias Baumann[b], Tobias Kuemmerle[b,e,*]

[a]*Department of Environmental Science Policy and Management, University of California–Berkeley, 130 Mulford Hall #3114, Berkeley, CA 94720-3114, USA*
[b]*Geography Department, Humboldt University Berlin, Under den Linden 6, 10099 Berlin, Germany*
[c]*Department of Applied Economics 200A Ballard Extension Hall, Oregon State University, Corvallis, Oregon 97331-3601, USA*
[d]*SILVIS Lab, Department of Forest and Wildlife Ecology, University of Wisconsin-Madison, 1630 Linden Drive, Madison, WI 53706, USA*
[e]*Integrative Research Institute on Transformations of Human-Natural Systems (IRI THESys), Humboldt University Berlin, Under den Linden 6, 10099 Berlin, Germany*

Received 22 August 2016; accepted 20 January 2017
Available online 2 February 2017

Abstract

Many systems and processes in ecology cannot be experimentally controlled, either because the temporal and spatial scales are too broad, or because it would be unethical. Examples include large wildfires, alternative conservation strategies, removal of top predators, or the introduction of invasive species. Unfortunately, many of these phenomena also do not occur randomly in time or space, and this can lead to different biases (selection bias, unobserved variable bias) in statistical analyses. Economics has evolved largely without experiments, and developed statistical approaches to study "quasi-experiments", i.e., situations were changes in time or space reveal relationships even in the absence of a controlled experiment. The goal of our paper was to compare and evaluate four quasi-experimental statistical approaches commonly used in economics, (1) matching, (2) regression discontinuity design, (3) difference-in-differences models, and (4) instrumental variables, in terms of their relevance for ecological research. We contrast the strengths and weaknesses of each approach and provide a detailed tutorial to demonstrate these approaches. We suggest that quasi-experimental methods offer great potential for investigating many phenomena and processes in ecological and coupled human-natural systems. Furthermore, quasi-experimental methods are common in environmental policy research and policy recommendations by ecologists may be more valuable when based on these methods.

Keywords: Observational data; Natural experiments; Econometrics; Matching; Difference-in-differences; Regression discontinuity design; Instrumental variables

*Corresponding author at: Geography Department, Humboldt University Berlin, Under den Linden 6, 10099 Berlin, Germany.
Tel. : +49 (0) 30 2093 9372; fax: +49 (0) 30 2093 6848.
E-mail address: tobias.kuemmerle@hu-berlin.de (T. Kuemmerle).

http://dx.doi.org/10.1016/j.baae.2017.01.005
1439-1791/

Introduction

Ecologists and conservation biologists alike strive to understand causal relationships such as the effects of alternative land management strategies on biodiversity (Macchi, Grau, Zelaya, & Marinaro 2013; Phalan, Onial, Balmford, & Green 2011) or whether protected areas are effective in slowing deforestation (Andam, Ferraro, Pfaff, Sanchez-Azofeifa, & Robalino 2008; Arriagada, Ferraro, Sills, Pattanayak, & Cordero 2012). Experiments are the gold standard for scientific discovery, and there is a rich history of experimentation in ecology. However, experiments are often infeasible or impractical (Diamond 2001; Sagarin & Pauchard 2010), raising the question of how to gain the strongest possible inference from observational data. This question is far from trivial because many of the phenomena that cannot be experimentally controlled also do not occur randomly and thus are problematic for many forms of statistical analysis.

Ecologists have developed and adopted numerous statistical techniques to analyze non-random applications of treatments in ecological settings such as hierarchical modeling (Bolker et al. 2009; Cressie, Calder, Clark, Hoef, & Wikle 2009; Royle & Dorazio 2008) and structural equation models (Grace et al. 2012; Shipley 2002; Wootton 1994). Model selection (Burnham & Anderson 2002) can help when selecting between competing models in non-experimental settings, and there is generally high awareness of issues such as pseudo-replication (Hurlbert 1984) and spatial autocorrelation (Legendre 1993) in both experimental and observational data. Likewise, many ecologists are familiar with and have adopted theory and techniques for causal inference from foundational works by Pearl (2009) and Pearl, Glymour, and Jewell (2016). However, other approaches developed outside of ecology could complement current ecological methods.

Economics evolved largely without experiments and developed "quasi-experimental", "causal" or "micro-econometric" methods to analyze observational data (Cameron & Trivedi 2005) in labor economics (Ashenfelter 1987), development economics (Strauss 1986), environmental economics (Greenstone & Gayer 2009), health economics (Jones 2007), and education economics (Angrist & Krueger 1991). Given the success of economics to develop techniques to gain stronger inferences when analyzing observational data, we suggest these techniques may be useful for ecological questions as well. We focus here on the four most common quasi-experimental techniques in economics: matching estimators, regression discontinuity design, difference-in-differences modeling, and instrumental variables.

The key to quasi-experimental methods is to identify valid control and treatment groups and account for potentially unobserved correlation between treatment assignment and outcome. This can be done by matching observations with similar characteristics but different treatments (matching estimators), by exploiting break points in data series (discontinuity design), by observing changes in time and treatment (difference-in-differences modeling), or by using predicted values generated by exogenous covariates as substitutes for endogenous covariates (instrumental variables).

Quasi-experimental approaches may be useful to answer ecological questions (Creel & Creel 2009) because many ecological processes take place in settings where non-random treatments and unobservable variables may lead to bias. For example, bird species distributions may depend on factors such as understory vegetation or the availability of snags for which broad-scale data is not available (Beaudry et al. 2010). Similarly, natural disturbances such as fires, windfall, or landslides, are often correlated with land-use legacies of past disturbances that researchers cannot account for (Foster, Fluet, & Boose 1999; Seidl, Schelhaas, & Lexer 2011). Whenever ecological processes are non-random treatments, there is a strong possibility that the treatment and outcome are correlated in ways that are unobservable, and quasi-experimental approaches offer a way to reduce this potential bias (See Table 1 for a list of studies and suggestions for applying these methods).

Our goal here was to provide a non-technical comparison and evaluation of the strengths and weaknesses of four quasi-experimental analysis methods, namely (1) matching, (2) regression discontinuity design, (3) difference-in-differences models, and (4) instrumental variables, in terms of their relevance for ecological research. In addition, we provide a simulated dataset underlying the examples we use, along with the code (for both R and STATA) to generate these data and solve the models (see Appendices A and B in Supplementary materials).

Four quasi-experimental approaches

General setup

The approaches presented here are designed to quantify the impact of a treatment (e.g., a new protected area, a wildfire, or a change in climate) on an outcome (e.g., deforestation, invasive species spread, or biodiversity change). The challenge is to gain the strongest inference possible – or the estimated parameter that is the closest possible estimate of the true effect – when analyzing observational data where the treatment was not random.

To demonstrate the applicability of these approaches we apply them to the hypothetical problem of assessing the impacts of a large wildfire (i.e., treatment) on plant species richness (i.e., outcome). Field data from a survey of plots within (treated group) and outside the fire perimeter (control group) are simulated for species richness, elevation, slope and distance to water (Fig. 1). Since an estimated treatment effect may differ from the true effect simply due to random chance in any particular dataset, we simulate 1000 realizations of this dataset and evaluate the *average* treatment effect across the 1000 realizations to assess the bias from an estimator. The

Table 1. Data requirements, possible ecological applications, and key references for each of the four quasi-experimental methods described in paper. As general references we suggest (Angrist & Pischke, 2008; Cameron & Trivedi, 2005; Wooldridge, 2011).

Method	Possible ecological applications	Key references
Matching	Wherever treatments are non-random but all relevant variables are observed by researchers. Examples may include wildfire, landslides, windfall, bleached corals, invasive species spread, and areas with a unique land use.	Alix-Garcia et al. (2012), Andam et al. (2008), Andam, Ferraro, and Hanauer (2013), Caliendo and Kopeinig (2005), Ferraro, McIntosh, and Ospina (2007), Heckman, Ichimura, Smith and Todd (1996), and Rosenbaum and Rubin (1983a, 1983b)
Regression discontinuity design	Wherever there is a sharp break in treatments. Examples include breaks in soil types, administrative borders, protected areas borders, fishing zones and land use zones, species ranges.	Chay and Greenstone (2004), Grout et al. (2011), Imbens and Lemieux (2008), Lee and Card (2008), and Lee and Lemieux (2009)
Difference-in-differences	Where there are observations before and after a treatment. Examples include population dynamics before and after the introduction of a pathogen or the extirpation of a top predator, protected areas if pre-treatment data is available, wars, socioeconomic shocks.	Athey and Imbens (2006), Bertrand et al. (2004), and Horsch and Lewis (2009)
Instrumental variables	Where the treatment assignment is simultaneously determined with the outcome.	Angrist and Krueger (2001), Busch and Cullen (2009), Creel and Creel (2009); Miguel, Styanath, and Sergenti (2004), and Sims (2010)

advantage of a simulated dataset is that the researcher knows the true parameter values, and so can directly assess the bias arising from alternative estimators.

Terminology: selection bias, omitted variables, and unobservables

Ecologists and economists sometimes use the terms selection bias, omitted variable bias, and unobservables, in slightly different ways. We therefore here introduce these terms before introducing the methods. The term *selection bias* stems from the social sciences and medical literature where participants "self-selected" into treatment groups, such as, participation in a job training program. When that is the case, program participants will likely have characteristics that were not observed by the researcher, but that impacted participation and the outcome. For example, participants in a job training program are likely more motivated to find work than non-participants. This unobserved difference is referred to as self-selection bias. In ecology, the phrase selection bias is less intuitive since a place does not self-select to be burned, or protected, and an individual does not self-select to be infected or otherwise treated.

However, with the exception of matching methods, where selection is motivated by observed components, selection bias can also stem from *omitted variable* bias, where an omitted variable is correlated with both treatment assignment and the outcome. In the example of the job training program, if the researcher could observe the omitted variable, in our case motivation, then selection bias could be controlled for. Likewise, in the wildfire example presented here, we could correct for selection bias by having information on the omitted variables (e.g., land use history) that influence plant species richness and wildfire. Economists often use the term *unobservable* as a proxy for omitted variables, especially if the exact nature of the omitted variables is not known. That is, we may know that unobservables impact fire occurrence and plant diversity, without being able to say exactly if the most important unobservable is past land use history, small-scale geographic variation or some other variable for which there is no data. This is why the term *unobservables* is often preferred over *omitted variables*, since we do not know precisely what the omitted variables are.

Matching estimators to accounting for selection bias

One way to account for selection bias is to select treated and control observations where the likelihood of treatment is similar (as is the case in a randomized trial). This can be accomplished via matching estimators (Rosenbaum & Rubin 1983a, 1983b; Abadie & Imbens 2006; Ho, Imai, King, & Stuart 2006). Treated and control observations can either be matches based on their covariates (e.g., Mahalanobis matching, covariate matching), or their probability of treatment (e.g., propensity score matching).

In our wildfire example, a classical regression analysis would lead to biased coefficients if characteristics that deter-

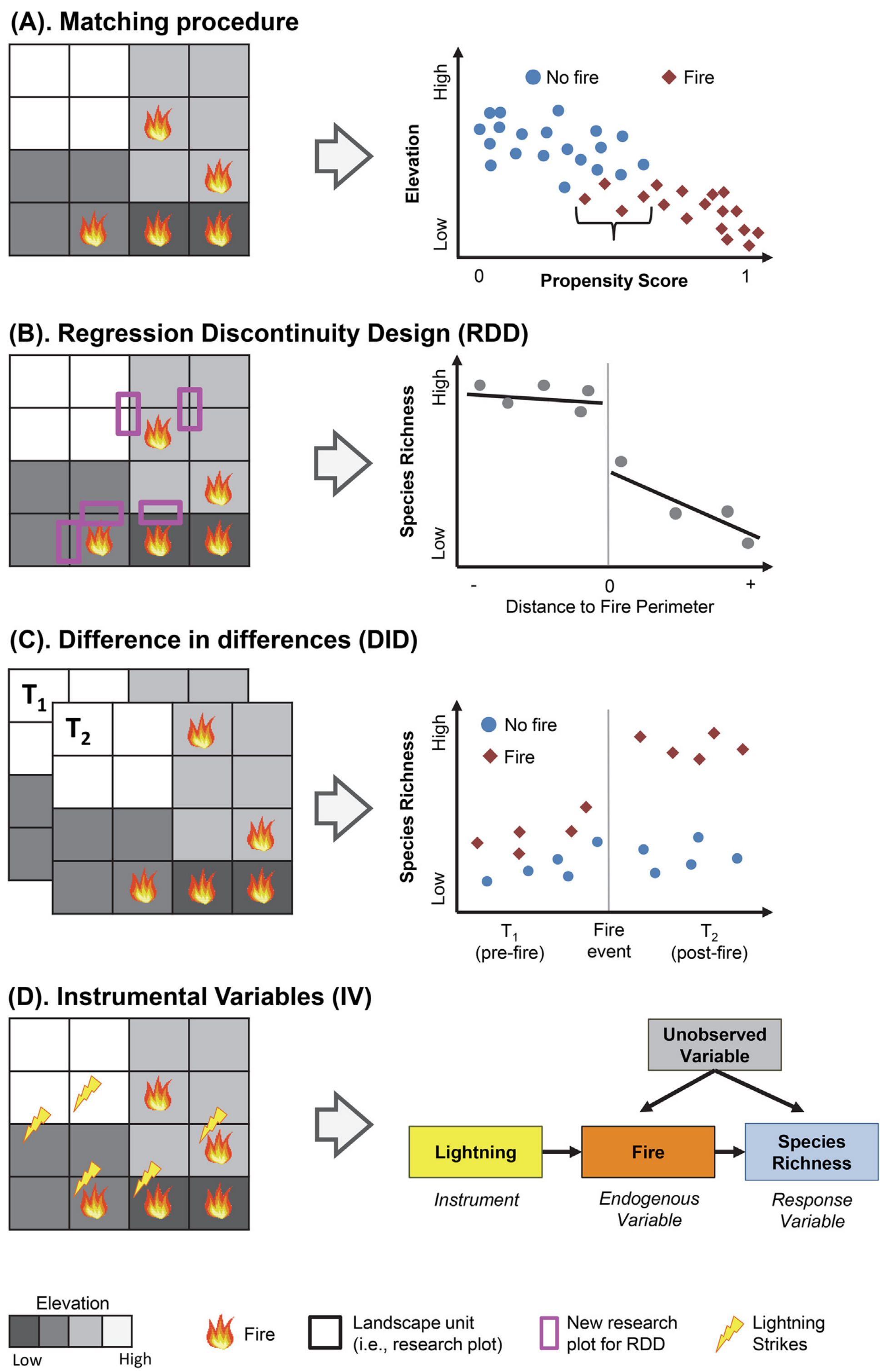

Fig. 1. Visualization of four methods described in the manuscript. (A) *Matching procedures*. The right panel shows a landscape, gridded into sample plots which vary in elevation and the presence of fire. Because fire is correlated with lower elevations, most of the treated observations are outside the range of the control observations in our example. The left panel displays the range of propensity scores in relation to elevation and treatment. Matches are chosen from within the bracketed area, and the tails of the distribution are excluded from the analysis. (B) *Regression discontinuity design.* In the left panel new research plots are selected at the edge of burned parcels, where both observed and unobserved covariates may be similar. The right panel shows species richness on each side of the discontinuity. The large jump of the function at the discontinuity suggests that there is a difference in species richness between the burnt and unburnt areas. (C) *Difference in differences*. This panel represents a landscape before and after a burn event. The right panel depicts the change in species richness after the fire which can be interpreted as the treatment effect. (D) *Instrumental Variables*. We use lightning as an instrument for fire because it is correlated with fire,

mine species richness are also strongly correlated with the likelihood of fire and if this relationship is not precisely specified. For example, elevation and slope in all likelihood influence both fire and richness, making it difficult to separate fire effects from those of slope and elevation, especially when the type of the exact form of the functional relationship between variables (e.g., linear vs. non-linear) is not known.

To demonstrate this effect, we simulated a dataset (code available in Supplementary materials) in which the probability that a plot is burned (fire) is a positive function of continuous measures of both elevation and slope. Further, suppose the "true" effect of fire was to increase species richness by five. Slope affects species richness through a non-linear quadratic relationship, while all other covariates had a linear relationship with species richness, and there was no correlation between fire and any unobservable covariates of species richness. For this dataset, we estimated the impact of fire on species richness using both ordinary least square (OLS) regression, in which slope was deliberately mis-specified as a linear relationship, and a matching estimator. Specifically, we calculated the propensity score (i.e., the probability of treatment) using a probit regression, for each plot, and matched plots inside and outside the fire area with similar propensity scores.

We compare OLS and matching estimators for 1000 simulated datasets. On average, OLS overestimated the fire effect at 7.14, while matching estimates the effect to be 4.91, very close to the true effect of 5 (See Supplementary materials for more details). Furthermore, matching nearly eliminated differences in the means of the covariates of the treated and the untreated observations (Table S6). Intuitively, matching identified valid controls (i.e., observations with similar characteristics) to compare the treated observations with, and made inference stronger by reducing the difference between the estimated treatment effect and the true effect.

The systematic bias from OLS arises from two facts. First, the true functional form relationship of slope on species richness is quadratic, and omitting the squared component of slope in OLS regression biased the parameter estimate on slope. Second, because fire is correlated with slope, the bias from mis-specifying the functional form for slope biases the OLS estimate of the treatment effect. Matching intuitively breaks the correlation between fire and covariates, making estimates less sensitive to errors in the functional form of covariates. Importantly, matching allows the researcher to closely estimate the true effect without correctly specifying the functional relationship of other variables.

However, matching statistics are still subject to omitted variable bias. If there are unobserved covariates (e.g., land-use history) correlated with both treatment and outcome, then matching estimates will still be biased. Furthermore, matching estimators only provide an accurate estimate for plots that were selected as matches, and the treatment effect may differ for the whole population.

Regression discontinuity design

Regression discontinuity design (RDD) accounts for omitted variable bias by exploiting a discontinuity in either space, time, or policy to separate observations into treatment and control groups (Imbens & Lemieux 2008; Lee & Lemieux 2009; Grout, Jaeger, & Plantinga 2011). The key assumption is that at the discontinuity, unobservables are equal between treated and control observations. This is similar to the intuition of a "natural experiment" where a natural break on the treatment leads to an experiment like condition (Diamond 2001).

In our wildfire example, a strong discontinuity is the boundary of the wildfire. If the underlying covariates, both observable (slope, elevation, distance to water) and unobservable (land use history) are similar on both sides of the edge of the fire, then the average difference in outcomes between treated and control observations provides an estimate of the treatment effect. However, the treatment effect is only unbiased if land-use history did not determine the fire boundary.

Again, we estimated the impact of fire on plant richness using linear regression and regression discontinuity design with simulated data. We assumed low correlation between fire, elevation and slope so that fire takes place more equally across the landscape than in the previous section. However, in contrast to the dataset in the previous section, we assumed that land-use history influences species richness strongly (i.e., if a plot is in an area that has historically been farmland, species richness increases by 5) and, importantly, land-use history is not known to the researcher (i.e., it is an unobserved variable) and also correlated with current fire risk. In other words, our simulated dataset has a strong omitted variable. Over our 1000 simulations, the average OLS estimated treatment effect is only 2.75 while the RDD estimate was 5.20, close to the true effect of 5.

In practice, RDD can be implemented with either parametric or non-parametric regression models (we used a local polynomial regression). One question is how far from the discontinuity to use data. This is an important question if the regression is non-linear and there may be a bias-efficiency trade-off between larger and smaller distances, although data-driven methods can estimate optimal choice in many settings (Lee & Lemieux 2009). A nice property of RDD is that the method lends itself to graphical analysis, since there is typically a visual jump in the outcome at the discontinuity if there is a non-zero treatment effect. In our wildfire example, this jump would manifest itself in the number of

but the causal relationship is only one way – lightning causes fire. We assume some unobserved correlation between treatment and outcome such that there is correlation between the error term and the treatment. The instrument can then be used to break this correlation.

species at the edge of the burned area. However, RDD provides treatment effects potentially only for a small range of the data.

Difference-in-differences modeling

Data from before and after a treatment can be exploited to control for omitted variable bias using difference-in-differences (DiD) estimation (Bertrand, Duflo, & Mullainathan 2004; Donald & Lang 2007). DiD estimation is conceptually similar to the before-after-control-impact (BACI) experimental design (Stewart-Oaten & Bence 2001). In DiD models omitted variable bias is eliminated (1) even if treatment is correlated with time-invariant unobservables (which are effectively differenced out of the model), and (2) if treatment is uncorrelated with time-varying unobservables. However, the two key identifying assumptions of DiD estimation are (1) that the treatment and time-invariant unobservables are additively separable, and (2) the parallel-trends assumption, which states that in the absence of treatment, the treatment and control group would have experienced similar changes in the outcome.

Conceptually, the DiD estimator calculates the treatment effect as the difference between the differences of the treated and non-treated observations before and after treatment. In our wildfire example, this is the difference between post and pre-treatment plant species richness in the burned plots, minus the difference between post and pre-treatment species richness in the non-burned plots. Again, we used a simulated dataset to compare OLS with a DiD estimator. In this dataset, the true treatment effects of both the fire effect and the land-use history effect were 5. For the DiD estimation, the simulated data for this example contains two periods, before fire, and post-fire. Land-use history is time-invariant, correlated with fire, and is again unobserved. For 1000 simulated datasets, the average estimate of the fire treatment effects from the OLS regression was 2.71 (S.D. 0.13), substantially lower than the true effect of 5, whereas the mean DiD estimate was exactly equal to 5 (S.D. 0.19).

Most DiD models are estimated using regression techniques, but non-parametric approaches or matching techniques can also be used (Athey & Imbens 2006), and can be superior (Heckman, Ichimura, & Todd 1998) if there are distributional differences in the underlying covariates that the regression cannot control for. Likewise, the DiD estimator can be expanded to incorporate multiple time periods and multiple treatments. The validity of the DiD estimator is based on the parallel-trends assumption, which is not directly testable (Bertrand et al. 2004), but can be loosely evaluated. Graphing the outcomes over time allows a researcher to visually examine parallel trends and if a DiD analysis on two periods before the treatment shows a treatment effect, then the parallel trends assumptions do not hold.

Instrumental variables

In cases where there is no discontinuity in the treatment assignment and where data is only available for one period, another way to address the omitted variable bias problem is the instrumental variables (IV) approach. The basic strategy of this approach is to introduce an instrumental variable that is correlated with the treatment, but otherwise independent of the potential outcome. This instrumental variable can then be used to eliminate the correlation between the treatment and unobservables that are affecting the outcome. The use of an instrumental variable can – in principle – reduce the omitted variable bias term to zero (Angrist & Krueger 2001).

An IV estimation is most commonly done by parameterizing a two-stage least-squares regression. The first-stage regression predicts the level of the treatment, using the instrumental variable and all other exogenous variables as predictors. The predicted treatment, rather than the actual treatment, is then used in the second-stage regression to estimate the treatment effect. Including predicted-treatment rather than actual treatment in the second stage introduces a new set of variation into estimation that requires standard error adjustment (Wooldridge 2011). In practice, it is common to use several instrumental variables alternatively.

The key to IV estimation is to identify an instrumental variable that is both correlated with the treatment and yet uncorrelated with the unobservable characteristics. In our wildfire case, a good instrument would be one that is correlated with the treatment (fire), but uncorrelated with plant species richness except through fire. One option may be lightning strikes. We assume that lightning strikes only affect plant species diversity by starting fires, not through any other mechanism, and that lightning strikes are uncorrelated with unobserved past land use histories on a plot.

In our simulated data, fire is again correlated with unobserved land-use history and with a binary variable indicating the presence of a lightning strike, while unobserved land-use history and lightning strikes are uncorrelated. We use a two-stage least-squares estimator to calculate the impact of fire on species richness. In the first-stage regression, we predict the level of fire by regressing lightning strikes, elevation, distance to stream and slope on fire. We then substitute these predicted values for the actual value of fire in the second stage regression. Across 1000 simulated datasets, the mean OLS estimate of the fire effects is 2.98 (S.D. 0.32), while the mean IV estimate was equal to 4.97 (S.D. 0.70), much closer to the true effect.

One limitation of the IV approach is that it is impossible to test directly whether there is correlation between the outcome unobservables and the instrument. However, there are several tests that can indirectly examine the validity of this assumption. One common approach is to regress the instrument on the observed covariates. If the instruments are unrelated, then it is more likely that the unobservables are also unrelated. Another good indicator can be the correlation between the instrument and potential outcomes in a place or time where there should be no such relationship.

When to use which quasi-experimental technique

Which quasi-experimental technique to use for a given question is a key choice a researcher makes when modeling the impact of a treatment on an outcome. This decisions is often influenced heavily by data availability. While there can, and should be, a great deal of nuance in making this choice, we here provide a few tips on how different data structures can influence what choice of technique is most useful.

A first question to ask is if there are data that one could obtain that impacts both treatment and outcome? If the answer is no, the system in question may be free of omitted variable bias. Using matching may be a good choice in this case because it can balance covariates when regression fails, especially if there are questions about the functional form of the relationships. If however, one can think of variables that impact both treatment and outcome, and that one does not have access to, then an analysis would most likely suffer from omitted variable bias and alternatives to matching (or OLS regression) that can control for some types of omitted variable bias should be used.

A second key data structure to look for are breakpoints in the data. These breakpoints may be spatial, for example the boarder of a wildfire or protected area, or numerical, for example if a treatment might be applied or below a given threshold (i.e., pesticide is only applied if at least 20 pest are present. In the presences of such breaks, regression discontinuity design may be well suited, especially if one thinks that the omitted variable bias is similar on either side of the break points.

Third, if there is temporal variation in the treatment, and data is available both before and after treatment, difference-in-differences modeling may be a good option. This is because when data before and after a treatment are available, difference-in-difference modeling can control of time invariant unobservable. Difference-in-difference modeling can also be combined with matching if one thinks that there is additional unbalance in the covariates that cannot be controlled for by regression alone.

Finally, if the data is such that it is difficult to tell whether the treatment is a cause or an effect, instrumental variables may be appropriate. However, in such settings additional data is needed, namely, data on a valid instrument. Finding such an instrument may take creativity on the part of the researcher, as there is no single method to discover a valid instrument.

Discussion

Understanding the relationships among the different elements of ecological systems is the goal of ecological science and necessary for environmental policy making. Experimentation, controlled and replicated, is the best way to infer causal relationships, and provides the strongest evidence possible. However, quasi-experimental techniques may be useful to ecologists who want to gain the strongest inference possible from observational data obtained in non-experimental settings. Here, we compared four techniques – matching, regression discontinuity design, difference-in-differences models, and instrumental variables – in terms of their relevance for ecological research, and we suggest that all four can be valuable for answering ecological questions. Econometric texts provide more in-depth, mathematical descriptions of these techniques (Angrist & Pischke 2008; Cameron & Trivedi 2005; Wooldridge 2011).

The heart of all four techniques is the identification of valid controls and treatment groups from data where the treatment is not assigned randomly. The reason why so much emphasis is placed on valid controls is that non-random assignment can bias results in two ways. First, non-random assignment can lead to treated and control groups having different underlying distributions of covariates. In cases where the overlap in these distributions is poor, even including these covariates in a regression may not remove bias from coefficients. The solution to this problem is to recreate the control and treated groups by selecting observations where the covariates are similar. Matching techniques are well-suited to do this.

The second problem relates to non-random assignment and occurs when an unobserved variable co-varies with both the treatment and the outcome. In this case, simply creating treated and controlled samples based on observed covariates will not solve the omitted variable problem. Instead, it is necessary to control for omitted variable bias along with observed differences in covariate distributions. One can do this by taking advantage of breakpoints in the data (regression discontinuity design), examining changes over time (difference-in-differences models), or using variables that are unrelated to the outcome, but correlate with the covariate of interest (instrumental variables approach). Which technique is most appropriate for a given question depends upon the problem at hand and the nature of the available data (See Supplementary materials for a discussion on when to use what technique).

As answering many ecological questions requires studying empirical data that contain potential biases due to non-random treatments and unobservable variables, the quasi-experimental methods discussed here may be of great value to fundamental and applied ecologists. Perhaps the most prominent example of where these methods have recently been applied in ecology is the assessment of protected area effectiveness (Andam et al. 2008; Honey-Rosés, Baylis, & Ramírez 2011; Sims 2010). Protected areas are often placed on lands with low threat of deforestation (Joppa & Pfaff 2009), introducing bias into direct comparisons of deforestation inside and outside of protected areas because the underlying drivers of deforestation are fundamentally different (Andam et al. 2008). Matching estimators reduce this bias by identifying those observations that are valid counterfactuals. Similar statistical problems exist in a host of policy questions that may be of interest to natural scientists, including species protection (Greenstone & Gayer 2009), the effectiveness of payments for ecosystem services (Alix-Garcia, McIntosh, Sims, & Welch 2013), pollution mitigation (Chay & Greenstone 2004; Greenstone & Gallagher 2008), and land-use change (Alix-Garcia, Kuemmerle, & Radeloff 2012; Bento, Towe, & Geoghegan 2007; Lewis, Provencher, & Butsic 2009).

When correctly parameterized and applied, these quasi-experimental techniques provide strong evidence for, and better estimates of, the relationships between elements in ecological systems. However, they do not replace controlled experiments. Indeed, each technique is subject to a set of unique assumptions that must be satisfied, and often these assumptions are not testable. Therefore, as in experimentation, successful implementation of these techniques places a premium on research design. In the best case, a well-designed research project using observational data combined with quasi-experimental techniques can lead to estimates that closely identify the true relationships between a treatment and an outcome.

Acknowledgements

This research was supported by the Alexander von Humboldt Foundation, the Einstein Foundation Berlin, Germany, and the National Science Foundation Coupled Natural Human Systems program grant BCS-0814424. We thank three reviewers and the editor for constructive and very helpful comments on earlier manuscript versions.

Appendix A. Supplementary data

Supplementary data associated with this article can be found, in the online version, at http://dx.doi.org/10.1016/j.baae.2017.01.005.

References

Abadie, A., & Imbens, G. W. (2006). Large sample properties of matching estimators for average treatment effects. *Econometrica*, *74*(1), 235–267. http://dx.doi.org/10.1111/j.1468-0262.2006.00655.x

Alix-Garcia, J., Kuemmerle, T., & Radeloff, V. C. (2012). Prices, land tenure institutions, and geography: A matching analysis of farmland abandonment in post-socialist Eastern Europe. *Land Economics*, *88*(August), 425–443.

Alix-Garcia, J., McIntosh, C., Sims, K. R. E., & Welch, J. R. (2013). The ecological footprint of poverty alleviation: Evidence from Mexico's oportunidades program. *Review of Economics and Statistics*, *95*(2), 417–435. http://dx.doi.org/10.1162/REST_a_00349

Andam, K., Ferraro, P., Pfaff, A., Sanchez-Azofeifa, G., & Robalino, J. (2008). Measuring the effectiveness of protected area networks in reducing deforestation. *Proceedings of the National Academy of Sciences of the United States of America*, *105*(42), 16089–16094.

Andam, K. S., Ferraro, P. J., & Hanauer, M. M. (2013). The effects of protected area systems on ecosystem restoration: A quasi-experimental design to estimate the impact of Costa Rica's protected area system on forest regrowth. *Conservation Letters*, http://dx.doi.org/10.1111/conl.12004, online only

Angrist, J. D., & Krueger, A. B. (1991). Does compulsory school attendance affect schooling and earnings? *The Quarterly Journal of Economics*, *106*(4), 979–1014.

Angrist, J. D., & Krueger, A. B. (2001). Instrumental variables and the search for identification: From supply and demand to natural experiments. *Journal of Economic Perspectives*, *15*(4), 69–85. http://dx.doi.org/10.1257/jep.15.4.69

Angrist, J. D., & Pischke, J.-S. (2008). *Mostly harmless econometrics: An empiricist's companion*. Princeton University Press.

Arriagada, R. A., Ferraro, P. J., Sills, E. O., Pattanayak, S. K., & Cordero, S. (2012). Do payments for environmental services reduce deforestation? A farm level evaluation from Costa Rica. *Land Economics*, *88*(2), 382–399.

Ashenfelter, O. (1987). (5th ed.). *Handbook of labor economics* (Vol. 1999) Elsevier.

Athey, S., & Imbens, G. W. (2006). Identification and inference in nonlinear difference-in-differences models. *Econometrica*, *74*(2), 431–497.

Beaudry, F., Pidgeon, A. M. A., Radeloff, V. V. C., Howe, R. W. R., Mladenoff, D. J. D., & Bartelt, G. G. A. (2010). Modeling regional-scale habitat of forest birds when land management guidelines are needed but information is limited. *Biological Conservation*, *143*(7), 1759–1769. http://dx.doi.org/10.1016/j.biocon.2010.04.025

Bento, A., Towe, C., & Geoghegan, J. (2007). The effects of moratoria on residential development: Evidence from a matching approach. *American Journal of Agricultural Economics*, *89*(5), 1211–1218. http://dx.doi.org/10.1111/j.1467-8276.2007.01086.x

Bertrand, M., Duflo, E., & Mullainathan, S. (2004). How much should we trust differences-in-differences estimates? *The Quarterly Journal of Economics*, *119*(1), 249–275. http://dx.doi.org/10.1162/003355304772839588

Bolker, B. M., Brooks, M. E., Clark, C. J., Geange, S. W., Poulsen, J. R., Stevens, M. H. H., et al. (2009). Generalized linear mixed models: A practical guide for ecology and evolution. *Trends in Ecology & Evolution*, *24*(3), 127–135. http://dx.doi.org/10.1016/j.tree.2008.10.008

Burnham, K. P., & Anderson, D. R. (2002). *Model selection and multimodel inference* (2nd ed.). New York: Springer.

Busch, J., & Cullen, R. (2009). Effectiveness and cost-effectiveness of yellow-eyed penguin recovery. *Ecological Economics*, *68*(3), 762–776. http://dx.doi.org/10.1016/j.ecolecon.2008.06.007

Caliendo, M., & Kopeinig, S. (2005). Some practical guidance for the implementation of propensity score matching. *Journal of Economic Surveys*, *22*(1588), 31–72. http://dx.doi.org/10.1111/j.1467-6419.2007.00527.x

Cameron, A. C., & Trivedi, P. K. (2005). *Microeconometrics: Methods and applications*. Cambridge: Cambridge University Press.

Chay, K. Y., & Greenstone, M. (2004). Does air quality matter? Evidence from the housing market. *SSRN Electronic Journal*, http://dx.doi.org/10.2139/ssrn.544182

Creel, S., & Creel, M. (2009). Density dependence and climate effects in Rocky Mountain elk: An application of regression with instrumental variables for population time series with sampling error. *Journal of Animal Ecology*, *78*(6), 1291–1297.

Cressie, N., Calder, C. A., Clark, J. S., Ver, Hoef J. M., & Wikle, C. K. (2009). Accounting for uncertainty in ecological analysis: The strengths and limitations of hierarchical statistical modeling. *Ecological Applications*, *19*(3), 553–570. http://dx.doi.org/10.1890/07-0744.1

Diamond, J. (2001). Ecology. Dammed experiments!. *Science*, *294*(5548), 1847–1848. http://dx.doi.org/10.1126/science.1067012

Donald, S. G., & Lang, K. (2007). Inference with difference-in-differences and other panel data. *Review of Economics and Statistics*, *89*(2), 221–233. http://dx.doi.org/10.1162/rest.89.2.221

Ferraro, P. J., McIntosh, C., & Ospina, M. (2007). The effectiveness of the US endangered species act: An econometric analysis using matching methods. *Journal of Environmental Economics and Management*, *54*(3), 245–261. http://dx.doi.org/10.1016/j.jeem.2007.01.002

Foster, D. R., Fluet, M., & Boose, E. R. (1999). Human or natural disturbance: Landscape-scale dynamics of the tropical forests of Puerto Rico. *Ecological Applications*, *9*(2), 555–572.

Grace, J. B., Schoolmaster, D. R., Guntenspergen, G. R., Little, A. M., Mitchell, B. R., Miller, K. M., et al. (2012). Guidelines for a graph-theoretic implementation of structural equation modeling. *Ecosphere*, *3*(8), art73. http://dx.doi.org/10.1890/ES12-00048.1

Greenstone, M., & Gallagher, J. (2008). Does hazardous waste matter? Evidence from the housing market and the superfund program. *Quarterly Journal of Economics*, *123*(3), 951–1003.

Greenstone, M., & Gayer, T. (2009). Quasi-experimental and experimental approaches to environmental economics. *Journal of Environmental Economics and Management*, *57*(1), 21–44. http://dx.doi.org/10.1016/j.jeem.2008.02.004

Grout, C. A., Jaeger, W. K., & Plantinga, A. J. (2011). Land-use regulations and property values in Portland, Oregon: A regression discontinuity design approach. *Regional Science and Urban Economics*, *41*(2), 98–107. http://dx.doi.org/10.1016/j.regsciurbeco.2010.09.002

Heckman, J. J., Ichimura, H., Smith, J., & Todd, P. (1996). Sources of selection bias in evaluating social programs: An interpretation of conventional measures and evidence on the effectiveness of matching as a program evaluation method. *Proceedings of the National Academy of Sciences of the United States of America*, *93*(23), 13416–13420, http://doi.org/VL - 93.

Heckman, J. J., Ichimura, H., & Todd, P. (1998). Matching as an econometric evaluation estimator. *Review of Economic Studies*, *65*(2), 261–294.

Ho, D. E., Imai, K., King, G., & Stuart, E. A. (2006). Matching as nonparametric preprocessing for reducing model dependence in parametric causal inference. *Political Analysis*, *15*(3), 199–236. http://dx.doi.org/10.1093/pan/mpl013

Honey-Rosés, J., Baylis, K., & Ramírez, M. I. (2011). A spatially explicit estimate of avoided forest loss. *Conservation Biology*, *25*(5), 1032–1043. http://dx.doi.org/10.1111/j.1523-1739.2011.01729.x

Horsch, E. J., & Lewis, D. J. (2009). The effects of aquatic invasive species on property values: Evidence from a Quasi-experiment. *Land Economics*, *85*(3), 391–409.

Hurlbert, S. H. (1984). Pseudoreplication and the design of ecological field experiments. *Ecological Monographs*, *54*(2), 187–211. http://dx.doi.org/10.2307/1942661

Imbens, G. W., & Lemieux, T. (2008). Regression discontinuity designs: A guide to practice. *Journal of Econometrics*, *142*(2), 615–635.

Jones, A. (2007). *Applied econometrics for health economists: A practical guide*. Radcliffe Publishing.

Joppa, L. N., & Pfaff, A. (2009). High and far: Biases in the location of protected areas. *PLoS One*, *4*(12), e8273. http://dx.doi.org/10.1371/journal.pone.0008273

Lee, D. S., & Card, D. (2008). Regression discontinuity inference with specification error. *Journal of Econometrics*, *142*(2), 655–674.

Lee, D. S., & Lemieux, T. (2009). *Regression discontinuity designs in economics*. National Bureau of Economic Research (14723).

Legendre, P. (1993). Spatial autocorrelation: Trouble or new paradigm? *Ecology*, *74*(6), 1659–1673. http://dx.doi.org/10.2307/1939924

Lewis, D. J., Provencher, B., & Butsic, V. (2009). The dynamic effects of open-space conservation policies on residential development density. *Journal of Environmental Economics and Management*, *57*(3), 239–252. http://dx.doi.org/10.1016/j.jeem.2008.11.001

Macchi, L., Grau, H. R., Zelaya, P. V., & Marinaro, S. (2013). Trade-offs between land use intensity and avian biodiversity in the dry Chaco of Argentina: A tale of two gradients. *Agriculture, Ecosystems & Environment*, *174*, 11–20. http://dx.doi.org/10.1016/j.agee.2013.04.011

Miguel, E., Styanath, S., & Sergenti, E. (2004). Economic shocks and civil conflict: An instrumental variables approach. *Journal of Political Economy*, *112*(4), 725–753.

Pearl, J. (2009). *Causality: Models, reasoning, and inference*. Cambridge [England]; New York: Cambridge University Press.

Pearl, J., Glymour, M., & Jewell, N. P. (2016). *Causal inference in statistics: A primer*. Chichester, West Sussex: Wiley.

Phalan, B., Onial, M., Balmford, A., & Green, R. E. (2011). Reconciling food production and biodiversity conservation: Land sharing and land sparing compared. *Science (New York, NY)*, *333*(6047), 1289–1291. http://dx.doi.org/10.1126/science.1208742

Rosenbaum, P. R., & Rubin, D. B. (1983a). Assessing sensitivity to an unobserved binary covariate in an observational study with binary outcome. *Journal of the Royal Statistical Society. Series B (Methodological)*, *45*(2), 212–218.

Rosenbaum, P. R., & Rubin, D. B. (1983b). The central role of the propensity score in observational studies for causal effects. *Biometrika*, *70*(1), 41–55. http://dx.doi.org/10.1093/biomet/70.1.41

Royle, A. J., & Dorazio, R. M. (2008). *Hierarchical modeling and inference in ecology: The analysis of data from populations, metapopulations and communities [hardcover]* (1 edition). Academic Press.

Sagarin, R., & Pauchard, A. (2010). Observational approaches in ecology open new ground in a changing world. *Fron-*

tiers in Ecology and the Environment, *8*(7), 379–386. http://dx.doi.org/10.1890/090001

Seidl, R., Schelhaas, M.-J., & Lexer, M. J. (2011). Unraveling the drivers of intensifying forest disturbance regimes in Europe. *Global Change Biology*, *17*(9), 2842–2852. http://dx.doi.org/10.1111/j.1365-2486.2011.02452.x

Shipley, B. (2002). *Cause and correlation in biology: A user's guide to path analysis, structural equations and causal inference [paperback]* (1 edition). Cambridge University Press.

Sims, K. R. E. (2010). Conservation and development: Evidence from Thai protected areas. *Journal of Environmental Economics and Management*, *60*(2), 94–114. http://dx.doi.org/10.1016/j.jeem.2010.05.003

Stewart-Oaten, A., & Bence, J. R. (2001). Temporal and spatial variation in environmental impact assessment. *Ecological Monographs*, *71*(2), 305–339. http://dx.doi.org/10.1890/0012-9615(2001)071[0305:TASVIE]2.0.CO;2]

Strauss, J. (1986). Does better nutrition raise farm productivity? *Journal of Political Economy*, *94*(2), 297–320.

Wooldridge, J. (2011). *Introductory econometrics: A modern approach* (5th ed.). Cengage South-Western.

Wootton, J. T. (1994). Predicting direct and indirect effects: An integrated approach using experiments and path analysis. *Ecology*, *75*(1), 151. http://dx.doi.org/10.2307/1939391