SANDEE

Working Paper, No 70–12

Analyzing Institutions in Resource and Development Econometrics: Recognizing Institutions, Exploring Levels and Querying Causes

Vikram Dayal

ECONOMICS & THE ENVIRONMENT

Published by the South Asian Network for Development and Environmental Economics (SANDEE)
PO Box 8975, EPC 1056, Kathmandu, Nepal.
Tel: 977-1-5003222 Fax: 977-1-5003299

SANDEE research reports are the output of research projects supported by the South Asian Network for Development and Environmental Economics. The reports have been peer reviewed and edited. A summary of the findings of SANDEE reports are also available as SANDEE Policy Briefs.

National Library of Nepal Catalogue Service:

Vikram Dayal
Analyzing Institutions in Resource and Development Econometrics: Recognizing Institutions, Exploring Levels and Querying Causes

(SANDEE Working Papers, ISSN 1893-1891; WP 70–12)

ISBN: 978-9937-8521-9-7

Key words:
Institutions
Multilevel modelling
Causality
Resources
Development

SANDEE Working Paper No. 70–12

Analyzing Institutions in Resource and Development Econometrics: Recognizing Institutions, Exploring Levels and Querying Causes

Vikram Dayal
Associate Professor, Institute of Economic Growth
New Delhi, India

August 2012

South Asian Network for Development and Environmental Economics (SANDEE)
PO Box 8975, EPC 1056, Kathmandu, Nepal

SANDEE Working Paper No. 70–12

The South Asian Network for Development and Environmental Economics

The South Asian Network for Development and Environmental Economics (SANDEE) is a regional network that brings together analysts from different countries in South Asia to address environment-development problems. SANDEE's activities include research support, training, and information dissemination. Please see www.sandeeonline.org for further information about SANDEE.

SANDEE is financially supported by the International Development Research Center (IDRC), The Swedish International Development Cooperation Agency (SIDA), the World Bank and the Norwegian Agency for Development Cooperation (NORAD). The opinions expressed in this paper are the author's and do not necessarily represent those of SANDEE's donors.

The Working Paper series is based on research funded by SANDEE and supported with technical assistance from network members, SANDEE staff and advisors.

English Editor
Carmen Wickramagamage

Comments should be sent to
Vikram Dayal, Institute of Economic Growth, New Delhi, India
email: vikday@iegindia.org

Contents

Abstract

In this paper, I propose three strategies to advance institutional analysis in resource and development econometrics: (1) recognition of institutional variables; (2) use of multilevel thinking and estimation; and (3) the adoption of causal graphs. I illustrate each strategy with examples from three previously published studies: (a) biomass extraction from Ranthambhore National Park by Dayal (2006), (b) air pollution in Goa by Das et al. (2009), and (c) trade-offs and synergies between carbon storage and livelihood benefits from forest commons by Chhatre and Agrawal (2009). As I see it, while there is no explicit institutional content in studies (a) and (b), by recognizing caste as a social norm and the role of norms in household decisions it is possible to further integrate an institutional perspective into these studies. My analysis also demonstrates how multilevel thinking is intrinsic to institutional thinking. Since there can be data at different levels within a study, using study (c) where forests are nested in countries, I show how multilevel statistics can help unpack variation in the data at forest and country levels. Similarly, since causality is vital to policy though extremely difficult to establish with observational data, a phenomenon that leads to disagreements among scholars, using examples from the three studies I show how causal graphs can help separate the disagreements between scholars into disagreements about the underlying causal structure and the correspondence between an agreed to causal structure and the data on hand. The paper examines specific strategies to include institutional analysis in resource econometrics.

Key Words: Institutions, Multilevel modelling, Causality, Resources, Development

Analyzing Institutions in Resource and Development Econometrics: Recognizing Institutions, Exploring Levels and Querying Causes

1. Introduction

Many resource and development economists use econometrics and microeconomic theoretical models but do not incorporate perspectives from institutional economics into their work because often their formal training, like mine, does not include institutional economics. Therefore, while they may, like me, have read institutional scholarship and feel that institutions are important, they may resemble the economists in the following quotation from Robert Solow (1990), where he draws attention to the fundamental importance of norms in the labor market—norms which have implications not just for the labor market but for the wider economy:

> "*I want to make the case that the labor market really is different … it cannot be understood without taking account of the fact that participants, on both sides, have well developed notions of what is fair.… Among economists, it is not obvious at all that labor as a commodity is sufficiently different from artichokes (pp. 3 & 4)*."

Evidently, economists who wish to be more inclusive therefore need to do things a little differently from the typical economist that Solow describes above. According to Schmid (2004), who discusses different ways of doing institutional analysis, econometrics and case studies are equally valid in institutional analysis, a view shared by Rodrik (2008). In this paper, I concentrate on certain issues at the intersection of resource economics, institutional analysis and econometric analysis. According to Ostrom (2005), a starting point for an institutional analysis is recognizing institutions. But econometric analysis requires further conversion of institutions into a variable format, which may be imperfect. Among institutional scholars, Ostrom is one who has stressed multiple levels throughout her career while the paper by Gibson et al. (2000) details questions of scale and associated levels in different disciplines. In this paper, I link multilevel concepts with the statistical technique of multilevel modeling. Although economists frequently use panel data econometrics, the use of multilevel modeling with cross-section data is still very limited. However, recent years have seen a great renewal in questions of causality, with economists coming to rely more and more on experiments and quasi-experiments instead of the structural equation models which were a key area in traditional econometrics. The present study draws on recent developments in causal graphs (Pearl, 2000 and Spirtes et al., 1993) that have been exposited by Morgan and Winship (2007) in order to discuss causal issues.

I suggest the following three strategies, which I discuss sequentially in the paper, to further integrate institutional analysis into resource and development econometrics: (1) recognition of institutional variables; (2) use of multilevel thinking and estimation; and (3) the adoption of causal graphs. I will illustrate these strategies with examples from the following three studies: (a) what I will refer to as the "firewood extraction study" (Dayal, 2006); (b) what I will refer to as the "air pollution study" (Das et al., 2009); and (c) what I will refer to as the "carbon and livelihoods study" (Chhatre and Agrawal, 2009). I was the author of the first study while being a co-author in the second.

In the "firewood extraction study," Dayal (2006) focused on the extraction of forest biomass by village households and developed a microeconomic model of the decisions regarding levels and sources of extraction of forest biomass. The study, which was an empirical examination of the biomass extraction behavior in a sample of 227 households living in, and close to, Ranthambhore National Park, India, attempted to arrive at an empirical measure of the spatial aspects of extraction. The study found village location, ownership of biogas, and caste to be key explanatory variables of forest biomass extraction.

In the "air pollution study," Das et al. (2009) developed a theoretical model to guide the empirical examination of data on time allocation and exposure to air pollution in different microenvironments (outdoors, indoors, cooking, and workplace) and their effects on the respiratory health of people. The sample was drawn from different clusters with different histories and different degrees of mining in the state of Goa, India. The study found age and gender to play significant roles in time allocation. The results of the study revealed that only those households that used biomass fuels alone were likely to experience indoor concentrations of air pollution that were greater than outdoor concentrations. The factors affecting fuel choice were education, wealth and availability of commercial fuels. After accounting for endogeneity and measurement errors, the researchers found biomass fuel and smoking to be statistically significant factors in explaining chronic respiratory symptoms, even after factoring in exposure to air pollution indoors, outdoors and in the workplace.

In the "carbon and livelihoods study," Chhatre and Agrawal (2009) focused on factors that affected tradeoffs and synergies between the level of carbon storage in forests and their contributions to livelihoods. The authors found that larger forests were more effective in enhancing both carbon and livelihoods outcomes, particularly when local communities also had high levels of rule-making autonomy. Their study drew on data from different countries gathered by the International Forestry Resources and Institution's (IFRI) research program.

In this paper, I will be intuitive and informal, and will discuss the different theory and data interactions related to integrating institutional analysis in resource and development econometrics.

2. Recognizing Institutions

Vatn (2005) distinguishes between three different kinds of institutions: conventions, norms and formally sanctioned rules. But Vatn (2005) sees a tension between the individualist and social constructivist positions in the social sciences with regard to the approach to institutions—a tension that can be plotted in terms of the direction of causality between individuals and institutions. As Hodgson (2000) puts it: "In the writings of Veblen and Commons there is both upward and downward causation; individuals create and change institutions, just as institutions mold and constrain individuals" (p.326). Ostrom (2005) points out that institutional analysis requires "digging deeper than markets and hierarchies" (p.819), the analyst having to recognize institutions that are often intangible, and not easily measurable. Alston (1996) also draws attention to the difficulty of observing institutions quantitatively:

> "*Institutional changes usually have some unique features limiting the data points and thus generally preventing conventional statistical analysis.... Frequently, quantitative measures of the causes or consequences of institutional change are simply not available; even when they are available, better evidence may come from the qualitative historical record (pp. 29 &.30).*"

Of the three case-studies under consideration in this paper, the first two studies, the fuelwood case study (Dayal, 2006) and the air pollution case study (Das et al., 2009), did not have institutional analysis as an objective whereas the carbon storage and livelihood study aimed to study institutions (Chhatre and Agrawal, 2009) using variables such as rule-making autonomy and ownership. Since rule-making autonomy is difficult to observe quantitatively, Chhatre and Agrawal (2009) used perceptions of rules by users and assumed that when it was "about the right level of conservation", the variable "AUTONOMY" equaled one whereas it was zero otherwise. However, Ternstrom et al. (2010) have objected to the use of the term "autonomy" for this perceived phenomenon, arguing instead that the variable underlying autonomy appeared to reflect sustainability. In their response to this objection, Chhatre and Agrawal (2010) recognized the underlying variable to be the perception of rules, which I consider to be a crucial variable. However, the debate between Tersntrom et al. (2010) and Chhatre and Agrawal (2010) also draws attention to a feature in regression analysis where we are often forced to collapse something varied and complex into something binary. Hence, when analyzing institutions quantitatively, we often translate a subtle social process into a variable whereas if prior thought about institutions goes into the research design a more nuanced approach is possible. This approach is evidenced in the study by Chopra and Gulati (1998) where they captured the evolution of institutions over time by making distinctions in dummy variables.

Although institutional analysis was not the stated objective of the fuelwood case study (Dayal, 2006), it nevertheless contains important institutional elements. The study examines the fraction of a household's extraction of fuelwood from Ranthambhore National Park, recognizing that a household can choose to source its gathered fuelwood from private land, village common land and the national park. The fraction of a household's extraction of fuelwood from Ranthambhore National Park is modeled in the microeconomic theoretical component (first stage: level; second: source/where) that guided the empirical analysis.

In this study, the author regressed the fraction of fuelwood sourced from Ranthambhore National Park on the following explanatory variables:

- Caste dummy (Brahmin), statistically significant at 1% level;
- Village dummy (Indala), also statistically significant at 1% level;
- Quantity of land, statistically significant at 10% level;
- House type (a proxy for wealth), household size, fraction of males, cattle owned, and goats owned, which were not statistically significant.

These results encourage a closer examination of the role of caste in biomass extraction despite the inherent difficulties in analyzing its role. As the above study shows, an economist using a microeconomic model and econometric analysis is not at ease with the bewildering variety of castes encountered in the field, the multivariate regression analysis employed using a dummy for only one or two castes (here Brahmin and low caste). In consequence, the analysis is coarse in its treatment of the variety of castes in a given Indian village. In another paper, however, Dayal (2008) conducted a non-parametric small sample investigation of diverse castes in one village. In this paper where the investigator did not have a formal model and did not use sophisticated econometrics, he showed that there was a great deal of heterogeneity among the different castes in their biomass extraction patterns, i.e., fuelwood, fodder and grazing (see Figure 1).

Conversations that I had in the field (in 2008) with key respondents confirmed the important role caste plays in biomass extraction. One respondent asserted that Gujjars were very comfortable with rearing goats while a Brahmin respondent thought that he knew more than the other castes about cattle-rearing because his cows gave him high yields. But while caste has no role in the microeconomic model of Dayal (2006), it plays a role in the econometric model because applied economists control for 'household characteristics'. But it is a slow moving variable, with zero measurement error, in contrast to other variables, and that partly explains its statistical significance. With the econometric evidence thus goading me, I therefore ventured into social ecology, and found Gadgil and Malhotra's fascinating case study (1998) set in the Western Ghats, which provides a functional interpretation of caste and its ecological significance:

> *"The relatively simple society of the high-rainfall tracts near the crest of the Western Ghats is largely made up of small, often single-clan settlements of Kunbis and Gavlis. Here the Kunbis lived ... in the lower valleys, while the Gavlis lived ... on the upper hill terraces. ... Thus, the cultivation of valleys and lower hill slopes was restricted to Kunbis and that of hill terraces to Gavlis; maintenance of livestock and use of fodder and grazing resources was largely with Gavlis, while Kunbis had the monopoly of hunting wild animals.* (Gadgil and Malhotra, p.31)"

The above shows how caste and ecology have co-evolved over centuries in much of India and the dominant role that caste as a social norm plays in the traditional rural economy. As Dasgupta (2012) says, "Social norms are not constructed out of thin air; they evolve" (p.186). In the case of caste, he suggests that "exploitation can masquerade as cooperation" (p.180). These reflections on caste help those of us in resource economics to appreciate the tension in the social sciences between the individualist and social constructivist positions (Vatn, 2005) and to recognize the role that caste plays as an institutional norm in Indian society.

3. Using Multilevel Thinking and Estimation

According to Luke (2004), context plays a role in virtually any situation studied by researchers–for example, a child learning or couples avoiding divorce. As he puts it, "characteristics or processes occurring at a higher level of analysis are influencing characteristics or processes at a lower level" (p.1).

A recurring characteristic of Elinor Ostrom's work is the attention to levels. In a 2000 study, Gibson, Ostrom and Ahn (2000) define scale as the spatial, temporal, quantitative or analytical dimensions used to measure and study any phenomenon. They define levels as the units of analysis located at the same position on the scale while making a distinction between conceptual levels and spatial levels (see Figure 2). For example, while international treaties are at the conceptual level of constitutional choice and spatially at the international level, buying land is at the conceptual level of operational choice and spatially at the level of the household. According to hierarchy theorists, understanding a complex system requires zooming in and out to a higher and lower level. Thus, we can enrich economic analysis by examining institutions at different levels.

In the case studies in use in this paper, the units of analyses are nested at higher levels. Hence, while households are nested in villages in the fuelwood paper, in the air pollution paper, individuals are nested in households that are nested in villages, which in turn are nested in mining clusters. In the paper on carbon and livelihoods, on the other hand, forests are nested in countries.

The fuelwood case study in Ranthambhore National Park, for instance, offers us levels related to caste and the Wildlife Act (Figure 3), where the legal rule, the Wildlife Act, determines the boundary of the Park. In the case of the air pollution study that takes the mining regions of Goa as its study site, the institutions are the mining regulations and the gender norms related to cooking (see Figure 3).

Statistical multilevel modeling is one method to examine institutions at different levels. Multilevel modeling is a generalization of regression methods that can improve prediction considerably (Gelman, 2006). Intuitively, at its simplest, one can think of a simple regression of y on x, with an intercept. While the observations on x may come from different groups, we could run a simple pooled regression where we ignore the grouping. We could then use a simple multilevel model where we let the intercept of the simple regression vary by group. Using data on carbon storage from Chhatre and Agrawal (2009), I illustrate the model by regressing carbon storage on the perception of rules. Figure 4 shows the pooled (over different countries) regression, and Figure 5 show the regression once we allow the intercept to vary with country. As evident from Figure 5, the country intercept for India and Kenya respectively is considerably higher than in the pooled regression.

Multilevel modeling also helps us unpack variations between and within groups. While we could do an anova for this, Figure 6 shows that the variation in carbon storage between countries is substantial in comparison with the variation within countries. The bold line in the boxes shows the median. Thus, this clearly shows that there is a large difference between the median for India and for Mexico. Since the width of the box shows the interquartile range, we can see graphically that the difference in values between India and Mexico is substantial relative to the variation within these countries.

Although more sophisticated and detailed multilevel models can be used, I will confine myself to one example from Luke (2004): with the reading score of a child as the dependent variable (Y) and the study time put in by the child (X) and size of class (W) as independent variables. Based on these, a possible multilevel model would be:

Level 1: $Y_{ij} = \beta_0 + \beta_{1j} X_{ij} + r_{ij}$

Level 2: $\beta_{oj} = \gamma_{00} + \gamma_{01} W_j + u_{0j}$

$\beta_{1j} = \gamma_{10} + \gamma_{11} W_j + u_{1j}$

But the slopes and intercepts in the Level 1 equation can vary across classrooms. According to Luke (2004), there are theoretical and statistical reasons for using multilevel variables. Because phenomena are multilevel and depend on context, there is a theoretical justification; because regression coefficients can vary across contexts and errors can be correlated within contexts, there is a statistical justification.

However, all three case studies under consideration in this paper—the fuelwood study, the air pollution study, and the carbon storage and livelihoods study—do not use multilevel modeling. As Luke (2004) points out, "Despite the importance of context, throughout much of the history of the health and social sciences, investigators have tended to use analytic tools that could not handle these types of multilevel data and theories" (p.2). While the carbon storage and livelihoods study has 10 groups and is therefore amenable to multi-level modeling, where the number of groups at the higher level is small, as in the fuelwood study (where it is four villages), multilevel modeling is not useful. Therefore, while we can get richer predictions with multilevel modeling, Gelman (2006) cautions that "the 'direct' and 'contextual' effects cannot necessarily be interpreted causally for observational data" (p.434)

4. Drawing Causal Graphs

4.1 Causality and Causal Diagrams

Causality is important because it is implicit in assertions on, or advocacy of, policies. We can distinguish between regression that is predictive or descriptive (conditional expectation) and causal. Most econometric textbooks do not discuss causality. Among the exceptions are Goldberger (1998), who clarifies and emphasizes the conditional expectation interpretation, and Gelman and Hill (2006) and Stock and Watson (2010), who make a clear distinction between models for prediction and those for causes.

Economists however often think of the causal issue in terms of the identification of the parameters of a structural equation. Manski (2006) provides an example of an identification problem that is relevant to this paper. People in groups tend to behave similarly; this could be because of group norms (i.e., the endogenous effects) or because they have similar characteristics and environment (i.e., the correlated effects). To achieve identification in the case of such problems, empirical observations by themselves are not enough. We can divide the problem of inference into statistical and identification components. The identification component consists of the conclusions that can be drawn with unlimited data. The statistical component is about how sampling variability affects conclusions. Manski (2006) however added that he came to the conclusion that "identification is the more fundamental problem" (p.5) in the course of his long research career.

Causal graphs help us see causal issues, which are easier to see but harder to resolve. Some statisticians like Freedman (1991) have warned us about the pitfalls of using regression analysis for causal inference. So Freedman (1991) praised John Snow's careful work on cholera, which used "shoe leather and logic" (p.300) (i.e., field work and logic) rather than statistical techniques to arrive at causal inferences.

The difficulty in dealing with causal issues is related to the nature of the research questions and the complexity of the systems. For example, we can contrast the following two questions: (1) which of the following interventions has a greater effect on educational outcomes, school bag or blackboard, versus (2) what are the causes of deforestation in different countries? The causal issue is easier in the second than in the first.

According to Greenland and Brumback (2002), there are four major types of causal models that are complementary. First, graphical models can illustrate qualitative population assumptions and sources of bias. Second, sufficient-component cause models can illustrate hypotheses about mechanisms of action. Third, potential-outcome (or counterfactual) models can provide a basis for the quantitative analysis of effects. Fourth, structural-equations models can provide a basis for quantitative analysis of effects.

Greenland et al. (1999) advocate the use of causal diagrams since they "do not incorporate the strong parametric assumptions of conventional models; instead, they display assumptions about the web of causation that are not captured by conventional models" (p. 38). I too prefer causal graphs because they help me 'see'. They are qualitative, so I don't get swamped by statistical detail, which is not to say that that it is unimportant. However, since I have used system dynamics type simulation, I am aware of causal feedback loop diagrams. Moreover, I think causal graphs can foster communication between different groups of scholars.

Spirtes, Glymour and Scheines (1993) and Pearl (2000) are key publications in the development of modern causal graphs, more specifically, of directed acyclic graphs used for exploring causality. Pearl (1999) emphasizes the distinction between causal thinking and probability theory:

> "*The word cause is not in the vocabulary of standard probability theory. It is an embarrassing yet inescapable fact that probability theory, the official language of many empirical sciences, does not permit us to express sentences such as 'Mud does not cause rain'; all we can say is that the two events are mutually correlated, or dependent – meaning that if we find one, we can expect to encounter the other. Scientists seeking causal explanations for complex phenomena or rationales for policy decisions must therefore supplement the language of probability with a vocabulary for causality, one in which the symbolic representation for the causal relationship 'Mud does not cause rain' is distinct from the symbolic representation for 'Mud is independent of rain'. (p.1)*"

To summarize, we use probability thinking for empirical work, but causal thinking requires us to put in a directional arrow: rain → mud.

Two relatively accessible and applied expositions move between the models of causality that Greenland and Brumback (2002) discuss. While Shipley (2000) moves between structural equation models and causal graphs in biology, Morgan and Winship (2007) move between the potential outcome model and causal graphs, with examples from social science. In this paper, I draw heavily on Morgan and Winship (2007) in intuitively conveying what causal graphs are and, later, when I discuss the interpretation of an instrumental variable.

A regression Y = a + b X1 + c X2 + e can be interpreted predictively or causally. Predictively, we say the regression gives us the conditional expectation of Y. If we interpret it causally, we have in mind an implicit causal graph (as in Figure 7) while the correctness of this causal interpretation depends on the correctness of the implicit causal graph for the situation that we are studying. If the causal pathways do not correspond with what appears in Figure 7, then the causal interpretation is problematic.

If we want to explore the effect of X on Y, following Winship and Morgan, we can list the key ingredients of causal graphs as follows:

- Common cause (controlling for M, the common cause, will block path): X ← M → Y
- Mediator (controlling for Z, will block path): X → Z → Y
- Collider (controlling for B, opens path): X → B ← Y

Whether we should control for a variable depends on whether it is a common cause, a mediator, or a collider. However, we need to see the causal graph as a whole. Figure 8 illustrates three specification options for studying the effect of X on Y, given the causal graph.

The intuition for excluding the collider is best given with an example. Assume a situation where a car fails to start in cold weather where there is petrol in the tank. While petrol in the tank of a car and very cold weather are on the face of it uncorrelated, they can be common causes of a car failing to start in cold climates. The car not starting is the collider. Whether there is petrol in the car or not tells me nothing about cold weather. But if my car fails to start, and there is petrol in the tank, I would make an educated guess that it is cold weather that is the cause.

Thus, directed acyclic graphs can be used to understand specification, the use of instrumental variables, the design of experiments, etc. However, they are not cyclic, i.e., they do not include instantaneous feedback, with feedback complicating causal enquiry by this means (Shipley, 2000). This is a serious limitation, for according to Dasgupta (2009), feedback is central to resource and development economics: "Theoretical considerations and empirical evidence show that the persistence of poverty in a world of economic progress should be traced to socioeconomic, metabolic, and ecological processes involving positive feedback" (p.7). Structural equations modeling can handle instantaneous feedback. Feedback is central to systems dynamics, which simulates causal stock and flow structures, with the simulations based on a spectrum of information–case studies, experiments etc. Timmins and Schlenker (2009) argue that structural modeling and reduced form modeling are complementary with one advantage of the former over the latter being its ability to engage with feedback processes. Chopra and Gulati (1998) have used a structural equations system to capture some of the feedback aspects.

However, we can overcome this problem of representing feedback if we can make the time ordering explicit and can slice time. In such situations, the instantaneous cyclic representation can be converted into an acyclic one: that is, instead of X ← → Y we can have X1 → Y2 and Y1 → X2, where 1 and 2 represent time.

4.2 Causal Issues in Chhatre and Agrawal

Chhatre and Agrawal (2009) appear to be using a causal interpretation of regression since they make reference to causal pathways. Heuristically, a causal graph can help us see what difficulties we face when, like Chhatre and Agrawal (2009), we use cross-section data to model a dynamic, path-dependent process. Although Ternstrom et al. (2010) raise the issue of the dynamics of the resource, Chhatre and Agrawal, as evident from their response (2010), are quite aware of the issue, pointing to data limitations such as the lack of panel data that are common to quantitative studies of the commons. Carbon storage and livelihoods have a dynamic element in that carbon storage and livelihoods can both be high today but fall tomorrow (Ternstrom et al., 2010). Even with a very simplistic causal graph of carbon storage and livelihoods, which abstracts from such biophysical details as fire and grazing and institutional details, the causal graph brings out our difficulties (Figure 9).

We only observe values for the last time-period in a long process with path dependence. However, a cross-section will sometimes have a large variation in a slow-moving variable while high frequency time series will give us variation in fast-moving variables. And institutions are usually slow-moving variables. We can use such cross-sectional data in descriptive regression and speculate about possible causal mechanisms, aware that the regression is only a shadow of these.

In an empirical enquiry we could think of a causal graph, but where would that come from? It could be from a microeconomic model or theory. However, different disciplines have different styles of theorizing, each discipline concentrating more often than not on certain kinds of causal pathways. Nevertheless, scholars from different disciplines could brainstorm and think about possible pathways. Again, by way of illustration, I refer to the paper by Chhatre and Agrawal (2009), my paper (Dayal 2006), and to my knowledge of similar economic studies (for example, Pattanayak et al., 2004), while drawing also on Gibson, McKean and Ostrom (2000). As Figure 10 makes evident, there are different categories of pathways: institutional, economic, sociological, contextual and biophysical. The causal link (I1 in Figure 10) between rules in use and extraction and, thereby, forest stock is one of the key insights of the volume titled *People and Forests* edited by Gibson, McKean and Ostrom (2000). Schweik (2000) found that institutional factors—such as the caste system and the status of rule monitoring—along with factors such as distance and geographical barriers played a role in the spatial distribution of *shorea robusta* in the Chitwan District of Nepal while Agrawal (2000), who engaged in an empirical study of village forest councils in Almora district in India, found that smaller forest councils were less successful in collective action than larger forest councils—suggesting the causal paths I5 and I6 in Figure 10. Economic studies such as those by Pattanayak et al. (2004) and Dayal (2006) emphasize the causal pathway E1. The variables that are enclosed in Figure 10 are observed in the Chhatre and Agrawal (2009) dataset. We could also decide to focus on how carbon stock is affected by rule perception. So carbon stock is in a circle. We can then focus on pathways relating to these variables. It might also be fruitful to do a causal EDA, i.e., explore data with a causal graph such as Figure 10.

Apart from the number of causal pathways, there are other issues that need attention such as the issue of complicated feedbacks, the possibility that some causal arrows might be time- and context-specific, and the fact that there might be path dependence and evolution.

At times, some pointed potential outcome type studies can be done as, for example, in Somanathan et al. (2009); these focus on potential outcomes that are abstracted from the diverse causal processes. We may cite what happens in a clinical trial by way of example. In clinical trials of a new pill, when the pill is ingested, something happens inside the treated person over a period of time, which may or may not cure the person's headache, on the basis of which the researcher is able to arrive at the causal inference that the pill on average cures headaches or not.

4.3 Reinterpreting a Causal Tool (Instrumental Variable)

I now offer a re-interpretation of the instrumental variable that is meant to unpack causality using the air pollution study by Das et al. (2009) as my example. The authors of the study had worked several summers trying to examine the effect of total exposure to air pollution (i.e., indoor plus outdoor pollution, etc., weighted by the time spent). We were struggling because it was not 'right' in our judgment since we knew we had a measurement error and we suspected endogeneity. As seen in Figure 11a, exposure is a cause of sickness which is a cause of the inability to work, which in turn affects the allocation of time to different activities and exposure to air pollution. In Figure 11b, the disturbance term is the common cause of both exposure and sickness. However, the paper by Pitt et al. (2005) on the health effects of indoor air pollution, which uses female hierarchy as an instrumental variable, helped us to establish the causal effect of exposure on sickness as evident in Figures 11a and 11b. In the theoretical model that we used, as with Pitt et al. (2005), we assumed that the household was optimizing utility, which guided the empirical research. Since this assumption solved the problem at hand, we were able to conclude the research on a triumphant note.

However, in a detailed discussion, Agarwal (1997) points out that while economists have improved on unitary models by developing bargaining models of the household, they have left out social norms from the account. Social norms, Agarwal (1997) maintains, are exogenous in the short run but endogenous in the long run and, therefore, (1) set limits to the domain of bargaining, (2) determine bargaining power, and (3) influence the conduct of bargaining. However, Agarwal (1997) offers a methodological suggestion to the perceived problem:

> "*I both use and emphasize the usefulness of what I term 'analytical description' for capturing the complexity and historic variability of gender relations in intra- and extra-household dynamics. By analytical description I mean a formulation that seeks to comprehensively spell out both qualitative and quantitative factors that might impinge on outcomes, without being pre-constrained by the structure that formal modeling imposes, or by data limitations. (p.2)*"

Similarly, Akerlof and Kranton (2010), who argue that gender matters in the workplace in the United States, suggest that careful observations rather than statistical tests are more enlightening in studying identity. Following Friedman's (1953) ideas about "The Methodology of Positive Economics" broadly, Das et al. (2009) did not worry much about whether the household was actually optimizing; what mattered to them most was the prediction or getting the regression 'right'. But the efficacy of the instrumental variable–female hierarchy, in this instance– should make us pause because it lends support to a view of the household which is in tune with the role of norms and identity.

The disadvantage of analytical description is that we may be uncomfortable about its validity. Therefore questions of validity should inform the design of further studies that seek to go beyond analytical description although this may be difficult. In a perceptive paper on 'Behavior, Environment and Health in Developing Countries: Evaluation and Valuation,' Pattanayak and Pfaff stress the importance of social interactions and the difficulty in identifying them empirically (2009).

5. Conclusions and Recommendations

Though the use of experiments and quasi-experiments is increasing rapidly in the field of economics, observational data is rich with different processes and slow-moving variables might have the time to leave their imprints on cross-sectional data. In 1989, Norgaard had pleaded for methodological pluralism while Barrett and Carter, in a 2010 discussion of experiments, maintain that 'development economics requires healthy methodological pluralism that recognizes all identification strategies' intrinsic limitations' (p.543). It is therefore evident that we need to recognize institutional variables, which may not exist in our theory, and may be hard to measure. Where institutional variables may be there in our regressions, we need to interpret them carefully. We can balance the tension between macro and micro, agency and structure, to some extent, with multilevel modeling. We can describe data with regression, and graphs and multilevel models can help in doing so. We can use causal graphs to help further understand causal paths. We can exploit thick descriptive historical case studies (when we expect path dependence), sociological studies (that may report on social construction), potential outcome studies, and simulation studies of social ecological systems (that may make strong causal assumptions), and synthesize from these. While Ostrom used a variety of methods, including meta-analysis of case studies with the Institutions and Development Framework (see Ostrom, 2009), Dasgupta (2003) made use of 'ideas and information culled from a number of disciplines: anthropology, demography, ecology, economics ...' (p.ix), anchored by short appendices with theoretical models, in order to establish a 'composite characterization' in his 'Inquiry Into Well-Being and Destitution.'

In sum, the discussion above strives to demonstrate how resource economists can move on the path of methodological pluralism in institutional analysis by recognizing institutions, using multilevel models, and adopting causal graphs.

Acknowledgements

I am grateful to the South Asian Network for Development and Environmental Economics (SANDEE) for the Partha Dasgupta Fellowship which allowed me to visit the Workshop in Political Theory and Policy Analysis in Bloomington, Indiana. I am also grateful to Professor Karl-Goran Maler for helping me visit Bloomington. I am grateful to the Institute of Economic Growth, my employer. The Workshop in Political Theory and Policy Analysis hosted me and provided a most conducive environment to learn and write. I thank Professor Kanchan Chopra and Professor E Somanathan for their excellent comments.

References

Agarwal, B (1997) '"Bargaining" and gender relations, within and beyond the household'. *Feminist Economics* 3(1): 1-51

Agrawal, A (2000) 'Small is beautiful, but is larger better? forest-management institutions in the Kumaon Himalaya, India'. In Gibson, GC; McKean, MA; Ostrom, E (eds.) *People and Forests: Communities, Institutions and Governance*, London: MIT Press, pp. 57-85

Akerlof, GA; Kranton, RE (2010) *Identity Economics: How Our Identities Shape Our Work, Wages, and Well-being*. Oxford: Princeton University Press

Alston, LJ (1996) 'Empirical work in institutional economics: an overview'. In Alston, LJ; Eggertson, T; North, DC (eds.) *Empirical Studies in Institutional Change*. New York: Cambridge University Press, pp. 25-30

Barrett, CB; Carter, MR (2010) 'The power and pitfalls of experiments in development economics: some non-random reflections'. *Applied Economic Perspectives and Policy* 32 (4): 515–548

Chhatre, A; Agrawal, A (2009) 'Trade-offs and synergies between carbon storage and livelihood benefits from forest commons'. *PNAS* 106 (42): 17667-17670

Chhatre, A; Agrawal, A (2010) 'Reply to Ternstrom et al.: challenges of studying multiple outcomes simultaneously'. *Proceedings of The National Academy of Sciences – PNAS* 107 (14): E53-E53

Chopra, K; Gulati, SC (1998) 'Environmental degradation, property rights and population movements: hypotheses and evidence from Rajasthan (India)'. *Environment and Development Economics* 3: 35-58

Das, S; Dayal, V; Murugesan, A; Uma, R; Sehgal, M; Chhabra, SK (2009) 'An integrated empirical model of health effects of air pollution: the case of mining in Goa, India'. IEG Discussion Paper, Institute of Economic Growth, Delhi

Dasgupta, P (2003) *An Inquiry into Well-Being and Destitution.* Oxford: Oxford University Press

Dasgupta, P (2009) 'Poverty traps: exploring the complexity of causation'. In Braun, J von; Hill, RV; Rajul Pandya-Lorch (eds.) *The Poorest and the Hungry: Assessments, Analyses, and Actions*, Washington. DC: IFPRI, pp.107-118

Dasgupta, P (2012) Dark matters: exploitation as cooperation'. *Journal of Theoretical Biology* 299: 180-187

Dayal, V (2006) 'A micro-econometric analysis of household extraction of forest biomass goods in Ranthambhore National Park, India'. *Journal of Forest Economics* 12: 145-163

Dayal, V (2008) 'Inter- and intra-village diversity among forest dependent households in Ranthambhore National Park, India'. *International Journal of Ecological Economics and Statistics* 10: 27-39

Eggertsson, T (2005) *Imperfect Institutions: Possibilities and Limits of Reform.* Ann Arbor, MI: University of Michigan Press

Freedman, DA (1991) 'Statistical models and shoe leather'. *Sociological Methodology* 21: 291-313

Friedman, M (1953) 'The methodology of positive economics'. In Friedman, M (ed.), *Essays in Positive Economics*. Chicago: University of Chicago Press, pp.3-43

Gadgil, M; Malhotra, KC (1998) 'The ecological significance of caste'. In Ramachandra, G (ed.) *Social Ecology*. New Delhi: Oxford University Press, pp. 27-41

Gelman, A (2006) 'Multilevel (hierarchical) modeling: what it can and cannot do'. *Technometrics* 48 (3): 432-435

Gelman, A; Hill, J (2006) *Data Analysis Using Regression and Multilevel/Hierarchical Models*. Cambridge: Cambridge University Press

Gibson, GC; McKean, MA; Ostrom, E (eds) (2000) *People and Forests: Communities, Institutions and Governance*. London: MIT Press

Gibson, CC; Ostrom, E; Ahn, TK (2000) 'The concept of scale and the human dimensions of global change: a survey'. *Ecological Economics* 32: 217-239

Goldberger, AS (1998) *Introductory Econometrics.* London: Harvard University Press

Greenland, S; Pearl, J; Robins, JM (1999) ‘Causal diagrams for epidemiologic research’. *Epidemiology* 1(10): 37-48

Greenland, S; Brumback, B (2002) ‘An overview of relations among causal modeling methods’. *International Journal of Epidemiology* 31: 1030-1037

Hodgson, GM (2000) ‘What is the essence of institutional economics?’. *Journal of Economic Issues* XXXIV(2): 317-329

Luke, DA (2004) *Multilevel Modeling*. New Delhi: Sage University Press

Manski, CF (2006) ‘Identification problems in the social sciences and life’. Available at: http://www.economia.uniroma2.it/lhc/Manski/lectio.pdf, (Accessed on 22 October 2011)

Morgan, SL; Winship, C (2007) *Counterfactuals and Causal Inference.* Cambridge: Cambridge University Press

Norgaard, RB (1989) ‘The case for methodological pluralism’. *Ecological Economics* 1: 37-57

Ostrom, E (2005) ‘Doing institutional analysis: digging deeper than markets and hierarchies’. In Menard, C; Shirley, MM (eds.) *Handbook of New Institutional Economics*. Dordrecht: Springer, pp. 819-848

Ostrom, E (2009) ‘Beyond markets and states: polycentric governance of complex economic systems’. Nobel Prize Lecture, 8 December 2009, Stockholm, Sweden

Pattanayak, SK; Sills, E; Kramer, R (2004) ‘Seeing the forests for the fuel’. *Environment and Development Economics* 9(2): 155–179

Pattanayak, SK; Pfaff, A (2009) ‘Behavior, environment, and health in developing countries: evaluation and valuation’. *Annual Review of Resource Economics* 1: 183-217

Pearl, J (1999) Graphs, causality and structural equation models’. In Ader, HJ; Mellenbergh, GJ (eds.) *Research Methodology in the Social, Behavioral and Life Sciences*. London: Sage University Press

Pearl, J (2000) *Causality: Models, Reasoning, and Inference*. Cambridge: Cambridge University Press

Pitt, MM; Rosenzweig, MR; Hassan, MN (2005) ‘Sharing the burden of disease: gender, the household division of labor and the health effects of indoor air pollution’, CID Working Paper No. 119, Center for International Development, Harvard University, Boston, MA

Rodrik, D (2008) *One Economics, Many Recipes: Globalization, Institutions, and Economic Growth*. New Jersey: Princeton University Press

Schmid, AA (2004) *Conflict and Cooperation: Institutional and Behavioral Economics*. Oxford: Wiley-Blackwell

Schweik, CM (2000) ‘Optimal foraging, institutions, and forest change: a case from Nepal’. In Gibson, GC; McKean, MA; Ostrom, E (eds.) *People and Forests: Communities, Institutions and Governance*’. London: MIT Press, pp. 99-134

Shipley, B (2000) *Cause and Correlation in Biology: A User’s Guide to Path Analysis, Structural Equations and Causal Inference*. Cambridge: Cambridge University Press

Solow, RM (1990), *The Labor Market as a Social Institution*. Cambridge: Basil Blackwell

Somanathan, E; Prabhakar, R; Mehta, BS (2009) ‘Decentralization for cost-effective conservation’. *PNAS* 106 (11): 4143-4147

Spirtes, P; Clark, G; Scheines, R (1993) *Causation, Prediction and Search*. New York: Springer-Verlag

Stock, JH; Watson, MW (2010) *Introduction to Econometrics*. Boston: Addison-Wesley

Ternstrom, I; Mukhopadhyay, P; Ghate, R (2010) ‘Sustainability, autonomy, and benefits from forest commons’. *Proceedings of The National Academy of Sciences - PNAS* 107 (14): E52-E52

Timmins, C; Schlenker, W (2009) ‘Reduced-form versus structural modeling in environmental and resource economics’. *Annual Review of Resource Economics* 1: 351-380

Vatn, A (2005) *Institutions and the Environment*. Cheltenham: Edward Elgar

Figures

Figure 1: Caste and Use of Park (Fariya Village)

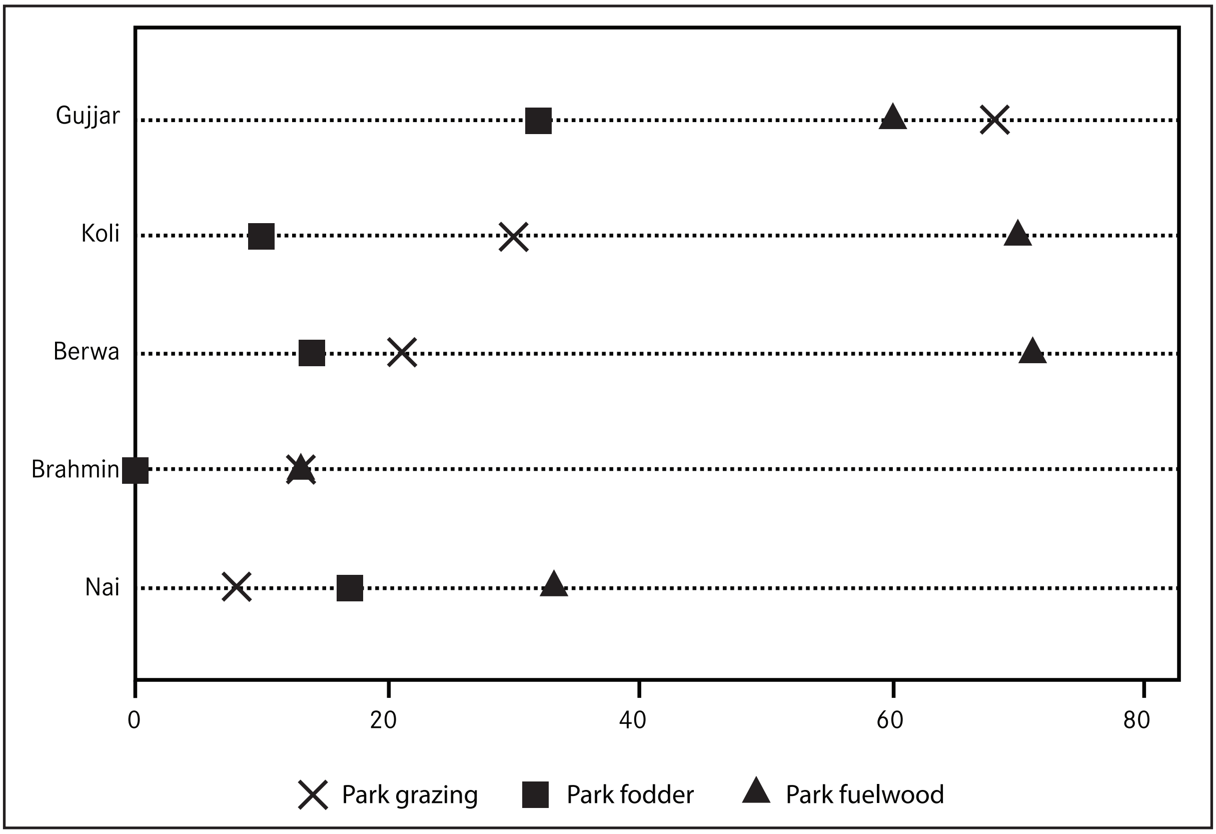


Figure 2: Conceptual Levels and Spatial Levels

	Conceptual levels		
Spatial level	Constitutional choice	Collective choice	Operational choice
International	International treaties	Policy making	Managing supervising projects
Household	Marriage	Policies family member	Buying land
Source: Gibson, Ostrom and Ahn (2000)			

Figure 3: Levels and Institutions in the Fuelwood and Air Pollution Case Studies

Biomass extraction in
Ranthambhore National Park (Dayal

Levels:

Formal rule:
Wildlife Act

1. Inside and outside village
2. Household

Social norm:
Caste

Indoor and outdoor air pollution
in Goa (Das, Dayal, et al.)

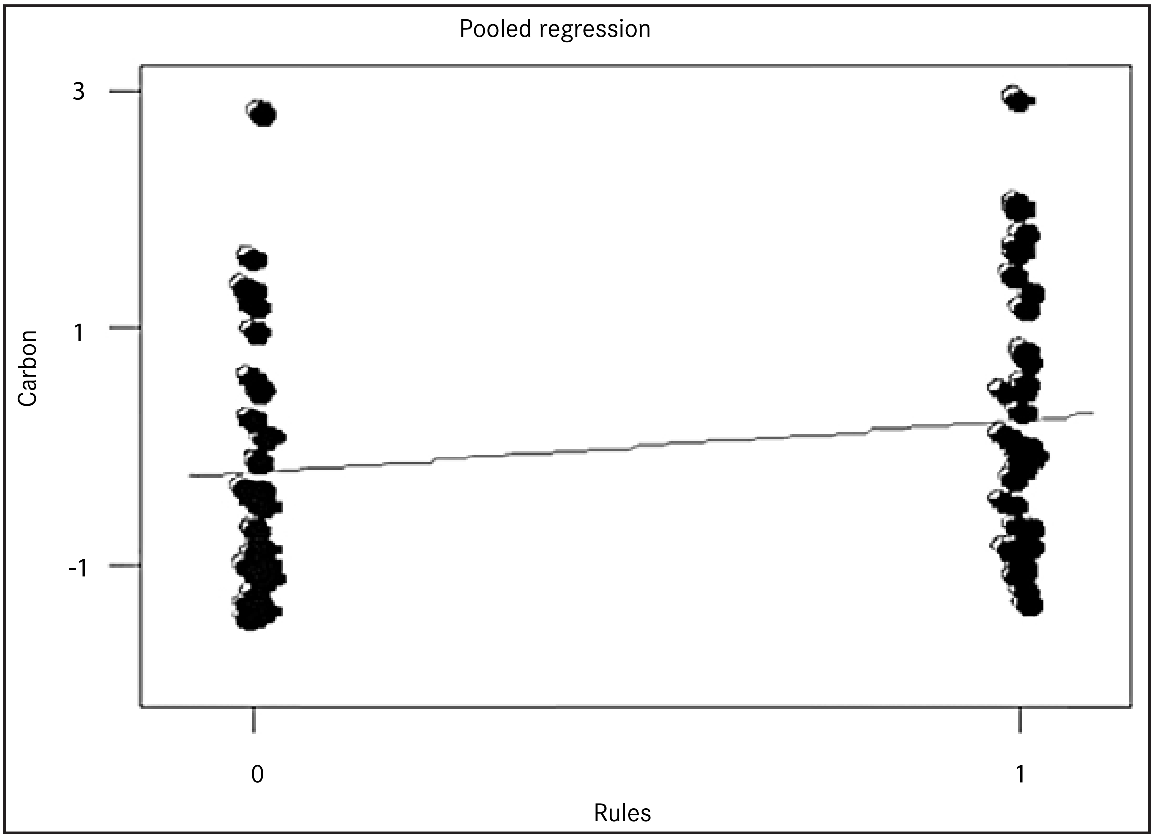
Levels:

Formal rule:
Mining regulation

1. Mining cluster
2. Village
3. Household
4. Indiidual

Social norm:
Gender

Figure 4: Pooled (over different countries) Regression



Based on Data in Chhatre and Agrawal (2009)

Figure 5: Multilevel Model Based on Data in Chhatre and Agrawal (2009)

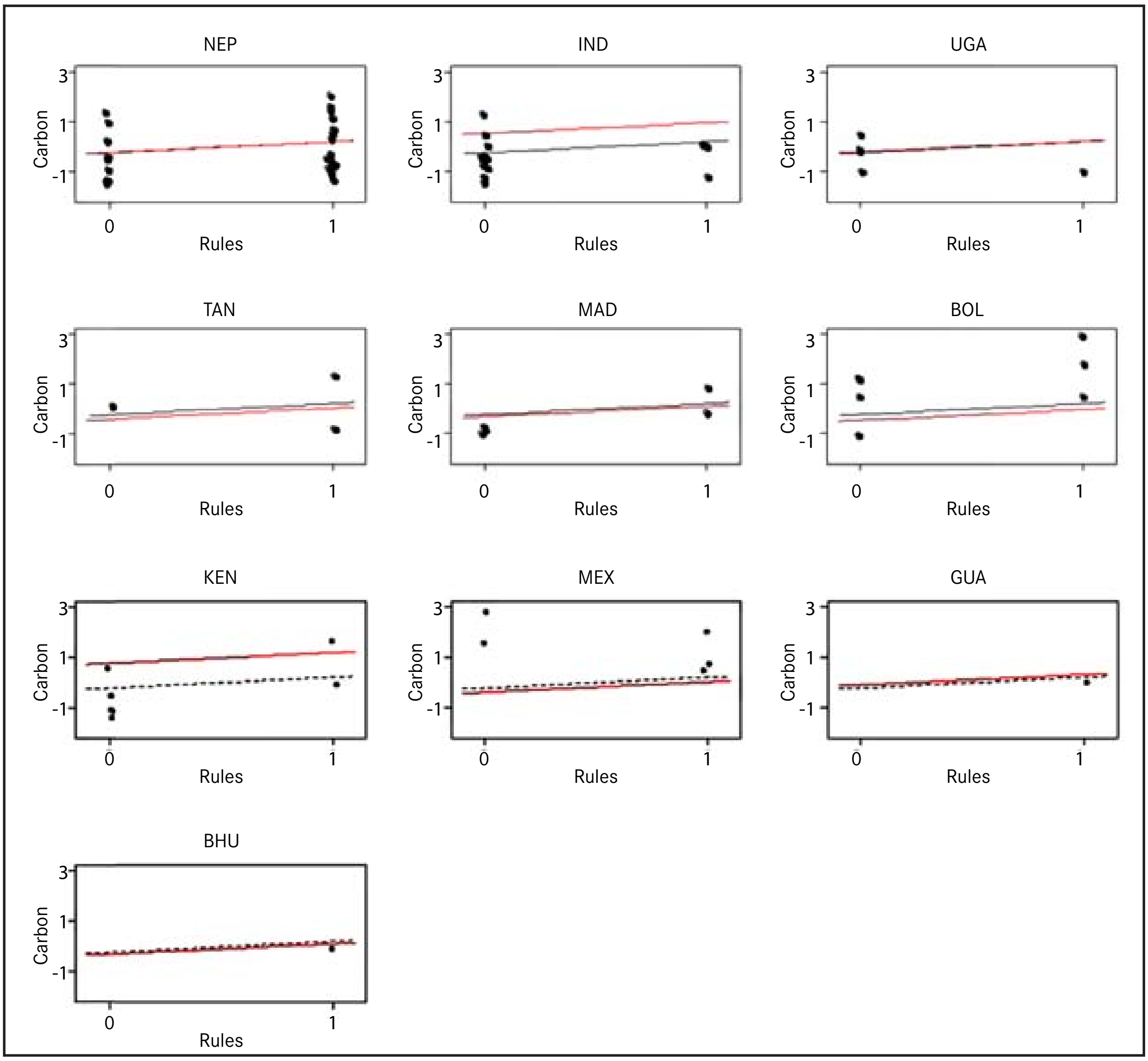
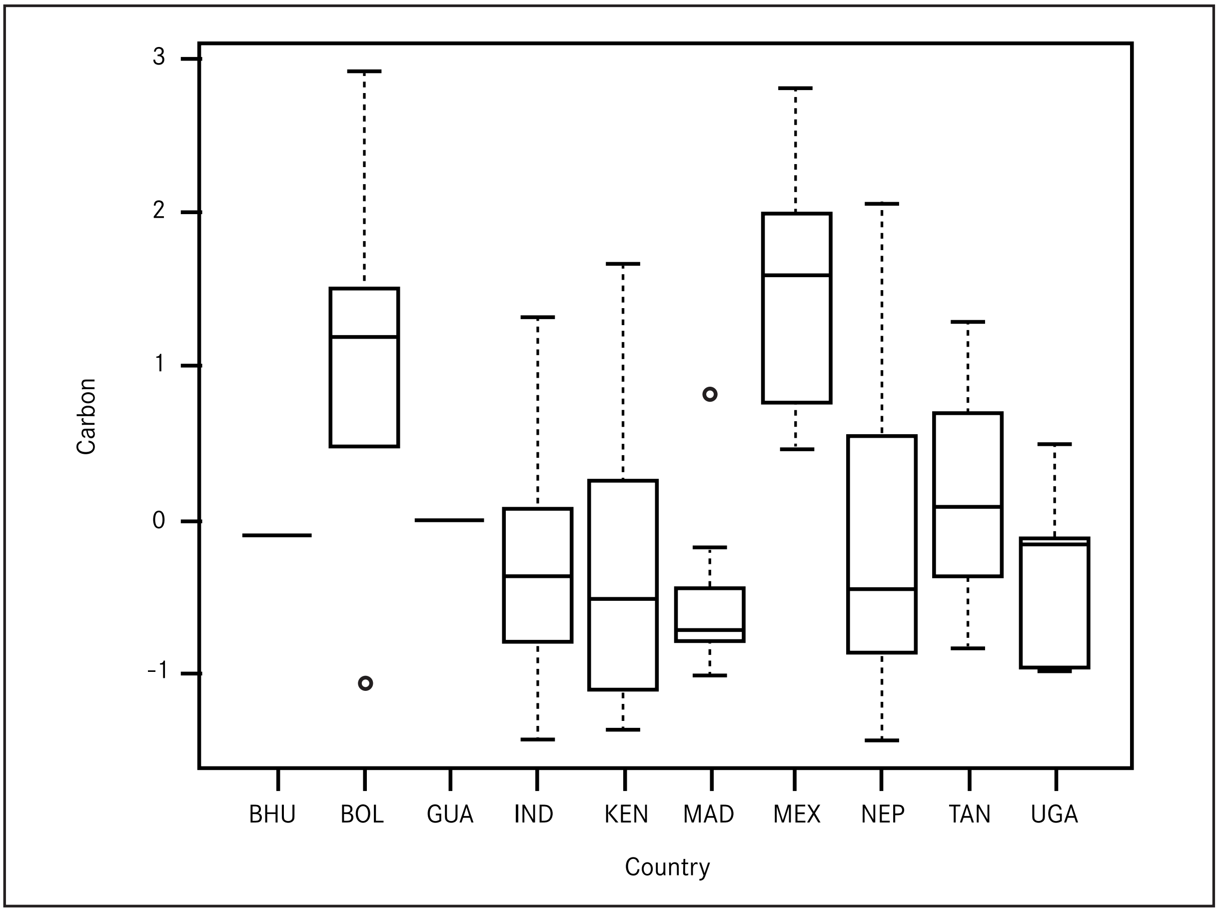


Figure 6: Box Plots of Carbon Storage by Country



Using Data from Chhatre and Agarwal (2009)

Figure 7: Implicit Causal Graph for Regression Y = a + b X1 + c X2 + e

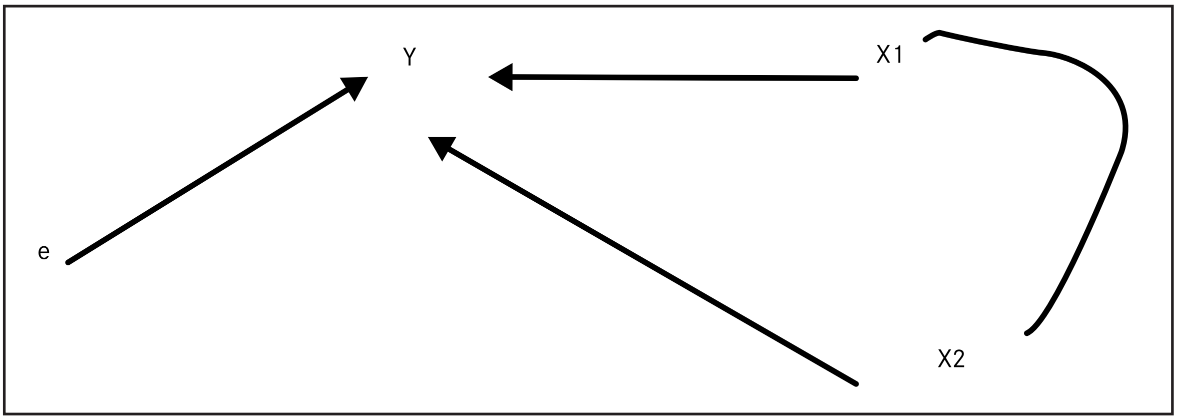


Figure 8: Causal Graph and Regression Options

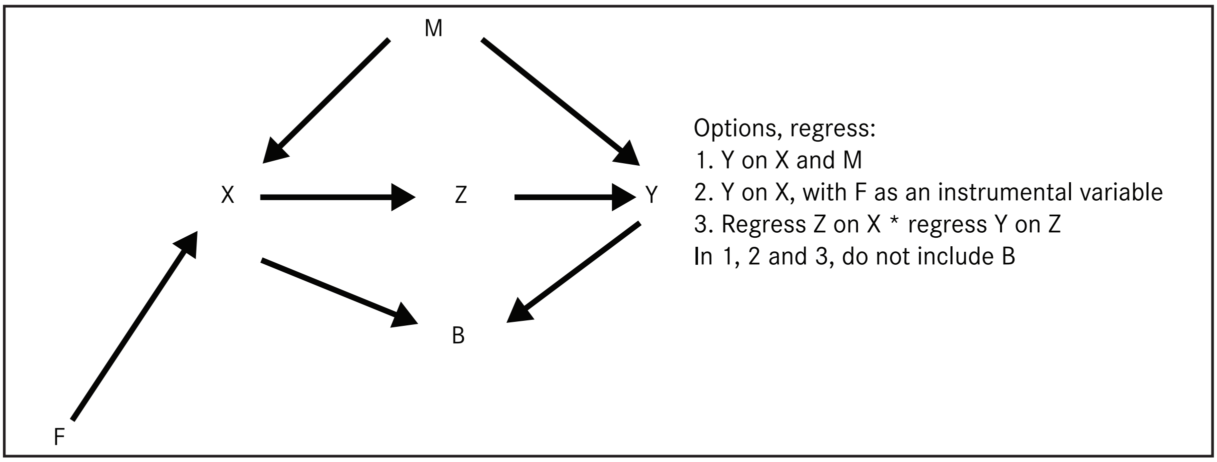


Figure 9: Causal Graph of Carbon Storage and Livelihoods

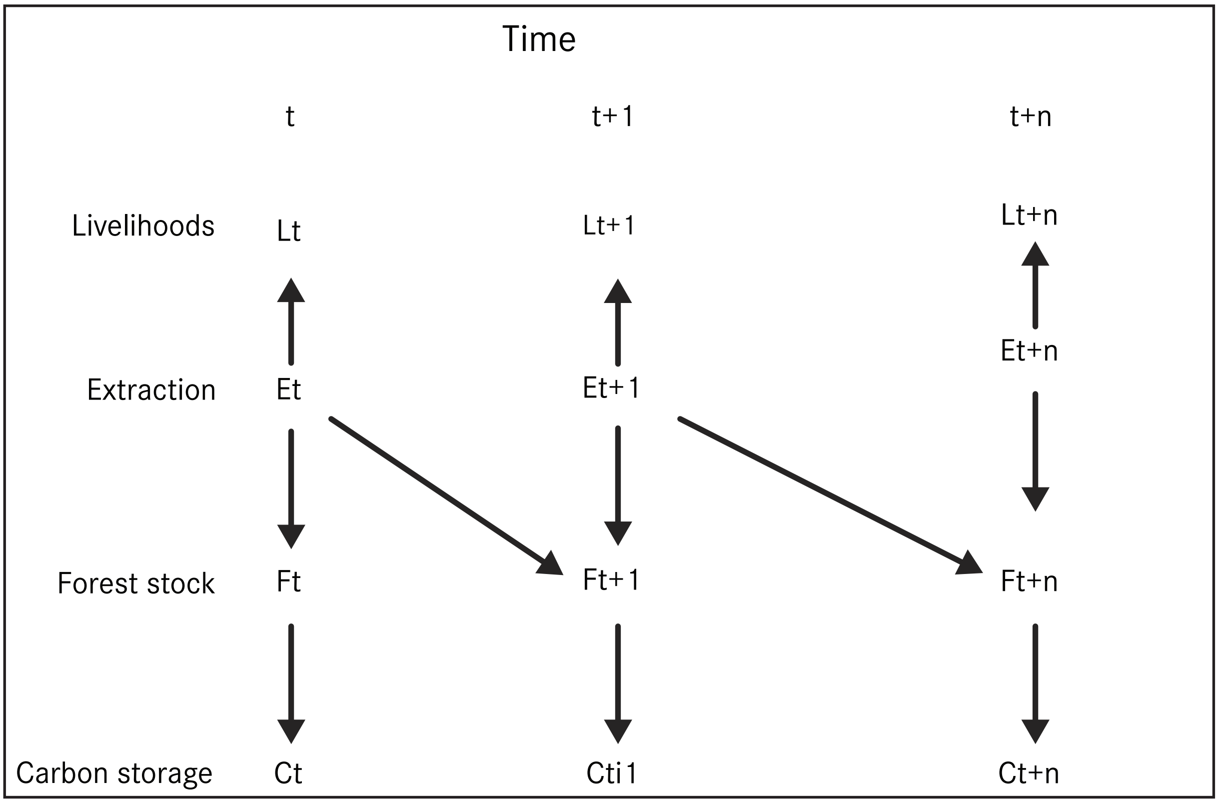


Figure 10: Causal Pathways for Carbon Stock and Livelihoods

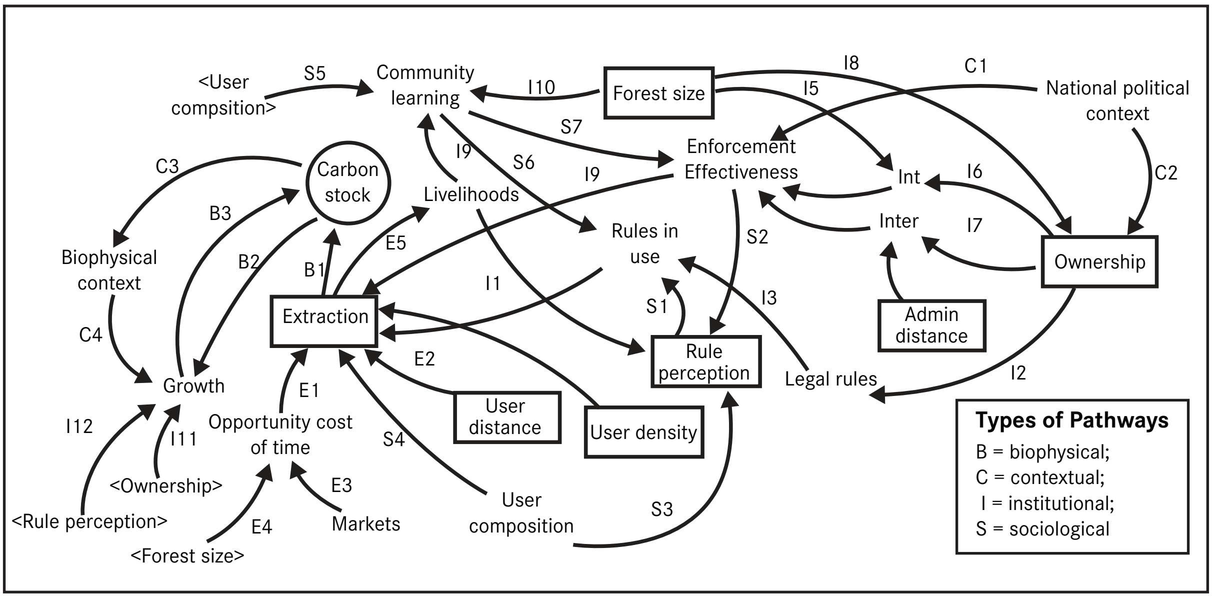
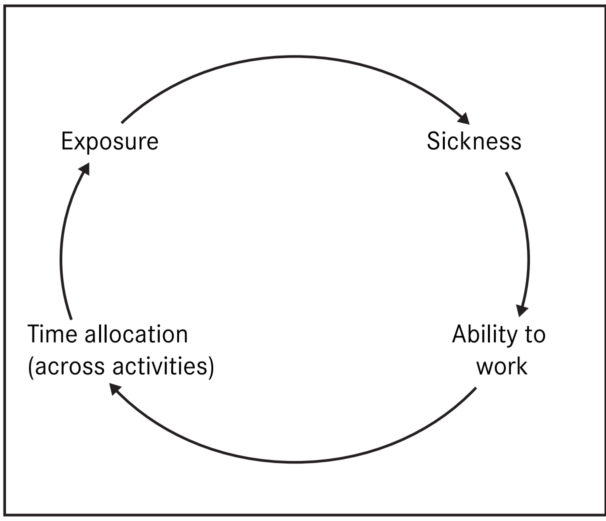
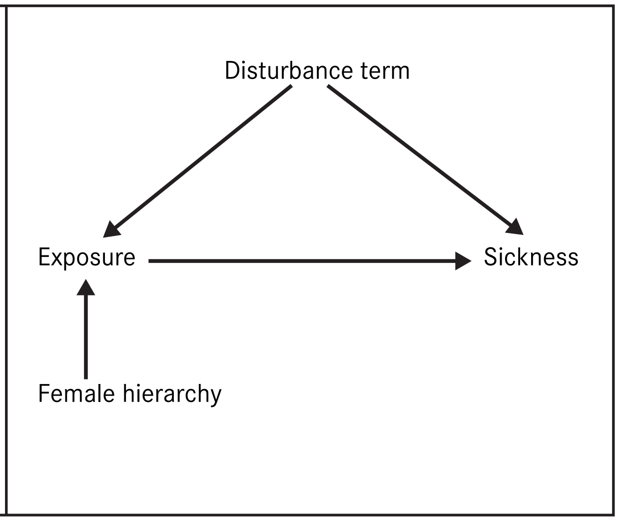


Figure 11a: Endogeneity of Exposure and Sickness

Figure 11b: Causal Graphs for Exposure, Sickness and Female Hierarchy





SANDEE
P.O. Box 8975, E.P.C 1056, Lalitpur, Nepal
Street address: c/o ICIMOD, Khumaltar, Lalitpur, Nepal
Tel: 977 1 5003222, **Fax:** 977 1 5003299, **Email:** info@sandeeonline.org **Web:** www.sandeeonline.org

SANDEE Sponsors