DISCUSSION PAPER

March 2012 ■ RFF DP 12-13

Ex Post Evaluation of Forest Conservation Policies Using Remote Sensing Data

An Introduction and Practical Guide

Allen Blackman

1616 P St. NW
Washington, DC 20036
202-328-5000 www.rff.org

RESOURCES FOR THE FUTURE

Ex Post Evaluation of Forest Conservation Policies Using Remote Sensing Data: An Introduction and Practical Guide

Allen Blackman

Abstract

Rigorous, objective evaluation of forest conservation policies in developing countries is needed to ensure that the limited financial, human, and political resources devoted to these policies are put to good use. Yet such evaluations remain uncommon. Recent advances in conservation best practices, the widening availability of high-resolution remotely sensed land-cover data, and the dissemination of geographic information system capacity have created significant opportunities to reverse this trend. This paper provides a nontechnical introduction and practical guide to a relatively low cost method that relies on remote sensing data to support ex post analysis of forest conservation policies. It describes the defining features of this approach, catalogues and briefly reviews the studies that have used it, discusses the requisite data, explains the principal challenges to its use and the empirical strategies to overcome them, provides some practical guidance on modeling choices, and describes in detail two recent case studies.

Key Words: forest conservation policy, evaluation, literature review, REDD

JEL Classification Numbers: Q23, Q28, Q56, Q57

Contents

Ex Post Evaluation of Forest Conservation Policies Using Remote Sensing Data: An Introduction and Practical Guide

Allen Blackman*

1. Introduction

According to United Nations Food and Agriculture Organization, the rate of deforestation in tropical countries remains "alarmingly high." For example, in both Latin America and Africa, it averaged 0.5 percent per year over the past decade, five times the global rate (FAO 2011). This deforestation, along with equally important forest degradation, has contributed to a host of local environmental problems, including biodiversity loss, soil erosion, and aquifer depletion, and has adversely affected some forest communities (Chomitz 2007).

Deforestation and degradation in developing countries also are a leading cause of climate change, arguably the most serious global environmental problem. Tropical deforestation accounts for a fifth to a quarter of total anthropogenic emissions of greenhouse gases (Houghton 2005; IPCC 2007). As a result, policies aimed at reducing emissions from deforestation and degradation (REDD) and capturing attendant local environmental and socioeconomic co-benefits (REDD+) have attracted considerable attention.

But even with the accelerating pace of investment in REDD+ activity, forest conservation policymakers in developing countries still have limited financial, human, and political resources. Therefore, it is important that their initiatives be effective and efficient. Ensuring that, in turn, requires objective, rigorous evaluations of the extent to which forest conservation policies achieve their aims. Such evaluations allow stakeholders to modify existing policies and shape future ones to maximize "bang for the buck."

Until recently, however, objective, rigorous evaluations were uncommon, for at least two reasons. One is that historically, the demand for rigorous evaluation of forest conservation policy has been limited (Ferraro and Pattanayak 2006; Millennium Ecosystem Assessment 2005). Second, collecting and analyzing the requisite data have been costly. Customarily, evaluations of

---

* Allen Blackman, Resources for the Future, Washington, DC, blackman@rff.org, 202-328-5073; and Environment for Development Center for Central America, Turrialba, Costa Rica.

Funding for this research was provided by the InterAmerican Development Bank and the Swedish Research Council Formas' COMMONS program. I'm grateful to Leonardo Corral, Nancy McCarthy, Juha Siikämaki, and Paul Winters for helpful comments and suggestions. Remaining errors are my own.

forest conservation policies have relied on expensive field measurements. Over the past two decades, however, publicly available, high-resolution remotely sensed (mostly satellite) data on deforestation and degradation—along with the geographic information system (GIS) capacity needed to analyze the data—have dramatically reduced evaluation costs. These advances in conservation best practices, data availability, and GIS capacity have created significant new opportunities to enhance our understanding of the effectiveness and efficiency of forest conservation policy.

The objective of this paper is to provide a nontechnical introduction and practical guide to the rigorous ex post evaluation of forest conservation policies using remotely sensed data. The principal target audiences are forest conservation policymakers, practitioners, and academics with limited exposure to this method.

2. Empirical Approach

This section describes the broad empirical approach to using remotely sensed data for evaluating conservation efforts, including the defining characteristics and associated terminology.

2.1. Defining Characteristics

The approach to evaluating forest conservation policy discussed in this paper—which for convenience we refer to as spatial evaluation—has five defining characteristics.

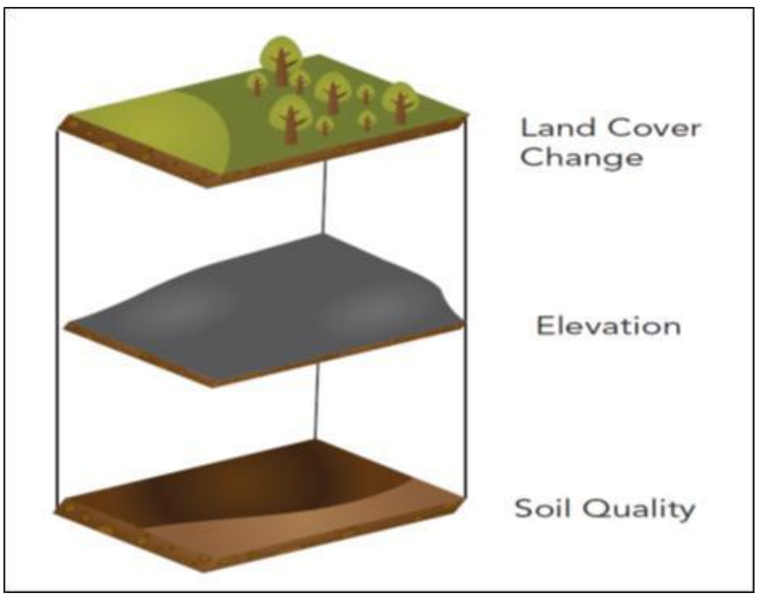
Land-Cover Change Effects. The goal of the analysis is to measure the causal effect of a forest conservation policy on land-cover change, including deforestation, which, as noted above, triggers an array of local and global environmental problems (Chomitz 2007). Note that similar techniques can be used to measure the causal effects of such policies on the socioeconomic status of local communities (e.g., Sims 2010; Ferraro and Hanauer 2011). However, we limit the scope of this paper to the analysis of land-cover change effects.

Remote Sensing Data. Remote sensing data are used to measure land-cover change. Such data are derived from sensors aboard aircraft or satellites, including passive sensors such as conventional cameras that detect reflected natural sunlight and other electromagnetic radiation and active sensors such as RADAR and LiDAR that emit and detect artificial radiation. Passive satellite sensors are most commonly used.

Spatial Variation. Causal effects are measured by analyzing spatial variation in (i) land-cover change (some areas are deforested and others are not); (ii) forest conservation policies (some areas are subject to the policy and others are not); and (iii) other geophysical and socioeconomic drivers of land-cover change (some areas are closer to cities, others are farther; some have high rainfall, some do not, etc.). When panel data are available, both temporal and spatial variation may be analyzed.

Geographic Information System. Relational data derived from a geographic information system (GIS) is used to analyze spatial (and sometimes temporal) variation. The GIS is constructed by collecting, georeferencing, and compiling spatial data on the three categories of variables just mentioned. The unit of analysis in the relational database is defined by the evaluator. It can be a pixel, or "plot" in the land-cover change map (typically, a $30m^2$ cell for maps derived from Landsat images); a cell in a user-defined grid (e.g., a $2km^2$ cell); or an administrative unit (e.g., a county or township). (The advantages and disadvantages of each option are discussed in Section 7.2, below). Each unit in the GIS—or a subsample of these units—represents a single observation in the relational database that has all the data contained in the various layers of the GIS. Figure 1 illustrates the data framework in the case of a cell-level relational database.

Figure 1. Data Framework



Secondary Data. Aside from possibly classifying raw remote sensing images, the evaluations described in this document rely solely on secondary data collected from statistical and geospatial data agencies and other sources (discussed in Section 4, below). This feature of the method makes it relatively inexpensive and quick to implement. Although studies such as Arriagada et al. (In Press) and Sierra and Russman (2006), which combine secondary data with survey data and other field measurements, are quite useful, we will focus on studies that use simpler, less expensive methods that are likely to be more realistic for evaluators with limited resources.

2.2. Terminology

Given those defining characteristics, the type of empirical study described in this paper amounts to a statistical analysis of relational data. In discussing these data, we will rely on the terminology used in the project evaluation literature. The *unit of analysis* is a spatial entity, such as a plot, cell, or administrative division. The *treatment* variable measures the extent to which the unit was exposed to the forest conservation policy. The *outcome* variable measures land-cover change on that unit. The *control* variables measure other drivers of land-cover change, such as soil quality and travel time to population centers. For example, Blackman et al. (2011) aim to measure the effect on deforestation of Mexican protected areas created prior to 1993. To do that, the authors construct a GIS comprising spatial data on 1993–2000 land-cover change (derived from Landsat satellite images), the location of protected areas, and six other drivers of land-cover change—elevation, slope, rainfall, soil quality, travel time to population centers, land tenure, and indigenous population. Their unit of analysis is a $30m^2$ plot. They use the GIS to create a relational database of a sample of such plots, each of which includes all the information in the various layers of the GIS. In their statistical analysis, the outcome variable is a dichotomous dummy indicating whether the plot was deforested between 1993 and 2000. The treatment variable is a dichotomous dummy indicating whether the plot was located in a pre-1993 protected area. The control variables are the other drivers of land-cover change listed above.

3. Literature

Tables 1–3 summarize the salient features of the studies that have used the broad empirical approach described above, including study area, treatment, specific empirical approach, methods used to control for spillovers (if any), methods used to construct the sample, land-cover data (years, resolution, and land-cover categories), and control variables. Each table summarizes studies using a different unit of analysis: the plot (Tables 1A and 1B), the polygon (Table 2), and the cell (Table 3). We reference these tables in the remainder of this paper.

Table 1A. Plot-Level Spatial Evaluations Focused on Evaluating Forest Conservation Policy

Article	*Study area*	*Treatment*	*Empirical approach*	*Control for spillovers?*	*Sampling*	*Years*	*Land-cover data resolution*	*Categories*	*Controls*
Andam et al. (2008)	Costa Rica	Protected areas (~150)	Covariate matching	Covariate matching to measure leakage	Random	• 1960 • 1986 • 1997	3ha (173m^2)	• Forested • Cleared	• Land-use productivity • Distance to forest edge, roads, cities
Blackman et al. (2011)	Mexico	Protected areas (57)	Propensity score matching	Propensity score matching to measure leakage	2km grid	• 1993 • 2000	30m^2	• Forested • Cleared	• Elevation • Slope • Rainfall • Rainfall squared • Soils (18) • Travel times to pop. centers • Indigenous • Communal tenure • Regions (9)
Joppa and Pfaff (2010a)	Global	Protected areas	Covariate matching	None	Stratified random (1:4 treatment: control)	• 2000 • 2005	1km^2	• Forested • Cleared	• Elevation • Slope • Distance to road • Distance to cities • Ecoregion • Ag. suitability
Müller and Munroe (2005)	Central Vietnam	• Agricultural intensification • Protected areas	Multinomial logit	Spatial lag vars.	200m grid	2000	50m^2	• Nonag. • Ag. • Paddy	• Elevation • Slope • Slope (lag) • Soils (2) • Distances to roads, neighbors • Population • Ethnic minority • Village dummy • Irrigated area • Shape complexity
Nelson and Chomitz (2011)	Global (tropical forest biome in LAC, Africa, Asia)	• Protected areas	Covariate matching	None	Random	2000–2008†	1km^2	Dependent variable is incidence of forest fires, not land-cover change	• Travel time to major cities • Distance to cities, roads • Precipitation • Elevation • Slope

Nelson et al. (2001)	Darién, Panama	• Indigenous reserves (2) • Protected areas	Multinomial logit	• Lat.-long. fixed effects • Spatial lag vars.	None (includes all plots)	• 1987 • 1997	$500m^2$	• Forest (3 types) • Marsh • Cleared	• Elevation • Slope • Slope (lag) • Temperature • Soil index • Soil index (lag) • Concession • Travel cost to pop. centers (IV)
Pfaff et al. (2008)	Costa Rica	Payments for environmental services	• Covariate matching • Propensity score matching	None	Random	• 1987 • 1997 • 2000	$28m^2$	• Forested • Cleared	• Slope • Precipitation • Holdridge life zones (7) • Distances (9) • Road density • Close to sawmill • Close to main town • Close to school • Percentage cleared • Percentage cleared^2
Robalino et al. (2008)	Costa Rica	Payments for environmental services	• Covariate matching • Propensity score matching	None	Random	• 2000 • 2005	$28m^2$	• Forested • Cleared	• Slope • Precipitation • Holdridge life zones (2) • Distances (7) • Distance to rivers • Percentage cleared before 1997

IV = instrumental variable; NS = not specified in article; SE = standard error.
†Dependent variable is incidence of forest fire on sample plot, a proxy for land-cover change in humid tropics (where nonanthropogenic forest fires are very rare).

Table 1B. Plot-Level Spatial Evaluations Focused on Drivers of Land Use and Land-Use Change but Including Some Policy Regressors

Article	*Study area*	*Treatment*	*Empirical approach*	*Control for spillovers?*	*Sampling*	*Years*	*Land-cover data resolution*	*Categories*	*Controls*
Chomitz and Gray (1996)	Southern Belize	• Tenure types (national, forest reservation)*	Multinomial logit	Bootstrapping to correct SEs	1km grid	1994	1:50,000	• Ag. (2 types) • Natural vegetation	• Slope • Soils (5) • Distance to market • Flood hazard • Rainfall
Cropper et al. (2001)	Northern Thailand	• Protected areas (NS) • Wildlife sanctuaries (NS)*	Bivariate probit with IV (near river)	Spatial lag vars. as robustness check	5km grid	1986	1:1 million	• Forested • Cleared	• Elevation • Slope • Soils (12) • Travel cost to pop. centers • Watershed • Pop. density • Province fixed effects (17)
Deininger and Minten (2002)	Southern Mexico	• Protected area • Irrigation • Access to public extension	Probit	None	1km grid	• 1980 • 1993	$100m^2$	• Forested • Cleared	• Elevation • Slope • Rainfall • Soils (13) • Tenure • Indigenous pops • Distance to road • Pop. density (IV) • Poverty (IV) • Subsidized credit
Mertens et al. (2002)	Northeast (Pará), Brazil	• Protected areas • Land-use zoning	Multinomial logit, by type of farmer	Increasing grid size	NS	• 1986 • 1992 • 1999	$30m^2$	• Forested • Cleared	• Distances to town, village, dairy industry, main road, secondary road, river, forest edge
Mertens et al. (2004)	Eastern (Santa Cruz), Bolivia	• Protected areas • Forest concessions • Land-use zoning	Logit	None (test rejects spatial autocorrelation)	Random	• 1989 • 1994	$1km^2$	• Forested • Cleared	• Precipitation • Distances to nonforest, roads, trails, cities • Soils (7)

IV = instrumental variable; NS = not specified in article.
*Also includes evaluation of impact of roads.

Table 2. Polygon-Level Spatial Evaluations of Forest Conservation Policy

Article	*Study area*	*Treatment*	*Empirical approach*	*Control for spillovers?*	*Sampling*	*Years*	*Land-cover data resolution*	*Categories*	*Controls*
Alix-Garcia et al. (In Press)	Mexico	Payments for environmental services	• Covariate matching • Regression w/ matched sample • Controls drawn from rejected applicants	Covariate matching to measure leakage	Parcels enrolled in PES program	2003–2006	NS	• Forested • Cleared	(Drawn from rejected applicants) • Area of parcel • *Ejido* tenure • Avg. slope • Avg. elevation • Type of forest • Percentage forest • Road density • Regions (4)
Honey-Roses et al. (2011)	Central Mexico	Payments for environmental services	• D-i-D • Covariate matching • Covariate matching with spatial weights variant	See empirical approach	Land management units (*ejidos, communidades*, state-owned land)	• 1993 • 2009	NS	• Forested • Cleared	• Percentage conserved forest • Avg. slope • Avg. elevation • Distance to roads • Aspect • Type of forest • Indigenous • Tenure • State fixed effects • Type of protected area • Multiple spatial lags
Sims (2010)	Thailand	Protected areas	Regression • Cross-sectional data (2000) and IV (in priority watershed) • Panel data (5 years) and locality fixed effects	None	Locality	• 1967 • 1973 • 1985 • 1992 • 2000	$30m^2$-$60m^2$	• Forested • Cleared	For cross-sectional model: • Avg. slope • Max. slope • Avg. elevation • Maximum elevation • Distance to Thai border • Distance to large river • Distance to mineral deposits • Ecoregion • Avg. temperature • Avg. rainfall • Distance to railroad • Upper watershed dummy • Touristic waterfall

									dummy • Historical forest cover • Historical distance to roads • Historical distance to major city • Irrigation potential • District fixed effects

IV = instrumental variable; NS = not specified in article.

Table 3. Cell-Level Spatial Evaluations of Forest Conservation Policy

Article	*Study area*	*Treatment*	*Empirical approach*	*Control for spillovers?*	*Sampling*	*Years*	*Land-cover data resolution*	*Categories*	*Controls*
Gaveau et al. (2009)	Sumatra, Indonesia	Protected areas	Propensity score matching	Compare deforestation in buffers and adjacent unprotected areas	Random	1990–2000	5km^2 cells	• Forested • Cleared	• 1990 forest cover • Slope and elevation • Forest edge • Travel times to roads, logging roads
Sánchez-Azofiefa et al. (2007)	Costa Rica	Payments for environmental services	Ordinary least squares	None	None	• 1986 • 1997 • 2000	5km^2 cells	Percentage cleared	• Slope • Holdridge life zones (2) • Distance to cities (3) • Percentage cleared before 1997

4. Data

This section provides an overview of the data typically used to analyze forest conservation policies using the approach described in this paper.

4.1. Outcome Variables

As noted above, a defining feature of the spatial evaluation method described here is reliance on land cover and land-cover change maps derived from remote sensing images. An entire discipline focuses on the science of generating such maps, and the associated literature is vast. The review by Fagan and DeFries (2009) is good entry point. Depending on the unit of analysis, various types of outcome variables are used. Plot-level analyses typically use dichotomous variables indicating whether or not the plot was cleared during the study period. They also use categorical variables indicating the specific land cover substituted for forest. Cell- or administrative unit-level analyses typically rely on continuous variables indicating the percentage cleared.

4.2. Treatment Variables

Only a handful of polices have been analyzed. The bulk of the studies examine protected areas. Other treatments examined include payments for environmental services, tenure reform, agricultural intensification programs, and land-use zoning (Tables 1–3, column 3). As in the case of outcome variables, depending on the unit of analysis, various types of treatment variables are used. Plot-level analyses typically use dichotomous variables indicating whether or not the plot was treated during the study period while cell- or administrative unit-level analyses usually rely on continuous variables indicating the percentage of the cell or unit that was treated. For example, a plot-level analysis of the effect of protected areas on land cover change might use a dichotomous variable indicating whether or not the plot was located inside a protected area while a cell-level analysis might use a continuous variable indicating the percentage of the cell that was inside a protected area.

4.3. Control Variables

The following geophysical, institutional, and socioeconomic characteristics of land units have been used to control for the effects of drivers of land-cover change other than forest conservation policies (Table 1–3, column 10).

Geophysical Characteristics. The following variables are most commonly used: terrain characteristics including elevation, slope, and aspect (directional orientation); climate measures including precipitation, temperature, proximity (Euclidian distances or travel times) to roads, population centers, markets, forest edges, processing facilities (dairy, sawmill, etc.); soil characteristics; and ecoregion. In addition, a few studies have included flood hazard, watershed or irrigated area.

Socioeconomic Characteristics. Socioeconomic variables are often omitted from spatial evaluations of forest conservation policies, presumably because they may be endogenous—that is, because they may be influenced by the outcome variable or pick up the effects of unobserved drivers of land-cover change. When studies do include these variables, they tend to use either lagged values or instrumental variables (discussed in Section 6.2.2, below) to control for potential endogeneity (e.g., Cropper et al. 2001; Deininger and Minten 2002). Socioeconomic control variables have included: population density, poverty, and indigenous population.

Administrative Unit. Variables that indicate the administrative unit (province, county, regulatory region) for each unit of analysis are included in some studies (e.g., Blackman et al. 2011; Cropper et al. 2001; Sims 2010). These "fixed effects" can pick up the influence of unobserved variables that are correlated with administrative units (see Section 6.2.3, below).

Spatial Lags. Finally, spatial lag variables, which measure average levels of geophysical characteristics on neighboring land units, have been incorporated in a few regression analyses (Cropper et al. 2001; Müller and Munroe 2005; Nelson et al. 2001). Such variables include slope and elevation.

4.4. Data Sources

Typically, data on outcomes—that is, land cover and land-cover change maps—are the most difficult to acquire. Increasingly, however, data needed to construct outcome, treatment, and control variables are publicly available, often on the Internet. This section lists a number of publicly available sources for such data.

Forest Conservation Policies. Global data on protected areas are available from the UN Environment Programme World Database on Protected Areas (UNEP-WCMC 2007).

Geophysical Characteristics. Terrain (elevation, slope, and aspect) data are available from NASA Shuttle Radar Topography Mission digital terrain elevation maps (Farr et al. 2007). Location of population centers is available from gridded global population maps, such as Gridded Population of the World (CIESIN 2005) and LandScan global population data (Dobson et al 2003). Soils data are available from the UN Food and Agriculture Organization (FAO 2003).

Precipitation data are available from WorldClim (Hijmans et al. 2005). And ecological zone data are available from FAO and the World Wide Fund for Nature (Fischer et al. 2002; Olson et al. 2001). Travel times to population centers generally must be estimated using standard iterative routines in ArcGIS that take into account terrain and road characteristics.

Socioeconomic Characteristics. Spatial data on socioeconomic characteristics are available on a case-by-case basis, mostly from census records maintained by statistical agencies.

Administrative Unit. Data on administrative boundaries are available from a variety of sources, including metasources such as Gridded Population of the World (CIESIN 2005).

5. Evaluation Challenges

This section discusses the main challenges to evaluating forest conservation policy using the broad approach outlined above.

5.1. Estimating Counterfactuals

The overarching challenge in evaluating forest conservation policy is the same as in the evaluation of any type of policy intervention, such as an antimalarial vaccination program. Ideally, for each unit of analysis (hectare, person), the causal effect of the policy (protected area, vaccine) would be measured by comparing the outcome (land-cover change, health status) with the treatment and without it. However, we never actually observe both. For treated units, we observe the outcome with treatment, but not without it. And for untreated units, we observe the outcome without treatment, but not with it. In each case, the unobserved outcome is the *counterfactual*. The overarching challenge in evaluating the policy is that counterfactual outcomes are (by definition) not observed and therefore must be estimated. Generally, this problem is simplified by focusing only on treated units. Hence, causal effects of the policy are measured as the difference between (i) the average outcomes with treatment, which are observed, and (ii) the average outcomes without treatment, which are estimated. In policy evaluation terminology, this effect is known as the average treatment effect on the treated (ATT).

Unfortunately, the three broad strategies most commonly used to estimate conservation policy counterfactuals are flawed and prone—if not likely—to generate biased results.[1] In what follows, we first define each strategy and then explain why it is problematic.

[1] See Joppa and Pfaff (2010b) for a detailed review of the literature on protected area evaluation.

Pretreatment Outcomes for Treated Units. A common approach to estimating counterfactuals for treated units is to use observed outcomes for these units before treatment (e.g., Liu et al. 2001; Gaveau et al. 2009). The causal effect of the policy is measured as the difference in average outcomes before treatment and after treatment. That is, the estimate of the effect is based on a before-and-after comparison. For example, say the unit of analysis is a $30m^2$ plot, the treatment is the creation in 2004 of a protected area on a subset of plots in a study area, the outcome is the deforestation rate, and the outcome period is 2005–2010. Deforestation rates on plots in the protected area prior to the establishment of the protected area, say from 1998–2003, would be used to estimate missing counterfactuals—that is, what 2005–2010 deforestation rates on these plots would have been had they not been protected. The causal effect of the policy would be measured as the difference between the average deforestation rate for protected plots over the period 1998–2003 and the average rate over the period 2005–2010. This empirical strategy depends on the assumption that any and all changes in the outcome variable during the study period are attributable to treatment.

Untreated Outcomes. A second common approach to estimating counterfactuals for treated units is to use outcomes for untreated units during the outcome period (e.g., DeFries et al. 2005; Olivera et al. 2007). The causal effect of the policy is measured as the difference in the average outcome for treated units and for untreated units during the (same) outcome period. That is, the estimate of the effect is based on a with-and-without comparison. Continuing the above example, 2005–2010 rates of deforestation on plots outside the protected area would be used to proxy for the missing counterfactual. The causal effect of the policy would be measured as the difference between the average 2005–2010 deforestation rate for protected plots and the average 2005–2010 deforestation rate for unprotected plots. This empirical strategy depends on the assumption that, aside from being subjected to the treatment, there are no systematic differences between the treated and untreated units that affect outcomes.

Changes in Untreated Outcomes. A third common approach to estimating counterfactuals for treated units is to use *changes* in outcomes for untreated units during an outcome period that starts before treatment is implemented and ends afterwards (e.g., Honey Roses et al. 2011). The causal effect of the policy is measured as the difference in the changes in average outcomes for treated units and untreated units during the outcome period. That is, the effect is measured using a “difference-in-differences” comparison (known in natural sciences as a “before-after-control-impact” comparison) that essentially combines the two approaches described above. Continuing the above example, the measure of interest would be the change in deforestation rates between 1998–2003 and 2005-2010. The change during this period on plots outside the protected area would be used to proxy for the missing counterfactual. The causal effect of the policy would be measured as the difference between changes for protected and unprotected plots. This empirical strategy depends on the assumption that, were it not for the treatment, outcomes on the treatment and control plots would exhibit the same trend over time.

Outcomes in Nearby Areas. As discussed below, a problem with using outcomes for untreated units to proxy for the counterfactual is that these units may be different from treated units in ways that affect the outcome. For example, they may be closer to or farther away from urban sprawl. A common approach to addressing this problem is to use outcomes for treated units located near the treated units (Mas 2005; Nagendra et al. 2004). The causal effect of the policy is measured as the difference in average outcomes for treated units and for nearby untreated units during the same outcome period. That is, the estimate of the effect is based on another type of with-and-without comparison. Continuing the above example once again, 2005–2010 deforestation on plots in a 10km buffer zone outside the protected area would be used to proxy for the missing counterfactual. The causal effect of the policy would be measured as the difference between the average 2005–2010 deforestation rate for protected plots and the average 2005–2010 deforestation rate for unprotected plots in a 10km buffer. This empirical strategy also depends on the assumption that, aside from being subjected to the treatment—here, being adjacent to a policy area—there are no systematic differences between the treated and untreated units that might affect outcomes.

5.2. Challenges to Estimating Credible Counterfactuals

In evaluating forest conservation policies, three factors complicate the task of estimating the missing counterfactual for treated units.

Exogenous Temporal Variation in Deforestation and Degradation.

Rates of deforestation and degradation vary over time for reasons that have little or nothing to do with conservation policies, including changing macroeconomic conditions, demographics, technology, climate, and ecology. Evaluations based on before-after comparisons that do not control for these exogenous factors will conflate their effects with those of the conservation policy.

Nonrandom Siting of Forest Conservation Policies. The areas targeted by forest conservation policies are generally not randomly located. Rather, to minimize implementation costs and/or conflicts with economic development, they are often sited in relatively remote areas where competition for land is limited and deforestation rates are low to begin with (Joppa and Pfaff 2009; Millennium Ecosystem Assessment 2005). In other cases, they are areas with particular ecological characteristics that protect biodiversity or ecosystem services. Evaluations that do not control for this nonrandom siting will conflate its effects with those of the conservation policy itself.

Spillovers. Conservation policies targeting one set of land plots can affect land use on neighboring plots, a phenomenon known as spillover. Spillovers can be either negative (i.e., conservation policies exacerbate deforestation and/or degradation on neighboring plots) or positive (i.e., conservation policies stem deforestation and/or degradation on neighboring plots). Negative spillovers can occur for several reasons (Alix-Garcia et al. In Press). First, policies such as legal protections that inhibit deforestation and degradation on one plot may displace it to neighboring plots, an effect known as *substitution*. For example, the creation of a well-enforced protected area may cause shifting agriculture, logging, and fuelwood collection to migrate to neighboring areas. Second, policies that dampen the supply of timber or nontimber forest products on one set of plots may reduce market supply of these products, which increases their prices in local markets. That, in turn, spurs increased production on neighboring plots, an effect known as price *slippage*. Finally, conservation policies that increase wealth on one set of plots can cause them either to expand production on neighboring plots or to increase their demand for forest products, a phenomenon known as a *wealth effect*. Conservation policies can also generate positive spillovers—that is, they may stem deforestation and/or degradation on neighboring plots. This can occur if they spur private conservation measures (e.g., from ecotourism) or improve regulatory monitoring and enforcement of land-use restrictions in neighboring areas. In any case, evaluations that do not control for spillovers may either overestimate or underestimate the net effect of conservation policies.

It is easy to see how those three complications—exogenous temporal variation in deforestation and degradation, nonrandom siting, and spillovers—could seriously bias the results of evaluations that rely on the three strategies discussed above for estimating counterfactuals. We will consider each strategy in turn.

Pretreatment Outcomes for Treated Units. Evaluations that rely on pretreatment outcomes for treated units to estimate the counterfactual are undermined when outcomes change over time because of exogenous factors. Continuing with the example above, say a study of the effectiveness of a protected area that was created in 2004 estimates the counterfactual as the 1998–2003 rate of deforestation on protected plots. Further, say a leading cause of deforestation in the study area is conversion to pasture and that deforestation declined precipitously between 2005 and 2010 because of a crash in the market for cattle. Hence, the average post treatment (2005–2010) rate of deforestation was much lower than the average pretreatment (1998–2003) for reasons unrelated to the protected area. The study would erroneously attribute this decline in deforestation rates to the protected area, thereby significantly overestimating its causal effect.

Untreated Outcomes. Evaluations that use outcomes on untreated units to estimate the counterfactual are undermined when the conservation policy is nonrandomly sited. The general technical term for this problem is selection bias, which refers to nonrandom selection of units into the treatment. Continuing with our protected area example, say the protected area was sited in a remote location with minimal economic activity or deforestation. As a result, the deforestation rate for treated units is much lower than untreated units for reasons that have nothing to do with the legal protections afforded by the policy. The study would erroneously attribute this difference in deforestation rates to the protected area, thereby significantly overestimating its causal effect.

Changes in Untreated Outcomes. Evaluations that use the before-after change in outcomes on untreated units to estimate the counterfactual also are undermined when the conservation policy is nonrandomly sited. Continuing with our protected area example, say the protected area was sited in a remote location with minimal economic activity or deforestation. As a result, the before-after change in the deforestation rate for treated units is much lower than for untreated units for reasons that have nothing to do with the legal protections afforded by the policy. The study would erroneously attribute this difference to the protected area, thereby significantly overestimating its causal effect.

Outcomes in Nearby Areas. Evaluations that use outcomes on untreated units close to treated units to estimate the counterfactual are undermined when the conservation policy spurs spillovers. Continuing with our protected area example, say the protected area displaces illegal logging to adjacent areas. As a result, the average deforestation rate for treated units is much lower than that on nearby untreated units. The study would erroneously attribute the entire difference in deforestation rates to the conservation policy without taking into account that the policy was responsible for spurring deforestation in nearby areas as well as stemming it inside the protected area. As a result, the study would significantly overestimate the policy's causal effect.

6. Empirical Strategies

In the literature on ex post evaluation of forest conservation policies, two empirical strategies have been used to address the evaluation challenges discussed above: matching and regression analysis. These are often referred to as quasi-experimental methods because they mimic the result of a experiment in which the treatment is randomly assigned in order to avoid selection bias. Although a detailed treatment of these methods is beyond the scope of this document, we briefly describe each below, paying particular attention to those aspects most relevant to the evaluation of forest conservation policy, and providing citations for readers interested in details.

6.1. Matching

The guiding principle in matching is constructing a counterfactual for the treatment units—that is, an estimate of what the deforestation rate on these units would have been absent the conservation policy—using the deforestation rate on “matched” control units that have not been subjected to the policy but that are otherwise very “similar”—specifically, similar in terms of confounding variables that affect both selection into the treatment (i.e., policymakers’ choices about which land units to target for conservation) and the outcome (land-cover change) (Caliendo and Kopeinig 2008; Morgan and Harding 2006). For example, say the treatment in question is a protected area and the outcome is the rate of deforestation during a defined study period. Furthermore, say policymakers tend to choose remote, high-elevation plots for legal protection, and such plots also tend to have relatively low deforestation rates. Matching requires identifying remote, high-elevation unprotected plots to serve as matched controls. Once matched control sites have been identified, the effect of the conservation policy is estimated as the difference between the average deforestation rates on treatment sites and on matched control sites. Technically, this difference is the average treatment effect on the treated—the ATT. A similar method can be used to account for leakage. It is estimated as the difference between the average deforestation rate on sites adjacent to treatment sites and a second set of matched control sites not adjacent to these sites.

Ideally, one would find matched control units that have the exact same observable characteristics (e.g., distance to cities, elevation, soil quality) as treatment units. However, when the number of treated units and the number of observable characteristics are large, this may not be possible. For example, if data on land units include 30 variables, for every treatment unit, it may not be possible to find a control unit with the exact same configuration of these 30 variables. Two approaches are available for circumventing this problem. Both collapse the difficult problem of matching all observable characteristics to a much simpler one of matching a single index of these characteristics.

6.1.1. Propensity Score Matching

Propensity score matching (PSM) pairs treatment and control units based on their propensity score—the probability of the unit’s being treated, as predicted by a dichotomous choice (usually probit or logit) regression model. Propensity scores can be interpreted as an index of the characteristics of the unit of analysis (plot, cell, administrative unit) weighted by their importance in predicting treatment (Rosenbaum and Rubin 1983).

In practice, PSM entails several steps. First, a dichotomous choice regression model is estimated that explains which units were treated on the basis of their observable characteristics. The regressors should include control variables that affect both selection into the treatment (the policymaker's choice about which units of analysis to target for the conservation policy) and the outcome (land-cover change). In addition, these variables should be unaffected by selection into the treatment or anticipation of it (Caliendo and Kopeinig 2008). Next, the estimated coefficients from this first-stage regression model are used to generate propensity scores for each treatment unit and each control unit, and the propensity scores are used to match treatment units with control units. Finally, the effect of the treatment is estimated as the difference in the average outcome for the treated and matched control units. A difference-in-means test is used to determine whether this average treatment effect on the treated is statistically significant.

Various methods are available to match treatment and control units based on propensity scores (Caliendo and Kopeinig 2008; Morgan and Harding 2006). The most straightforward is nearest neighbor 1-to-1 matching, wherein each treated unit is matched to the untreated unit with the closest propensity score. An alternative is nearest neighbor 1-to-*n* matching, wherein each treated unit is matched to the *n* untreated units with the closest propensity scores and the counterfactual outcome is the average of the outcomes for these *n* units. Another alternative is kernel matching, wherein a weighted average of all untreated units is used to construct the counterfactual outcome. The weights are based on similarity as measured by propensity scores.

In theory, standard errors of PSM estimators are potentially biased because, among other things, they do not take into variation stemming from the estimation of propensity scores (Caliendo and Kopeinig 2008; Heckman et al. 1998). A common solution is to calculate PSM standard errors by bootstrapping, an iterative technique that entails repeated resampling. However, recent work has demonstrated that this approach also generates biased standard errors in the case of 1-to-*n* nearest neighbor matching estimators (Abadie and Imbens 2008). As a result, the matching literature increasingly relies on covariate matching, for which methods are available to calculate unbiased standard errors (Abadie and Imbens 2006; Table 1A, below).

The effectiveness of matching in controlling for the nonrandom siting of conservation policies depends on the untestable assumption that the first-stage regression includes all confounding variables that affect both selection into the treatment and the land-cover change. Sensitivity analysis (e.g., Rosenbaum bounds) can be used to test the extent to which a violation of this assumption is likely to drive matching results (Caliendo and Kopeinig 2008; Rosenbaum 2002).

6.1.2. Covariate Matching

As noted above, propensity scores can be interpreted as an index of similarity between two land units that have several relevant observable characteristics. In more general and technical terms, it can be seen as a scale-invariant measure of the distance between two points in multidimensional space. Covariate matching uses alternative measures of such distances, the most common of which is Mahalanobis distance (Abadie and Imbens 2011).

As in the case of PSM, in practical terms, covariate matching entails first pairing treatment and control observations and then estimating treatment effects as the difference in the average outcome for the treatment units and matched control units. And as in the case of PSM, a variety of methods are available to match treatment and control units, including 1-to-1, 1-to-*n*, and kernel matching. As noted above, methods are available to calculate unbiased estimates of standard errors from covariate matching (Abadie and Imbens 2006).

6.2. Regression

6.2.1. Simple

Early evaluations of conservation policies using the broad empirical approach discussed in this paper relied on relatively simple regression analysis (Table 1B). The dependent variable is a measure or index of land cover or land-cover change, the key independent variable is a dichotomous *treatment dummy* that indicates whether the unit was treated, and additional independent variables are controls. Specifications include probit and logit models when the dependent variable is dichotomous (deforested versus no change) and multinomial logits when the dependent variable is polychotomous (e.g., pasture, row agriculture, forest).

For example, Deininger and Minten (2002) examine the effect of protected areas (among other policies) on land-cover change in southern Mexico between 1980 and 1993. They use a plot-level probit model in which the dependent variable is a dichotomous dummy that indicates whether the plot was cleared during this period, a independent treatment dummy indicates whether the plot was in a protected area, and controls measure plot characteristics that affect land-cover change, including elevation, slope, rainfall, soil characteristics, distance to roads, distance to population centers, population density, poverty, indigenous population, and access to credit. The magnitude and significance of the marginal effect on the protection dummy is used to measure the effect of protected areas on deforestation.

A weakness of this approach is that it generates biased treatment effect estimates if the treatment dummy picks up unobserved factors that influence the outcome, namely nonrandom siting of protected areas. That is, the treatment dummy may be endogenous. Such methods also generate biased results if a significant portion of the control units are not similar to the treated units in terms of characteristics that affect the outcome (Ho et al. 2007; Morgan and Winship 2007).

To be fair, the main aim of many studies that use a simple regression approach is not to evaluate a forest conservation policy but to identify the drivers of land-cover change, among which is forest conservation policies. Studies that focus specifically on evaluating the effect of conservation policies tend to use other methods that generate more reliable treatment effect estimates.

6.2.2. Instrumental Variables

The instrumental variable (IV) method takes advantage of known correlations between the treatment (exposure to a conservation policy) and instruments, which are characteristics of treated units that plausibly affect the probability of treatment but do not directly affect outcomes (Morgan and Winship 2007; Wooldridge 2009). In the context of forest conservation policy evaluation, instruments do not affect land-cover change except through the probability of treatment. For example, if conservation policies tend to target areas close to rivers to ensure the continued provision of hydrological services, and this locational characteristic has little or no direct effect on land-cover change, proximity to rivers or location in a priority watershed might have use as an instrument (e.g., Cropper et al. 2001; Sims 2010). In any case, instruments can be used to control for nonrandom siting of conservation polices.

In practice, the IV approach entails several steps: identifying an instrument or set of instruments; estimating a first-stage regression, wherein the treatment is the dependent variable, and the instrument (or set of instruments) along with the other variables are independent variables; using this first-stage regression to predict values of the treatment dummy; using these predictions to substitute for the treatment variable in a second-stage land-cover change regression; and finally correcting the standard errors to account for the fact that the treatment variable is estimated rather than measured. This can be done using a sequential procedure in which each stage is estimated separately or one in which both stages are estimated simultaneously using maximum likelihood.

A drawback of the IV approach is that credible instruments are not easy to identify. A second limitation is that this method measures only the effect of the conservation policy on the subset of units for which siting was affected by the instrument. As a result, this method estimates an effect of the conservation policy—technically, a local average treatment effect—that is conceptually different from estimates derived through other methods (Morgan and Winship 2007).

6.2.3. Fixed Effects

A second approach to controlling for the nonrandom siting of conservation policies is to use fixed effects—dummy variables that aim to pick up the effect of unobserved confounding variables that vary within (but not across) the unit of analysis chosen for that fixed effect. Fixed effects are practical when the analysis uses panel data. Typically, the fixed effects match the level of analysis. For example, Sims (2010) estimates fixed effects models to measure the effect of protected areas on land-cover change in Thailand using locality-level (i.e., township level) panel data covering four consecutive multi-year periods between 1967 and 2000. Her fixed effects are at the locality-level, that is, she includes in her regression model a separate dummy variable for each of the more than 4000 localities in her data set. In a plot-level analysis with tens of thousands of plots, however, plot-level fixed effects may not be feasible. An option is to include fixed effects at a lower level of spatial resolution—for example, at the township level.

6.3. Combining Matching and Regression

As discussed above, matching can be used to generate treatment effect estimates by calculating the difference between average outcome for the treated observations and matched control observations, and then testing for statistical significance. But matching can also be used to "preprocess" data to reduce or ideally completely remove selection bias, and then use a parametric regression model—including, for example, an ordinary least squares, probit, duration, or difference-in-differences regression—to estimate the treatment effect (Caliendo and Kopeinig 2008; Ho et al. 2007). In practice, this approach entails matching treatment and control observations, dropping unmatched observations from the data set, and then running the parametric regression. Hence, matching can be combined with common parametric techniques to analyze a wide variety of data and problems. Several of the conservation policy evaluations summarized in Tables 1-3 use this approach (e.g., Alix-Garcia et al. In Press and Honey-Roses et al. 2011).

7. Modeling Choices

This section provides some brief guidance on the modeling choices that need to be made in spatial forest conservation evaluations.

7.1. Data Assembly

It probably goes without saying that having high-quality geospatial data is a necessary condition for a high-quality spatial forest conservation policy evaluation. Although these data will inevitably be less than ideal, it is helpful to characterize the ideal.

Land-Cover Data. Ideally, the cross-sectional land-cover maps used to estimate land-cover change would be based on compatible classification strategies, to ensure that the maps are comparable. They would have a fine resolution so as to be able to pick up small-scale, fragmented land-cover change (the standard are maps created from Landsat imagery, which typically have a resolution of $30m^2$). In addition, they would measure forest degradation, which is the predominant type of land-cover change in some areas, as well as clearing. Finally, they would cover a period before and after the start of the policy's implementation. Such panel data facilitate measurement of treatment effects based on panel data approaches, which are widely seen as more reliable and credible than cross-sectional methods.

Data on Land Characteristics. Ideally, this information also would be high-resolution panel data. Probably more important, as discussed in Section 6.1, one would measure or proxy for all the important observable confounding factors that affect both selection into the treatment (the policymaker's choice of which units of analysis to target for the conservation policy) and the outcome (land-cover change). Toward that end, it is important to develop an understanding of the factors that that drove land cover change and the siting of the conservation policy. Such an understanding can help determine what observable confounding factors should be included in the analysis. For example, if protected areas were intentionally sited in high biodiversity locations, it would be important to include a measure of biodiversity. An understanding of the siting of the policy also can help identify instrumental variables and potential biases due to omitted variables.

7.2. Unit of Analysis

As discussed in Section 2.1, evaluators must choose a plot, cell, or administrative division as the unit of analysis. Each has advantages and disadvantages.

In principle, the unit of analysis should comport with the decision making processes being analyzed. For example, in most programs involving payment for ecosystem services (PES), participants are land managers who decide whether to participate and also whether to clear tree cover on their land. In such cases, the level of analysis should be the land manager. Among other advantages, this approach would facilitate the use of land manager-level control variables (e.g., household size, education).

Unfortunately, however, in many developing countries digital cadastral data (shape files) defining the land management units—particularly those not subjected to the policy—are hard to come by. For example, for a study of a PES program, digitized data on participating land management units are typically collected by the agency administering the program and are publicly available. Data on land management units not participating in the program, however, are less likely to be available. A plot- or cell-level study would circumvent this problem.

Another advantage of a plot-level analysis is that it simplifies the definition of the control variables based on high-resolution data. For example, say the data in question describe slope, elevation, and aspect. Each of these variables takes a single value at the plot level (provided plots are relatively small compared with the digital elevation map) but many more values at the cell or administrative division level, in which case summary statistics (mean, median, maximum, minimum, or variance) must be used to create a single value for the entire unit. These summary statistics may introduce aggregation bias. That is, they may be too coarse to capture correlations with treatment and outcome variables. The same problem arises with distance and travel time variables.

An advantage of using cells or administrative divisions as units of analysis is that they facilitate the use of ordinary least squares regression techniques (because the dependent variable is the percentage of the cell or administrative division cleared), which are relatively well understood, simple, and flexible and impose a minimal computational burden.

An advantage of using administrative divisions is that data on control variables may be easily obtained at the same level. For example, census data are often available at the county level.

Finally, a disadvantage of using administrative divisions as the unit of analysis is that their boundaries and size may depend on factors that affect outcomes. For example, administrative units are often subdivided because of population growth, and as a result, smaller units tend to be more densely populated. Depending on the methods used, such correlations can bias estimates of treatment effects.

7.3. Sampling in Plot- and Cell-Level Analyses

Plot- and cell-level analyses typically use a subsample of the population of plots in a land-cover or land-cover change map for the actual empirical analysis (Tables 1 and 3, column 6). One reason is that using the population of plots or cells in a map that has a high spatial resolution and/or a broad geographic scope often generates data sets so large as to be computationally impractical. Sampling circumvents the problem. In addition, sampling can be used to create a data set with plots or cells separated by more than a minimum distance, a feature that can reduce or even eliminate bias due to spatial spillovers.

Two methods have been used to generate subsamples. One is to randomly select them using a function in ArcGIS, sometimes eliminating plots that are less than a minimum distance from each other to reduce bias from possible spatial spillovers (e.g., Andam et al. 2008; Nelson and Chomitz 2011). An alternative, used in plot-level analyses, is to overlay a rectangular grid on the land-cover map and select those plots where gridlines cross (Table 1A). For example, Blackman et al. (2011) use a 2km grid, and Müller and Munroe (2005) use a 200m grid.

Typically, the definition of the sample used in the plot-level empirical analysis entails a second step—dropping all plots that were not forested in an initial period (e.g., Blackman et al. 2011; Andam et al. 2008; Deininger and Minten 2002; Robalino et al. 2008). The purpose is to more accurately identify the effect of the policy in stemming forest cover loss.

7.4. Study Area Definition

In defining a study area, it is important to keep in mind that the evaluation approach described here measures the causal effects of a forest conservation policy by analyzing spatial variation in land cover change, the policy, and other geophysical and socioeconomic drivers of land cover change. Therefore, the study area must be large enough encompass significant variation in these factors. To be more precise, it must encompass both untreated areas that are similar to treated areas (in terms of characteristics affecting land cover change) and untreated areas that are dissimilar. When that is not possible, perhaps because all treated and untreated areas are similar (e.g., flat, remote, in a single ecoregion, etc.), the broad evaluation approach described in this paper may simply not be feasible.

7.5. Measuring and Correcting for Spillovers

Studies have used a variety of strategies to measure or control for spatial spillovers (Tables 1–3, column 5). The most common approach is to use matching to measure spillovers (Andam et al. 2008; Blackman et al. 2011; Joppa and Pfaff 2010a; Nelson and Chomitz 2011; Alix-Garcia et al. In Press; Gaveau et al. 2009). Plots in a buffer zone (typically 10km wide) outside of policy areas (typically protected areas) are designated as treatment plots. Spillovers are measured by comparing outcomes for these treated plots with outcomes for matched control plots outside both the policy and buffer areas. A positive and significant treatment effect (i.e., the average outcome on the treated plots minus the average outcome on the control plots) indicates that the policy has positive spillovers and a negative and significant treatment effect indicates the opposite. In either case, these effects must be taken into account in assessing the policy's net effect.

Several studies that use regression models rely on spatial lag variables—typically mean slope or elevation on neighboring land units—to control for spillovers (Müller and Monroe 2005; Nelson et al. 2001; Cropper et al. 2001). These variables are analogues to temporal lags in time series models. They purport to pick up the effect that land use on one plot has on neighboring plots. Although this approach may help to control for spillovers, its ability to accurately measure them is dubious due to endogeneity (Robalino et al. 2006).

In addition to these two common approaches, several studies use ad hoc strategies to control for spillovers. A study that relies on regression analysis uses latitude and longitude fixed effects (Nelson et al. 2001); one bootstraps to correct standard errors potentially biased by spatial spillovers (Chomitz and Gray 1996); one increases the size of the sampling grid to reduce spillovers (Mertens et al. 2002); one tests for spillovers and finds none (Mertens et al. 2004); and one that relies on covariate matching matches on spatial lag variables (Honey-Roses et al. 2011).

7.6. Empirical Strategy

The choice among different empirical strategies briefly described above is the topic of a literature, and summarizing it is beyond the scope of this paper. That said, we make three quick points. First, to ensure robustness, it is generally advisable to use, and report results from, a variety of methods for estimating treatment effects. Indeed, this is the standard approach in the literature (Andam et al. 2008; Blackman et al. 2011; Robalino et al. 2008; Sims 2010).

Second, the approach chosen obviously depends on the policy being analyzed and the available data. IV approaches are feasible only when good instruments can be identified, and fixed effects approaches are feasible only when panel data are available.

Third, matching approaches generate reliable results only if the matching analysis includes the important confounding factors that affect both selection into the treatment and the outcome, and if matches are "high quality"—that is, if treatment and matched control groups are, on average, indistinguishable after matching. A growing set of papers provides practical advice for ensuring matching quality (Caliendo and Kopeinig 2008; D'Agostino 1998; Ho et al. 2007; Morgan and Harding 2006). In part, high quality matching will depend on meeting the criteria discussed in Section 7.4—defining a study area with significant heterogeneity.

8. Examples

This section describes in detail two recently published ex post analyses of forest conservation policies that use the methods discussed above.

8.1. Protected Areas in Costa Rica

Andam et al. (2008) examine the causal effect on deforestation of establishing protected areas in Costa Rica. As the authors point out, the country is widely viewed as a regional, if not global, leader in forest conservation policy. Roughly a third of its forest cover was cleared in the 1960s and 1970s for cattle ranching, coffee, and bananas. However, this trend slowed dramatically in subsequent decades. Proponents say that Costa Rica's 150 protected areas, which cover almost half of the country's remaining forests, deserve at least some of the credit. This assertion motivates the authors' evaluation.

The study uses three years of land-cover data: 1960, 1986, and 1997. The first land cover is derived from aerial photographs and the second two from Landsat images. The studies do not use land-cover data after 1997 to avoid conflating the effect of protected areas and the national PES program that was established in 1997.

The study's unit of analysis is a one-hectare ($173m^2$) plot. Starting with all plots in the entire country that had forest cover in 1960, the authors draw a random sample of 20,000. After eliminating plots in wetlands and indigenous reserves (which are governed differently than protected areas), as well as plots with data problems (e.g., cloud cover), they arrive at a national sample of 15,283 plots.

Of these 15,283 plots, 2,711 are located in one of the country's 150 protected areas, and 12,572 are located in areas that have never been legally protected. Hence, the treatment group comprises 2,711 plots and control plots are drawn from the remaining 15,283 plots.

The outcome variable is a dichotomous dummy that indicates whether deforestation occurred during a defined outcome period that depends on the vintage of the protected area. The study splits the sample of protected areas into two subsamples: those created before 1979 and those created between 1979 and 1996. For protected areas in the first subsample, the outcome period is 1960–1997 and the authors use only treatment and control plots that were forested in 1960. For protected areas in the second subsample, the study period is 1986–1997 and the authors use only plots that were forested in 1986. Hence, for each subsample, all plots used in the analysis were forested at the beginning of the outcome period, and the treatment effect depends on the relative rate of clearing on treatment versus control plots.

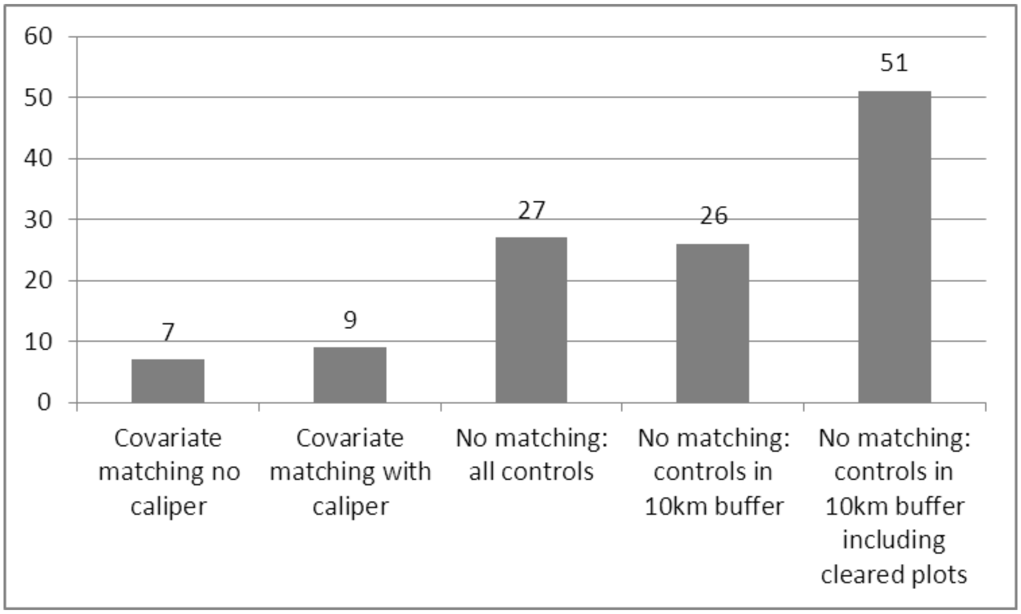
The study uses four sets of control variables. The first is land-use capacity, a categorical variable based on climate, soil type, and slope. The remaining variables are distances, specifically distances to the nearest cleared plot, to roads, and to large cities. (The authors also experimented with additional plot characteristics, including distances to railroads and rivers, population density, educational levels, immigration, and fuelwood use).

The study uses matching to control for the nonrandom siting of protected areas. Specifically, it uses 1-to-2 covariate matching based on Mahalanobis distance, both with and without a caliper. A caliper defines a criterion for judging the quality of the matches. Treated plots for which a match that meets this criterion cannot be found are eliminated from the sample. For the sake of comparison, the study reports results from three "naïve" estimators likely to generate biased results: the difference between the average deforestation rate inside protected areas and outside (i.e., all unmatched control plots are used to estimate the counterfactual); the difference between the average deforestation rate inside protected areas and in a 10km buffer area outside (i.e., unmatched control plots in the buffer area are used to estimate the counterfactual); and the difference between the average deforestation rate inside protected areas and in a 10km buffer area outside without first dropping plots cleared at the beginning of the outcome period. The study uses a Rosenbaum bounds procedure to test the sensitivity of the results to influence of unobserved variables that affect the decision to protect certain areas and not others, and deforestation.

Finally, to determine whether protected areas caused positive or negative spillovers in adjacent areas, the study compares rates of deforestation in a 10km buffer area around protected areas with deforestation on a matched sample of unprotected areas outside the buffer zones. Again, 1-to-2 covariate matching using Mahalanobis distance, both with and without calipers, is used to pair treatment and control observations.

Figure 2 presents the study's findings for the subsample of protected areas created before 1997. The naïve estimators that use questionable methods to generate a counterfactual suggest that protected areas created before 1979 were very effective: they reduced deforestation by 26–51 percentage points. By contrast, the two matching estimators suggest that protected areas were much less effective: they reduced deforestation by only 7–9 percentage points. All five treatment effects are statistically significant. The Rosenbaum bounds test indicate that the two matching results are robust to bias from unobserved factors. Hence, these results clearly illustrate that the conventional methods of generating counterfactuals (discussed in Section 5.1) can result in overly optimistic evaluations of the effectiveness of conservation policies.

Figure 2. Estimates of Pre-1979 Costa Rican Protected Areas' Causal Effects on Deforestation Using Five Estimators



(Source: Andam et al. 2008)

8.2. Protected Areas in Thailand

Sims (2010) examines the causal effect on deforestation (and local socioeconomic conditions, in which we are not interested here) of establishing protected areas in northern Thailand, which has 88 protected areas and the majority of the country's protected forests. The study uses land-cover data from five years—1967, 1973, 1985, 1992, and 2000—to create a panel covering four periods. Land-cover maps from the last four years are derived from Landsat images and have a resolution of $30m^2$ to $60m^2$.

In contrast to Andam et al. (2008), who use plots as a unit of analysis, a dichotomous variable to measure outcomes (forested versus not forested), and matching to control for nonrandom siting, Sims uses an administrative district as the unit of analysis, a continuous variable to measure outcomes, and regression analysis with fixed effects and instrumental variables to control for nonrandom siting. Sims's unit of analysis is a Thai "locality," with an average size of $82km^2$ and an average population of 5,000. She uses all 4,113 localities in the study area in the analysis.

Sims's outcome variable is the percentage of the locality cleared in a period defined by the intervals between her five land covers. The treatment is the percentage of the locality that is protected in a given period.

The study uses ordinary least squares regressions along with fixed effects and instrumental variables to measure the effect of protected areas on deforestation. The principal regression uses the panel data along with locality fixed effects to control for the nonrandom siting of protected areas. The model is specified as

$$y_{ijt} = \gamma x_{ijt} + \alpha_i + \beta_{jt} + \varepsilon_{ijt}$$

where

i is a locality index;
j is a district index (districts are the administrative unit above locality);
t is a period index;
y_{ijt} is the percentage forest cover in locality i, district j, period t;
x_{ijt} is the percentage in protected area in locality i, district j, period t;
γ is parameter;
α_i is a locality fixed effect;
β_{jt} is year/district fixed effect; and
ε_{ijt} is random error term

As a robustness check, the model is estimated for the full sample of 4,113 localities and two subsamples—(i) all localities with more than 10 percent forest cover in the initial baseline year (1967); and (ii) all localities with more than 50 percent forest cover in this year—the idea being that localities with higher initial percentage of forest cover may serve as a better indicator of the effectiveness of protection.

In addition to this fixed effects panel regression, the study also estimates a cross-sectional regression with an IV to control for nonrandom siting of protected areas. The instrument is a dummy variable that indicates whether the locality intersects a major tributary river. As per Cropper et al. (2001), the logic is that localities in important watersheds are more likely to have been selected for protection by policymakers seeking to preserve hydrological services. The assumption is that proximity to a major tributary river affects which localities were protected but does not have an important independent effect on the probability of clearing. The IV regression is specified as

$$y_{ij} = \gamma \hat{\mathbf{x}}_{ij} + \psi \mathbf{z}_i + \beta_j + \varepsilon_{ij}$$

where

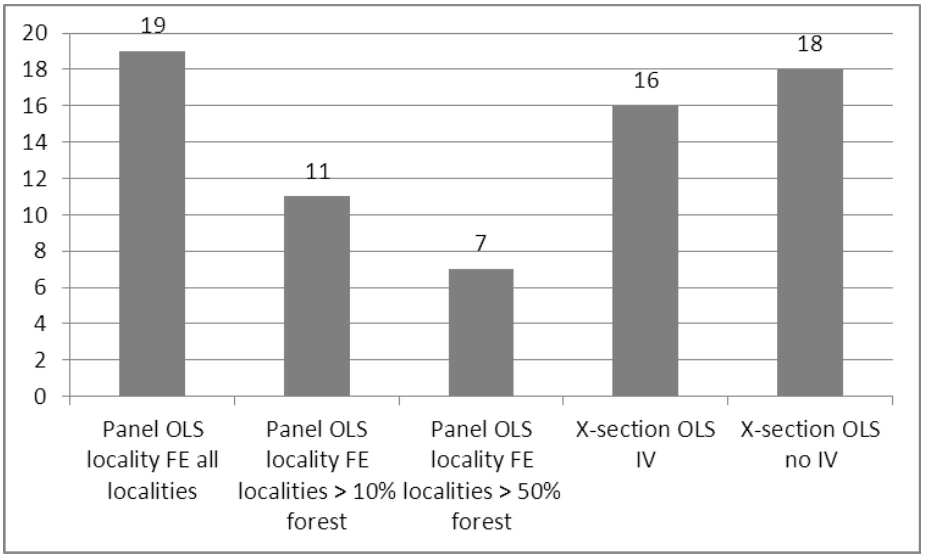
i is a locality index;
j is a district index (districts are the administrative unit above locality);
y_{ijt} is the percentage forest cover in locality i, district j;

$\hat{x}_{ij}$ is an instrument for the percentage in protected area in locality i, district j;
z is a vector of locality characteristics;
γ is parameter;
ψ is parameter;
β_j is a district fixed effect; and
ε is random error term

The vector of locality characteristics includes average slope, maximum slope, average elevation, maximum elevation, distance to the Thai border, distance to the nearest large river, distance to the nearest mineral deposits, ecoregion, average temperature, average rainfall, distance to the nearest major railroad line, an upper watershed dichotomous dummy, a touristic waterfall dichotomous dummy, the historical forest cover, historical distances to major and minor roads, distance to the nearest major city, and irrigation potential. Finally, for the sake of comparison, the study includes this same regression without an instrument for the treatment variable. The study does not attempt to measure or control for spillovers.

Figure 3 presents the major results, which indicate that protected areas have significant effects on deforestation in northern Thailand. The treatment effects estimates range from 7 to 19 percentage points. Among the panel effects models, estimated effects are larger for models that include localities with a relatively small percentage of forest cover in the initial period. Among the cross-sectional models, the one that uses an IV approach to control for nonrandom siting generates a smaller treatment effect, as would be expected if protected areas are sited in places with relatively low deforestation rates.

Figure 3. Estimates of Protected Areas' Causal Effects on Deforestation in Northern Thailand



FE = fixed effects; IV = instrumental variable
(Source: Sims 2010)

9. Planning Ahead

This report focuses on *ex post* evaluation of forest conservation policies that already have been implemented. However, planning an evaluation in concert with implementation can improve its effectiveness and efficiency. This section provides brief recommendations for such planning.

First, as noted in Section 7.1, it is important to identify and collect data on factors affecting land cover change in the study area and factors affecting the siting of the conservation policy, both of which should be treated as confounders in the statistical analysis. Anticipating the benefits of identifying and collecting information on siting *before* the conservation policy is implemented creates an opportunity to directly observe the siting decision making process.

It also creates an opportunity to influence this process. Although outside of this paper's focus on quasi-experimental methods that correct for selection bias after it occurs, it bears mentioning that planning an evaluation in concert with the implementation of the conservation policy makes possible avoiding such bias before it occurs by randomly assigning the conservation policy across space. For example, a payments for ecosystem services or land titling project might be rolled out so that it affects some randomly selected communities before others. Given sufficient lags in implementation, the areas affected later can serve as controls for areas affected earlier. In principle, this type of randomization greatly simplifies and strengthens impact evaluation (Ferraro 2009).

Third, advance planning creates opportunities to ensure that geolocator information are collected for treatment and control areas. The former data are, of course, a prerequisite for spatial evaluation. While not absolutely necessary, the latter data can be extremely useful. For example, for an evaluation of payments for ecological services or land titling that targets farms and communities, such data would facilitate a farm- or community-level evaluation instead of a plot-level evaluation, with the attendant benefits described in Section 7.2.

A related opportunity is collecting geolocator information for areas that applied for but were ultimately not included in the conservation program, which can be used as controls. For example, in an evaluation of the deforestation effects of Mexico's national payments for hydrological services program, Alix Garcia et al. (In Press) uses as controls communities that unsuccessfully applied to participate.

Fifth, advance planning makes it possible to ensure up-to-date land cover change data are readily available by commissioning experts to create them from raw remote sensing data before the evaluation begins. Along with geolocator information for policy areas, land cover change usually are the lynchpin of spatial evaluation, but unfortunately, are publicly available only with a significant lag. In part, that is because creating high-resolution maps from raw remote sensing data is a time-consuming process that generally entails checking predicted land use classifications against first-hand field observations. In principle, such groundtruthing can be incorporated into policy implementation. Commissioning land cover change maps ahead of time also makes possible tailoring them to the evaluation. For example, an evaluation of a project aimed at stemming selective logging as well as clear cutting could commissions high resolution maps needed to accurately measure this activity.

Finally, although again beyond this paper's focus on analysis using secondary data, advance planning makes it easier to combine secondary data on confounding factors with primary survey or field data collected in the course of implementation on, for example, land manager characteristics (e.g., Arriagada et al. In Press).

10. Conclusion

This paper provides an introduction to a relatively low cost method for ex post analysis of forest conservation policies, based on remote sensing data. We have described the defining features of this approach, catalogued and briefly reviewed the studies that use it, discussed the requisite data, explained the principal challenges to using this method and the empirical strategies to overcome them, provided some practical guidance on modeling choices, and described in detail two recent case studies. For uninitiated readers interested in applying these methods, the recommended next step is to read those studies listed in Tables 1–3 that appear most relevant to their purpose.

Looking ahead, there is tremendous scope for further applications of this approach to evaluating forest conservation policies. As Tables 1–3 indicate, only a few studies examining a few policies in a few countries have appeared. Prospects for further analysis are particularly bright in light of the widening availability of high-resolution remote sensing data and corresponding rapid advances in geospatial analysis; these changes will soon make high-resolution panel data on land-cover change available for entire continents (Fagan and DeFries 2009).

References

Abadie, A., and G. Imbens. 2006. Large sample properties of matching estimators for average treatment effects. *Econometrica* 74: 235–67.

———. 2008. On the failure of the bootstrap for matching estimators. *Econometrica* 76(6): 1537–57.

———. 2011. Bias-corrected matching estimators for average treatment effects. *Journal of Business and Economic Statistics* 29(1): 1–11.

Alix-Garcia, J., E. Shapiro, and K. Sims. In Press. Forest conservation and slippage: Evidence from Mexico's National Payments for Ecosystem Services Program. *Land Economics.*

Andam, K., P. Ferraro, A. Pfaff, G. Sanchez-Azofeifa, and J. Robalino. 2008. Measuring the effectiveness of protected area networks in reducing deforestation. *Proceedings of the National Academy of Sciences* 105(42): 16089–94.

Arriagada, P. Ferraro, S. Pattanayak, E. Sills, and S. Cordero. In Press. Do payments for environmental services reduce deforestation? A farm-level evaluation from Costa Rica. *Land Economics* 88.

Blackman, A., J. Robalino, and A. Pfaff. 2011. Paper park performance: natural protected areas in Mexico. Working paper. Washington, D.C.: Resources for the Future.

Caliendo, M., and S. Kopeinig. 2008. Some practical guidance for the implementation of propensity score matching. *Journal of Economic Surveys* 32: 31–72.

Center for International Earth Science Information Network (CIESIN), Columbia University; and Centro Internacional de Agricultura Tropical (CIAT). 2005. Gridded population of the world, Version 3 (GPWv3). Palisades, N.Y.: Socioeconomic Data and Applications Center (SEDAC), Columbia University. Available at http://sedac.ciesin.columbia.edu/gpw.

Chomitz, K. 2007. A*t loggerheads? Agricultural expansion, poverty reduction and environment in the tropical forests.* Washington, D.C.: World Bank.

Chomitz, K., and D. Gray. 1996. Roads, land use, and deforestation: A spatial model applied to Belize. *World Bank Economic Review* 10(3): 487–512.

Chomitz, K., and D. Gray. 1995. Roads, land markets, and deforestation. Policy Research Working paper 1444. Washington, D.C.: World Bank.

Cropper, M., C. Griffiths, and M. Muthukumara. 2001. Predicting the location of deforestation: The role of roads and protected areas in north Thailand. *Land Economics* 77(2): 172–86.

D'Agostino, R. 1998. Tutorial in biostatistics propensity score methods for bias reduction in the comparison of a treatment to a non-randomized control group. *Statistics in Medicine* 17: 2265–81.

DeFries, R., A. Hansen, A. Newton, and M. Hansen. 2005. Increasing isolation of protected areas in tropical forests over the past twenty years. *Ecological Applications* 15: 19–26.

Deininger, K., and B. Minten. 2002. Determinants of deforestation and the economics of protection: An application to Mexico. *American Journal of Agricultural Economics* 84(4): 943–60.

Dobson, J. E., E. A. Bright, P. R. Coleman, and B.L. Bhaduri. 2003. LandScan: A global population database for estimating populations at risk. pp. 267-281 in V. Mesev (ed.) *Remotely Sensed Cities.* Taylor and Francis: London.

Fagan, M.E., and R.S. DeFries. 2009. Measurement and monitoring of the world's forests: A review and summary of remote sensing technical capability, 2009–2015. In The world's forests: Design and implementation of effective measurement and monitoring. Washington, D.C.: Resources for the Future.

Farr, T.G., et al. 2007. The shuttle radar topography mission. *Reviews of Geophysics* 45. Available at http://www2.jpl.nasa.gov/srtm.

Ferraro, P., and M. Hanauer. 2011. Protecting ecosystems and alleviating poverty with parks and reserves: 'Win-win' or tradeoffs? *Environmental and Resource Economics* 48: 269–28.

Ferraro, P. 2009. Counterfactual thinking and impact evaluation in environmental policy. In M. Birnbaum and P. Mickwitz (eds.) *Environmental program and policy evaluation. New Directions for Evaluation, 122,* 75–84.

Ferraro, P., and S. Pattanayak. 2006. Money for nothing? A call for empirical evaluation of biodiversity conservation investments. *PLoS Biology* 4(4): 482–88 .

Fischer, G., H. van Velthuizen, F. Nachtergaele, and S. Medow. 2002. Global agro-ecological zones (Global-AEZ). Food and Agriculture Organization/International Institute for Applied Systems Analysis (FAO/IIASA). http://www.iiasa.ac.at/Research/LUC/GAEZ/index.htm.

Food and Agriculture Organization (FAO). 2003. World reference base map of world soil resources. Available at http://www.fao.org/ag/agl/agll/wrb/soilres.stm.

———. 2011. State of the world's forests 2011. Rome: FAO.

Gaveau, D., J. Epting, O. Lyne, M. Linkie, I. Kumara, et al. 2009. Evaluating whether protected areas reduce tropical deforestation in Sumatra. *Journal of Biogeography* 36: 2165–75.

Heckman, J., H. Ichimura, and P. Todd. 1998. Matching as an econometric evaluation estimator. *Review of Economic Studies* 65(2): 261–94.

Hijmans, R., S. Cameron, J. Parra, P. Jones, and A. Jarvis. 2005. Very high resolution interpolated climate surfaces for global land areas. *International Journal of Climatology* 25: 1965–78. Available at http://www.worldclim.org.

Ho, D., K. Imai, G. King, and E. Stuart. 2007. Matching as nonparametric preprocessing for reducing model dependence in parametric causal inference. *Political Analysis* 15: 199–236.

Honey-Roses, J., K. Baylis, and I. Ramirez. 2011. Do our conservation programs work? A spatially explicit estimator of avoided forest loss. *Conservation Biology* 25(5):1032-1043.

Houghton, R.A. 2005. Tropical deforestation as a source of greenhouse gas emission. In P. Moutinho and S. Schwartzman (eds.), *Tropical deforestation and climate change.* Washington, D.C.: Amazon Institute for Environmental Research.

Intergovernmental Panel on Climate Change (IPCC). 2007. *Climate change 2007: The physical basis.* Working Group I contribution to the fourth assessment report. New York: Cambridge University Press.

Joppa, L., and A. Pfaff. 2009. High and far: Biases in the location of protected areas. *PLoS ONE* 4(12): e8273.

———. 2010a. Global protected area impacts. *Proceedings of the Royal Society B.*

———. 2010b. Reassessing the forest impacts of protection. *Annals of the New York Academy of Sciences* 1185: 135–49.

Liu, J., M. Linderman, Z. Ouyang, L. An, J. Yang, and H. Zhang. 2001. Ecological degradation in protected areas: The case of Wolong Nature Reserve for giant pandas. *Science* 292: 98–101.

Mas, J. 2005. Assessing protected areas effectiveness using surrounding (buffer) areas environmentally similar to the target area. *Environmental Monitoring and Assessment* 105: 69–80.

Mertens, B., D. Kaimowitz, A. Puntdewo, J. Vanclay, and P. Mendez. 2004. Modeling deforestation at district geographic scales in time periods in Santa Cruz, Bolivia. *International Regional Science Review* 27(3): 271–96.

Mertens, B., R. Poccard-Chapuis, M.-G. Piketty, A.-E. Lacques, and A. Ventureri. 2002. Crossing spatial analyses and livestock economics to understand deforestation processes in the Brazilian Amazon: The case of São Félix do Xingú in South Pará. *Agricultural Economics* 27: 269–94.

Millennium Ecosystem Assessment. 2005. *Ecosystems and human wellbeing: Policy responses.* Washington, D.C.: Island Press.

Morgan, S., and D. Harding. 2006. Matching estimators of causal effects: Prospects and pitfalls in theory and practice. *Sociological Methods and Research* 35(1): 3–60.

Morgan, S., and C. Winship. 2007. Counterfactual and causal inference: Methods and principles for social research. New York: Cambridge University Press.

Müller, D., and D. Munroe. 2005. Tradeoffs between rural development policies and forest protection: Spatially explicit modeling in the Central Highlands of Vietnam. *Land Economics* 81(3): 412–25.

Nagendra, H. 2008. Do parks work? Impact of protected areas on land cover clearing. *Ambio* 37: 330–37.

Nelson, A., and K. Chomitz. 2011. Effectiveness of strict vs. multiple use protected areas in reducing tropical forest fires: A global analysis using matching methods. *PLOS One* 6(8): e22722.

Nelson, G., V. Harris, and S. Stone. 2001. Deforestation, land use, and property rights: Empirical evidence from Darién, Panama. *Land Economics* 77(2): 187–205.

Oliveira, P., G. Asner, D. Knapp, A. Almeyda, R. Galván-Gildemeister, S. Keene, R. Raybin, and R. Smith. 2007. Land-use allocation protects the Peruvian Amazon. *Science* 317: 1233–36.

Olson, D., et al. 2001. Terrestrial ecoregions of the world: A new map of life on earth. *BioScience* 51: 933–38.

Pfaff, A., J.A. Robalino, and G.A. Sanchez-Azofiefa. 2008. Payments for environmental services: Empirical analysis for Costa Rica. Terry Sanford Institute of Public Policy Working Paper Series SAN08-05.

Robalino, J., A. Pfaff, and A. Sánchez. 2006. Estimating spatial interactions in deforestation decisions. In A. Kantolean, U. Pascual, and T. Swanson (eds.), *Biodiversity economics: Principles, methods and applications.* Cambridge University Press.

Robalino, J., A. Pfaff, G.A. Sanchez-Azofeifa, F. Alpízar, C. León, and C.M. Rodríguez. 2008. Changing the deforestation impacts of ecopayments: Evaluation (2000–2005) in Costa Rica's PSA program. Discussion Paper 08-24. Turrialba, Costa Rica: Environment for Development.

Rosenbaum, P. 2002. *Observational studies*, 2nd ed. New York: Springer-Verlag.

Rosenbaum, P., and D. Rubin. 1983. The central role of the propensity score in observational studies for causal effects. *Biometrika* 70: 41–55.

Sanchez-Azofeifa, G., A. Pfaff , J. Robalino, and J. Boomhower. 2007. Costa Rica's payment for environmental services program: Intention, implementation and impact. *Conservation Biology* 21(5): 1165–73.

Sierra, R., and E. Russman. 2006. On the efficiency of environmental service payments: A forest conservation assessment in the Osa Peninsula, Costa Rica. *Ecological Economics* 59(1): 131–41.

Sims, K. 2010. Conservation and development: Evidence from Thai protected areas. *Journal of Environmental Economics and Management* 60: 94–114.

United Nations Environment Program, World Conservation Monitoring Center (UNEP-WCMC). 2007. World Database on Protected Areas. Cambridge, UK: World Conservation Union (IUCN) and UNEP-World Conservation Monitoring Center. Available at http://www.wdpa.org.

Wooldridge, J. 2009. *Introductory econometrics: A modern approach*, 4th ed. Cincinnati: South-Western College Publishing.