3 Estimating spatial interactions in deforestation decisions

Juan A. Robalino, Alexander Pfaff, Arturo Sanchez-Azofeifa

1 Introduction

Ongoing decreases in the stock of tropical forest have long been a major concern, due to their implications for biodiversity loss and provision of ecosystem services. Ecological research also provides evidence that even if the stock is held constant, the spatial pattern of forest affects the level of services generated (McCoy and Mushinsky 1994; Twedt and Loesch 1999; Diaz *et al.* 2000; Parkhurst *et al.* 2002; Coops *et al.* 2004; Scull and Harman 2004). A highly fragmented forest made up of small patches may not provide the minimum habitat size that some organisms require. Thus it may offer less protection for species than the same amount of unfragmented forest. It is then important to understand the effects of human activities that fragment standing forest and, as a result, alter the size, the shape, and also the spatial arrangement of habitat. These properties of habitat affect extinction rates of local populations.

Standard economic models of rural land use (e.g. agriculture/forest frontiers) will generate predictions of spatial pattern down to the level of detail that their data permit. However, a focus on spatial pattern highlights a question these models do not address: are there spatial dynamics *per se*? If we look behind observed spatial correlation, do one's land-use choices actually have any causal impacts upon those made by one's neighbours? This chapter presents a model of such spatial interactions and then discusses a method to empirically test for their presence using observed deforestation behaviour. Their existence has implications for the stock of forest, its pattern and the effect of policies on forests.

Acknowledgements: we would like to thank Geoffrey Heal, Malgosia Madajewicz, Ernst Nuppenau, Cristian Pop-Eleches, Rajiv Sethi, Arthur Small, Chris Timmins and Eric Verhoogen for their valuable comments. We thank for generous support the following institutions: Center for Economics, Environment and Society at the Earth Institute and the Institute of Social and Economic Research and Policy at Columbia University; NASA; Tinker Foundation Inc.; the National Center for Ecological Analysis and Synthesis; and the National Science Foundation. All errors are our own.

P1: FQF/FGC P2: FQF
9780521866835c03.xml CUUK868B-Kontoleon March 27, 2007 23:8

Such research builds upon a number of existing literatures. Empirical economic analysis of deforestation, one piece of the much broader literature on the economics of land use, has provided evidence of significant effects on land-use and land-cover outcomes of biophysical and socio-economic factors that one might expect to affect the relative profitability of competing land-use types (Panoyotou and Sungsuwan 1989; Rudel 1989; Stavins and Jaffe 1990; Cropper and Griffiths 1994; Pfaff 1999). Recently, much of this work has employed spatially specific data, making use of geographic information systems (GIS). Predictions are then more spatially detailed (Chomitz and Gray 1996; Geoghegan *et al.* 2001; Kok and Veldkamp 2001; Serneels and Lambin 2001; Walsh *et al.* 2001; Irwin and Bockstael 2002). In this chapter we add 'neighbour effects', or spatial interactions, to this strand of the literature.

Neighbour effects form one part of a broad set of studies of 'social interactions' between agents that need not be spatial. Two spatial examples of such interactions that concern land-use and agriculture are: (i) externalities in US residential development, which have been analysed both theoretically (Turner 2005) and empirically (Irwin and Bockstael 2002), and (ii) adoption of agricultural technologies that affects neighbours' adoption decisions (Case 1992). This chapter brings methods from the social interactions literature to tropical deforestation. The results of this blend can be integrated with many existing discussions of land use and habitat conservation.

For example, rules for selecting habitat for reserves have often been suggested from purely ecological perspectives focused upon where species may exist (Tubbs and Blackwood 1971; Kirkpatrick 1983; Cocks and Baird 1989; Polasky *et al.* 2000), but they could reflect land use as well. This has happened in the consideration of land costs that vary across sites (Ando *et al.* 1998; Polasky *et al.* 2000) and of the threat of clearing that also varies (Pfaff and Sanchez 2004; Costello and Polasky 2004). This chapter suggests another land-use consideration for where public actors should focus, the spatial spillovers to neighbouring land-use decisions from reserves. Results concerning spatial interactions may suggest varying the intensity of such conservation actions over space.

Here we develop a model of neighbours' interactions that builds upon work by Brock and Durlauf (2001a). The key to the empirical application of this model is the use of instrumental variables to identify the magnitude of the effect individuals in a neighbourhood have on each other's choices.[1]

[1] Neighbourhoods and neighbours are defined in the following section. However, a neighbourhood can be seen as an area of land. Two individuals are neighbours if the land they manage is located in the same neighbourhood. While this paper uses a specific definition of neighbourhood and therefore neighbours, it is up to the researchers to redefine the concept of neighbourhood according to their needs.

We rely heavily upon the GIS for many calculations of the neighbours' characteristics, including the biophysical characteristics of neighbouring plots that serve as instruments.

The use of an instrumental variable addresses simultaneity and the presence of spatially correlated unobservable effects. Simultaneity rises when the explanatory variable not only affects but also is affected by the dependent variable.[2] In this case, neighbours' deforestation decisions (the explanatory variable) affect the individual's deforestation decision (the dependent variable) and simultaneously the individual's deforestation decision affects neighbours' deforestation decisions. Since the individual's decisions do not, though, affect the biophysical characteristics of neighbours' plots, using those characteristics to instrument for neighbouring deforestation choices addresses the simultaneity problem. The same reasoning applies to spatial correlation between the unobservable factors that affect an individual's decision and the unobservable factors that affect their neighbours' decisions.[3] A correlation between the two sets of unobservable factors implies correlation between individual and neighbours' behaviours that does not indicate causality. The same correlation does not, though, imply a correlation between an individual's choice and the biophysical characteristics of neighbours' plots. If the instrument is correctly chosen, it addresses these two major issues for the estimation of such interactions (Moffitt 2001).

If positive interactions exist in deforestation, as suggested by the example here that makes use of Costa Rican data, three important consequences arise. Neighbours' decisions reinforcing each other will generate more homogenous forest outcomes within neighbourhoods, i.e. highly fragmented forest patterns are less likely. Also, changed incentives to deforest in one location (e.g. from land policies) spill over to affect areas nearby. Finally, interactions generate the possibility that a given region could end up with significantly different deforestation outcomes (multiple equilibria) due simply to changes in beliefs about what neighbours will do. This depends upon the magnitude of the interactions.[4] Thus, projecting the effects of policies based on extrapolations from past equilibria could miss the possibility that a policy could induce another equilibrium. An agency could implement a policy with the expectation of small increases in deforestation, based on low clearing rates in the past equilibrium, and

[2] See Greene (2003) and Maddala (1983) for more details and examples.
[3] An example of a spatially correlated unobservable effect is the effect on deforestation decisions caused by a soil characteristic that is similar among neighbours, observable by all individuals but unobservable to the researcher.
[4] See Cooper and John (1988) and Brock and Durlauf (2001a) for the role of the magnitude of the interactions in the existence of multiple equilibria.

be surprised by its impact when changed beliefs about neighbours' deforestation behaviour amplifies the policy's impact. Another implication is that it may be desirable to intervene in the interest of a new equilibrium. Suboptimal equilibria can be maintained if individuals have self-fulfilling expectations of suboptimal actions by their neighbours, such as all clearing even though all would be better off by conserving.

This chapter is structured as follows. Section 2 describes a simple model of interactions in the context of deforestation, based on an equilibrium in beliefs about the neighbours' actions. Section 3 discusses empirical issues in measurement of interactions and the benefits of using an instrumental variable approach. Data requirements for analysing neighbours' interactions in deforestation decisions are discussed in section 4. Finally, results for two regions within Costa Rica, as well as discussion of how to obtain the equilibria once the parameters of the model are estimated, are presented in section 5.

2 A model of interactions in deforestation

Social interactions exist when an individual's decision is affected by decisions of other individuals. Models with social interactions can be divided into global interaction models and local interaction models (Blume and Durlauf 2001; Brock and Durlauf 2001a; Glaeser and Scheinkman 2001). Global interaction models are those in which individuals are affected by the decisions of the entire population (see Brock and Durlauf 2001a) and local interaction models are those in which individuals are affected only by the decisions of neighbouring individuals (Schelling 1971; Blume 1993; Ellison 1993). This chapter addresses the modelling of local interactions in the context of deforestation, applying the concepts discussed by Brock and Durlauf (2001a).

Empirical economic models of land use without interactions, applied to deforestation, study how the relative profitability of agricultural and forest land uses is determined by a set of exogenous factors. Some of these models use continuous dependent variables such as county-level deforestation (Stavins and Jaffe 1990, Cropper and Griffiths 1994; Pfaff 1999; Pfaff and Sanchez 2004). Other models use discrete dependent variables, as implied by the observation of binary plot-level deforestation decisions (Chomitz and Gray 1996; Geoghegan *et al.* 2001; Robalino and Pfaff 2005).

Here we develop a discrete dependent model with interdependent individual deforestation decisions. The model assumes a forested area divided in n plots. Each plot is managed by one individual and each individual manages only one plot. Each individual faces a decision between conserving forest in the plot, f, or clearing the plot to engage in an alternative land use, c. In addition, decisions are assumed to be based on the maximisation of profits. Therefore, individuals clear their forest if the profit of any of the non-forest land uses is larger than the profits to be gained by conserving the forest.

Profits are affected by three different factors: a vector of observable plot characteristics, neighbours' deforestation decisions and a random profit shock. As in other standard deforestation models, the effect of the vector of the individual i's observable characteristics, x_i, on profits depends on the action taken by the individual. High levels of rainfall in a plot, for example, increase profits if the plot is deforested and used for agriculture. However, it may decrease profits if the individual decides to conserve the forest for tourism activities. Tourists prefer visiting sunny areas with low levels of rainfall. Therefore, individual i obtains $x_i\beta_c$ when he/she clears but he/she gets $x_i\beta_f$ when he keeps forest, where β_c and β_f are two vectors of parameters that linearly map plots' characteristics into profits.

Standard econometric deforestation models also allow for the existence of unobservable elements that affect profits and thus deforestation decisions. The random profit shock represents the magnitude by which i's profits are affected by these characteristics in ways observed only by that same actor i.

Privately observed characteristics' effects on profits also depend on the action taken by i. For instance, each individual possesses skills in working the land that are unknown to the rest of the agents. A highly skilled individual would obtain greater profits if he decides to engage in agriculture but no particular gain if the decision is to conserve the forest. Therefore, individual i receives an additional $\varepsilon_i(c)$ if clearing occurs but an additional $\varepsilon_i(f)$ if the forest is conserved.

Finally, neighbours' decisions also affect individual profits, unlike in standard empirical models of deforestation by individuals. The individual i's neighbourhood is defined as the area, outside i's plot, covered by forest within a distance r of any point inside i's plot. The set of i's neighbours, N_i, contains the individuals with plots that intersect the neighbourhood of i. It can be assumed that neighbours' decisions affect the profits of clearing based on the fraction of the neighbourhood being deforested, m_i.

Furthermore, it can be assumed that neighbours' effects on individual i's profits also depend on the action taken by i. If a fraction m_i of the neighbourhood is cleared and i also decides to clear, he receives $\rho_{cc}m_i$ for mimicking neighbours' behaviour and $\rho_{cf}(1-m_i)$ for deviating from the neighbours' behaviour. If he conserves his plot of forest he gets $\rho_{fc}m_i$ for deviating from his neighbours' behaviour and $\rho_{ff}(1-m_i)$ for mimicking his neighbours' behaviour. The parameters ρ_{cc}, ρ_{cf}, ρ_{fc} and ρ_{ff} map neighbours' deforestation decisions into profits.

However, deforestation decisions are simultaneous. Hence, individuals form beliefs or expectations about the fraction of the neighbourhood that his neighbours would deforest, m_i^e. Therefore, i clears if expected profits of clearing, $\pi_i(c, m_i^e)$, are larger than the expected profits of conserving his forest, $\pi_i(f, m_i^e)$. Formally, the individual clears if

$$x_i\beta_c + \rho_{cc}m_i^e + \rho_{cf}\left(1 - m_i^e\right) + \varepsilon_i(c) > x_i\beta_f + \rho_{fc}m_i^e + \rho_{ff}m_i^e + \varepsilon_i(f). \tag{1}$$

If the distribution of the difference of the shocks, $\varepsilon_i(c) - \varepsilon_i(f)$, is independent, identically and normally distributed across individuals, the probability, $p_i \in [0, 1]$, that agent i clears is:

$$p_i = \Phi\left(x_i\beta + (\rho_c + \rho_f)m_i^e - \rho_f\right) \tag{2}$$

where Φ represents the standard normal distribution function, ρ_c represents $\rho_{cc} - \rho_{fc}$, ρ_f represents $\rho_{ff} - \rho_{cf}$, and β represents $\beta_c - \beta_f$.

Under rational expectations, individuals compute the probability of their neighbours' clearing, based on which they form beliefs about the fraction of their neighbourhood that will be deforested. Putting this formally,

$$m_i^e = \sum_{j \neq i} w_{ij}\Phi\left(x_i\beta + (\rho_c + \rho_f)m_j^e - \rho_f\right) \tag{3}$$

where w_{ij} is the fraction of land managed by agent j in i's neighbourhood. The equilibrium in expectation is the vector, $(p_1, p_2, \ldots, p_n) \in [0, 1]^n$, that solves the set of equations:

$$p_i = \Phi\left(x_i\beta + (\rho_c + \rho_f)\sum_{j \neq i}\left(w_{ij}p_j\right) - \rho_f\right) \quad \forall i. \tag{4}$$

This set of equations has at least one solution, p^* (Brock and Durlauf 2001b; Robalino and Pfaff 2005). A solution is an equilibrium that generates self-consistent believes. In equilibrium, all individuals' beliefs about their neighbours' actions equals their neighbourhood expected deforestation. Empirically, this allows the neighbourhood's actual deforestation to be used to estimate the interaction coefficient ρ defined as: $\rho_c + \rho_f$.

In fact, there could be more than one vector of probabilities of deforestation that satisfy the system of equations (4). The number of equilibria depends on the magnitude of the interaction coefficient, ρ, and on the plots' observable characteristics.[5] Once the parameters have been

[5] Brock and Durlauf (2001b) and Brock and Durlauf (2001a), for instance, show how the magnitude of the interaction coefficient affects the number of equilibria assuming specific observable characteristics of the agents under a different assumption about the distribution function of the shocks.

P1: FQF/FGC P2: FQF
9780521866835c03.xml CUUK868B-Kontoleon March 27, 2007 23:8

estimated, computational procedures can be used to search for the number of vectors (equilibria) that satisfy the equations of the spatial deforestation model.

The potential existence of multiple equilibria has important implications. Different deforestation outcomes could arise in the same region. Given the irreversibility of deforestation decisions, such effects can last over time. There could also be a Pareto dominant equilibrium outcome. In such a case, decentralised decisions do not assure the best outcome and government intervention is then justified.

3 Estimation strategy

The identification of interaction effects has been widely discussed in economics (e.g. Brock and Durlauf 2001a; Glaeser and Scheinkman 2001; Moffitt 2001; Conley and Topa 2002; Beyer and Timmins 2003) and in modelling land use in particular (Irwin and Bockstael 2002). A number of alternatives have been proposed, but consensus is that the best solution depends on the application (Glaeser and Scheinkman 2001; Moffitt 2001).[6] Simultaneity and the presence of spatially correlated unobservable variables are among the most important sources of bias that should be addressed.

Simultaneity is present in the estimation of interaction coefficients in any application. If individual i is affected by individual j, then individual i also affects j's decision (if j belongs to i's neighbourhood, i must belong to j's neighbourhood). This two-directional process biases the estimation. Without this potential bias being addressed, the estimate of the interaction coefficient would reflect not only the effect of agent j's action on i's decision but also the effect of i's decision on j's action.

Another critical issue is that only limited information in terms of individual and plot characteristics can be observed. Many other driving factors of deforestation end up in the errors of the regression equation. This is especially important since some of those other factors are also spatially correlated. The estimation, then, of the interaction term ρ by only using the neighbourhood deforestation rate, m_i, is inconsistent. What appear to be effects of neighbouring choices on individuals' choices could be the result of spatial correlation between unobserved deforestation drivers.

Some estimation techniques can address simultaneity, others can address spatially correlated errors and some can address both.

[6] In each application, one specific econometric problem might be more severe than another. Therefore, the best strategy of estimation will vary according to the application and data availability.

3.1 *Spatial econometric approach*

Anselin's (1988) Spatial Autoregressive (SAR) model has been used for the estimation of local interactions. The SAR model deals with simultaneity by solving the econometric equation for the dependent variable present in the right- and left-hand sides of the equation and then estimating the non-linear resulting functional form of the parameter via maximisation of a likelihood function.[7]

Anselin's Spatial Error model addresses the correlation of the errors. However, this approach strongly depends on the assumption of the spatial relation of the errors. If the spatial error structure is not correctly determined, some of the unobservable variables will still affect the estimates of the interaction coefficient. Knowing the correct spatial structure of the unobservable variables that affect deforestation is by definition impossible. This is also true for the Anselin's General Spatial model that considers spatial correlation both of the errors and, as in the SAR, of the dependent variable.

3.2 *Instrumental variable approach*

Simultaneity and the presence of spatially correlated unobservable factors can be addressed using instrumental variable techniques (Evans *et al.* 1992; Moffitt 2001). The ideal instrumental variables are exogenous neighbours' characteristics that explain neighbours' deforestation decisions and that are not correlated to the unobservable shocks that affect individuals' deforestation decisions. If these conditions hold, the variation in the instruments can be used to infer the effects of neighbours' deforestation in individuals' deforestation decisions.

Using neighbours' characteristics to infer the interaction effect avoids simultaneity. In this case, the individual decision does not affect the exogenous neighbours' characteristic. Therefore, the feedback effect of the individual deforestation decision on neighbours' decisions does not affect the estimation.

Additionally, the instrumental variable approach addresses the effects of spatially correlated unobservable factors in the estimation. These factors do in part drive the deforestation decisions of neighbours, but do not affect their exogenous characteristics. By using exogenous neighbours' characteristics, the correct estimate of the interaction parameter can still be accomplished. This is true as long as the exogenous neighbours' characteristics are uncorrelated with the individual's unobservable shocks.

[7] The application of these models causes significant computational demands, limiting the possibility of using large data sets.

The key condition of the IV strategy is, therefore, that the instruments not be correlated with the unobservable factors that drive individuals' decisions. For example, the average of the neighbours' minimum distance to a local road can be used as instrument. This variable affects neighbours' deforestation decisions but also could reflect unobservable abundance of roads in the area, something which while unobserved may also directly lead the individual to clear forestland. This choice of instrument would reflect the effects of interactions and the direct effects of the abundance of roads in the interaction coefficient jointly, which clearly would bias the interaction coefficient.

Neighbouring ecological characteristics and topography may be uncorrelated to the unobservable individuals' characteristics. It follows that deforestation decisions may be affected by the individuals' own ecological and topographic characteristics, but not by their neighbours' ecological and topographic characteristics. These are the instruments that we use in the model that we show in this chapter in the context of Costa Rica.

However, there could still be unobservable variables that affect individuals' deforestation and that are correlated with such exogenous neighbouring characteristics. One response to such potential issues is to absorb possible unobservable plots' characteristics that could be correlated with the instrument in the deforestation equation itself with control variables. For instance, controlling for the density of local roads in the neighbourhood would reduce the bias when using neighbours' minimum distance to the roads as the instrument. In general, controlling for spatially explicit variables can minimise any possible correlation between the instrument and unobservable plot characteristics that directly affect plot deforestation.

3.3 Discreteness and the Two Stage Probit Least Squares

The Two Stage Probit Least Squares (2SPLS) method (Maddala 1983) is available in order to implement instrumental variable techniques in a discrete dependent variable approach. This method involves two steps. The first step consists of regressing neighbours' deforestation on the instruments and the rest of exogenous variables that explain the individuals' deforestation decision. The second stage consists of using the predicted values of the first stage regression to estimate the interaction effect in the individual's deforestation equation.

In the first stage the instruments and exogenous individuals' characteristics[8] are used to predict neighbourhood deforestation using a linear

[8] Adding exogenous individuals' characteristics improves efficiency in the estimation.

specification,

$$m_i = \Pi_1 \sum_{j \in N_i} (w_{ij} \bar{x}_j) + \Pi_2 x_i + \mu_i \tag{5}$$

where Π_1 and Π_2 are the reduced-form coefficients to be estimated, $\bar{x}_j$ are the exogenous characteristics that affect the decision of only the individual j, for all j in N_i and does not affect the decision of individual i. Therefore, the instrument is

$$\sum_{j \in N_i} (w_{ij} \bar{x}_j), \tag{6}$$

which represents the value of the average exogenous characteristics in the neighbourhood of i.[9] The estimated reduced-form coefficients, $\hat{\Pi}_1$ and $\hat{\Pi}_2$ in equation (5), are used to predict neighbourhood deforestation, $\hat{m}_i$.

In the second stage, neighbourhood deforestation is substituted for using its predicted values. Then, the interaction coefficient ρ can be estimated from

$$\Pr(y_i = 1) = \Phi(x_i \beta + \rho \hat{m}_i - \rho_f) \tag{7}$$

by standard likelihoods methods,[10] where the dependent variable, y_i, is discrete and reflects the observation of whether plot i has been deforested, 1, or not, 0, in a specific period.

4 Data requirements and GIS

The estimation of the parameters of the model requires information on the individual's deforestation decision, y_i, the individual's vector of observable characteristics, x_i, and the individual's neighbourhood deforestation, m_i. Additionally, some of the observable characteristics of the individual's neighbours should satisfy certain conditions, discussed below, to construct the instrument. If these conditions hold, the instrument can be used to estimate correctly the interaction parameter.

Geographic information systems can be used to process spatially specific data to produce the variables required for the analysis. Recently, GIS has been used to analyse deforestation (see Chomitz and Gray 1996; Pfaff 1999; Kok and Veldkamp 2001; Serneels and Lambin 2001; Walsh *et al.* 2001; Irwin and Bockstael 2002).

The number of observations available for analysis is extremely large when using spatially explicit or pixel-level forest information. If

[9] Note that $i \notin N_i$.
[10] We follow the standard normalisation assumption that the variance of the privately observed shocks, σ, is one as in Brock and Durlauf (2001).

computations for either variable creation or estimation become difficult when using all of the point observations, a valid alternative is drawing random samples of pixels or locations from the maps. This can simplify the calculations and speed up these processes. This section discusses how the required variables are obtained, suggests what variables can be used to build the instrument, and conveys the role that GIS can play in this approach to estimation.

4.1 Deforestation decisions

The object of study is the analysis of deforestation decisions in privately owned land during a period of time. Therefore, the analysis should be focused on those plots that are covered by forest. Furthermore, we exclude land within national parks as these are owned by the government and decisions about the management of the land in these areas are not based on individuals' profit calculations as assumed in the model. If plot i has been deforested by the end of the period, y_i the dependent variable is assumed to have value 1 but if the plot is still covered by forest at the end of the period, y_i, it is given the value 0.

Forest satellite pictures can be used to obtain deforestation dynamics (e.g. Chomitz and Gray 1996; Pfaff 1999; Kok and Veldkamp 2001; Serneels and Lambin 2001; Walsh *et al.* 2001). Data from Costa Rica are used to illustrate the estimation procedure. Forest satellite pictures taken in 1986 and 1997 and developed by the Tropical Scientific Center in Costa Rica are used to describe the presence of forest in 30 m^2 pixels across Costa Rica. In this study, 10,000 randomly drawn pixels across Costa Rica serve as plot observations to analyse deforestation. From these pixels, only those that are in privately owned forest are considered for the analysis. The dependent variable, in this case, is obtained as follows. Pixels covered by forest in 1986 that are deforested by 1997 are associated with value 1 and pixels covered by forest in 1986 that are still covered by forest in 1997 receive the value 0.

4.2 Neighbourhood deforestation

The hypothesis being tested is whether the fraction of the neighbourhood that is deforested, m_i, affects the individual's deforestation decision, y_i. Therefore, the information about m_i is required. One of the advantages of using GIS is that it is possible to calculate the actual fraction of the neighbourhood that is deforested during the period of study. Another alternative is using the randomly drawn sample of pixels in the neighbourhood and calculating the fraction that is deforested during the period. Brock and Durlauf (2001b) discuss the use of the sample to infer the fraction of neighbours that take a specific decision.

In order to calculate deforestation within the neighbourhood, the concept of neighbourhood should be well defined. Definitions of neighbourhoods and neighbours in the literature are as numerous as the type of interactions that have been studied. It is common to define neighbourhoods using political divisions such as provinces, counties or districts.[11] However, neighbourhoods can also be defined by distances alone, regardless of political boundaries. Here we follow the second approach. That is, neighbourhoods are defined based on distances and such an approach is used to estimate interactions in our Costa Rica example. More specifically, the neighbourhoods are defined as the areas covered by forest within a 10 km radius. Any two plots separated by a distance smaller than 10 km, covered by forest, are considered to be neighbouring plots. Figure 3.1, for instance, shows the location of a plot represented by a star, the 10 km radius neighbourhood, represented by the large circle, sampled neighbouring plots, represented by triangles, and the rest of the sampled plots (observations), represented by dots. Forest satellite pictures are used to calculate the deforested fraction of these neighbourhoods between 1986 and 1997 in Costa Rica.

4.3 Observable drivers of deforestation

Observable characteristics that are commonly used in deforestation models are those that describe the socio-economic conditions in the plot, such as population (Cropper and Griffiths 1994; Anderson *et al.* 2002), local wages (Pfaff 1999; Anderson *et al.* 2002), distance to markets (Pfaff 1999; Geoghegan *et al.* 2001; Serneels and Lambin 2001; Anderson *et al.* 2002), distance to roads (Chomitz and Gray 1996; Pfaff 1999; Sernels and Lambin 2001; Anderson *et al.* 2002; Geoghegan *et al.* 2001), and those that describe the ecological conditions in the plot, including vegetation type (Serneels and Lambin 2001; Pfaff and Sanchez 2004), the slope of the terrain (Chomitz and Gray 1996) and soil type (Cropper and Griffiths 1994; Chomitz and Gray 1996; Pfaff 1999; Geoghegan *et al.* 2001).

An example of a set of plot characteristics that would control effectively for factors that might be correlated to the instrument and affect individuals' deforestation in Costa Rica is presented in Table 3.1. These variables are also calculated using GIS.

[11] Others have also defined neighbourhoods based on social connections (e.g. see Conley and Udry 2001 for an application in agriculture). The literature on 'networks' also defines neighbourhoods from different perspectives.

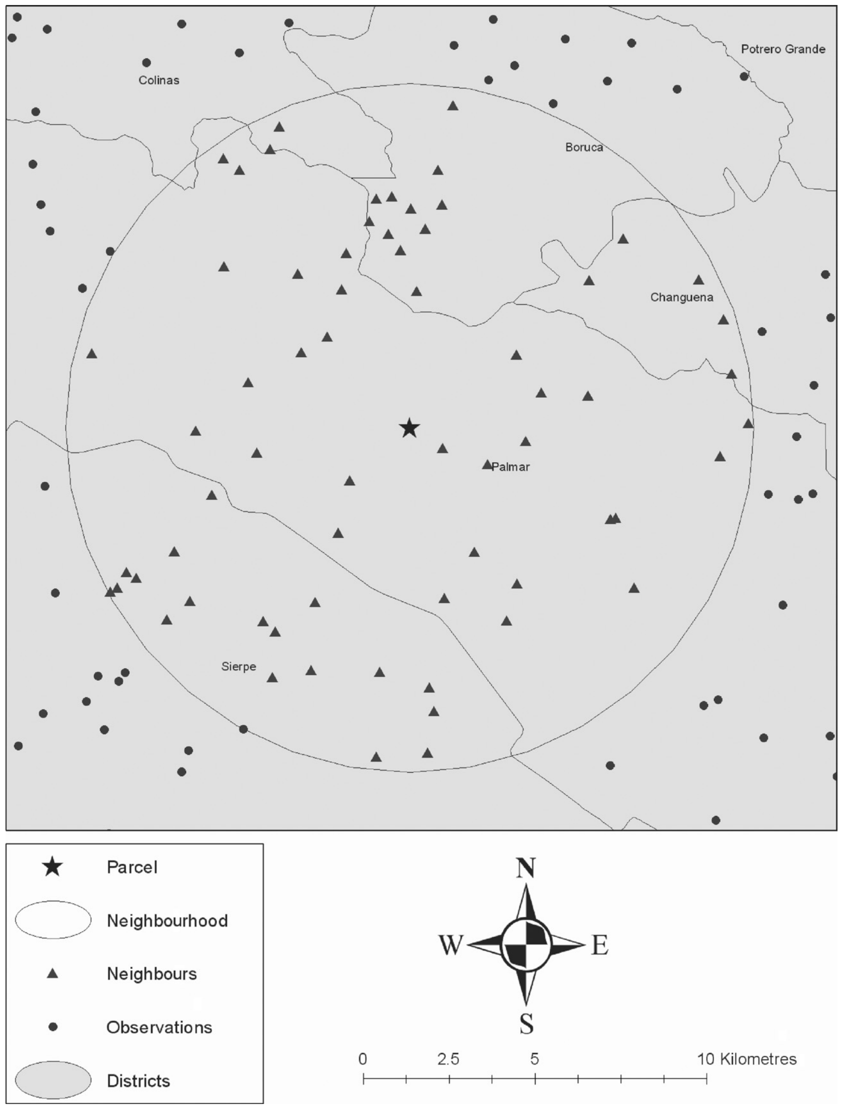

Figure 3.1. Illustration of the observations, neighbourhoods and neighbours

Table 3.1 *List of plot characteristics*

Type	Characteristics	
Distances	Distance to the city (San José)	DSJ
	Distance to a port (Caldera)	DTC
	Distance to a port (Limon)	DTL
	Distance to national roads	DNR
	Distance to local roads	DLR
	Distance to sawmills	DTS
	Distance to schools	DSC
	Distance to cleared areas	DCA
	Distance to towns (county capital)	DMT
Natural characteristics	Slope of the terrain Life Zones	SLO
Characteristics of areas around the plot	Length of national roads at 10 km radius	LNR
	Length of local roads at 10 km radius	LLR
	number of sawmills at 10 km radius	NSM
	Number of towns at 10 km radius	NMT
	Number of schools at 10 km radius	NSC
	Percentage of cleared area at 10 km radius	CLP

4.4 Instrumenting neighbourhood deforestation

As noted above, an instrument should satisfy two conditions. First, it should explain the neighbours' deforestation decisions. Second, it should not be correlated to unobservable characteristics that affect i's decision. The first condition suggests that the characteristics from the vector x_j that affect j's decision, where j represents those individuals in i's neighbourhood, should be considered as instruments. However, not all of j's observable characteristics can be used as instruments. Some of these characteristics, as discussed before, are correlated to unobservable characteristics that affect i's deforestation decision, which violates the second condition.

Plots' characteristics determined by nature can satisfy these conditions. Natural characteristics reflect a source of exogenous variation that can be useful in identification of social interaction processes (e.g. Chaudhuri 1999; Munshi 2003).

Two proposed instruments are neighbours' slopes of the terrain and neighbours' ecological characteristics.[12] These are chosen as they do not affect the individuals' deforestation decisions directly. Moreover,

[12] The classification of the plots' ecological characteristics is based on Holdridge Life Zones.

individuals' deforestation decisions are affected by their own slopes and own ecological characteristics. Computing these instruments is a simple task. Each plot i has a set of neighbouring plots. Each neighbouring plot has its characteristics, such as slope of the terrain or ecological characteristics. Therefore, the instruments can be easily computed by calculating the average of the neighbouring plot characteristics.

5 Results and equilibria

Two techniques are used to estimate the interaction parameter: standard probit and 2SPLS. The 2SPLS uses neighbours' slopes as the instrument for neighbours' deforestation. These techniques are applied to two different regions in Costa Rica shown in Figure 3.2. The regions were chosen based on their quantity of forest and their ecological importance. The area that was left out of the analysis does not have enough deforestation to perform the analysis.

Descriptive statistics of the characteristics of the plots in each region are shown in Table 3.2. The classification of the regions was accomplished by regrouping the government's planning sectors. Specifically, Region 1 contains Huetar Norte and Huetar Atlantica, Region 2 contains Brunca, the Central Area and the Central Pacific, and finally, the region left out is Chorotega. We divide Costa Rica in these two regions in order to test whether the level of interactions could vary across space. If so, that could generate different policy implications for the different areas. Regions are grouped according to their characteristics and location.

Estimates of the interactions parameter are presented in Table 3.3.[13] In Region 1, the standard probit estimate suggests that interactions are positive and significant. However, standard probit estimates are upward biased due to simultaneity and the presence of spatially correlated unobservable factors. Unbiased estimates can be found by using a 2SPLS technique. The 2SPLS estimates show insignificant neighbourhood effects. These two results show that if simultaneity and the presence of spatially correlated effects are not addressed in the empirical approach to measuring interactions, one might conclude wrongly that interactions in Region 1 exist when there is no evidence for that.

However, using 2SPLS can also lead the researcher to conclude that interactions exist. In Region 2, probit and 2SPLS estimates of the interaction parameter are positive and significant and their magnitude is similar.[14] Standard errors under the 2SPLS however are larger than the standard errors from the probit estimates. This difference arises as

[13] In the Appendix, complete regression results are presented.
[14] A statistical test can not reject the null hypothesis that these estimates are equal.

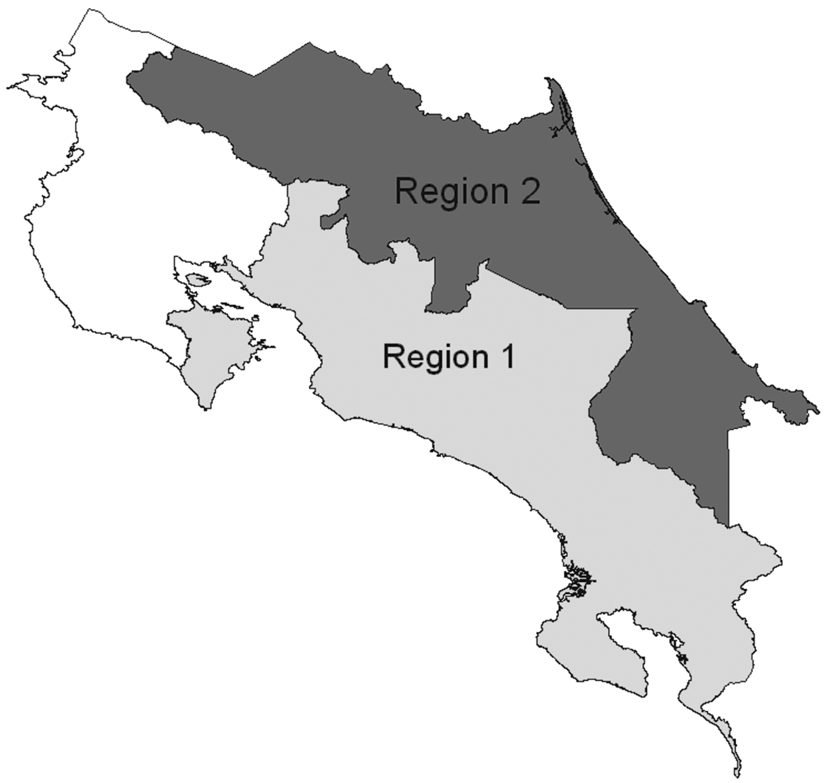

Figure 3.2. Region 1 and Region 2*

a consequence of the presence of simultaneity and spatially correlated unobservable effects. Moreover, the difference in the 2SPLS estimates between Region 1 and Region 2 shows that the presence of interactions might vary across regions.

Once the parameters of the model have been estimated, a numerical procedure can be used to search for the equilibrium outcome, p^*. The probabilities of deforestation in equilibrium can be computed by an iterative process. A set of initial beliefs, $p^{(1)}$, generates a second set of beliefs, $p^{(2)}$, using[15]

$$p_i^{(2)} = \Phi\left(x_i\hat{\beta} + \hat{\rho}\sum_{j\neq i}\left(w_{ij}\,p_i^{(1)}\right) - \hat{\rho}_f\right) \quad \forall i \tag{8}$$

* Area's outside the regions do not have enough deforestaation in the sample

15 Note that ρ_f cannot be identified from the model when among the set of individual's characteristics a constant term is present. The process can still go on since the estimated constant term would contain both effects.

Table 3.2 *Descriptive statistics for Region 1 and Region 2*

		Region 1 (obs. 637)				Region 2 (obs. 810)			
Variables		Mean	S. D.	Min	Max	Mean	S. D.	Min	Max
Deforestation									
Def. decision: y	Dummy	0.08	0.28	0	1	0.20	0.40	0	1
Neighbours' def.	%	0.07	0.07	0	0.57	0.20	0.13	0.01	0.70
Instrument									
Neighbours' slopes	Degrees	11.25	5.59	0	25.79	2.64	3.08	0	12.78
Controls									
SLO	Degrees	11.04	8.01	0	32.05	2.66	4.48	0	26.56
Bad Life Zones	Dummy	0.21	0.40	0	1	0.52	0.50	0	1
Good Life Zones	Dummy	0.17	0.37	0	1	0.01	0.11	0	1
DSJ	Km	89.06	53.38	0	223.04	77.57	34.83	1.48	174.48
DTL	Km	131.22	45.51	44.65	234.19	108.59	68.35	0.99	273.33
DTC	Km	126.12	74.35	9.75	278.94	119.25	39.94	35.91	239.69
DLR	Km	2.37	2.30	0.00	21.37	2.54	2.08	0.00	10.26
DNR	Km	3.87	3.77	0.01	31.09	5.17	4.33	0.01	22.08
DTS	Km	18.60	10.70	0.98	60.20	20.31	11.99	1.29	71.46
DSC	Km	12.29	7.42	0.32	53.65	19.39	11.24	0.99	54.79
DCA	Km	0.22	0.29	0.00	2.60	0.27	0.38	0.00	3.02
DMT	Km	15.50	7.99	0.76	54.82	27.44	15.13	1.89	71.46
LNR	Km	38.93	32.61	0	284.92	28.36	23.93	0	159.58
LLR	Km	52.88	41.12	0	248.91	50.00	34.47	0	234.05
NSM	Units	0.45	1.50	0	16	0.53	1.47	0	11
NMT	Units	0.36	0.81	0	9	0.19	0.61	0	7
NSC	Units	1.19	4.01	0	66	0.51	1.47	0	17
CLP	%	0.57	0.26	0	0.98	0.44	0.21	0	0.98

Table 3.3 *Estimates of the interactions parameter (ρ)*

	Region 1		Region 2	
	2SPLS	Probit	2SPLS	Probit
$\hat{\rho}$	0.40	6.49**	3.31**	3.06**
Standard errors	12.28	1.37	1.31	0.50
N	637	637	810	810
-Log likelihood	147	147	353	353

** indicates significance at 99%

The iterative process consists in computing $(p^{(1)}, p^{(2)}, \ldots, p^{(k)})$, until $p^{(k)}$ equals $p^{(k+1)}$. The set of probabilities of deforestation, $p^{(k)}$, is an equilibrium because it satisfies the set of simultaneous equation (4). Formally,

$$p_i^{(k)} = \Phi\left(x_i\hat{\beta} + \hat{\rho}\sum_{j\neq i}\left(w_{ij}\,p_i^{(k)}\right) - \hat{\rho}_f\right) \quad \forall i. \tag{9}$$

This procedure finds only stable equilibria. It has been argued, though, that this type of equilibrium is more likely to be observed in the long run. This is because if the system is in an unstable equilibrium, then small changes in the beliefs of the agents can shift the system to a stable equilibrium. By increasing the number of initial conditions considered, the probability of finding all of the equilibria increases.

Using the set of equations (9), it can be seen that changes in the vector of individual characteristics x_i affect the probabilities of deforestation of other individuals. This effect depends on the magnitude of the interaction, ρ. A change in characteristics of an individual i affects the probability that i clears, which in turn affects the probability that j clears, given that i is j's neighbour. This second effect depends on the magnitude of the interaction coefficient, ρ. This example shows how policy interventions that affect only individual i could end up affecting all of the individuals that have i as a neighbour.

6 Conclusion

The dependency of the provision of ecosystem services on the stock and the spatial distribution of forest is leading researchers to focus on the spatial dynamics of forest. This chapter has discussed a method to empirically test one of the key factors that shape the stock and spatial pattern of forest: neighbours' interactions in deforestation decisions. The methodology applied, based upon the use of instrumental variables, could be used in different regions where different species reside but are threatened by deforestation and in other land-use contexts, such as settings where reforestation is occurring in cleared areas.

An illustration of the approach proposed here was presented for two regions in Costa Rica. In one of the regions, it is shown that there is no evidence for interactions using the instrumental variable approach, contradicting the result of a standard approach. In the other region, using instruments positive spatially reinforcing interactions are found.

Such interactions have important implications. Positive interactions reduce forest fragmentation within neighbourhoods and imply that policies which alter incentives to deforest in one location have spillover effects in neighbouring locations. Further, they create the possibility of multiple equilibria. The potential for multiple equilibria implies that projections of the effects of new policies which are based on extrapolations from past equilibria could be missing the possibility that a policy could induce another equilibrium.

Further research could focus on identifying impacts of spillover effects and multiple equilibria on the supply of environmental services. Ecological results or new research can link the quantity of forest and its spatial structure with the supply of environmental services. Analysis of overall impacts could be accomplished by generating simulations in an integrated model using interaction parameters of different magnitudes.

REFERENCES

Anderson, L., Granger, C., Reis, E., Weinhold, D. and Wuder, S. 2002. *The Dynamics of Deforestation and Economic Growth in the Brazilian Amazon*. Cambridge: Cambridge University Press.

Ando, A., Camm, J., Polasky, S. and Solow, A. 1998. Species distributions, land values, and efficient conservation, *Science*. **279** (5359). 2126–2128.

Anselin, L. 1988. *Spatial Econometrics: Methods and Models*. Boston: Kluwer Academic Publishers.

Bayer, P. and Timmins, C. 2003. Estimating equilibrium models of sorting across locations. Yale University. Economic Growth Center. Discussion Paper No. 862. 1–31.

Blume, L. E. 1993. Statistical mechanics of strategic interaction, *Games and Economic Behaviour*. **5**. 387–424.

Brock, W. A. and Durlauf, S. N. 2001a. Discrete choice with social interactions. *Review of Economic Studies*. **68** (2). 235–260.

Brock, W. A. and Durlauf, S. N. 2001b. Interactions-based models. In J. J. Heckman. and E. Leamer (eds.). 3329–3371.

Case, A. 1992. Neighborhood influence and technological change. *Regional Science and Urban Economics*. **22** (3). 491–508.

Chaudhuri, S. 1999. Forward-looking behavior, precautionary savings, and borrowing constraints in a poor, agrarian economy: tests using rainfall data. *Working Paper 9899–10*. Columbia University.

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

Cocks, K. D. and Baird, I. A. 1989. Using mathematical-programming to address the multiple reserve selection problem – an example from the Eyre peninsula, South-Australia. *Biological Conservation*. **49** (2). 113–130.

Conley, T. G. and Topa, G. 2002. Socio-economic distance and spatial patterns in unemployment. *Journal of Applied Econometrics*. **17** (4). 303–327.

Cooper, R. and John, A. 1988. Coordinating coordination failures in Keynesian models. *Quarterly Journal of Economics*. **103** (3). 441–463.

Coops, N. C., White, J. D. and Scott, N. A. 2004. Estimating fragmentation effects on simulated forest net primary productivity derived from satellite imagery. *International Journal of Remote Sensing*. **25** (4). 819–838.

Costello, C. and Polasky, S. 2004. Dynamic reserve site selection. *Resource and Energy Economics*. **26** (2). 157–174.

Cropper, M. and Griffiths, C. 1994. The interaction of population-growth and environmental-quality. *American Economic Review*. **84** (2). 250–254.

Diaz, J. A., Carbonell, R., Virgos, E., Santos, T. and Telleria, J. L. 2000. Effects of forest fragmentation on the distribution of the lizard psammodromus algirus. *Animal Conservation*. **3**. 235–240.

Ellison, G. 1993. Learning, local interaction, and coordination. *Econometrica*. **61** (5). 1047–1071.

Evans, W. N., Oates, W. E. and Schwab, R. M. 1992. Measuring peer group effects – a study of teenage behavior. *Journal of Political Economy*. **100** (5). 966–991.

Geoghegan, J., Villar, S. C., Klepeis, P., Mendoza, P. M., Ogneva-Himmelberger, Y., Chowdhury, R. R., Turner, B. L. and Vance, C. 2001. Modeling tropical deforestation in the southern Yucatan peninsular region: comparing survey and satellite data. *Agriculture Ecosystems and Environment*. **85**. 25–46.

Glaeser, E. and Scheinkman, J. 2001. Measuring social interactions. In S. N. Durlauf. and H. P. Young. (eds.). 83–132.

Greene, W. H., 2003. *Econometric Analysis*. Upper Saddle River, NJ: Prentice Hall.

Irwin, E. G. and Bockstael, N. E. 2002. Interacting agents, spatial externalities and the evolution of residential land use patterns. *Journal of Economic Geography*. **2** (1). 31–54.

Kirkpatrick, J. B. 1983. An iterative method for establishing priorities for the selection of nature reserves – an example from Tasmania. *Biological Conservation*. **25** (2). 127–134.

Kok, K. and Veldkamp, A. 2001. Evaluating impact of spatial scales on land use pattern analysis in Central America. *Agriculture Ecosystems and Environment*. **85**. 205–221.

Maddala, G. S. 1983. *Limited-dependent and Qualitative Variables in Econometrics*. Cambridge, New York: Cambridge University Press.

McCoy, E. D. and Mushinsky, H. R. 1994. Effects of fragmentation on the richness of vertebrates in the Florida scrub habitat. *Ecology*. **75** (2). 446–457.

Moffitt, R. 2001. Policy Interventions, low-level equilibria, social interactions. In S. N. Durlauf. and H. P. Young. (eds.). 45–82.

Munshi, K. 2003. Networks in the modern economy: Mexican migrants in the U.S. labour market. *Quarterly Journal of Economics*. **118** (2). 549–599.

Nelson, G. C. and Hellerstein, D. 1997. De roads cause deforestation? Using satellite images in econometric analysis of land use. *American Journal of Agricultural Economics*. **79**. 80–88.

Panoyotou, T. and Sungsuwan, S. 1989. An econometric study of the causes of tropical deforestation: the case of northeast Thailand. *Development Paper 284*. Harvard Institute for International Development.

Parkhurst, G. M., Shogren, J. F., Bastian, C., Kivi, P., Donner, J. and Smith, R. B. W. 2002. Agglomeration bonus: an incentive mechanism to reunite fragmented habitat for biodiversity conservation. *Ecological Economics*. **41** (2). 305–328.

Pfaff, A. S. P. 1999. What drives deforestation in the Brazilian Amazon? Evidence from satellite and socioeconomic data. *Journal of Environmental Economics and Management*. **37** (1). 26–43.

Pfaff, A. S. P. and Sanchez-Azofeifa, G. A. 2004. Deforestation pressure and biological reserve planning: a conceptual approach and an illusative application for Costa Rica. *Resource and Energy Economics*. **26** (2). 237–254.

Polasky, S., Camm, J. D., Solow, A. R., Csuti, B., White, D. and Ding, R. G. 2000. Choosing reserve networks with incomplete species information. *Biological Conservation*. **94** (1). 1–10.

Rudel, T. K. 1989. Population, development, and tropical deforestation – a cross-national-study. *Rural Sociology*. **54** (3). 327–338.

Schelling, T. C. 1971. Dynamic models of segregation. *Journal of Mathematical Sociology*. **1** (2). 143–186.

Scull, P. R. and Harman, J. R. 2004. Forest distribution and site quality in southern Lower Michigan, USA. *Journal of Biogeography*. **31** (9). 1503–1514.

Serneels, S. and Lambin, E. F. 2001. Proximate causes of land-use change in Narok District, Kenya: a spatial statistical model. *Agriculture Ecosystems and Environment*. **85**. 65–81.

Stavins, R. N. and Jaffe, A. B. 1990. Unintended impacts of public-investments on private decisions – the depletion of forested wetlands. *American Economic Review*. **80** (3). 337–352.

Tubbs, C. and Blackwood, J. 1971. Ecological evaluation of land for planning purposes. *Biological Conservation*. (3). 169–172.

Turner, M. A. 2005. Landscape preferences and patterns of residential development. *Journal of Urban Economics*. **57** (1). 19–54.

Twedt, D. J. and Loesch, C. R. 1999. Forest area and distribution in the Mississippi alluvial valley: implications for breeding bird conservation. *Journal of Biogeography*. **26** (6). 1215–1224.

Walsh, S. J., Crawford, T. W., Welsh, W. F. and Crews-Meyer, K. A. 2001. A multiscale analysis of LULC and NDVI variation in Nang Rong district, northeast Thailand. *Agriculture Ecosystems and Environment*. **85** (1–3). 47–64.

Table 3.A1 *Regression results: probit estimates and second-stage estimates from 2SPLS*

	Dependent variable deforestation decisions 86–97			
	Region 1		Region 2	
	Probit	2SPLS	Probit	2SPLS
NDE (ρ)	6.498	0.401	3.065	3.314
	(1.378)	(12.28)	(0.508)	(1.312)
GLZ	−0.614	−0.513	0.129	0.110
	(0.285)	(0.280)	(0.178)	(0.199)
BLZ	0.275	0.198	−3.509	−3.513
	(0.270)	(0.260)	(38.05)	(62.50)
DSJ	0.008	0.007	0.013	0.013
	(0.009)	(0.012)	(0.008)	(0.008)
DLI	−0.016	−0.016	−0.012	−0.010
	(0.006)	(0.007)	(0.005)	(0.005)
DCA	−0.010	−0.007	−0.016	−0.015
	(0.006)	(0.012)	(0.009)	(0.009)
DLR	0.016	0.036	−0.019	−0.007
	(0.046)	(0.047)	(0.036)	(0.035)
DNR	0.050	0.036	−0.010	−0.007
	(0.043)	(0.042)	(0.021)	(0.021)
DTS	−0.016	−0.011	0.008	0.007
	(0.014)	(0.013)	(0.008)	(0.008)
DTH	−0.023	−0.031	−0.001	−0.004
	(0.024)	(0.023)	(0.007)	(0.007)
PTC	−2.806	−2.582	−1.613	−1.592
	(0.795)	(0.762)	(0.339)	(0.336)
DMT	0.016	0.029	0.004	0.003
	(0.021)	(0.034)	(0.006)	(0.006)
SDA	−0.053	−0.053	−0.006	−0.005
	(0.014)	(0.014)	(0.017)	(0.019)
LNR	0.007	0.014	−0.002	−0.002
	(0.006)	(0.012)	(0.005)	(0.004)
LLR	−0.003	−0.002	−0.001	−0.001
	(0.004)	(0.004)	(0.002)	(0.002)
NSM	−0.196	−0.120	−0.049	−0.055
	(0.119)	(0.125)	(0.056)	(0.056)
NMT	0.193	0.028	−0.007	0.011
	(0.216)	(0.278)	(0.192)	(0.196)
NHS	−0.040	−0.070	0.074	0.073
	(0.055)	(0.076)	(0.065)	(0.063)
CLP	1.314	1.050	−0.160	−0.416
	(1.483)	(1.468)	(1.168)	(1.150)
CLP2	−1.975	−1.335	0.634	0.751
	(1.384)	(1.472)	(1.127)	(1.092)
Constant	1.738	1.386	0.892	0.673
	(1.052)	(1.169)	(1.131)	(1.135)

In parenthesis standard errors

Table 3.A2 *Regression results: first stage*

Dependent variable is neighbours' deforestation 86–97		
	Region 1	Region 2
Instrument		
Neighbours' sloes (Π_1)	−0.0019	−0.0234
	(0.0007)	(0.0017)
Controls for efficiency (Π_2)		
GLZ	−0.0057	−0.0764
	(0.0074)	(0.0102)
BLZ	0.0010	0.0151
	(0.0078)	(0.0328)
DSJ	−0.0006	−0.0013
	(0.0002)	(0.0004)
DLI	0.0000	0.0003
	(0.0002)	(0.0003)
DCA	0.0006	0.0010
	(0.0001)	(0.0005)
DLR	0.0011	0.0029
	(0.0012)	(0.0021)
DNR	−0.0004	0.0036
	(0.0010)	(0.0012)
DTS	−0.0002	−0.0021
	(0.0004)	(0.0004)
DTH	0.0004	−0.0009
	(0.0005)	(0.0004)
PTC	−0.0139	−0.0363
	(0.0090)	(0.0104)
DMT	0.0021	−0.0019
	(0.0005)	(0.0003)
SDA	0.0000	0.0000
	(0.0004)	(0.0010)
LNR	0.0008	−0.0009
	(0.0001)	(0.0003)
LLR	0.0000	0.0011
	(0.0001)	(0.0001)
NSM	0.0041	−0.0011
	(0.0027)	(0.0031)
NMT	−0.0150	−0.0513
	(0.0069)	(0.0112)
NHS	−0.0019	0.0006
	(0.0012)	(0.0041)
CLP	−0.0011	0.0250
	(0.0459)	(0.0645)
CLP2	0.0315	0.0500
	(0.0430)	(0.0650)
Constant	−0.0066	0.3140
	(0.0329)	(0.0636)

In parenthesis standard errors