

# MIT Joint Program on the Science and Policy of Global Change



## What Drives Deforestation in the Brazilian Amazon?

Alexander S.P. Pfaff

Report No. 16  
December 1996

The MIT Joint Program on the Science and Policy of Global Change is an organization for research, independent policy analysis, and public education in global environmental change. It seeks to provide leadership in understanding scientific, economic, and ecological aspects of this difficult issue, and combining them into policy assessments that serve the needs of ongoing national and international discussions. To this end, the Program brings together an interdisciplinary group from two established research centers at MIT: the Center for Global Change Science (CGCS) and the Center for Energy and Environmental Policy Research (CEEPR). These two centers bridge many key areas of the needed intellectual work, and additional essential areas are covered by other MIT departments, by collaboration with the Ecosystems Center of the Marine Biology Laboratory (MBL) at Woods Hole, and by short- and long-term visitors to the Program. The Program involves sponsorship and active participation by industry, government, and non-profit organizations.

To inform processes of policy development and implementation, climate change research needs to focus on improving the prediction of those variables that are most relevant to economic, social, and environmental effects. In turn, the greenhouse gas and atmospheric aerosol assumptions underlying climate analysis need to be related to the economic, technological, and political forces that drive emissions, and to the results of international agreements and mitigation. Further, assessments of possible societal and ecosystem impacts, and analysis of mitigation strategies, need to be based on realistic evaluation of the uncertainties of climate science.

This report is one of a series intended to communicate research results and improve public understanding of climate issues, thereby contributing to informed debate about the climate issue, the uncertainties, and the economic and social implications of policy alternatives.

Henry D. Jacoby and Ronald G. Prinn,  
*Program Co-Directors*

For more information, contact the Program office:

MIT Joint Program on the Science and Policy of Global Change

Postal Address: 77 Massachusetts Avenue  
MIT E40-271  
Cambridge, MA 02139-4307 (USA)

Location: One Amherst Street, Cambridge  
Building E40, Room 271  
Massachusetts Institute of Technology

Access: Telephone: (617) 253-7492  
Fax: (617) 253-9845  
E-mail: [globalchange@mit.edu](mailto:globalchange@mit.edu)  
Web site: <http://web.mit.edu/globalchange/www>

# WHAT DRIVES DEFORESTATION IN THE BRAZILIAN AMAZON?

## Evidence from Satellite and Socioeconomic Data

Alexander S.P. Pfaff\*  
*Columbia University*

Submission as a working paper to the  
M.I.T. Joint Program on the Science and Policy of Global Change  
October 5, 1996

### ABSTRACT<sup>1</sup>

This paper analyzes the determinants of deforestation in the Brazilian Amazon. From a model of optimal land use, it derives and then estimates a deforestation equation on county-level data for the period 1978 to 1988. The data include a deforestation measure from satellite images, which is a great advance in that it allows improved within-country analysis. Evidence exists that: increased road density in a county leads to more deforestation in that county and in neighboring counties; government-subsidized development projects increase deforestation; greater distance from markets south of the Amazon leads to less deforestation; and better soil quality leads to more deforestation. The results for government provision of credit are mixed across specifications. The population density, although the primary explanatory variable in most previous empirical work, does *not* have a significant effect when all the variables motivated within the model are included. However, a quadratic specification yields a more robust population result: the first few people entering an empty county have significantly more impact than the same number of people added to a densely populated county. This result suggests the importance of the spatial distribution of population.

---

\* Address correspondence to: Alex Pfaff Columbia University, Department of Economics, 420 W. 118th St., Room 1018 IAB, New York, NY 10027. Phone: (212) 854-4190 (-8059, fax). Email: ap196@columbia.edu.

<sup>1</sup> I want to thank in particular Richard Schmalensee, Robert Stavins, Robert Solow and Jon Gruber for continuing support of and advice on this project. I also want to thank Jim Poterba, Kaivan Munshi, Amy Ando, Maureen Cropper, Richard Eckaus, David Genesove, Donald Marron, Bill Miracky, Chris Sanchirico, Robert Schneider, Matt Slaughter, and Jeff Vincent for helpful discussions, and seminar participants at MIT, Harvard, Columbia, and the NBER for useful comments. I gratefully acknowledge funding from: an NSF graduate fellowship, the Schultz Fund, the University of Florida/FLAS, the Social Science Research Council's International Predissertation Fellowship Program, the M.I.T. World Economy Lab, and the M.I.T. Joint Program on the Science and Policy of Global Change. On another front, after much searching I discovered that David Skole at the University of New Hampshire had collected state of the art satellite vegetative land cover data, and that Eustaquio Reis at the Institute for Applied Economic Research in Rio de Janeiro (IPEA/Rio) had collected a fantastic databank for the Amazon. Both during my extended visit at IPEA/Rio and during my work at MIT, their efforts and those of their colleagues have helped me greatly in using these datasets. Among their colleagues, I want to thank in particular Walter Chomentowski at UNH, and Maria Jose Silveira Pessoa and the computational staff at IPEA/Rio. In making their data available to me and in their ongoing contributions, these two groups have been absolutely crucial in making this paper possible.

## 1. INTRODUCTION

The depletion of rainforests has demanded the attention of policymakers during the 1980's and 90's. Initial concern about potential extinction of species has been joined by alarm about possible future global warming caused by atmospheric accumulation of "greenhouse gases" such as the carbon dioxide released by deforestation. Policymakers must understand what drives deforestation if they are to respond appropriately to such concerns. However, important questions remain about why rainforests are being cut down and whether public policies can affect the rate at which deforestation takes place. It is these questions that this paper seeks to address. It is important to note that despite the attention given to Amazon rainforest depletion, over eighty percent of this rainforest remains. Thus these are not merely historical questions, but questions whose answers should inform policies which will significantly affect global rainforest survival.

Much has been written about these questions, but economic understanding remains rudimentary at best. Existing empirical research has focused heavily on population. This paper advances beyond previous empirical analyses in large part because of the innovative merging of state-of-the-art satellite data on deforestation with an outstanding, county-level dataset for the Amazon. In addition, the empirical work is motivated by a systematic presentation of a relevant economic framework which encompasses not only population but also other factors such as roads, credit, and soil quality. The equation for estimation is derived from this optimal land use model.

A number of significant effects are found: increased road density in a county leads to more deforestation in that county and in neighboring counties; government development projects lead to greater deforestation; confirming the importance of transport costs, greater distance from markets south of the Amazon is associated with less deforestation; and confirming economic intuition, greater soil quality is associated with more deforestation. Also of policy interest, the evidence on government provision of subsidized credit is mixed across specifications.

A result which has characterized much of the existing empirical literature is that population is the primary explanatory variable for deforestation. Many of those results, however, arose in regressions in which the population density was one of very few or even the only explanatory variable. In that form, those results are confirmed here: when population is the sole explanatory variable, it has a significant positive effect on deforestation (although little explanatory power). However, when other variables motivated within the land use model are included, the population density is no longer significant. This result suggests that the earlier results for population may be significantly affected by the omission of important variables.

In principle, though, a large, exogenous increase in population density should affect output demand and input supply conditions and thus affect land clearing. Thus while in this result the population density was dominated by other explanatory factors, population deserved further attention. Along these lines, a quadratic population specification is more robust to the inclusion of other variables. The first few people entering an empty county have significantly more impact on forest clearing than the same number of people added to an already densely populated county. This additional result suggests the importance of the spatial distribution of population.

The paper is organized as follows: Section 2 provides background. Section 3 reviews the existing literature. Section 4 presents a model of optimal land use and derives first a plot-level land allocation

decision rule and then a county-level deforestation equation for estimation. Section 5 describes the data and presents priors and issues for empirical implementation. Section 6 presents the results. Finally, Section 7 concludes and presents potential extensions.

## 2. BACKGROUND

### 2.1 Geography

The Legal Amazon<sup>2</sup> is an immense area, most of which was covered by forest at one time,<sup>3</sup> and most of which remains forested today.<sup>4</sup> Bordering a number of countries in the northwest corner of Brazil (see Figure 1), it contains 5 million of Brazil's total area of 8.5 million square kilometers; the latter area is larger than the continental United States. Rivers permeate the region (see again Figure 1), including the Amazon River, which traverses the region from west to east.<sup>5</sup>

### 2.2 History

Since at least the 1960's, occupation and use of the Amazon region has been at some level a national policy goal. The military government officially in power from the 1960's into the 1980's promoted occupation of the region.<sup>6</sup> Many felt that such empty land was an ideal 'release valve' for the pressures arising from a growing population.<sup>7</sup> Finally, many also felt that the region offered boundless resources and great hopes for progress, and those in power apparently shared those visions of progress and/or were happy to make use of such hopes.<sup>8</sup>

---

<sup>2</sup> The Legal Amazon is made up of all of the North region (the states Acre, Amapa, Amazonas, Para, Rondonia, Roraima, and Tocantins) plus parts of the states of Maranhao, Mato Grosso, and Goias. The southern edge is the 16th parallel, and the eastern edge is the 44th meridian.

<sup>3</sup> Although it is difficult to determine what the truly "original" vegetation was in any location, in particular given any history of human habitation, a best guess (see Skole(1993)) is that all of the area was forested except for about one sixth of the region which was covered with a scrubby vegetation called *cerrado* and about three percent of the region which is *varzea*, or seasonally flooded land near rivers (see, for example, Goulding(1993)).

This figure of one sixth may appear to contradict Table 1, which gives a mean of 0.30 for the fraction of Amazon county area covered by *cerrado*. However, that thirty percent figure is an unweighted average of the county *cerrado* densities. Only an area-weighted average would reproduce the one sixth figure that is correct for the region. The fact that not all land was originally forested affects the definition of the dependent variable below.

<sup>4</sup> Exactly how much deforestation has taken place is disputed (see, for example, Skole and Tucker (1993), Fearnside *et al.* (1990b), and INPE (1992)). However, the region is at most ten to fifteen percent deforested.

<sup>5</sup> This massive river is the confluence of runoff from higher areas to the south, west, and north of its basin.

<sup>6</sup> Hecht and Cockburn (1990) provide the following quotation from General Castello Branco, in 1964: "Amazonian occupation will proceed as though we are waging a strategically conducted war". As part of the 'military philosophy of and strategy for regional development' that they cite, motivations may have included the desire to discourage both incursions from bordering countries and the formation of domestic guerrilla opposition.

<sup>7</sup> For example, Hecht and Cockburn (1990) provide the famous citation from General Emilio Medici, who offered to provide "a land without men for men without land". They also quote General Golbery de Couto de Silva as referring to "the vast hinterlands waiting and hoping to be aroused to life and to fulfill their historic destiny".

<sup>8</sup> For example, Hecht and Cockburn (1990) cite the ideology of modernization, as in the phrase "*Isto é um país que vai pra frente*" (which might be translated: "This is a country that moves (or, is moving) forward"). They also quote President Getulio Vargas, from 1940: " -- the highest task of civilizing man: to conquer and dominate the valleys of the great equatorial torrents, transforming their blind force -- into disciplined energy".

To open the region, roads were built, accompanied by colonization and titling projects. Subsidized credit was offered, and income taxes were forgiven if the funds went to approved development projects. Dams were constructed, and a free trade zone was created in Manaus.<sup>9</sup>

### 2.3 A Few Statistics

The actions taken appear to have stimulated occupation of the Amazon (although correlation may not indicate causality). Table 1 presents a few county-level statistics which document significant changes.<sup>10</sup> Kilometers of roads per county grew significantly over the decade 1975-1985 (see also Figure 2). Total population more than doubled from 1970 to 1991, and urban population more than tripled. Finally, cleared forest area increased significantly.<sup>11</sup>

## 3. REVIEW OF EXISTING LITERATURE

While many have previously considered either tropical deforestation in general or deforestation in the Brazilian Amazon, little empirical work of the sort presented in this paper has been done. Here I briefly review cross-country analyses, within-country analyses of other countries, and within-country analyses of Brazil, and then indicate the ways in which the empirical work in these existing analyses differs from the empirical work below.<sup>12</sup>

### 3.1 Cross-Country Analyses

A number of attempts have been made to correlate factors of interest with national measures of deforestation. These include: Lugo *et al.* (1981) using population, energy use, and terrain variables; Allen and Barnes (1985) and Palo *et al.* (1987), using population, land use and output measures; Rudel (1989); Cropper & Griffiths (1994), using population as well as income levels; and Deacon (1994), using population and measures of government weakness or instability. A number of

---

<sup>9</sup> It should be noted that the push into the Amazon region also appears to have involved factors other than public actions. For instance, droughts in the northeast made that region sufficiently inhospitable to cause significant migration into the Amazon (an inhospitable environment). Also, a shift into more capital-intensive, mechanized agriculture in the South is alleged to have created a significant pool of landless unemployed, to whom migration to the Amazon may have looked relatively promising. Further mention of this phenomenon is found below.

<sup>10</sup> All data are for the 1970 *município*, or county, structure of 316 counties. See section 5 for explanation.

<sup>11</sup> The means for fraction cleared in Table 1 (e.g., 0.16 in 1988) are unweighted averages of the fractions cleared for each of the counties in the region. Only a weighted average using county areas as weights would reproduce the value obtained by dividing cleared area for the whole region by the total area of the whole region.

<sup>12</sup> This section addresses only related empirical work. At least three other literatures are of interest although not directly comparable with this paper. First is theoretical discussion of deforestation, for example Ehui & Hertel (1989), Vincent (1990), Nerlove (1991), Hyde & Sedjo (1992), Jones & O'Neill (1994, 1995, and others), Deacon (1995), and Hyde, Amacher and Magrath (1996). Second is a group of analyses of relevant Brazilian government policies, such as Hecht (1985), Browder (1985), Repetto (1988), Binswanger (1989), and Mahar (1992), which contain much policy information but little formal analysis of the determinants of deforestation. Third is field studies, of the Brazilian Amazon and/or deforestation, which address a wide range of relevant issues, but which differ in focus from this work. These include Hecht (1982), Fearnside (1986, 1990a), Smith (1982), Uhl *et al.* (1991), Bunker (1985), Moran (1981, 1990), Branford & Glock (1985), Schmink & Wood (1992), Jones *et al.* (1992), Vincent & Binckley (1992), Amacher, Hyde & Joshee (1992), Schneider *et al.* (1993) and Scherr (1995).

results are of interest, such as Cropper & Griffiths' 'stage of development' interpretation of the significance of income levels, and Deacon's government variables. The dominant result, though, is that population (or its density) is the most significant factor in explaining deforestation. This is partially explained by the fact that such analyses often use few explanatory variables (and, in the extreme, population alone). While cross-country results should be applied to a given country only with caution, the studies that make up this sizable part of the existing empirical literature provide an important point of reference for this paper.

### **3.2 Within-Country Analyses Of Other Countries**

Sungsuwan (1985) and Sungsuwan & Panoyotou (1989) find that deforestation in Thailand is driven by population density, wood price, income, and distance to Bangkok. Southgate *et al.* (1991) also focus in part on population, in Ecuador's Amazon region: they first explain population with variables expected to affect "the prospect of capturing agricultural rents", and then explain deforestation with population and other factors. Harrison (1991), studying Costa Rica, questions the simplest focus on population, allowing different effects in different regions, and questions whether population is a cause or a "shared symptom". Finally, Kummer (1991) is one of few to find only a small role for population growth in deforestation.

### **3.3 Within-Country Analyses Of Brazil**

Almeida (1992) provides a great deal of information at the level of the entire Amazon region. However, the book tries mainly to answer the question "Was agricultural colonization of the Amazon worth its cost, and the best option?". Thus it provides more measurement and aggregation than testing of the importance of given factors in deforestation. Reis & Margulis (1990) and Reis & Guzman (1992) present econometric analyses of deforestation in the Brazilian Amazon. They find population density, road density, and crop area to be important determinants of their deforestation measure.

The obvious difference between the analyses of other countries and this paper is that they do not address Brazil, which is home to such an enormous amount of tropical rainforest.<sup>13</sup> Reis' work leads into this paper, including the use here of some of his dataset. However, this paper advances beyond these works in two principal ways: first, it motivates the empirical work by systematically laying out the suggested economic framework; second, it makes use of the innovative merging of state-of-the-art satellite data (containing multiple observations of deforestation over time for all counties in the region) with an outstanding county-level dataset for the Amazon (including new data on credit and development project policies).<sup>14</sup>

---

<sup>13</sup> Skole & Tucker (1993)'s figures indicate that Brazil contains thirty percent of the world's forested area.

<sup>14</sup> Since this paper was first circulated, two other papers have appeared using similar methodology with respect to these two points. These are Chomitz & Gray (1996), on Belize, and Cropper & Griffiths (1996), on Thailand.

Note that at county-level, a satellite measure of deforestation may be no better than a census measure. However, satellite data may allow for additional geographic precision in the analyses, as the satellite measurements can be for units much smaller than, for instance, an average Amazon county. Chomitz & Gray's analyses make use of this additional precision. Satellites may also provide additional observations over time at lower cost.

## 4. THE CONCEPTUAL FRAMEWORK

Underlying the empirical analyses below is an optimal land use framework. Land is allocated between alternative uses in order to obtain the greatest expected total return. Dynamic issues fit naturally into this framework, as expected total return includes not only all immediate returns but also expected future returns. For example, the likelihood of obtaining title through current land clearing would be expected to affect the current land use decision.<sup>15</sup> However, here I present the optimal land use framework in a simple, static fashion. I do so in order to focus on the elements of this problem which can be incorporated into the empirical work presented below. Data requirements for incorporating dynamic issues are discussed in the Data section below.

After setting out the basic framework, I derive the deforestation equation for estimation in two steps: first, I derive a plot-level, optimal, land-allocation decision rule as a function of variables which are in principle observable; and second, I adapt the derivation to generate a county-level, optimal, land-allocation decision rule which implies a deforestation equation which can be estimated with the existing, county-level data.

### 4.1 An Optimal Land Use Model

Here, I assume that land will be allocated to maximize immediate returns. Further, for any given land use, the immediate return will be defined as the current profits obtained:

$$B_{ijt}^l = P_{ijt} \$ Q_{ijt}^l(I_{ijt}, Z_{ijt}) - R_{ijt} \$ I_{ijt}, \quad (1)$$

where:

$l$  = a given land use

$i$  = county

$j$  = a plot of land within county  $i$

$t$  = the year

$B_{ijt}^l$  = profits obtained from land use  $l$  on plot  $ij$  in year  $t$

$P_{ijt}$  = the vector of prices for outputs  $Q$  (all possible outputs from any land use);  
these are fixed, 'farm gate' prices, differing by time and by plot but not by  $Q$

$Q_{ijt}^l$  = the vector of all outputs produced (potentially including shelter);  
the production functions  $Q^l(!)$  are different for different land uses,  
and output is a function not only of chosen inputs  $I$  but also of fixed factors  $Z$

$I_{ijt}$  = the vector of inputs to production for land use  $l$  on the plot;  
the plot size is small and fixed, so input choice does not include amount of land

$R_{ijt}$  = the vector of prices for inputs  $I$  (all inputs to all possible outputs);  
these are fixed, 'farm gate' prices, differing by time and by plot but not by  $I$

While (1) abstracts away many of the variables often linked with deforestation, it is temporarily useful in providing a simple view of the structure of the optimal land decision. Given the definitions

---

<sup>15</sup> Alston, Libecap and Schneider (1995) examine property rights in the Amazon. Interestingly, their preliminary findings included no significant effect of land title on clearing (personal communication, 2/95).



above, choosing the greatest immediate return implies the land use decision problem:

$$\max_{l, l'} B^l_{ijt} \quad (2)$$

In a general model, many land uses might be differentiated. In this paper, only two land use states are possible: cleared and uncleared.<sup>16</sup> Cleared land is the link to deforestation: if land was originally forested, cleared land is deforested land. To simplify notation below, the following definition of  $V^l_{ijt}$ , the maximum value of land in land use  $l$ , is useful:

$$V^l_{ijt} = \max_{l' | l} B^l_{ijt}, \text{ i.e. } V^l_{ijt} \text{ includes optimal input choice given } l. \quad (3)$$

This definition, and the fact that there are only two land use states, imply the decision rule:

$$\text{Choose } l_t = \text{cleared iff: } V^{\text{cleared}}_{ijt} > V^{\text{uncleared}}_{ijt} \quad (4a)$$

$$\text{or, Choose } l_t = \text{cleared iff: } V^{\text{cleared}}_{ijt} - V^{\text{uncleared}}_{ijt} > 0 \quad (4b)$$

A land use decision rule of this sort will lead to the equation to be estimated. However, the components of this expression need to be rewritten for their empirical implications to be clear.

#### 4.2 Potentially Observable Variables and A Plot-Level Decision Rule

The model above lacks ties to many variables which come to mind as possible determinants of deforestation. Here, variables are suggested in relation to the expressions above:

$$P_{ijt} \text{ are functions of: } \quad p_{it}, p_t; n_{it}; h_{ijt}, m_{ijt}; d_{it}, c_{it} \quad (5a)$$

$$R_{ijt} \text{ are functions of:}^{17} \quad r_{it}, r_t; n_{it}; h_{ijt}, m_{ijt}; c_{it}; v_{ijt} \quad (5b)$$

$$\text{Fixed factors } Z_{ijt} \text{ include: } \quad q_{ijt} \quad (5c)$$

$p_{it}$  = county-level output prices

$p_t$  = national-level output prices

$n_{it}$  = county population

$h_{ijt}$  = plot access to roads and rivers

$m_{ijt}$  = plot distances from important markets

$d_{it}$  = county development projects

$c_{it}$  = county (subsidized) credit infrastructure

$r_{it}$  = county-level input prices

$r_t$  = national-level input prices

$v_{ijt}$  = plot vegetation type

$q_{ijt}$  = plot soil quality

<sup>16</sup> Further differentiation of land uses would be not only more general, but also of empirical interest. For instance, the effects of independent variables of interest may well vary across land uses which involve cleared land.

<sup>17</sup> The plot-level, farmgate prices  $P_{ijt}$  and  $R_{ijt}$  are not directly observed.

These relations merit some explanation. Plot-level prices for both outputs and inputs should be functions of whatever the relevant market prices may be (county, national, or world) plus transport costs, which should be affected by distances as well as by access to roads and rivers. Population may affect output prices as a demand shifter, and may affect input prices as labor supply. Development projects may serve to stimulate demand as well, while credit may both lower rental rates and release suppressed demand for outputs. Finally, vegetation density may indicate clearing costs of land, while soil quality should provide increased productivity. Note that variables may play multiple roles even with respect to a single land use. If these imply opposing effects, the direction of the variable's net effect will be ambiguous.

Working from (4b) and the relations (5a-c), and collapsing to a more reduced form expression (again, a single variable may play several roles with respect to a single land use), the optimal land use decision rule can be restated in the following form:<sup>18</sup>

$$\begin{aligned} & \text{Choose } l_t = \text{cleared iff:} \\ & V_{ijt}^{\text{cleared}}(\mathbf{p}_{it}, \mathbf{p}_t; \mathbf{r}_{it}, \mathbf{r}_t; n_{it}; \mathbf{h}_{ijt}, \mathbf{m}_{ijt}; d_{it}, c_{it}; v_{ijt}; q_{ijt}) > \\ & V_{ijt}^{\text{uncleared}}(\mathbf{p}_{it}, \mathbf{p}_t; \mathbf{r}_{it}, \mathbf{r}_t; n_{it}; \mathbf{h}_{ijt}, \mathbf{m}_{ijt}; d_{it}, c_{it}; v_{ijt}; q_{ijt}) \end{aligned} \quad (6)$$

The effects of these variables on the absolute levels of the returns to cleared and uncleared land uses will not be considered. Rather, the variables are distinguished only by whether a change in a given variable raises or lowers the difference between the gains from cleared land uses and those from uncleared land uses. This motivates the following decision rule:

$$\begin{aligned} & \text{Choose } l_t = \text{cleared iff:} \\ & D_{ijt}^{\text{cleared}} > 0, \text{ where } D_{ijt}^{\text{cleared}} = V_{ijt}^{\text{cleared}} - V_{ijt}^{\text{uncleared}} \end{aligned} \quad (7)$$

Recall, it is a land use decision rule such as (4) or (7) which leads to the equation to be estimated below. However, the omission of factors from years other than  $t$  must be explained before (7) can be used as a basis for estimation. Perhaps the most significant omission in (7) from the determinants of current land use is past land use. Forest regrowth is not 'instantaneous' (on an annual timescale). Thus if a plot of land is cleared in year  $t-1$ , it may be impossible for that plot to be uncleared again by year  $t$ , no matter what parameter values obtain. Empirically, though, the clearing observations are separated by 10 years. Thus I will assume here, as in all such static models, that forest regrowth is sufficiently fast that uncleared land is a viable option even for plots which were cleared in the previous observation.<sup>19</sup>

Other obvious omissions from (7) are of factors affecting future returns. While issues such as property rights and forest regrowth rates are left for future treatment of dynamic issues, as

<sup>18</sup> From this point on, the variables within parentheses may be suppressed in the notation.

<sup>19</sup> The possible error here is in assuming that a plot which is currently cleared was worth clearing because of the current state of the independent variables when in fact it was not, but is instead cleared because of what was optimal in the past. Note that if in the process of frontier development "the right-hand side is always rising", i.e. if pressures to clear land generally grow stronger over time, in practice this situation may only rarely arise.

However, it must be conceded that without more data over time, it is hard to address dynamics completely.

Finally, it should be noted that the forest which grows back may differ significantly from what was there.

discussed below, one interpretation of (7) does include expected future returns in a particular way. If expectations of future parameter values are formed on the basis of current (year  $t$ ) values, then the variables in (7) may, in an additional role, imply consideration of the future.

Thus I arrive at a plot-level decision rule for optimal land allocation between potential uses. Expression (7) motivates an estimation of the effects of the variables listed above on land use choice. Such an estimation would use plot-level data: first, a discrete dependent variable indicating whether a plot  $ij$  is cleared or uncleared in year  $t$ ; and second, plot-level independent variables, such as distance from a plot to the nearest paved road. However, I am unable to estimate such an equation, for lack of plot level observations.<sup>20</sup>

### 4.3 County-Level Observations and A County-Level Equation

As no variables are observed at the plot level, expression (7) must be adapted to the existing observations, which are at the *municipio*, or county level. Working from (7), one possible adaptation of the model to this data limitation would be to assume that  $D_{ijt}^{cleared} = D_{it}^{cleared}$  for all plots  $j$  in each county  $i$ .<sup>21</sup> However, the model would then operate solely at a county level, as all plots within a county would be identical. Thus it would predict that at some threshold level of the appropriately weighted combination of factors affecting land use choice, a whole county would shift in land use from uncleared to cleared, or vice versa. That would be a clear problem with the model, as in fact, most counties contain both cleared and uncleared plots.

Thus I assume, along the lines of Stavins & Jaffe (1990), the existence of within-county, pay-off-relevant, unobserved, plot-level characteristics. More specifically, I define  $_{ijt}$ , distributed across plots  $j$  within county  $i$  in year  $t$ . For a plot,  $_{ijt}$  is the difference between additional maximum profits (added to  $V_{it}^{uncleared}$ ) attainable from plot  $ij$  in year  $t$  when uncleared and additional maximum profits (added to  $V_{it}^{cleared}$ ) attainable from that plot when cleared. Thus:

$$D_{ijt}^{cleared}(!) = D_{it}^{cleared}(!) - ,_{ijt} \quad (9)$$

From (7), given (9), if for a given plot a net positive influence of the observed determinants which may affect  $D_{it}^{cleared}(!)$ , i.e. may favor cleared land use for the county, overcome a net influence of the unobserved determinants which may affect  $_{ijt}$ , i.e. may favor uncleared use for the plot, then plot  $ij$  is allocated for cleared use in year  $t$ .<sup>22</sup> This implies:

$$\begin{aligned} & \text{Choose } l_t = \text{cleared iff:} \\ & D_{it}^{cleared}(\mathbf{p}_{it}, \mathbf{p}_t; \mathbf{r}_{it}, \mathbf{r}_t; n_{it}; \mathbf{h}_{it}, \mathbf{m}_{it}; d_{it}, c_{it}; v_{it}; q_{it}) > ,_{ijt} \end{aligned} \quad (10)$$

<sup>20</sup> Most limiting is the availability of plot-level independent variables. The clearing observations in the satellite data are more easily obtained for greater geographic disaggregation.

<sup>21</sup> Where, as in the definition of  $D_{ijt}$  above:

$$D_{it}^{cleared} = V_{it}^{cleared}(\mathbf{p}_{it}, \mathbf{p}_t; \mathbf{r}_{it}, \mathbf{r}_t; n_{it}; \mathbf{h}_{it}, \mathbf{m}_{it}; d_{it}, c_{it}; v_{it}; q_{it}) - V_{it}^{uncleared}(\mathbf{p}_{it}, \mathbf{p}_t; \mathbf{r}_{it}, \mathbf{r}_t; n_{it}; \mathbf{h}_{it}, \mathbf{m}_{it}; d_{it}, c_{it}; v_{it}; q_{it}). \quad (8)$$

<sup>22</sup> For more discussion of this approach, see, for example, Green(1990), chapter 20. Note that the same within-county distribution, i.e. the same internal heterogeneity, is assumed for each observation (each county in each period).

Whatever the distribution of  $y_{ijt}$  within county  $i$  in year  $t$ , it follows that:

$$\%Cleared_{it} = F[D^{cleared}_{it}(!)], \quad (11)$$

where  $F[]$  is the cumulative distribution function of  $y_{ijt}$ .<sup>23</sup> It then follows that:

$$F^{-1}[\%Cleared_{it}] = D^{cleared}_{it}(!). \quad (12)$$

If  $y_{ijt}$  is distributed logistically, and  $\%Cleared_{it}$  is rewritten as  $y_{it}$ , the following holds:

$$\ln(y_{it} / (1-y_{it})) = D^{cleared}_{it}(!). \quad (13)$$

## 5. DATA, PRIORS, AND IMPLEMENTATION ISSUES

Here I present information relevant for moving from the conceptual framework to the results actually obtained. First, I present the raw data and the construction of the variables actually used in estimation. This motivates the specific versions of equation (13) which will be the basic equations for estimation. Then, for this equation, I present the priors for the estimation. Finally, I discuss potential concerns about multicollinearity and endogeneity. Regressions addressing these concerns are then simply included in the discussion of all results below.

### 5.1 Data And Variables

The data cover the entire Amazon region, at the *município*, or county, level. All variables are for the 1970 counties,<sup>24</sup> including those created with the 'neighbor matrix'<sup>25</sup>.

#### 5.1.1 Land Use and Land Characteristics Data

The land data include the satellite data, which come from specialists in state-of-the-art satellite measures at the University of New Hampshire (UNH).<sup>26</sup> The original units of observation by the satellite, much smaller than counties, are aggregated to county level. The data exist for 1975, 1978, and 1988. For those years, they give the breakdown of county area into three categories: standing forest, cleared forest, and "never-forested", i.e. *cerrado* (a scrubby vegetation). *Cerrado* is the main vegetation other than forest in the Brazilian Amazon, covering about one sixth of the total area. The *cerrado* areas are taken as constant. Thus the cleared forest measure omits any clearing which takes place in *cerrado* areas.

<sup>23</sup> Thus, for example, the vector of factors affecting relative land use returns may sweep through the within-county distribution over time. Eventually, additional increases have little effect, as all the land has been cleared.

<sup>24</sup> *Municípios* are subdivisions of states. The county structure in the Amazon changed over time. The number of counties increased, as old counties split into multiple new counties. In the analyses below, which incorporate observations from different years, the more recent observations have been aggregated backwards using the county-structure transformations. There were 316 counties in 1970, 336 in 1980, 399 in 1985, and 506 in 1991.

<sup>25</sup> This matrix of 1's and 0's, which indicates the other counties that border on any given county, is used to construct 'neighboring county' variables (averages for those counties). It is based on the 316 county, 1970 map.

<sup>26</sup> The references to consult concerning this data are Skole and Tucker (1993) and Skole *et al* (1994).

The deforestation variable is the fraction of the originally forested (i.e., non-*cerrado*) land in a cleared state in year  $t$ .<sup>27</sup> Thus it varies over both space and time. However, note that because of limited observations for the independent variables, only two of the three deforestation observations (1978 and 1988) are used in generating the basic results. The *cerrado* variable is the *cerrado* area over total county area. This varies only over space.

Also from UNH come the soil quality measures. These are densities of nitrogen and carbon in the soil. These vary only over space, as they are computed from a cross-sectional map of large areas viewed as homogeneous in soil type. The county density measures are weighted averages across those types, weighted by how much of the county falls in each soil-type region. The two soil variables available are almost exactly collinear, and thus only one is used.

### 5.1.2 Transport Data

Road lengths (paved and unpaved), river lengths, and distance measures come from maps provided by Brazilian government agencies. Road and river variables are used in the estimation as densities, i.e. as lengths divided by areas. Distances are used simply as lengths. Road observations exist for 1976 and 1986.<sup>28</sup> Thus roads vary over space and time, while rivers and distances vary only over space. The rivers included satisfy the "Class A navigability criterion" (they exceed a minimum depth for a minimum period of time during a typical year).

### 5.1.3 Government Actions other than Roads

Credit extension data come from the Banco do Brasil (BdB). They indicate how many BdB agencies existed in the county in 1985, as well as in what year the first BdB agency appeared. For estimation, this information was used to construct two variables which vary over space as well as in a particular way over time. For the 'credit agency density' variable, for 1985 the number of agencies was divided by the county area. For 1975, if the first agency had appeared by 1975, then the 1985 number of agencies was assigned to 1975 and divided by the county area, yielding the same value as for 1985. If the first agency had not yet appeared by 1975, a zero was assigned. For the 'credit existence' variable, an indicator variable for having more than zero agencies was used in 1985, and then the same procedure was followed for using the date of the appearance of the first agency in order to construct the 1975 indicator value.

Development projects information comes from SUDAM (Superintendency for the Development of Amazonia). For each project, quantitative measures are provided for 1985, and certain dates (e.g., first year of implementation) are also provided. A procedure like that just described was used for constructing the 1975 values for two variables, project area and project existence. Then the county 'project area density' and 'project density' variables are constructed by adding the areas of or totaling the number of projects in the county and dividing by county area. In total, the information available lists 247 development projects, yielding 234 observations after missing values, with a mean area of 330 km<sup>2</sup>. Yokomizo(1989) suggests that the bulk of these projects' impacts occurred in the southeast of the region (near the rest of the country).

---

<sup>27</sup> This definition was chosen because to label as 'deforested' a county which was never forested seemed wrong.

<sup>28</sup> It may be possible to further separate both paved and unpaved roads into federal and state subcategories.

I do not have data on colonization or land titling projects, although they are often discussed.<sup>29</sup> Also, to this point I do not use information on dams.<sup>30</sup> More generally, additional data on policies including those just discussed would be useful.

#### 5.1.4 Census Data

Rural and urban population come from the Brazilian Demographic Census, for 1970, 1980, and 1991.<sup>31</sup> Also from this source comes information on how many of the immigrants who arrived in a given Amazon county in a given decade came from each state within Brazil.

Because of concerns about endogeneity, the wage data collected is an average industrial wage. This should affect the wage paid for activities on cleared land, for instance, as it represents an 'outside option' for those involved in such activities. This industrial wage comes from the Industrial Census, for 1975, 1980, and 1985.

Local output prices were judged to have been too poorly measured to be used here.

#### 5.1.5 'Neighboring-county' Variables

In equation (13) and all discussion above, the following common assumption was implicit: the only factors that affect the land allocation decisions in a given county are the observed variables which describe that county. This assumption is worth testing. For instance, a paved highway running through a neighboring county might be expected to lower transport costs to a given county, even though it would not change the observed length of paved roads in that county. In order to test the 'spatial isolation' implications of this assumption, some regressions based on (13) will include "neighboring-county" versions of variables such as roads, population, projects, and credit. These are unweighted averages of the values for the variables described above for all the counties which share any common border with the county in question.<sup>32</sup>

---

<sup>29</sup> See, for example, Almeida(1992) on efforts by the colonization agency INCRA within the "national integration program". Whether the lack of this data implies that the key factors behind most immigration to the Amazon have been omitted from the analyses, however, remains in question, for these reasons: first, at least from general impressions, spontaneous immigration (responding to general conditions such as, e.g., soil quality and transportation) appears to greatly outnumber official, planned immigration in colonization programs; second, and applying even to official immigration, these programs could include (or be linked to) actions measured by data which are in these analyses, such as road construction or credit extension.

<sup>30</sup> The information available includes the name of the county in which the powerhouse is located, plus total inundated area, but not flooded areas in "non-powerhouse" counties. The inundated areas for the six dams listed in IBGE(1992) is 5,500km<sup>2</sup>. This is under 5% of total clearing in 1980. However, the construction of dams could well have a greater effect than the direct, one-for-one substitution of inundated area for forested area. For instance, an increased local supply of electricity and drop in the local price of electricity could act as a spur to local development which would lead to further deforestation. Such an overall effect could be seen in an elasticity resulting from the regressions below, but the dams are not yet included in the analyses, for lack of information on all the *municípios* which include inundated areas resulting from the dams.

<sup>31</sup> The definition of "urban" is not precisely laid out in Census books. They refer one to "municipal law".

<sup>32</sup> Using *PAVED<sub>i</sub>* for own paved roads and *PAVEDN<sub>i</sub>* for neighbor paved roads, one can see by rewriting  $(\$_1 * PAVED_i + \$_2 * PAVEDN_i)$  as  $\$ * (PAVED_i + **PAVEDN_i)$  that this variable inclusion effectively allows the data to determine the correct weighting of own and other roads in a "road access" variable.

### 5.1.6 Additional Data Needs for Dynamic Issues

As mentioned above, additional data would be necessary to address the empirical implications of a number of dynamic issues which might be incorporated into the model. Of course, the most basic need is for additional observations over time for all variables. However, other particular types of data might also be useful. For instance, some information on variation across counties in policies affecting property rights might allow useful analysis of the effects of changing property regimes. In addition, although their motivation remains speculative without a formal dynamic model and solutions (which are beyond the scope of this paper and will be addressed in future research), other variables might also be of interest. For instance, vegetation regrowth rates might affect optimal dynamic land use decisions, and could in principle be collected.

## 5.2 Priors For The Estimation

### 5.2.1 The Basic Equation for Estimation

Given the data available, the form of (13) which will be the basic equation for estimation is: <sup>33</sup>

$$\ln(y_{i,78}/(1-y_{i,78})) = D_{i,1}^{cleared}(n_{i,70}; h1_{i,76} h2_{i,76} h3_i, \mathbf{m}_i; d_{i,75}, c_{i,75}; w_{i,75}; v_i; q_i) + :_{i,78} \quad (14-1)$$

$$\ln(y_{i,88}/(1-y_{i,88})) = D_{i,2}^{cleared}(n_{i,80}; h1_{i,86} h2_{i,86} h3_i, \mathbf{m}_i; d_{i,85}, c_{i,85}; w_{i,85}; v_i; q_i) + :_{i,88} \quad (14-2)$$

$y_{it}$  = area of cleared forest over originally forested area

$n_{it}$  = density of county population

$h1_{it}$  = density of the length of paved roads in the county

$h2_{it}$  = density of the length of unpaved roads in the county

$h3_i$  = fixed density of the length of rivers in the county

$\mathbf{m}_i$  = fixed distances to the national and state capitals

$d_{it}$  = density of the area of the development projects in the county

$c_{it}$  = density of the number of Banco do Brasil branches in the county

$w_{it}$  = county average industrial wage

$v_i$  = fixed percentage of county area in cerrado

$q_i$  = fixed density of nitrogen in the soil

$:_{it}$  = effect of county-level unobserved factors

### 5.2.2 Priors

The following table gives the priors for the independent variables. A positive prior indicates that a higher value for the variable should lead to more cleared forest land. Explanations follow:

---

Effects between counties may arise not only because of observable determinants of deforestation, but also because of unobservable determinants. This idea throws into question the assumption of independence of error terms across counties. Neighbors might be expected to be more alike than randomly paired counties. If so, residuals should be corrected for potential non-sphericity in order to obtain proper inferences.

<sup>33</sup> In the runs in the tables below, the independent variables will be assumed to be linearly additive within  $D(!)$ . Also, the runs in the tables result from pooling all the observations. Other runs discussed use different specifications.

<u>VARIABLE</u>	<u>PRIOR</u>
Population	+
Roads	+
Rivers	+
Distances	-
Projects	+
Credit	+
Wage	-
<i>Cerrado</i>	-
Soil Quality	+

Population may increase demand and thus output prices, and by increasing labor supply may lower input prices. Roads and rivers are expected to lower transport costs, raising farmgate output prices and lowering farm-gate input prices. Distances, in contrast, raise transport costs. Development projects and credit infrastructure may increase demand and thus output prices, and credit should lower input costs. As an outside option, the industrial wage should raise the labor input price for use of cleared land. Soil quality may provide increased productivity particularly in uses of cleared land.<sup>34</sup> Finally, *cerrado* should have lower clearing costs. Recall, clearing in *cerrado* is not reflected in the dependent variable. The reasoning behind the prior, then, is that, all else equal, having more lower-clearing-cost *cerrado* in a county may draw clearing away from forested areas, such that a smaller fraction of the originally forested area is cleared.<sup>35</sup>

### 5.3 Multicollinearity and Endogeneity

The basic equations for estimation (14) assume that all the determinants of deforestation are taken as exogenous to the land use decision. However, the ongoing process of frontier development may be thought to involve a number of behaviors, such as clearing, migration, and government actions, all of which take each other into account. Thus it seems natural to ask whether some of the independent variables might be endogenous, at least for the purposes of the estimation. Here, I consider additional relationships which could imply statistical difficulties with estimating the basic equations.

---

<sup>34</sup> The expectation on soil's relative effect is based in part on the observation that the biological productivity in standing tropical rainforests is largely independent of the soil, because of efficient nutrient cycling processes.

<sup>35</sup> This claim requires justification. Here I suggest two rationales related to differences in clearing costs.

First, Brazilian law with respect to the Amazon suggests that landowners should keep some percentage of their land uncleared. Of course, this law may not be respected, although in their sample of small landholders in the Amazon, Alston, Libecap and Schneider (1995) find a mean fraction cleared of 40 percent (personal communication, 2/95). If the law is respected, and *cerrado* clearing costs are indeed lower than forest clearing costs, then all else equal, *cerrado* should be cleared and denser forest vegetation left uncleared in order to satisfy the law.

Second, increased economic activity within a county may in fact raise marginal input prices and lower marginal output prices (unlike in the model above, in which prices are given). If, all else equal, *cerrado* is used first because of lower clearing costs, then having more *cerrado* in a county raises the clearing cost for the plots of forested land considered for clearing, and should lower the fraction of originally forested area which is cleared.

There are also *cerrado* stories independent of clearing costs. For instance, *cerrado* might exist where soil is poor, or where little rain falls, i.e. in places where returns to cleared land are relatively low. However, at least the poor soil story appears wrong, as soil quality and *cerrado* density are not strongly correlated in this data. That result is perfectly plausible, as superb nutrient cycling systems allow lush rainforest to grow on quite poor soil.



Then I explore two reasons why endogeneity may be less of a problem for the estimation below than might be imagined: first, the existing data impose the use of lagged independent variables, which reduce the likely extent of endogeneity in the estimation; and second, some relationships imply not endogeneity but multicollinearity of independent variables. However, this potential problem of endogeneity will be given additional attention below.

An obvious candidate for endogeneity is output and input prices, which may be affected by the amounts of land use activity in a county. For this reason, as well as measurement problems, I do not use these prices (other than the industrial wage, a more exogenous proxy).<sup>36</sup>

Another common suggestion is that population is endogenous. However, to be truly endogenous in the basic equations, 1970 (and/or 1980) population would have to in some way result from 1978 (1988) clearing behavior, or of factors that jointly cause them both. This seems unlikely.<sup>37</sup> Further, population may not result from clearing itself. Rather, for example, it might respond to previous population or to access (e.g, roads, rivers, and distances). For example:

$$n_{it} = n_{it}(h1_{it}, h2_{it}, h3_i, m_i) \quad (15)$$

may be the relationship which in fact underlies the sense of 'endogeneity' here. This type of possible multicollinearity is considered in the estimation and interpretation below.<sup>38</sup> Finally, note that the expected bias from endogeneity would be positive, as, for instance, more clearing would be expected to lead to more population. However, in the results below we find a quite weak positive effect of population; this throws into question the existence of a large bias.

Another common suggestion is that government behavior may imply that policy variables such as roads, development projects, and credit infrastructure are caused by, or jointly caused along with, clearing behavior.<sup>39</sup> However, it is again important to note that the basic equations (14) indicate the use of lagged independent policy variables. In addition, the policy variables may be functions not of clearing itself, but of an independent variable such as population or roads. For example, an equation in the style of (15) might apply to credit infrastructure if governments find it worthwhile to establish a bank agency only if there exists a minimal population density. While such behavior would not imply endogeneity, the implied multicollinearity would indeed matter for interpretation

---

<sup>36</sup> Proxies for county output prices could in principle be constructed from national output prices, using information on the breakdown of total output from a county between distinct categories of outputs.

<sup>37</sup> It is possible that serially correlative processes could generate a link between these variables. However, the process would have to be quite strong in order to generate a significant correlation.

<sup>38</sup> In terms of the theoretical model, even should a form of (15) in which clearing were a determinant of population describe migration behavior and imply statistical endogeneity, it would not necessarily belong in the land model. If land decisionmakers do not take into account their effect on migration, they take  $n_{it}$  as exogenous. The same comment applies to taking into account other relationships, such as for government road construction behavior.

<sup>39</sup> Clearing behavior and policy variables could be jointly caused by unobserved factors. One such unobserved variable could be additional spatial heterogeneity in soil quality. Such heterogeneity could lead to clearing and also, independently, affect government actions like road location (assuming the government observes the heterogeneity).

Another type of unobserved variable might be announced policies. These may presage, for instance, a future road location (and thus 'cause' future roads). Further, if they lead to the purchasing of land in the future path of the road solely for speculative financial gains, then the transport cost interpretation of the observed road variable may be mistaken. However, despite stories about information leaks and speculation within government agencies in the Amazon, the scale of such land speculation seems likely to be relatively low.

of the credit and population results in the estimation below.<sup>40</sup> Such possibilities are considered below.<sup>41</sup>

For the reasons above, then, endogeneity may not be as severe a problem for the estimation of (14) as might be implied by the characterization of frontier development above. However, the road and population results are sufficiently central to the paper to warrant attempts to instrument for these variables. In this case, the instrument for roads will be lagged roads.<sup>42</sup>

The instruments for population will be conditions in the "weighted-average migrant state of origin" for an Amazon county. For example, if Rondonia gets most of its immigrants from the south, while Pará gets most of its immigrants from the northeast, then the flow of immigration to Rondonia should increase relative to that to Pará if wages in the south drop significantly while those in the northeast stay the same. As noted above, the breakdown by state of origin is known for the migrants to a given county in a given decade. This breakdown, i.e. the shares of migrants which come from each of the states, is not obviously related to clearing. However, it might be relevant for migration, as if it is relatively fixed over time, it shows how regional shocks are likely to affect differently the migration to different Amazon counties.

This breakdown generates weights for the computation for each county in the Amazon of weighted averages of state-level variables which are assumed to be relevant to the migration decisions: average industrial wage, employment rate, and labor intensity in agriculture. Higher wages and employment rates in migrants' states of origin are presumed to lead to lower levels of emigration from those states. A higher labor intensity in agriculture at the beginning of a decade is presumed to lead to a higher level of emigration from a state, as it implies a larger out-of-work labor pool should agricultural mechanization occur, as it often did during this period.<sup>43</sup>

## 6. RESULTS

---

<sup>40</sup> However, actual endogeneity remains a potential problem to be considered. For instance, the amount of credit disbursed by a credit agency might well respond to the opportunities for investment, which in turn might be a function of the current land use decisions being explained. Thus agency existence may be better than disbursements.

<sup>41</sup> Further, frontier development may be thought to involve more dynamic relations alongside (13), such as:

$$\begin{aligned} n_{i,t+1} &= n(h_{lit}, n_{it}) \quad \text{and} \quad h_{l,i,t+1} = h_l(h_{lit}, n_{it}), \text{ or} \\ n_{\text{NeighborCounty},t+1} &= n(h_{lit}, n_{it}) \quad \text{and} \quad h_{l,\text{NeighborCounty},t+1} = h_l(h_{lit}, n_{it}). \end{aligned}$$

These relationships describing migration and government behavior imply multicollinearity of population and roads in (13), as the two variables are caused by the same things. Further, the interpretation of coefficients in (13) would be less straightforward than in a static model, as the effects of past variables are more important. However, this would not change the underlying focus on effects of, for example, roads. Conclusions based on specific magnitudes of coefficients might change, but roads would be a mechanism for policy intervention. Note that Figure 3 appears to imply that roads grow up into the Amazon from where activity already exists. Interrelationships affecting public infrastructure have been studied explicitly in Binswanger *et al* (1993).

<sup>42</sup> One could imagine other factors affecting road location, such as the location of particular ecological or Indian reserves. However, it seems likely that the causation may run instead from road location to reserve location, as a number of reserves are quite recent, and they were most likely located where access was relatively difficult. Chomitz & Gray (1996) attempt to instrument for roads, and find that controlling for soil quality, it makes little difference.

<sup>43</sup> Particularly in the southern states, government programs encouraged a shift from a labor-intensive crop (coffee) to a more capital- and energy-intensive crop (soybeans). See, for example, World Bank (1982).

Recall from above that the basic equations for estimation are:

$$\ln(y_{i,78}/(1-y_{i,78})) = D_{i,1}^{cleared}(n_{i,70}; h1_{i,76}, h2_{i,76}, h3_i, \mathbf{m}_i; d_{i,75}, c_{i,75}; w_{i,75}; v_i; q_i) + :_{i,78} \quad (14-1)$$

$$\ln(y_{i,88}/(1-y_{i,88})) = D_{i,2}^{cleared}(n_{i,80}; h1_{i,86}, h2_{i,86}, h3_i, \mathbf{m}_i; d_{i,85}, c_{i,85}; w_{i,85}; v_i; q_i) + :_{i,88} \quad (14-2)$$

Here I discuss the results of estimation, including those presented in the tables below. Table 4 is for population only, Table 5 puts population and other factors together, and then Table 6 presents variations relative to Table 5 in how population is treated. All these runs involve pooling all the observations, and assume that  $D(!)$  is linearly additive. Finally, Table 7 presents regressions in which changes in land clearing over time are explained by changes in the variables from Table 5, or their levels if they do not change over time. In focusing on one category of effect at a time, the discussion of results will cut across tables (and runs not in the tables).

### 6.1 The Effects of Population on Deforestation

The first column of Table 4 confirms that here, as in the literature, if the population density itself is the only independent variable, then it appears to be a significant determinant of deforestation (although it does not by itself explain a great deal of the variation in deforestation). The second column's quadratic specification considers the intuitive hypothesis that an additional person in a population-dense county has less marginal impact on land clearing than one in a relatively empty area. The positive sign on the population density level and negative sign on its square support this hypothesis. However, the overall explanatory power is still not overwhelming.

The results just cited are quite robust to modifications of this basic specification such as the addition of time and state dummies, or working with cross sections or in differences. However, as indicated in the first column of Table 5, the linear population result does not stand up to the most important change in specification, i.e. the inclusion in the regression of other factors motivated by the optimal land use model. Further, the large drop relative to Table 4 in the population coefficients in both columns of Table 5 suggest that the earlier coefficients were picking up the effects of other, omitted factors. However, it is noteworthy that the nonlinear (concave) specification from Table 4 remains significant in Table 5; this specification, in which the marginal impact of additional people depends on how many people are already in a county, is more robust to the inclusion of other factors.

Table 6 makes two additional points about population. In the first column, the credit agency coefficient and standard error which result from dropping population indicate that the population and credit variables may well be multicollinear (as suggested in (15) above). This confounding of effects is consistent across specifications. The second column of Table 6 indicates perhaps most of all that the "conditions in the average migrant state of origin" instruments do not perform well in the first stage regression for population. As a result, the two-stage least squares regression yields significantly higher standard errors.

### 6.2 The Effects of Transport on Deforestation

The paved road coefficients in both columns of Tables 5 and 6 clearly support the intuition that

major roadways must be important in lowering the cost of access to the Amazon region.<sup>44</sup> The importance of transport costs is the dominant policy-relevant result in the empirical work<sup>45</sup>, and is robust to the inclusion of year and state dummies and working in cross-sections. It is worth noting that when working in differences, such as in Table 7, the neighboring county roads appear relatively more important, and neighboring unpaved roads appear to be significant. The results for the 'distance to the national seat' also confirm intuition about transport costs consistently across specifications. As distances are greater, transport costs are higher, lowering farm-gate output prices and raising farm-gate input prices, such that less land is cleared.

It should be noted, though, that not all results concerning transport are significant and in accord with the priors above. First, in the pooled cross-section regressions, the unpaved road coefficients are consistently not significant and in accord with the positive prior. This is the case even if paved road variables are not included. Consistent with this last result, a total roads variable is significantly weaker than simply a paved roads variable.<sup>46</sup>

Most surprisingly, while the prior for rivers was positive, rivers' effects are negative in Tables 5 and 6.<sup>47</sup> While the transport services provided by rivers and roads are not identical, and while rivers provide food through a production technology that does not require forest clearing, at the very least this result bears further investigation given the strong prior concerning transport costs. It is worth noting, though, that in Table 7 the fixed river density variable has a significant positive effect on the change in land clearing over time.

Finally, while the regressions in Tables 5 and 6 suggest that at a given point in a time a county further from markets south of the Amazon is likely to be less fully cleared, Table 7 suggests that the distance from markets south of the Amazon has a significant positive effect on the changes over time in fraction cleared. That is, the counties further from the rest of the country display larger changes over time in fraction of forest cleared. This more dynamic result would appear to indicate a movement of the development frontier up into the Amazon region.

### **6.3 The Effects of Development Projects and Credit Infrastructure on Deforestation**

Tables 5 and 6 show a significant, positive effect of development projects on clearing, in accord with the prior above. However, the neighbor version of the project variable is not significant. This might

---

<sup>44</sup> Note that the roads (or other) coefficients in the first column of Table 6 could be seen as a form of "full derivative" of deforestation with respect to roads if a relation such as (15) holds, such that roads are a factor in the level of population. However, note also that these coefficients are virtually identical to those in Table 4.

This is not surprising, as any indirect effect of roads through population must be small if population's effect is small.

<sup>45</sup> It may be worth noting that a specification using simply own-county paved roads had more than twice the explanatory power of the regression in the first column of Table 4. Also, a specification using simply own-county and neighboring-county paved roads provides more than twice the explanatory power of Table 4's second column.

<sup>46</sup> The distance to state seat result also contradicts the priors above. However, it is less clear, given that most major markets are outside of the Amazon, that this distance represents an increase in the costs of access to markets. In addition, while this bears further investigation, the main results do not depend on its inclusion. Further, it is worth noting that, unlike the distance from southern markets, this fixed variable has no effect in a differences regression.

<sup>47</sup> It is important to note that this result is not driven by the inclusion within the river variable of bodies of water such as small streams which could not be expected to really affect transport costs. All the rivers included here satisfy the "class A navigability criterion", and are thus significant transport options. In addition, the river result is not being driven by a significant correlation with the roads variables.

be taken to suggest that while these projects always involved some local clearing, the idea that projects would spur outwardly-spreading regional development was incorrect.

Additional analysis of project results may also be of interest. For instance, the inclusion of state dummies makes projects insignificant, and while projects have a significant effect in a first-observations cross-section, they are insignificant in the second cross-section. While the exact rationale for why this would be may not be clear, this result (confirmed in Table 7) that the additional projects created during this period did not affect land clearing is consistent with existing stories suggesting an early governmental push whose force diminished over the period.

Credit agency density is insignificant in Table 5. This could be the result of poor measurement. However, the first column of Table 6 provides an additional perspective: dropping population from the regression raises the credit coefficient and lowers its standard error, yielding significance of the credit variable. This may well result from the collinearity of credit agency location and population. That in turn might be just a statistical relationship, but might instead have meaningful interpretation in terms of government behavior. For example, as suggested in (15) above, the government's agency-location decision rule could be to place agencies where there are people to take advantage of them, i.e. in counties with sufficient population.

Additional analysis of the regressions in Table 5 yields further perspective. Credit has a significant effect in a first-observations cross-section, but not in the second cross-section, and not in the differences regressions in Table 7. As in the case of projects, additional information on policies may be necessary to fully understand why this might be the case, but this evidence of a single effective push followed by the lack of an effect is worth noting.<sup>48</sup>

#### **6.4 Other Effects on Deforestation**

Tables 5 and 6 show not only the expected positive coefficients but also consistent significance for soil quality as a determinant of cleared land. This result may be a little surprising, as the soil quality measure is rather coarse, with homogeneity asserted for regions large enough that they must in fact contain heterogeneous sub-regions. However, the result is robust even to the (noisy) instrumentation for the population variables in the second column of Table 6.

The results for *cerrado* density are similar, in that they are in keeping with priors but perhaps surprisingly strong. Such strong results may be evidence that clearing costs are a significant component of the costs of using cleared land. Further, they may be evidence that the law mandating that some portion of a landholder's property remain uncleared has effect.<sup>49</sup>

---

<sup>48</sup> This evidence could be explained in many ways, of course, between which I am unable to distinguish at this time. For instance, the first wave of projects locations could consciously have been in areas which possessed qualities beneficial for successful agricultural production but which I do not observe. The projects themselves might then actually have been redundant given the existing attraction of these locations. Then when later attempts to foster production and development were made in otherwise less beneficial locations, the true lack of an independent effect of projects was revealed. While this is certainly not necessarily the case, and many argue that projects themselves had a major effect, it is important to note that it is one possible explanation for what has been observed.

<sup>49</sup> It is important to note that while an idiosyncratic, non-uniform distribution of *cerrado* within the Amazon region could be behind this result, making it a spurious one, the facts do not support this. *Cerrado* is if anything more heavily concentrated in the south and east of the region, closer to the rest of the country, such that if anything one might expect that *cerrado* areas would be areas where conditions would lead to more clearing of forest.

## 7. CONCLUSIONS, EXTENSIONS, AND POLICY IMPLICATIONS

This paper has analyzed the determinants of deforestation in the Brazilian Amazon. A deforestation equation was derived from a model of optimal land use, and estimated using county-level data for the period 1978 to 1988. The empirical analysis advanced beyond previous analyses in large part because of the innovative merging of state-of-the-art satellite data on deforestation with an outstanding county-level dataset for the Amazon. In addition, the empirical work was motivated by a systematic presentation of a relevant economic framework which encompassed not only population but also other factors such as roads, credit, and soil quality.

Evidence was presented that: increased road density in a given county leads to more deforestation in that county and in neighboring counties; government-subsidized development projects lead to greater deforestation; confirming the importance of transport costs, greater distance from markets south of the Amazon leads to less deforestation; and confirming economic intuition, better soil is associated with more deforestation. The results for government provision of subsidized credit were mixed across specifications.

The population density, the primary explanatory in most previous empirical work, was indeed significant when it was the sole explanatory variable. However, its effect disappeared when all the variables motivated within the land use model were included. This result suggested that the earlier results for population may be significantly affected by the omission of important variables. In principle, though, a large, exogenous increase in population density should affect output demand and input supply conditions and thus affect land clearing. Thus while in this result the population density was dominated by other explanatory factors, population deserved further attention. Along these lines, a quadratic population specification was more robust to the inclusion of other variables. The first few people entering an empty county have significantly more impact on deforestation than the same number of people added to already densely populated county. This additional result suggested the importance of the spatial distribution of population.

There are a number of potential extensions of this work, many of which would involve additional data. The required data include more disaggregated land use data (for instance, disaggregating the cleared category to crops, ranching, and timber), as well as data which is further spatially disaggregated (i.e., at a sub-county level) and would thus allow better use of high-resolution satellite images. In addition, as mentioned above, dynamic issues such as property rights regimes could be incorporated into the empirical work through not only more data but also further formal modeling. The latter will be addressed in future research.

The mixed evidence for credit access, while also necessarily qualified by the extent of the existing data, raised the issue of endogenous policymaking in the interpretation of observed multicollinearity: the government may have located credit agencies where they would be likely to be used, for instance in population centers. More generally, policymakers themselves may respond to variables thought of as determinants of deforestation. Thus it is possible that, for instance, policy actions do further development in some way, but after deforestation has already been induced by other factors. This sort of speculation raises the more general point that the dynamics of frontier deforestation really have not been adequately addressed, suggesting again the need for additional observations over time, including for policy actions.

In terms of policy implications that emerge from the analyses, the result that regardless of other

factors included in the analyses, roads appear to be strong determinants of deforestation suggests one channel for affecting the rate of deforestation. Even the possibility that past government decisionmaking made roads partially a function of other variables might not matter for future policy purposes: even within a complex system of frontier expansion and development, as long as roads form one important causal link, they are a potential policy tool.

More speculatively, the road and quadratic population results could suggest that in order to achieve any given, desired level of regional development with minimal deforestation, the government might do better to build good roads to existing cities instead of roads out into sparsely populated areas (and perhaps use subsidies for urban employment instead of for rural agriculture). This suggestion arises from the results that roads channel development and that the per-person impact of population on deforestation is lower in areas of concentrated population.

Finally, it is important to note that the questions addressed in this paper are distinct from questions about socially optimal amounts of deforestation. Answering the latter type of question would require knowledge of the value of the standing forest. However, while these results do not address the question of whether an "optimal" rate of deforestation would be higher or lower than the current rate, they may suggest how best to go about effecting a given policy goal.

## Bibliography

Allen, Julia C. and Douglas F. Barnes (1985). "The Causes of Deforestation in Developing Countries". *Annals of the Association of American Geographers* 75:163-184.

Almeida, Anna Luiza Ozorio de (1992). *The Colonization of the Amazon*. Austin, University of Texas Press, 371p.

Alston, Lee J., Gary D. Libecap, and Robert Schneider (1995). "Property Rights and the Preconditions for Markets: The Case of the Amazon Frontier". *Journal of Institutional and Theoretical Economics* 151(1):89-107.

Amacher, Gregory S., William F. Hyde, and Bharat R. Joshee (1992). "The Adoption of Consumption Technologies under Uncertainty: A Case of Improved Stoves in Nepal". *Journal of Economic Development* 17(2):93-105.

Binswanger, Hans (1989). "Brazilian Policies That Encourage Deforestation in the Amazon." Environment Department Working Paper no.16, Washington D.C., The World Bank.

Binswanger, Hans P., Shahidur R. Khandker, and Mark R. Rosenzweig (1993). "How infrastructure and financial institutions affect agricultural output and investment in India". *Journal of Development Economics* 41:337-366, North-Holland.

Branford, S. and O. Glock (1985). *The Last Frontier: Fighting Over Land in the Amazon*. London, Zed Books.

Browder, John O. (1985). "Subsidies, Deforestation and the Forest Sector in the Brazilian Amazon." A Report to the World Resources Institute. Washington, D.C.

Bunker, Stephen G. (1985). *Underdeveloping the Amazon: Extraction, Unequal Exchange, and the Failure of the Modern State*. University of Illinois: Urbana.

Chomitz, Kenneth M. and David A. Gray (1996). "Roads, Land Use and Deforestation: A Spatial Model Applied to Belize". *World Bank Economic Review* 10(3):487-512.

Cropper, Maureen and Charles Griffiths (1994). "The Interaction of Population Growth and Environmental Quality". *American Economics Review* 84(2):250-254.

Cropper, Maureen, Charles Griffiths and Muthukumara Mani (1996). "Logging, Population Pressures and Deforestation in Thailand, 1973-1991". World Bank mimeo, presented in an Association of Environmental and Resource Economics session on 'Deforestation: Underlying Causes and Policy Implications', at the 1996 ASSA Winter Meetings, San Francisco.



Deacon, Robert (1994). "Deforestation and the Rule of Law in a Cross-Section of Countries." *Land Economics* 70(4):414-430.

Deacon, Robert (1995). "Assessing the Relationship between Government Policy and Deforestation". *Journal of Environmental Economics and Management* 28(1):1-18.

Detwiler, R.P. and Charles A.S. Hall (1988). "Tropical Forests and the Global Carbon Cycle". *Science*, 239:42-47.

Ehui, Simeon K. and Thomas W. Hertel (1989). "Deforestation and Agricultural Productivity in the Cote d'Ivoire". *American Journal of Agricultural Economics* 71(3):703-711.

Feamside, Philip M. (1986). *Human Carrying Capacity of the Brazilian Rainforest*. Columbia University Press, New York.

Feamside, Philip M. (1990a) "Environmental Destruction in the Brazilian Amazon", in Goodman, David and Anthony Hall, eds. *The Future of Amazonia: Destruction of Sustainable Development?*. MacMillan Press, Ltd., London.

Feamside, P.M., A.T. Tardin, and L.G.M. Meira (1990b). *Deforestation Rate in Brazilian Amazonia*. National Secretariat of Science and Technology, Brasilia, Brazil.

Feamside, Philip M. (1993). "Deforestation in Brazilian Amazonia: The Effect of Population and Land Tenure". *Ambio* 22(8):537-545.

Gasques, Jose Garcia and Clando Yokomizo (1985). "Avaliacao dos Incentivos Fiscais na Amazonia". IPEA, Brasilia.

Green, William H. (1990). *Econometric Analysis*. Macmillan Publishing Company, New York.

Goulding, Michael (1993). *Scientific American*, March 1993, pp.114-120.

Harrison, Susan (1991). "Population Growth, Land Use and Deforestation in Costa Rica, 1950-1984." *Interciencia* 16(2):83-93.

Hecht, Susanna (1982). *Cattle ranching development in the Eastern Amazon: evaluation of a development policy*. Ph.D. dissertation, University of California, Berkeley.

Hecht, Susanna B. (1993). "The Logic of Livestock and Deforestation in Amazonia". *Bioscience* 43(10):687-695.

Hecht, Susanna (1985). "Environment, development, and politics: capital accumulation and the livestock in eastern Amazonia". *World Development* 13(6):663-684.

Hecht, Susanna and Alexander Cockburn (1990). *The Fate of the Forest: Developers, Destroyers, and Defenders of the Amazon*. Harper Collins, Great Britain.

Houghton, R.A. *et al* (1987). "The flux of carbon from terrestrial ecosystems to the atmosphere in 1980 due to changes in land use: geographic distribution of the global flux". *Tellus*, 39B:122-139.

Hyde, William F. and Roger A. Sedjo (1982). "Managing tropical forests: reflections on the rent distribution discussion". *Land Economics* 68(3):343-350.

Hyde, William F., Gregory S. Amacher and William Magrath (1996 forthcoming). "Deforestation, Scarce Forest Resources, and Forest Land Use: Theory, Empirical Evidence, and Policy Implications". Forthcoming *World Bank Research Observer*.

IBGE (1992). *Anuario Estatístico 1992*. Instituto Brasileiro de Geografia e Estatística, Rio de Janeiro, Brazil.

INPE (1992). "Deforestation in Brazilian Amazonia". Instituto Nacional de Pesquisas Espaciais, São José dos Campos, Brazil.

Jones, Donald W., Virginia H. Dale, John J. Beauchamp, Marcos A. Pedlowski, and Robert V. O'Neill (1992). "Farming in Rondonia". Working Paper, Oak Ridge National Laboratory, Oak Ridge, Tennessee.

Jones, Donald W. and Robert V. O'Neill (1994). "Development Policies, Rural Land Use, and Tropical Deforestation". *Regional Science and Urban Economics* 24(6):753-771.

Jones, Donald W. and Robert V. O'Neill (1995). "Development Policies, Urban Unemployment, and Deforestation: The Role of Infrastructure and Tax Policy in a Two-Sector Model". *Journal of Regional Science* 35(1):135-153.

Kummer, David M. (1991). *Deforestation in the Postwar Philippines*. University of Chicago Geography Research Paper No. 234, The University of Chicago Press. Chicago, IL.

Kummer, David M. and B.L. Turner II (1994). "The Human Causes of Deforestation in Southeast Asia". *Bioscience* 44(5):323-328.

Lugo, Ariel E., Ralph Schmidt and Sandra Brown (1981). "Tropical Forests in the Caribbean". *Ambio*, 10:318-24.

Mahar, Dennis A. (1989) *Government Policies and Deforestation in Brazil's Amazon Region*. Washington, D.C., World Bank.

Mattos, Marli Maria and Christopher Uhl (1994). "Economic and Ecological Perspectives on Ranching in the Eastern Amazon." *World Development* 22(2):145-158.

Moran, Emilio (1981). *Developing the Amazon*. Indiana University Press, Bloomington.

Moran, Emilio (1990). "Private and Public Colonization Schemes in Amazonia", in Goodman, David and Anthony Hall, *The Future of Amazonia: Destruction or Sustainable Development?*. Macmillan Press, London.

Moran, Emilio F. (1994). "Integrating Amazonian Vegetation, Land-Use, and Satellite Data". *Bioscience* 44(5):329-338.

Nerlove, Marc L. and Efraim Sadka (1991). "Von Thunen's Model of the Dual Economy". *Journal of Economics* 54(2):97-123. Springer-Verlag, Austria.

Palo, Matti, Jyrki Salmi and Gerardo Mery (1987). "Deforestation in the Tropics: Pilot Scenarios Based on Quantitative Analyses." In *Deforestation or Development in the Third World*, edited by Matti Palo and Jyrki Salmi. Helsinki: Finnish Forest Research Institute.

Panayotou, Theodore and Somthawin Sungsuwan (1989). "An Econometric Study of the Causes of Tropical Deforestation: The Case of Northeast Thailand". HIID Development Discussion Paper No.284.

Reis, Eustaquio J. and Sergio Margulis (1991). "Options for Slowing Amazon Jungle Clearing", Dornbusch R. and Poterba J.(eds.) *Global Warming: The Economic Policy Responses*. Cambridge, MA, MIT Press, pp.335-375.

Reis, Eustaquio J. and Rolando M. Guzman (1992). "An Econometric Model of Amazon Deforestation". IPEA/Rio Working Paper, Rio de Janeiro, Brazil.

Repetto, Robert (1988). *The Forest for the Trees? Government Policies and the Misuse of Forest Resources*. World Resources Institute. Washington, DC.

Rudel, Thomas K. (1989). "Population, Development, and Tropical Deforestation: A Cross-National Study." *Rural Sociology* 54(3), pp.327-338.

Schmink, Marianne and Charles Wood (1992). *Contested Frontiers in Amazonia*. New York, Columbia University Press.

Scherr, Sara J. (1995). "Economic Factors in Farmer Adoption of Agroforestry: Patterns Observed in Western Kenya". *World Development* 23(5):787-804.

Schneider, Robert (1993). "Land Abandonment, Property Rights, and Agricultural Sustainability in the Amazon". World Bank, LATEN Dissemination Note #3, Washington, D.C..

Schneider, Robert, Gunars Platais, David Rosenblatt, and Maryla Webb (1993). "Sustainability, Yield Loss, and *Immediatismo*: Choice of Technique at the Frontier". World Bank, LATEN Dissemination Note #1, Washington, D.C., 11p.

Skole, David and Compton Tucker (1993). "Tropical Deforestation and Habitat Fragmentation in the Amazon: Satellite Data from 1978 to 1988." *Science*, 260:1905-1910.

Skole, David, W.H Chomentowski, W.A. Salas and A.D Nobre (1994). "Physical and Human Dimensions of Deforestation in Amazonia". *Bioscience* 44(5):314-322.

Smith, Nigel J.H. (1982). *Rainforest Corridors: The TransAmazon Colonization Scheme*. University of California Press, Berkeley, CA.

Southgate, Douglas and Rodrigo Sierra and Lawrence Brown (1991). "The Causes of Tropical Deforestation in Ecuador: A Statistical Analysis", *World Development*, Vol. 19, No. 9, pp.1145-1151.

Stavins, Robert N. and Adam Jaffe (1990). "Unintended Impacts of Public Investments on Private Decisions: The Depletion of Forested Wetlands". *American Economic Review*, 80(3):337-352.

Sungsuwan, Somthawin (1985). *A Study on the Causes of Deforestation in Northeast Thailand*. M.A. thesis, Thammasat University.

Uhl, Christopher *et al.* (1991) "Social, economic, and ecological consequences of selective logging in an Amazon frontier: the case of Tailandia". *Forest Ecology and Management* 46:243-273. Elsevier.

Vincent, Jeffrey R. (1990). "Rent Capture and the Feasibility of Tropical Forest Management". *Land Economics* 66(2):212-223.

Vincent, Jeffrey R. and Clark S. Binckley (1992). "Forest-Based Industrialization: A Dynamic Perspective", in Narendra P. Sharma, ed., *Managing the World's Forests: Looking For Balance Between Conservation and Development*. Kendall/Hunt, Dubuque, pp.93-137.

World Bank (1982). "Brazil: A Review of Agricultural Policies". Country Study, The World Bank, Washington, D.C.

Yokomizo, Clando. "Incentivos Financeiros e Fiscais na Pecuarizacao da Amazonia". Texto Para Discussao, no.22, Institute de Planejamento, Outubro 1989.

**TABLE 1****Road and River lengths (kilometers per county; means over all counties):**

	<u>1975</u>	<u>1985</u>
Paved Roads, length	16.7	36.7
Unpaved Roads, length	76.4	146.2
Rivers, length	44.8	44.8

**Population (number per county; means over all counties):**

	<u>1970</u>	<u>1980</u>	<u>1991</u>
Urban Population	8463	16180	28974
Rural Population	14524	19282	23519
Total Population	22987	35462	52493

**Fraction of a county which was never forested (from satellite data source).<sup>50</sup>**

	<u>Mean</u>	<u>Std. Dev.</u>	<u>Minimum</u>	<u>Maximum</u>
<i>Cerrado</i> density	.30	.42	0	1.0

**Fraction cleared in county (ratios of areas; mean over all counties):**

	<u>Mean</u>	<u>Std Dev</u>	<u>Minimum</u>	<u>Maximum</u>
Fraction Cleared, 1975	.01	.04	0	.37
Fraction Cleared, 1978	.04	.08	0	.45
Fraction Cleared, 1988	.16	.25	0	1.0

---

<sup>50</sup> The means for the *cerrado* and cleared fractions above are unweighted averages of the *cerrado* and cleared fractions for each of the counties in the region. Only a weighted average using county areas as weights would reproduce the value obtained by dividing, e.g., cleared area for the whole region by the total area of the region.

## Table 2 - Variable Names and Definitions

"*neighbor versions*" - for any variable, equally-weighted averages of the variable level in the surrounding counties

**% fraction cleared** - the ratio of sq.km of cleared forest area to sq.km. of the originally forested area

**Paved Road density** - the ratio of kilometers of paved road length to sq.km. of county area

**Paved Ngbr density** - the neighbor version of Paved Road density

**Unpvd Road density** - the ratio of kilometers of unpaved road length to sq.km. of county area

**Unpvd Ngbr density** - the neighbor version of Unpvd Road density

**River density** - the ratio of kilometers of river length to sq.km. of county area

**River Ngbr density** - the neighbor version of River density

**Population density** - the ratio of number of people to sq.km. of county area

**Squared Pop. density** - the square of Population density

**Popul.Ngbr density** - the neighbor version of Population density

**Squared Pop.Ngbr. density** - the square of Popul.Ngbr density

**Project Area density** - the ratio of sq.km. of projects' area to sq.km. of county area

**Project Ngbr density** - the neighbor version of Project Area density

**Credit Agency density** - the ratio of number of agencies to sq.km. of county area

**Credit Ngbr density** - the neighbor version of Credit Agency density

**Distance To State Seat** - kilometers to the state capital

**Distance To Nat'l Seat** - kilometers to Brasilia (representing the way to markets)

**Industrial Wage** - an average wage (salary expenditures over salaried employees)

**Nitrogen density** - a ratio of units of nitrogen to units of soil (representing typical soil quality)

**Carbon density** - a ratio of units of carbon to units of the soil (representing typical soil quality)

**Cerrado density** - the ratio of sq.km. of *cerrado* area to sq.km. of county area

**Origin Wage** - a weighted average of all states' average industrial wages, with weights equal to the fraction of immigrants to a county from the state (represents the attractiveness of the 'average state of origin' of immigrants)

**Origin Empl.Rate** - a weighted average of all states' average employment rates, with weights as for Origin Wage

**Origin Intensity** - a weighted average of all states' labor intensities in agriculture, with weights as for Origin Wage; represents the size of the pool of labor potentially released by agricultural mechanization

**TABLE 3 - County Level Descriptive Statistics**

	<u>Mean</u>	<u>Std Dev</u>	<u>Minimum</u>	<u>Maximum</u>
<u>Transport densities:</u>				
Paved Road 1975	.007	.015	0	.098
Paved Road 1985	.015	.031	0	.222
Unpaved Road 1975	.011	.016	0	.116
Unpaved Road 1985	.023	.030	0	.243
River	.006	.016	0	.140
<u>Development projects:</u>				
# per county 1975	.52	3.4	0	46
# per county 1980	.63	3.8	0	51
# per county 1985	.78	4.2	0	59
<u>Credit agencies:</u> (counties included here only if they have >=1)				
# per county 1975	1.4	1.14	1	7
# per county 1980	1.3	.93	1	7
# per county 1985	1.2	.75	1	7

**TABLE 4 - DEFORESTATION AND POPULATION**

<u>Variables</u>	<u>Linear Population, Isolated</u>	<u>Quadratic Population, Isolated</u>
Population density	.0058 ** (.0014)	.0230 ** (.0071)
Squared Pop. density		-.0018 ** (.0007)
Popul.Ngbr. density	.0164 ** (.0058)	.0344 ** (.0124)
Squared Pop.Ngbr. density		-.0067 * (.0031)
CONSTANT	- 5.140 ** (.3394)	-5.470 ** (.3516)
Adjusted R-Squared	0.048	0.078

Both regressions are pooled cross-sections, with 548 observations, from equations (14) [from which many variables are dropped].

In the parentheses are the corrected standard errors [ also ~, \*, and \*\* indicate significance at the 90, 95, and 99% levels].

I(II) : The coefficient and standard error are multiplied by 100(10,000).

**TABLE 5 - DEFORESTATION, POPULATION, AND OTHER FACTORS**

<u>Variables</u>	<u>Linear Population, with Other Factors</u>	<u>Quadratic Popul., with Other Factors</u>
Paved Road density	.0016 ** (.0003)	.0015 ** (.0003)
Paved Ngbr density	.0023 ** (.0009)	.0024 ** (.0009)
Unpvd Road density	- .0001 (.0004)	- .0002 (.0004)
Unpvd Ngbr density	.0010 (.0010)	- .0010 (.0010)
River density	.0006 (.0005)	.0006 (.0005)
River Ngbr density	- .0049 ** (.0017)	- .0051 ** (.0017)
Distance To State Seat	.0019 * (.0009)	.0020 * (.0009)
Distance To Nat'l Seat	- .0013 ** (.0005)	- .0013 ** (.0005)
Project Area density <sub>II</sub>	.0015 ** (.0003)	.0014 ** (.0004)
Project Ngbr density <sub>II</sub>	.0023 (.0080)	.0051 (.0090)
Credit Agency density	.0241 (.0197)	.0213 (.0208)
Credit Ngbr density	.0466 (.0420)	.0423 (.0424)
Industrial Wage	.0058 (.0140)	.0073 (.0139)
Nitrogen density	.0205 ** (.0043)	.0208 ** (.0043)
Cerrado density	- 5.111 ** (.6839)	- 5.019 ** (.6924)
Population density	.0009 (.0016)	.0100 * (.0044)
Squared Pop. density <sub>I</sub>		- .0009 * (.0004)
Popul.Ngbr. density	- .0002 (.0045)	- .0023 (.0134)
Squared Pop.Ngbr. density <sub>I</sub>		.0001 (.0025)
CONSTANT	- 2.912 ** (.9999)	- 3.046 ** (1.008)
Adjusted R-Squared	0.371	0.371

Both regressions are pooled cross-sections, with 480 observations, from equations (14).

In the parentheses are the corrected standard errors [ also ~, \*, and \*\* indicate significance at the 90, 95, and 99% levels].

I<sub>(II)</sub> : The coefficient and standard error are multiplied by 100(10,000).



**TABLE 6 - DEFORESTATION, OTHER FACTORS, AND POPULATION VARIATIONS**

<u>Variables</u>	<u>Other Factors, without Population</u>	<u>Other Factors, Instrumented Population</u>
Paved Road density	.0016 ** (.0003)	.0026 * (.0013)
Paved Ngr density	.0023 ** (.0009)	.0033 ~ (.0022)
Unpvd Road density	-.0001 (.0004)	.0004 (.0008)
Unpvd Ngr density	-.0009 (.0010)	-.0012 (.0016)
River density	.0007 ~ (.0004)	-.0028 (.0048)
River Ngr density	-.0050 ** (.0014)	.0039 (.0082)
Distance To State Seat	.0019 * (.0009)	.0016 * (.0009)
Distance To Nat'l Seat	-.0014 ** (.0005)	-.0017 ** (.0007)
Project Area density <sub>II</sub>	.0015 ** (.0003)	.0020 (.0021)
Project Ngr density <sub>II</sub>	.0020 (.0080)	-.0007 (.0188)
Credit Agency density	.0298 ** (.0114)	.0049 (.1323)
Credit Ngr density	.0467 (.0398)	.1872 ~ (.1342)
Industrial Wage	.0057 (.0142)	-.0158 (.0199)
Nitrogen density	.0205 ** (.0043)	.0177 ** (.0045)
Cerrado density	- 5.124 ** (.6873)	- 5.412 ** (.8555)
Population density		-.0439 (.0509)
Squared Pop. density <sub>I</sub>		.0057 (.0055)
Popul.Ngr. density		-.0494 (.0780)
Squared Pop.Ngr. density <sub>I</sub>		.0000 (.0179)
CONSTANT	- 2.901 ** (1.000)	- 1.974 ~ (1.239)
Adjusted. R-Squared	0.373	---

Both regressions are pooled cross-sections from equations (14) [w/dropped variables or 2SLS]. Column 1 has 480 obs's; Column 2 has 477.

In the parentheses are the corrected standard errors [ also ~, \*, and \*\* indicate significance at the 90, 95, and 99% levels].

<sub>I(II)</sub> : The coefficient and standard error are multiplied by 100(10,000).

**TABLE 7 - DEFORESTATION, OTHER FACTORS, AND POPULATION in differences**

<u>Variables</u>	<u>Linear Population, with Other Factors</u>	<u>Quadratic Population, w/Other Factors</u>
Paved Road density	- .0001 (.0004)	- .0001 (.0004)
Paved Ngr. density	.0011 * (.0006)	.0010 * (.0006)
Unpaved Road density	.0000 (.0003)	.0000 (.0003)
Unpaved Ngr. density	.0017 * (.0008)	.0013 ~ (.0009)
River density	.0017 * (.0009)	.0016 * (.0009)
River Ngr. density	- .0023 ~ (.0016)	- .0025 ~ (.0017)
Distance To State Seat	- .0005 (.0006)	- .0004 (.0006)
Distance To National Seat	.0010 ** (.0002)	.0010 ** (.0002)
Project Area density <sub>II</sub>	- .0004 (.0010)	- .0011 (.0010)
Project Ngr. density <sub>II</sub>	- .0053 (.0048)	- .0070 (.0050)
Credit Agency density	.0154 (.0196)	.0164 (.0199)
Credit Ngr. density	.0122 (.0239)	- .0073 (.0261)
Industrial Wage	- .0171 (.0211)	- .0163 (.0210)
Nitrogen density	- .0014 (.0065)	- .0009 (.0065)
Cerrado density	- .8415 * (.4613)	- .8034 * (.4602)
Population density	- .0031 (.0048)	.0068 (.0180)
Squared Pop. density <sub>I</sub>		- .0005 (.0010)
Ngr. Popul. density	.0084 (.0130)	.0680 * (.0371)
Squared Pop.Ngr. density <sub>I</sub>		- .0083 * (.0048)
CONSTANT	-----	-----
Adj. R-Sq.	0.101	0.107

Both regressions are pooled cross-sections from equations (14). Column 1 has 480 obs's; Column 2, 477.

Parentheses contain corrected standard errors [ ~, \*, \*\* indicate significance at 90,95,99%].

<sub>I (II)</sub>: the coefficient and standard error are multiplied by 100 (10,000).