What Drives Deforestation in the Brazilian Amazon?

Evidence from Satellite and Socioeconomic Data*

Alexander S. P. Pfaff

Department of Economics, School of International and Public Affairs, Center for Environmental Research and Conservation, Columbia University, New York, New York 10027

Received September 2, 1997; revised September 22, 1998

While previous empirical analysis of deforestation focused on population, this paper builds from a model of land use which suggests many determinants of deforestation in the Brazilian Amazon. I derive a deforestation equation from this model and test a number of those factors using county-level data for the period 1978–1988. The data include a satellite deforestation measure which allows improved within-country analysis. The major empirical finding is the significance of both land characteristics (such as soil quality and vegetation density) and factors affecting transport costs (such as distance to major markets and both own- and neighboring-county roads). Government development projects also appear to affect clearing, although credit infrastructure does not. However, as such policies themselves may be functions of other factors, estimated effects of policies must be interpreted with some caution. Finally, the population density does not have a significant effect on deforestation when many potential determinants are included. However, a quadratic specification reveals a more robust result: the first migrants to a county have greater impact than later immigrants. This implies that the distribution of population affects its impact.

1. INTRODUCTION

The depletion of rainforests has demanded the attention of policy makers in the 1980’s and 90’s. Initial concern about 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. Policy makers must understand the effects of the full set of potential drivers of 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. Despite the attention given to Amazon rainforest depletion, over 80% of this rainforest remains. Thus, these questions are not of

*This paper draws upon the first chapter of my dissertation Pfaff 25. While there I acknowledge additional thanks due, I want to thank in particular Richard Schmalensee, Robert Solow, and Robert Stavins, as well as Eustaquio Reis for both data and comments, David Skole for data, and three anonymous referees for comments. I gratefully acknowledge funding for this work from: an NSF graduate fellowship, the M.I.T. Schultz Fund, the University of Florida’s FLAS program, 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.

1 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). E-mail: ap196@columbia.edu.

0095-0696/99 $30.00
Copyright © 1999 by Academic Press
All rights of reproduction in any form reserved.
merely historical interest. Rather, their answers should inform policies which will significantly affect the global stock of rainforests.

Much has been written about these questions, but economic understanding remains relatively rudimentary. Existing empirical research has focused heavily on population. This paper advances beyond previous empirical analyses by motivating the empirical work through an economic framework that encompasses many factors, and by innovative merging of satellite data on deforestation with an outstanding data set for the Brazilian Amazon. This permits empirical testing of the effects of many potential determinants (although not all, given data limitations). While this approach addresses neither the optimal rate of deforestation nor why particular policies have been or will be chosen, it does address how best to effect a given policy goal.

The major empirical finding is the significance of a number of variables suggested by the land-use model, in particular both land characteristics (such as soil quality and vegetation type) and factors which affect transport costs (such as the density of paved roads in a county as well as in neighboring counties, and the distance to major markets). In addition, development project policies appear to have independent effects, although provision of credit infrastructure does not. However, if policies (such as the location of a bank branch) result from some sort of maximizing behavior by government agencies, the policies themselves may be functions of other factors, including other explanatory variables. The possibility of such links suggests some caution in interpreting the estimated effects of policies. Finally, the population density does not have a significant effect on deforestation when many potential determinants are included. However, a quadratic specification reveals a more robust result: the first migrants to a county have greater impact than later immigrants. Thus the impact of a given population depends on its distribution.

The paper is organized as follows: Section 2 provides background. Section 3 reviews previous empirical analyses. Section 4 presents a model of land use and derives first a plot-level and then a county-level deforestation equation for estimation. Section 5 describes the data and presents specification issues. Section 6 presents the results. Finally, Section 7 concludes.

2. BACKGROUND

The Legal Amazon is an immense area, most of which was covered by forest at one time, and most of which remains forested today. Bordering a number of countries in the northwest corner of Brazil, it contains five million of Brazil’s total

---

2 The Brazilian Legal Amazon is made up of all the states in the north region of Brazil (the states Acre, Amapa, Amazonas, Para, Rondonia, Roraima, and Tocantins) plus parts of the states of Maranhao, Mato Grosso, and Goias. Its southern edge is the 16th parallel, and its eastern edge is the 44th meridian.

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 (Skole and Tucker [30]) is that all of the area was forested except for about one sixth of the region which is covered with a scrubby vegetation called cerrado and about 3% of the region which is varzea, or seasonally flooded land near rivers (see, for example, Goulding [11]).

Exactly how much deforestation has taken place is disputed (see, for example, Skole and Tucker [30], Fearnside et al. [9], and INPE [16]). However, the region is at most 10–15% deforested.
area of 8.5 million square kilometers; the latter area is larger than the continental United States. Rivers permeate the region, including the Amazon River, which traverses the region from west to east.\(^5\)

Since at least the 1960’s, occupation and use of the Amazon has been a policy goal. The military government in power from the 1960’s to the 1980’s promoted occupation of the region.\(^6\) Many felt that such empty land was an ideal “release valve” for pressures arising from a growing population.\(^7\) Many also felt or hoped that the region offered boundless resources, and those in power apparently shared those visions of progress and/or were happy to make use of such hopes.\(^8\)

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.\(^9\)

The actions taken appear to have stimulated occupation of the Amazon (although correlation may not indicate causality). The road network expanded significantly over the decade 1975–1985.\(^10\) In addition, total population more than doubled between 1970 and 1991, and urban population more than tripled. Finally, cleared forest area increased significantly.

3. REVIEW OF PREVIOUS 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.\(^11\)

A number of cross-country analyses correlate factors of interest with national measures of deforestation. These include: Lugo et al. [19], Allen and Barnes [1], Hecht and Cockburn [14], and Almeida [2].

\(^5\)This massive river is the confluence of runoff from higher areas to the south, west, and north of its basin.

\(^6\)Hecht and Cockburn [14] provide the following 1964 quotation from General Castello Branco: “Amazonian occupation will proceed as though we are waging a strategically conducted war.” They cite a “military philosophy of and strategy for regional development.” Motivations for this may have included the desire to discourage both incursions from countries bordering the Amazon region and the formation of domestic guerrilla opposition.

\(^7\)For example, Hecht and Cockburn [14] 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.”

\(^8\)For example, Hecht and Cockburn [14] cite the ideology of modernization, as in the phrase “Isto e um pais que vai para frente” (which might be translated: “This country 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”.

\(^9\)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 (certainly a relatively inhospitable environment itself). 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. This is mentioned further below.

\(^10\)This growth of the road network is nicely depicted in maps 8–11 in Almeida [2].

\(^11\)This section addresses only such empirical work. Pfaff [25, 26] discuss a broader set of related works.
A number of different results are of interest, such as Cropper and Griffiths’ “stage of development” interpretation of the significance of income levels, and Deacon’s measurement and use of government weakness or instability. The dominant result is that population is the most significant factor in explaining deforestation (although some authors qualify this in varied ways). This is partially explained by the fact that such analyses often use few explanatory variables (in the extreme, simply population alone).

For Brazil, which possesses an enormous area of rainforest, Almeida [2] provides much information at the level of the entire Amazon region, but focuses more on measuring and aggregating costs and benefits of agricultural colonization than on testing the importance of particular determinants of deforestation. Reis and Margulis [28] and Reis and Guzman [27] 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. While Reis’ work is related to the analyses here (including through use of his data), this paper advances beyond those works in two principal ways: first, systematic motivation of the empirical work using an economic framework; second, innovative merging of state-of-the-art satellite data on land cover (which is capable of providing multiple observations of deforestation over time for the entire Amazon region) with Reis’ outstanding county-level data set for the Amazon.

4. THE CONCEPTUAL FRAMEWORK

Underlying the empirical analyses below is an economic land-use model which suggests many possible determinants of deforestation. The underlying premise is simple: land is allocated between alternative uses in order to maximize returns. From this model, I derive a plot-level, land-allocation decision rule. I then adapt the derivation to generate a county-level decision rule which implies a deforestation equation to be estimated with the existing, county-level data.

4.1. An Economic Land-Use Model

At any point in time, a plot of land is allocated between different land uses to maximize profit:

\[
\max_{i,t} \quad \pi_t = P_{ijt} \cdot Q_{ijt} - R_{ijt} \cdot I_{ijt},
\]

Some within-country analyses have been done for other countries: Panayotou and Sungsuwan [23] find that deforestation in Thailand is driven by population density, wood price, income, and distance to Bangkok; Southgate et al. [32], for Ecuador’s Amazon region, first explain population with variables expected to affect “the prospect of capturing agricultural rents,” and then explain deforestation with population and other factors; Harrison [13], for Costa Rica, suggests differing effects of population in different regions, and questions whether population is a cause or a “shared symptom”; Kummer [17] is one of few empirical studies of deforestation to find only a small role for population growth in deforestation; and Parks and Murray [24] consider the Pacific Northwest.

Figures in Skole and Tucker [30] indicate that Brazil contains 30% of the world’s forested area.

Since this work was done, other papers have appeared with similar approaches, in particular Chomitz and Gray [5] on Belize, Cropper et al. [7] on Thailand, Nelson and Hellerstein [21] on central Mexico, and Geoghegan et al. [10], on the Patuxent Watershed in the Baltimore, MD–Washington, D.C. area.
where \( l \) is a given land use, \( i \) is a county, \( j \) is a plot of land within county \( i \), \( t \) is the year, and

\[
P_{ijt} = \text{plot-level prices for the vector of possible outputs from any given land use } l,
\]

\[
Q_{ijt} = \text{the vector of all outputs produced from this land use } l \text{ (including shelter)},
\]

\[
I_{ijt} = \text{the vector of inputs used in all types of production},
\]

\[
R_{ijt} = \text{plot-level prices for the vector of inputs used.}
\]

Assuming two land uses (cleared and uncleared) and privately optimal input choice yields

\[
\max_{l(ii)} V_{ijt}^l \quad \text{where } V_{ijt}^l = \max_{(l|l)} \pi_{ijt}^l
\]

and the following simple but useful static view of the structure of the clearing decision:\(^15\)

\[
\text{Choose } l_{ijt} = \text{cleared} \quad \text{iff: } V_{ijt}^\text{cleared} > V_{ijt}^\text{uncleared}
\]

### 4.2. Observable Variables and a Plot-Level Decision Rule

The \( P_{ijt} \) and \( R_{ijt} \) plot-level output and input prices above are in principle directly observable. However, in practice they may not be observed. In that case, it is useful that \( P_{ijt} \) may be functions of national-level output prices \( (p_i) \), transport costs from important markets to the plot \( (h_{ij}) \), unpaved roads \( (h_{2ij}) \), and rivers \( (h_{3ij}) \), as well as of plot distances from state markets \( (m_{1ij}) \) and from national markets \( (m_{2ij}) \), and local output demand shifters such as county population \( (n_{ij}) \) and development projects \( (d_{ij}) \). Also, the \( R_{ijt} \) may be functions of national-level input prices \( (r_i) \), transport costs to the plot \( (a \text{ function of } h_{ij}, m_{ij}) \), local input supply shifters such as county population \( (n_{ij}) \) and county credit conditions \( (c_{ij}) \), and finally plot vegetation type \( (v_{ij}) \), which may in particular affect the cost of clearing the plot.

The effects of these variables merit discussion. Both higher \( P_{ijt} \) and lower \( R_{ijt} \) raise returns to land uses involving clearing and should lead to more clearing and deforestation. Thus greater road density, greater river density, and lesser distances to markets should increase deforestation (by lowering transport costs and thus raising \( P_{ijt} \) while lowering \( R_{ijt} \)). Increased population may increase output demand and labor supply, again raising \( P_{ijt} \) while lowering \( R_{ijt} \) and thus increasing deforestation. Development projects increase output demand, while credit infrastructure increases input supply; both lead to greater deforestation. Finally, increased soil

\(^{15}\) Dynamic elements of such decisions surely arise, for instance regarding the option to delay clearing or the optimal rotation of timber or crops. Here I present the land-use framework in a static fashion to motivate the basic empirical specifications below in a clear and simple way. Given specific assumptions relatively common in such empirical land-use analysis, the empirical specifications below are also consistent with certain dynamic models.

The static model would be more problematic, and timber price trends would be centrally important, if logging had been the dominant source of deforestation in the Brazilian Amazon. However, census data indicate that pasture and then crops were the dominant uses of cleared land. Timber often resulted from clearing for cattle or crops.
quality \((q_{ij})\) increases productivity particularly in uses of cleared land, while vegetation types which are easier to clear lower costs and increase returns; both of these yield more clearing.

Although any given variable may affect the absolute returns to all land uses, empirically only variables' effects on the difference between the gains from cleared land uses and those from uncleared land uses can be observed.\(^{16}\) This motivates the following plot-level decision rule:

\[
\text{Choose } l_{ijt} = \text{cleared } \iff D^\text{cleared}_{ijt}(p_i, r_i; h_i, m_i; n_i, d_i; c_{ij}; v_{ij}; q_{ij}) > 0,
\]

where \(D^\text{cleared}_{ijt}(\cdot) = V^\text{cleared}_{ijt}(\cdot) - V^\text{uncleared}_{ijt}(\cdot)\).

A land use decision rule such as (4) leads to the equation to be estimated below. This approach permits the testing of many potential determinants of deforestation and a number of possible determinants are explored empirically below. However, limitations on available data constrain the set of potential determinants that can be empirically tested. Considering some omissions from the model above helps to identify particular factors which merit additional investigation.

One omission from the model is tenure conditions, variation in which could influence land-use decisions. Were the model above forward-looking, tenure conditions for any plot would affect the influence that expected future returns on that plot would have on current decisions.\(^{17}\) By leaving this factor outside of the model above, I have implicitly assumed a regime common to the entire region and which affected all locations in the same way.\(^{18}\) However, tenure conditions do appear to have varied to some extent across the region, and to some effect.\(^{19}\)

Also lacking in (4) and the empirical work are factors describing relevant conditions in years other than \(t\). One obvious omission is past land use, since forest regrowth is clearly not “instantaneous” on an annual time scale. However, as the clearing observations are separated by 10 years, I assume (by assuming a static

\(^{16}\)For example, high soil quality is expected to lead to more clearing in part because it is not necessary for high biological productivity of standing tropical rainforests, given efficient nutrient cycling processes within rainforests.

\(^{17}\)For instance, the possession of title permits the sale of land cleared to later arrivals, which provides another mechanism by which a land user considering the decision to clear could capture future returns from use of the plot. Given that, while title appears neither in the static model nor in the empirical work presented, it could still be part of the interpretation of the empirical results. If land title's frequency rises with proximity and access to towns with major markets, then the effect of title may be part of the full interpretation of the empirical access variables. For example, if the dominant effect of title were to increase the relative returns in cleared land uses, the effect of more title as access increases would be consistent with the hypothesized positive effect of increased access on clearing.

\(^{18}\)The property rights regime for the Legal Amazon during the period studied here was the following: the land is originally owned by the government, but can be claimed by private settlers. However, to get title a settler must “improve” the land in question. “Improvement” effectively seems to mean clearing the land and producing on it.

\(^{19}\)For instance, Alston \textit{et al.} [3] study the effect of land title at a more disaggregate level, in particular using survey data for four sites in the state of Para. They find that land title promotes agricultural investment and land value. Within related literature, Deacon [8] looks for empirical implications of government weakness or instability. Also, Larson and Bromley [18] theoretically consider dynamic incentives under different property rights regimes, while Mueller [20] focuses theoretically on the role of property rights in dynamic frontier evolution. For this paper, it would be useful to have reliable region-wide observed variation in tenure conditions.
decision context) that regrowth is sufficiently fast that uncleared land is a viable option for plots which were cleared in the previous observation.\footnote{Footnote 42 considers an attempt at testing this assumption. This issue motivates more attention to dynamics.}

With these qualifications, the decision rule in (4) 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.\footnote{The satellite land-cover data are easier than explanatory factors to obtain for greater geographic disaggregation.}

4.3. A County-Level Decision Rule

As no variables are observed at the plot level, (4) must be adapted to the existing observations at the municipio, or county level.\footnote{Thus the $P_{ij}, R_{ij}$ become $P_i, R_i$, and plot-level variables listed above (access to transport, distances from markets, soil quality, vegetation type) will appear in the empirical work below at the county level (as $h_i, m, v, q$).} 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 $i$. However, if all plots within a county were identical, at some threshold level of the factors driving land use, a whole county would shift from uncleared to cleared, or vice versa. That would be an obvious problem with the model, as in fact most counties contain both cleared and uncleared plots.

Instead I assume within-county, payoff-relevant, unobserved plot-level characteristics. More specifically, I define $\xi_{ijt}$ distributed across plots $j$ within county $i$ in year $t$.\footnote{For more discussion of this approach, see, for example, Chap. 20 of Green [12]. Note that the same within-county distribution, i.e., the same internal heterogeneity, is assumed for each county in each period. This transformation to county-level, empirical, land-use implications is also along the lines of Stavins and Jaffe [33].} For a plot, $\xi_{ijt}$ is the difference between the additional maximum profits (i.e., the addition to $V_{ijt}^{uncleared}$) attainable if plot $ij$ is uncleared in year $t$ and the additional maximum profits (i.e., the addition to $V_{ijt}^{cleared}$) attainable if plot $ij$ is cleared in year $t$. Thus, while it is not observed, for each plot in $i$:

$$D_{ijt}^{cleared}(\cdot) = D_{it}^{cleared}(\cdot) - \xi_{ijt}, \quad (5)$$

Rewriting (4):

Choose $l_{ijt} = \text{cleared}$ \iff $D_{ijt}^{cleared}(\cdot) > \xi_{ijt}, \quad (6)$

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

$$\%\text{Cleared}_{it} = F(D_{it}^{cleared}(\cdot)), \quad (7)$$

where $F(\cdot)$ is the cumulative distribution function of $\xi_{ijt}$. If $\xi_{ijt}$ is distributed logistically, and the $\%\text{Cleared}_{it}$ variable is rewritten as $y_{it}$, then from inverting the cumulative distribution function:

$$\ln(y_{it}/(1 - y_{it})) = D_{it}^{cleared}(\cdot). \quad (8)$$
5. DATA AND SPECIFICATION

5.1. Data and Variables

The data are for all municipios in the Legal Amazon. All variables are for the 1970 counties.\footnote{Municipios, or counties, 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.}

5.1.1. Land-Cover and Land Characteristics

The land-cover data are satellite observations for 1975, 1978, and 1988 from the University of New Hampshire (UNH).\footnote{The references to consult concerning this data are Skole and Tucker [30] and Skole et al. [31].} The original units of observation are aggregated to county level. The data provide an exhaustive breakdown of county area into: once and still forest, once forest but cleared, and never forested. Cerrado, a scrubby vegetation, is the main vegetation other than forest, covering about one sixth of the Legal Amazon. Clearing cannot be observed in cerrado. The county deforestation variable ($y_{it}$) is the fraction of the originally forested i.e., non-cerrado land cleared in year $t$.\footnote{Thus a county which never had any forest could not be defined as “deforested.”} It varies over space and time, while cerrado varies only over space.

The county soil quality measures ($q_i$) (also from UNH) are estimated densities of nitrogen and carbon in the soil. These are computed as weighted averages based on the fraction of the county in each soil-type region, from a cross-sectional map of soil-type regions developed using the RADAM soil-sample points. Thus, they vary only over space. The two soil variables available are almost exactly collinear, so only one of the variables (nitrogen) is used in the regressions.

5.1.2. Transport Costs

County road densities [paved ($h_{1i}$) and unpaved ($h_{2i}$)]\footnote{It may be possible to further separate both paved and unpaved roads into federal and state subcategories.}, river densities ($h_{3i}$), and distances from county seat to state ($m_{1i}$) and national seats ($m_{2i}$) come from maps provided by Brazilian government agencies. Road observations exist for 1976 and 1986. Thus roads vary over space and time, while rivers and distances vary only over space. While the dimensions of the rivers included may vary significantly, the rivers included all satisfy the “Class A navigability criterion” (i.e., they exceed a minimum depth for a minimum period of time during a typical year).

5.1.3. Government Actions Other Than Roads

County credit data indicate how many Banco do Brasil (BdB) agencies existed in the county in 1985 and in what year the first BdB agency appeared. This was used to construct credit-agency density ($c_i$), which varies over space as well as in a
particular way over time.\textsuperscript{28} Country data from SUDAM (Superintendency for the Development of Amazonia) provide, for each development project, quantitative measures for 1985 plus certain dates (e.g., first year of implementation). The information available lists 247 projects, yielding 234 observations after missing values, with a mean area of 330 km\textsuperscript{2}. Yokomizo \textsuperscript{35} suggests that the bulk of these projects' impacts occurred in the southeast of the region. The 1975 values for projects' areas were constructed as was done for the credit agencies. Then a county-level development-project-area density ($d_{it}$) was constructed by adding the areas of all the projects in a given county and dividing by county area.

I do not have data on colonization or land titling projects, although they are often discussed.\textsuperscript{29} Also, to this point I do not use information on dams.\textsuperscript{30} Generally, additional systematic data for government actions including those just discussed would be useful.

5.1.4. Census Data

Plot-level output and input prices ($P_{ijt}$, $R_{ijt}$) are not directly observed. Also, concerns principally about quality but also about endogeneity ruled against using measures of county-level prices.\textsuperscript{31} The one county price used ($w_{it}$ of $R_{it}$) is an average industrial wage from the Industrial Census, for 1975, 1980, and 1985. This is an "outside option" for those working on cleared land.

\textsuperscript{28}For credit agency density, 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 was assigned to 1975 and divided by the county area, yielding the 1985 value. If the first agency had not yet appeared by 1975, a zero was assigned. This involves a stronger assumption about agencies in 1975 than would be required for a credit infrastructure existence indicator variable. However, this variable is used in the regressions to preserve the information on number of agencies.

\textsuperscript{29}Whether the lack of this data implies that key factors behind most immigration to the Amazon have been omitted from the analyses is debatable. One point is that, at least from general impressions, spontaneous immigration (responding to conditions such as soil quality and transportation) appears to have greatly outnumbered official, planned immigration within colonization programs. For further discussion of such programs, see, for example, Almeida \textsuperscript{2} on efforts by the colonization agency INCRA within the "national integration program."

\textsuperscript{30}The information available includes the name of the county in which the powerhouse is located and the total inundated area, but not inundated areas by county. The total inundated areas for the six dams listed in IBGE \textsuperscript{15} is 5500 km\textsuperscript{2}. This is under 5% of total clearing in 1980. However, the construction of dams could 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. If data by county existed, such an effect could be seen within an estimated elasticity.

\textsuperscript{31}Econometrically, I could regress county clearing on county output price, but if the county output supply curve shifts, then the observed price used to explain clearing may itself be affected by shifts in clearing and output supply. Thus I might instead instrument for the price with factors that shift demand but do not shift the supply curve. One set of exogenous demand shifts are those which drive prices in major distant markets. However, this variation exists only over time, and thus I cannot use it empirically as I have only two points in time. The conclusion includes a related comment regarding international timber prices. If distant market prices matter locally, then so do the transport costs from distant markets. I have included variables playing that role (distance as well as access to roads and rivers). Thus I can capture cross-sectional variation in the effective level of external demand for output. In case county-level prices are to some extent locally determined, I have also included local demand shifters (both population and the presence of projects). The concerns about the quality of the prices data dictate using these as proxies, not as instruments.
County population data \((n_c)\) come from the Brazilian Demographic Census, for 1970, 1980, and 1991. 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.

5.2. Specification Issues

5.2.1. "Neighboring-County" Variables

The following assumption is implicit above: the factors that affect land use in a given county are those which describe that county. As commonly interpreted, this assumes that factors describing neighboring counties do not matter. However, a paved highway running through one county might be hypothesized to affect transport access for plots in neighboring counties. In order to test this hypothesis, some regressions based on \(8\) will include "neighboring-county" versions of variables such as roads, population, projects, and credit. These are unweighted averages of the values for these variables for the counties which share a border with the county in question.\(^{32}\)

5.2.2. The Frontier Development Process

Ongoing frontier development may involve many decisions which affect each other, e.g., not only the clearing choices above but also migration decisions and government policy choices. An implication of relevance for the regressions below is that factors being tested as determinants of land use might themselves be in part reactions to other determinants or even to land use choices.

A common suggestion is that population is "endogenous." However, the existing data dictate the use of a lagged population measure, so that statistical endogeneity seems less likely. Further, many of the stories about the development process do not in fact imply that population results from clearing itself, but rather that it might respond to other explanatory factors, such as roads or rivers. Such a relationship would suggest not endogeneity but rather multicollinearity.

Another common suggestion is that agencies maximize in some way when making policy choices (such as road, project, or bank-branch locations), and as a result observed policies may be endogenous (e.g., caused by or jointly determined along with land clearing). However, regarding empirical work, note that the data dictate the use of lagged explanatory measures. Further, it is again important to distinguish between endogeneity and multicollinearity: for example, a Banco do Brasil administrator might deem a high number of transactions per bank branch as success, and thus attempt to maximize success by locating Amazon bank branches near concentrations of population; such a branch location rule would imply not endogeneity but rather a collinearity between a pair of explanatory variables, as branch location reacts to population.

\(^{32}\) Some determinants of land uses common to neighboring counties may be unobservable. If so, error terms may not be independent across counties. Neighboring counties might be expected to be more alike than randomly paired counties. If so, residuals should be corrected in order to obtain proper inferences. See, for instance, Anselin [4].
However, an attempt at formal instrumentation is warranted. One instrument for roads could be lagged roads. Instruments for population in an Amazon county are weighted averages across migrant-origin states (using as weights the share of the county's migrants from each state) of the industrial wage, employment rate, and labor intensity in agriculture in the states of origin. Higher wages and employment in origin-states are expected to lower levels of emigration to the Amazon, while higher labor intensity in agriculture is expected to raise emigration, as it implies a greater displacement of labor from shifts such as agricultural mechanization.

6. RESULTS

Putting the available data into the $D_{i,t}^{\text{cleared}}$ term in (8) yields the basic equations for estimation. The testable factors are paved road density, unpaved road density, river density, distances from county seat to the state and national seats, development project area density, Banco do Brasil branch density, average industrial wage, density of nitrogen in the soil, percentage of county area in cerrado, and population density. In this section, I discuss the results of estimation.

Table I presents the basic results from testing a number of potential determinants of deforestation. Table II presents two types of "extensions" of the basic results: the first column addresses two points about possible linkages between explanatory variables, and the second motivates further exploration of dynamic processes.

6.1. Principal Results and Interpretation

The major empirical findings in Table I confirm the prediction from the land-use model that a number of factors are potentially important determinants of deforestation. Both own- and neighboring-county paved roads are significant and positive (i.e., increase deforestation). The distance to major national markets is significant and negative, and land characteristics such as soil quality and vegetation type (i.e., cerrado) are significant and have the expected signs. These findings are robust to the inclusion of time and/or state dummies and to using cross-sections.

Some policies other than roads seem to have independent effects: development projects significantly increase deforestation, although infrastructure for distributing subsidized credit does not have a significant effect. However, for both policies, the possibility that government decision rules imply relationships between policies and other explanatory variables suggests looking beyond the basic empirical find-

33 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 and Gray [5] attempt to instrument for roads, and find that controlling for soil quality, it makes little difference.

34 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 [34].

35 The neighbor version of projects is not significant. This might be taken to suggest that while these projects involved local clearing, the idea that projects would spur outwardly-spreading regional development was incorrect.
TABLE I
Many Possible Determinants of Deforestation

<table>
<thead>
<tr>
<th>Variables</th>
<th>Coefficients</th>
<th>Marginal effects</th>
</tr>
</thead>
<tbody>
<tr>
<td>Paved-road density (km/sq km county area)</td>
<td>0.0015**</td>
<td>(0.0003)</td>
</tr>
<tr>
<td>Average neighboring-county paved-road density</td>
<td>0.0024**</td>
<td>(0.0009)</td>
</tr>
<tr>
<td>Unpaved-road density (km/sq km county area)</td>
<td>-0.0002</td>
<td>(0.0004)</td>
</tr>
<tr>
<td>Average neighboring-county unpaved-road density</td>
<td>-0.0010</td>
<td>(0.0010)</td>
</tr>
<tr>
<td>River density (km/sq km county area)</td>
<td>0.0006</td>
<td>(0.0005)</td>
</tr>
<tr>
<td>Average neighboring-county river density</td>
<td>-0.0051**</td>
<td>(0.0017)</td>
</tr>
<tr>
<td>Distance from county seat to state seat (km)</td>
<td>0.0020**</td>
<td>(0.0009)</td>
</tr>
<tr>
<td>Distance from county seat to national seat (km)</td>
<td>-0.0013**</td>
<td>(0.0005)</td>
</tr>
<tr>
<td>Project-area density (sq km/sq km county area)</td>
<td>0.0014**</td>
<td>(0.0004)</td>
</tr>
<tr>
<td>Average neighboring-county project-area density</td>
<td>0.0051</td>
<td>(0.0090)</td>
</tr>
<tr>
<td>Credit-agency density (#/sq km county area)</td>
<td>0.0213</td>
<td>(0.0208)</td>
</tr>
<tr>
<td>Average neighboring-county credit-agency density</td>
<td>0.0423</td>
<td>(0.0424)</td>
</tr>
<tr>
<td>Industrial wage</td>
<td>0.0073</td>
<td>(0.0139)</td>
</tr>
<tr>
<td>Density of nitrogen in soil</td>
<td>0.0208**</td>
<td>(0.0043)</td>
</tr>
<tr>
<td>Cerrado-area density (sq km/sq km county area)</td>
<td>-5.019**</td>
<td>(0.6924)</td>
</tr>
<tr>
<td>Population density (#/sq km county area)</td>
<td>0.0100*</td>
<td>(0.0044)</td>
</tr>
<tr>
<td>Squared population density</td>
<td>-0.009*</td>
<td>(0.0004)</td>
</tr>
<tr>
<td>Average neighboring-county population density</td>
<td>-0.0023</td>
<td>(0.0134)</td>
</tr>
<tr>
<td>Squared average neighboring-county</td>
<td>0.0001</td>
<td>(0.0025)</td>
</tr>
<tr>
<td>population density</td>
<td>0.0001</td>
<td>(0.0025)</td>
</tr>
<tr>
<td>Constant</td>
<td>-3.046**</td>
<td>(1.008)</td>
</tr>
<tr>
<td>Adjusted $R^2$</td>
<td>0.371</td>
<td></td>
</tr>
</tbody>
</table>

This regression uses two, pooled cross-sections and has 480 observations. Note the quadratic form of population. Parentheses contain corrected standard errors, and *, **, and *** indicate significance at the 90, 95, and 99% levels.

For this table, this variable's coefficient and standard error are multiplied by 10,000 (for ') or by 100 (for 'd').

Summarizing the discussion just below, I conclude that there is evidence that both policies significantly increase deforestation, but also that any conclusion of significant independent effects must be qualified given all the evidence.

The conclusion from Table I's pooled regression may be placed in question by the result that projects have a significant effect in the first cross-section, but are insignificant in the second. Perhaps projects do not have their own effect, and themselves reflect other factors. For instance, SUDAM may have attempted to maximize project success by placing the first wave of projects in the areas which most possessed a set of beneficial qualities which are unobserved here. If so, and if having those qualities leads to more clearing, the measured effects of early projects may in fact reflect the qualities of these choice locations. However, as the governmental impetus behind such projects is commonly said to have diminished over this period, and is likely to be poorly measured here, these results may reflect a significant effect of a strong push to develop projects and little effect of a later, weak impetus. Further, projects may have a significant effect on clearing but also lead to additional development; thus, other explanatory variables may respond to...
TABLE II
Linked Explanatory Variables and Dynamic Processesa

<table>
<thead>
<tr>
<th>Variables</th>
<th>Dropping population altogether</th>
<th>In differences (RHS when possibleb)</th>
</tr>
</thead>
<tbody>
<tr>
<td></td>
<td></td>
<td>0.0016** (0.0003)</td>
</tr>
<tr>
<td></td>
<td></td>
<td>0.0023** (0.0009)</td>
</tr>
<tr>
<td></td>
<td></td>
<td>0.0001 (0.0004)</td>
</tr>
<tr>
<td></td>
<td></td>
<td>0.0009 (0.010)</td>
</tr>
<tr>
<td></td>
<td></td>
<td>0.007 (0.0004)</td>
</tr>
<tr>
<td></td>
<td></td>
<td>0.0050**, (0.0015)</td>
</tr>
<tr>
<td></td>
<td></td>
<td>0.0191 (0.0009)</td>
</tr>
<tr>
<td></td>
<td></td>
<td>0.014**, (0.0005)</td>
</tr>
<tr>
<td></td>
<td></td>
<td>0.015**, (0.0003)</td>
</tr>
<tr>
<td></td>
<td></td>
<td>0.020, (0.0090)</td>
</tr>
<tr>
<td></td>
<td></td>
<td>0.0298**, (0.0114)</td>
</tr>
<tr>
<td></td>
<td></td>
<td>0.0467, (0.0398)</td>
</tr>
<tr>
<td></td>
<td></td>
<td>0.0057, (0.0142)</td>
</tr>
<tr>
<td></td>
<td></td>
<td>0.0205**, (0.0043)</td>
</tr>
<tr>
<td></td>
<td></td>
<td>5.124**, (0.6873)</td>
</tr>
<tr>
<td></td>
<td></td>
<td>0.0065, (0.0182)</td>
</tr>
<tr>
<td></td>
<td></td>
<td>0.0005, (0.0010)</td>
</tr>
<tr>
<td></td>
<td></td>
<td>0.0672*, (0.0375)</td>
</tr>
<tr>
<td></td>
<td></td>
<td>0.0082*, (0.0048)</td>
</tr>
<tr>
<td></td>
<td></td>
<td>2.901**, (1.000)</td>
</tr>
<tr>
<td></td>
<td></td>
<td>0.373</td>
</tr>
</tbody>
</table>

*Column 1 uses two, pooled cross-sections and has 480 observations. Column 2, in differences, has 218 observations. Parentheses contain corrected standard errors, and , *, and ** indicate significance at the 90, 95, and 99% levels.

For any variable, a a indicates that in column 2 the variable enters the regression as a level (i.e., not as a difference).

For this table, this variable’s coefficient and standard error are multiplied by 10,000 (for a ) or by 100 (for a ).

Projects. If so, at later dates an independent effect of projects may be indistinguishable.

Despite the fact that variation in the impetus from credit infrastructure is likely to be poorly measured here, the credit variable is significant in the first column of Table II, in which population has been dropped. The contrast with Table I appears to reflect the collinearity of bank branch locations and population. Such collinearity is not surprising if, as suggested above, Banco do Brasil attempted to maximize branch success by placing branches where there are people to take advantage of them, i.e., in counties with sufficient population. It does, however, muddy any conclusion regarding credit’s independent effect on deforestation.

Finally, population density and neighboring-county population density are not significant when they appear in a Table I-like regression as levels. However, the quadratic specifications for own and neighboring population in Table I show a significant effect: the first people entering an empty county have more impact than

---

The population and agency variables are significantly positively correlated (a correlation slightly above 0.6).
later immigrants, \(^{37}\) implying that the land-use impact of a given population depends on how that population is distributed.

### 6.2. Additional Results and Interpretation

The most surprising result in Table I is that for rivers, \(^{38}\) Given the prior, greater understanding is needed. Perhaps transport services from rivers and roads differ significantly. For example, the difficulty of creating fixed ports on the Amazon (with its vast floodplains) may dictate that the river affects migration but does not foster commercial traffic or affect returns. \(^{39}\) Interestingly, in the second column of Table II rivers have a strong effect of the expected sign (more below).

Another result worth mentioning, although not in the tables, is that instruments for population generated from conditions in migrants' states of origin perform poorly in a first-stage regression. The two-stage regression yields high standard errors, and little can be learned.

Cerrado's result, while expected, may also merit further discussion. \(^{40}\) The claim here is through low clearing costs cerrado can draw pressure away from forested areas, reducing forest clearing. This claim is justified in two ways. First, Brazilian law requires that landowners in the Amazon leave some percentage of their land uncleared, \(^{41}\) All else equal, given clearing costs we might expect the law to be satisfied by clearing cerrado and leaving denser forest uncleared. Second, the level of output produced in a county may raise local input prices and lower local output prices (unlike in the model, where all the prices are given). In that case, if lower-clearing-cost areas are used first, having more of them lowers the returns that apply to forest clearing.

\(^{37}\) For the own-county, significant population quadratic, these results imply that the marginal effect of additional population equals zero at a density of 556 people per square kilometer. In the 1980 demographic observations, only two (six) counties had densities over 200 (100). The median county population density was 4.7.

\(^{38}\) This result is not driven by the inclusion of small streams which could not be expected to affect transport costs. All rivers included here satisfy the "class A navigability criterion," and are thus significant transport options.

The distance to state seat may also be a surprise. However, it is less clear that this represents an increase in the overall costs of access to markets. In addition, the main results do not depend on its inclusion.

\(^{39}\) Further, although the varzea (flooded areas) may be nutrient rich when waters subside, they may not attract settlers given the flooding. While not likely to be the case for the Amazon's flat plains, in some locations rivers may be immediately surrounded by relatively steep slopes, which may serve to discourage land clearing for production. I am grateful to many for suggestions regarding the effects of rivers. Of course, all errors here are solely my own.

\(^{40}\) A nonuniform distribution of cerrado, e.g., location far from markets, could explain this negative result. However, cerrado is if anything more heavily concentrated in the south and east of the region, closer to the rest of the country (the cerrado variable is highly negatively correlated with distance to national seat, i.e., distance out into the frontier). Thus if anything such nonuniformity might be expected to produce a positive result for cerrado. Another interpretation might follow if this scrubby vegetation existed where soil is poor, or where little rain falls, i.e., in places where returns might be relatively low. However, the soil quality and cerrado density variables are essentially completely uncorrelated (note that nutrient cycling allows lush rainforest to grow on relatively poor soil).

\(^{41}\) Surely some violation occurs, but in a sample of 206 small landholders in Para, Alston et al. [3] find that the mean percentage of farm cleared is 40% (with a standard deviation of 24%).
Finally, given the theoretical limitations of the static model which motivated the analyses above, I consider the need to empirically explore dynamics. With observations for only two points in time, there are obvious empirical limitations. One question is whether the effects of changes in explanatory factors over time can be distinguished from the effects of cross-sectional differences among counties. While a differences regression does not permit estimation of such fixed effects, it should remove such cross-sectional differences in estimating the other effects.\(^{42}\)

In a differences regression, although the river, distances, soil and vegetation variables must drop out, the other basic results remain roughly the same. The second column of Table II presents a twist on this regression which to some extent tests for a simple dynamic process by re-introducing as levels the variables of interest that were dropped (all other variables remain in changes). An interesting result is that the distance from important national markets has a significant positive effect on the changes over time in fraction cleared. This does not necessarily contradict the evidence in Table I that distance raises transport costs and lowers clearing. Rather, it appears to be an additional, dynamic result indicating a movement of the development frontier up into the Amazon region. Alongside theoretical assertions concerning dynamics and the process of frontier development, this finding motivates further consideration of dynamics.

7. CONCLUSION

This paper analyzed the determinants of deforestation in the Brazilian Amazon. I derived a deforestation equation from an economic model of land use, and tested a number of potential factors using county-level data for the period 1978–1988. Previous empirical work has focused on population as the main determinant of deforestation. This work advanced beyond previous empirical analyses by motivating the empirical work through an economic framework that encompasses many factors, and by innovative merging of satellite data on deforestation with an outstanding data set for the Brazilian Amazon. This permitted empirical testing of the effects of many potential determinants (although not all, given data limitations). While this approach does not address either the optimal rate of deforestation or the political economic question of why certain policies are chosen, it does address how best to effect a given policy goal.

The major empirical finding was the significance of a number of variables suggested by the land-use model, in particular both land characteristics (such as soil quality and vegetation type) and factors which affect transport costs (such as the density of paved roads in a county as well as in neighboring counties, and distance to major markets). In addition, development project policies appeared to have independent effects, although provision of credit infrastructure did not. However, if policies (such as the location of a bank branch) result from some sort

\(^{42}\) If there are fixed cross-sectional differences, a second-period cross-section which uses first-period deforestation as an explanatory factor may offer some sense as to whether the fixed effects are significant explanatory factors. In fact, such a regression yields a highly significant effect of lagged clearing and greatly increased explanatory power. However, another interpretation of such a regression would be that prior clearing constrains the current land-use choices (even 10 years later), for instance due to slow regrowth. Such linkages would imply an irreversibility within the decision to clear, and are another motivation for further exploring the application of a dynamic model to data.
of maximizing behavior by government agencies, the policies themselves may be functions of other factors, including other explanatory variables. The possibility of such links suggests some caution in interpreting the estimated effects of policies. Finally, the population density did not have a significant effect on deforestation when many potential determinants were included. However, a quadratic specification revealed a more robust result: the first migrants to a county have greater impact than later immigrants. Thus, the impact of a given population depends on its distribution.

There are a number of potential extensions of this work, many of which would involve additional data. Such data includes disaggregated land-use information (e.g., on crop, ranching, and timber areas) as well as spatially disaggregated information (i.e., at a subcounty level) which would permit better use of high-resolution satellite images. As suggested above, it would be useful to have reliable, region-wide observed variation in tenure conditions. In addition, as both theory and some of the analyses above suggest that the dynamics of frontier expansion and deforestation merit further empirical work, additional observations for all variables over time could also be extremely beneficial. For instance, additional observations over time would facilitate investigation of the effects of determinants of deforestation which shift primarily or only over time, such as the world timber price or the national economic growth rate.

In terms of the policy implications of these analyses, the result that roads appear to be significant determinants of deforestation, regardless of what other factors are included, suggests one channel for affecting the rate of deforestation. Even the possibility that past government decision making implies that roads are partially a function of other variables might not matter much for future policy choices: 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, adding the quadratic population result (that the per-person impact of population on deforestation is lower in areas of concentrated population) to the roads result could suggest particular policies, conditional of course on policy objectives. Past policies may have been designed to distribute unclaimed government land to help the landless, and perhaps future policies will do the same. However, given external pressures and possible incentives, perhaps future policies will aim to achieve a desired level of regional development, output, or employment with minimal deforestation. Given the latter objectives, this paper’s results suggest that the government might do better to build good roads to existing cities instead of to sparsely populated areas, and to use subsidies for urban employment instead of for rural agriculture.

REFERENCES


