Underlying reasons and spatial heterogeneity of deforestation in Legal Amazon

Tomas Jusys
University of the Balearic Islands
Department of applied economics
PhD candidate in tourism and environmental economics

Jusys.tomas@gmail.com

March, 2013

Abstract. Deforestation in tropical rainforests became quite intensively researched topic among scholars. However, academic literature on forest clearings in Legal Amazon provide with conflicting findings. In their literature review paper on deforestation, Angelsen et al (1999) question some well known theses regarding deforestation (population, poverty, economic growth, timber market, tenure security and intensification theses). This study uses these theses as a guidance to form a linear model taking into consideration endogeneity issues and recognizes the possibly spatially heterogeneous nature of determinants affecting deforestation. Main contributions of this paper are twofold: it investigates empirically several most important theses of deforestation in Legal Amazon paying detailed attention to endogeneity problems, and, also, while exploring spatial heterogeneity, it additionally incorporates road and time distances into geographically weighted approach rather than only Euclidian distances and draws conclusions on whether distance measurement technique influences the results significantly. The results revealed 5 important contributors to deforestation in the global model – cattle ranching, agricultural activities, timber values, agricultural credit and road network. Most findings were in line with the observations made by Angelsen et al (1999). As for spatial approach, the study was unable to provide with the definite answer whether significant geographical differences exist, however, it suggests that cattle ranching, agricultural credit and road network variables may exhibit high variability in space. Finally, the most crucial finding is that distance measurement technique influences the distribution of local coefficients very significantly.

Keywords. Deforestation, Legal Amazon, endogeneity, two-stage least squares, spatial heterogeneity, geographically weighted regression.

Introduction

Deforestation in tropical rainforests is one of the major environmental hazards in today's world. It has far reaching consequences in many spheres of life, such as contribution to climate change, loss of biological diversity, loss of potential medical information contained in flora and fauna of tropical rainforests, soil erosion, and, most importantly for tourism businesses, loss of potential income from ecotourism.

Although tree cuttings started since ancient history of man, only in recent decades, thanks to advances in technology and evolution of complex economic ties, it became a worldwide problem. Every year deforestation claim around 13 million hectares of forests (Obersteiner et al, 2006), mostly affecting tropical rainforests in Southeastern Asia, Central Africa and Amazon basin.

This study is motivated by three goals. Firstly, it aims at testing if several theses, questioned by Angelsen et al (1999), namely, population, poverty, economic growth, timber market, tenure security and intensification theses, hold for Legal Amazon. Second interest is to find underlying causes of deforestation in Amazon basin. The third goal is to assess spatial heterogeneity in the region of Legal Amazon. The motivation to choose Legal Amazon as a study area lies in the fact that Amazon basin is the largest area of tropical rainforests, also in the richness of data provided by the government of Brazil and in conflicting findings of other authors, suggesting that further research is needed.

This paper will contribute to the existing literature in two ways. Firstly, it will be based on abovementioned theses which will guide which variables to consider as determinants of deforestation. In this way it is much more likely that the study will avoid omitted variable bias, potentially present in many other researches concerning deforestation in Legal Amazon. Also, endogenous nature of three variables will be challenged. Secondly, the paper will employ geographically weighed regression to assess spatial heterogeneity. Although this approach was already taken by Oliveira et al (2011), the authors used only Euclidian distances in computations of local coefficients. This study will consider Haversine (similar to Euclidian, the difference lies only in the fact that it considers Earth's curvature), road and time distances in the application of geographically weighted regressions and will draw conclusions on whether the choice of distance measurement method influences the results.

The results of this paper can equip Brazilian government with more useful information when applying laws or making other decisions regarding deforestation. The study should also be in the eyesight of tourism businesses, which have a great interest in saving Amazon forest stemming from increasing demand in recreational activities, such as wildlife watching. Finally, the findings should also be important for biologists, pharmacologists or climatologists, as well as environmental economists for gaining new ideas for their researches and for comparing the findings.

The paper will be organized as follows: it will commence with literature review, next section will present data and variables, further the paper will explain methodology, following by the analysis of the results and, finally, the last section will draw conclusions.

Literature review

This section will present several theses of deforestation emphasized by Angelsen et al (1999) and then will focus on reviewing the most relevant literature for this study.

Population thesis does not seem to hold for deforestation. Although the results in academic papers are quite conflicting (for instance, Hargrave (2009) finds that number of inhabitants has significant effect on deforestation, while Oliveira el at (2011) conclude that population density does not affect deforestation processes), Angelsen et al (1999) insist that those authors who find that population is a driving force of deforestation, failed to address the endogenous nature of population variables.

Poverty thesis is also under question and, according to Angelsen et al (1999), does not hold for deforestation. They argue that more affluent individuals are in better financial position to clear forests, since this activity requires investment. They also suggest treating poverty variables as endogenous in the model. Deforestation creates employment and generates income. Therefore, it may also be a determinant of affluence of individuals. Some studies consider environmental Kuznets curve approach. However, the findings are conflicting. Oliveira et al's (2011) conclusions are in line with Angelsen et al (1999). However, Araujo et al (2010) concluded the opposite – the results confirmed "U" shaped curve of GDP per capita, which affects deforestation.

Angelsen et al (1999) also argue that economic growth increase deforestation. They insist that augmentation in economy stimulates agricultural and timber prices thus boosting forest clearings. The findings by Hargrave et al (2012) support this view: the study found that changes in real GDP have a positive effect on changes in deforestation.

Timber market thesis lies on the notion that lower timber prices discourages proper forest management. However, the opponents of this view argue that lower timber prices should reduce logging activities and hamper agricultural expansion.

Tenure security thesis states that privatizing natural areas should protect them, since the owners should have economic incentives to manage their lands in a sustainable way. However, private lands also create easier access to forest recourses. In case of Legal Amazon, the empirical study conducted by Araujo et al (2010) suggest that land insecurity boosts deforestation, though the results are not rigid enough, as the conclusion depends on the model applied.

Intensification thesis is based on an idea that technological improvements in agriculture boost deforestation. However, Angelsen et al (1999) note that this relation may be both positive and negative. Technological advances in agriculture, such as acquisition of new tractors, may increase profitability of farming, thus creating more pressure on forests. On the other hand, if changes are more labour intensive, land management may become more productive, leading to reduction in the scale of loggings.

The lack of comprehensive theory leads to wide variety of determinants used by researchers to explain deforestation. Deforestation variable itself comes in different ways. Oliveira el at (2011) chose annual increments of deforested area, Araujo et al (2010) used natural logarithm of deforested area and Hargrave et al (2012) worked with differences in natural logarithms of cleared land.

As for independent variables, cattle ranching can be found almost in any paper that aims at identifying causes of deforestation. The variable, included in the models either directly as a number of cattle animals or though market mechanism as meat price, almost in all cases proved to be the key driver of deforestation. Academic literature tells that from 60 to 80 per cent of deforestation is caused solely by cattle business. Another common determinant present in many researches is agricultural activities. Though, the proxies differ. For instance, Oliveira el at (2011) use soybean and sugar cane acreages, Hargrave et al (2012) select soy price and Araujo et al (2010) – number of agricultural establishments. No matter which proxy is used, authors find that agricultural activities have positive and statistically significant effect on deforestation (except variable for sugar cane, which was found to be statistically insignificant). Next common variable concerns timber market. It is included either as wood price or as extraction of wood measured in cubic meters. Oliveira el at (2011) include two distinct variables to explain timber market: extraction of timber and forestry. Although timber market was found to affect deforestation, the findings are contradictory regarding sign of coefficients. Rural credit variable also often is in a list of determinants. It is included in the models as the amount of money granted, though Hargrave et al (2012) treat it as natural logarithm of real credit per square kilometer of municipality. The results are conflicting. Rural credit is either found to influence deforestation positively or not to have effect. One more commonly considered variable among researchers is population, measured as number or density of inhabitants. Most results state that population does not influence deforestation.

One interesting study regarding links between investments in roads and deforestation in Legal Amazon belongs to Pfaff et al (2007). The authors conclude that higher investment in road infrastructure boosts deforestation. However, Kaimowitz et al (1998) note that accessibility variables can be endogenous.

Hargrave et al (2012) also consider indigenous, protected and settlement areas as potential determinants of forest clearing. In this study they do not find any statistically significant effect on deforestation. However, in another paper Hargrave (2009) finds that protected areas tend to reduce deforestation. The same authors also take into consideration fine intensity variables, which proved to be statistically significant and reduce deforestation.

Some other less often used variables include amount of rainfall, slopes of terrain (though this variable is included only in papers which consider small geographical area for their study, since otherwise getting data is complicated) or the distance to capital city.

As for computational methods, almost all researchers employ either OLS or two-stage least squares regressions. Also, the majority of authors tried to tackle endogeneity problem. Araujo et al (2010) considered land insecurity to be endogenous and therefore applied two-stage least squares method with one instrument. Hargrave et al (2012) addressed the problem of endogeneity in two ways. Firstly, they chose environmental fining as endogenous variable and used one instrument to alleviate the problem (through two-stage least squares). Secondly, they applied general method of moments (in form of differences), though this time they treated many variables as endogenous, including environmental fining, wood prices, protection and settlement policies and economic activity. In the other paper Hargrave (2009) divided data on environmental fines by area of cleared land to avoid endogeneity. Pfaff et al (2007) used lagged variable of road investments to tackle the problem. The idea behind this notion is that deforestation this year cannot affect past investments into road infrastructure.

Oliveira el at (2011) suppose that endogeneity may exist between deforestation and income. To test it they apply Durbin-Wu-Hausman test. However, despite concluding that endogeneity is present, they do not try to tackle it.

Spatial heterogeneity is another important issue considered by several studies. Oliveira el at (2011) used geographically weighted regression. However, the model they applied assumes exogeneity in all independent variables. Additionally, the authors used GWR 3.0 software, which has a limitation of being incapable to perform computations with other than Euclidian distances. The authors find spatial variation in local coefficients of cattle ranching, soybean and sugar cane acreages and non timber products, although the latter variable was found to be statistically insignificant in global regression. Hargrave (2009) included state dummies to capture local characteristics. However, the results suggested no differences across states.

Although the literature regarding deforestation studies in Legal Amazon is abundant, there are some critical points to discuss. Most studies seem to suffer from omitted variable bias (unless they try to solve the problem by using instrumental variables), since some studies do not consider variables that other researchers find to be one of the drivers of deforestation in Legal Amazon. For instance, Oliveira el at (2011) do not include road variable in their model, Hargrave (2009) does not consider timber market and road infrastructure, while Araujo et al (2010) miss the data on credit, roads and timber. Another note is that many authors treat only one variable as endogenous in their models, though it is a very restrictive approach. All these different paths different authors take lead to conflicting findings. The only consensus across the literature is that cattle ranching and agricultural activities boost deforestation significantly and population does not play a role in explaining forest clearings.

Data and variables

The data used covers 503 municipalities in Legal Amazon. It is a cross sectional analysis, which uses data from year 2009, with exceptions for data of population (the data covers year 2010) and tenure (covers year 2006). The variables included in the model are chosen based on theses questioned by Angelsen et al (1999). Additionally, the model includes data on roads, following findings in the literature, especially Pfaff et al (2007), that road network is an important contributor to forest clearings. Few variables will be included in the analysis as instrumental variables to alleviate endogeneity issues. To be specific, exogenous variables are economic growth, cattle ranching, agricultural activities, timber values, rural credit for agriculture, tenure and population growth. Three variables – gross domestic product per capita, number of inhabitants and paved roads – will be treated as endogenous following recommendations by Angelsen et al (1999) and Kaimowitz et al (1998). Finally, four instrumental variables will be used to tackle endogeneity. Those include gross domestic product per capita in year 2008, gross domestic product in year 2009, number of inhabitants in year 1991 and in year 2000.

This study will consider whole agricultural area instead only soybean or sugar cane acreages, as many other studies did. As every temporary plant needs land to develop, considering only lands occupied by soybean would be too restrictive.

The model assumes little changes in population and tenure variables over time. This implies that, although data for year 2009 is not available, existing data should reflect the situation in year 2009 appropriately enough.

Data on roads was collected by the author. It was done by measuring the length of all roads within every municipality on computer screen (in pixels). The length of roads along the borders was divided equally between municipalities that share the border road. Roads under construction were not considered. However, the variable considers only the length of roads, but not distribution is space. In some occasions dense networks of roads are situated around the main city, and, although, the figure of road length is relatively high, such roads do not contribute to deforestation.

GDP per capita variable is used as a proxy for poverty. Rural credit for agriculture describes agricultural technology. The model assumes that all the money received as a credit for agricultural activities is used as investment in technological improvements, such as buying new tractors.

It is worth noting that the variables used in most cases have cumulative nature. For instance, the data on road network is data up to year 2009, not the increment in lengths of roads in year 2009. In case the data covered year 2009 only, the author checked whether temporal variations exist. In cases such variation was found, the author used cumulative data. To clarify, the agricultural credit variable is all the money received during the last decade of twentieth century. Since deforestation variable is area of forests cleared up to year 2009, it is crucial to regress it against cumulative variables (when temporal variations are high).

The model uses few instrumental variables. Adequate instruments should not correlate with dependent variable and should be strongly correlated with endogenous variable (have to be exogenous with respect to dependent variable and have to be excludable from the model without causing any important changes).

The most straightforward approach in searching such instruments is to use lagged variables of endogenous variables. In the model GDP per capita in year 2008 was used as an instrument for GDP per capita in year 2009, population in years 1991 and 2000 served as two instruments for population in year 2010. Further, all three instruments did not correlate with deforestation. GDP in year 2009, selected to instrument road variable, also had both desirable features – correlated strongly with road variable and was not correlated with deforestation.

Summary of variables used with some more information is presented below in table 1.

Table 1. Variables

Variable	Code name	Explanation	Source
DEPENDENT VARIABLE			
Deforestation	DEF_09	Deforested area in square kilometers	The Brazilian National

		until year 2009 inclusive	Institute for Space Research
EXOGENOUS VARIABLES			
Economic growth	EC_GR_00base	GDP growth in per cents between years 2000 and 2009	The Brazilian Institute of Geography and Statistics (SIDRA database) and author's calculations
Cattle ranching	CATTLE_09	Number of cattle (heads) in year 2009	The Brazilian Institute of Geography and Statistics (SIDRA database)
Agricultural activities	AGR_09	Area in square kilometers covered by temporal (yearly) agricultural plants in year 2009	The Brazilian Institute of Geography and Statistics (SIDRA database)
Timber	TIMBER_09	Value of timber in thousands of dollars in year 2009	The Brazilian Institute of Geography and Statistics (SIDRA database)
Rural credit for agriculture	CR_AGR_09	Cumulative rural credit for agriculture in dollars between years 2000 and 2009 inclusive	Central Bank of Brazil
Tenure	TENURE_06	Percentage of privately owned lands in year 2006	The Brazilian Institute of Geography and Statistics (SIDRA database)
Population growth	POP_GR_91base	Population growth in per cents between years 1991 and 2010	The Brazilian Institute of Geography and Statistics (SIDRA database) and author's calculations
ENDOGENOUS VARIABLES			
GDP per capita	GDP_PC_09	GDP per capita in dollars in year 2009	The Brazilian Institute of Geography and Statistics (SIDRA database)
Population	POP_10	Number of inhabitants in year 2010	The Brazilian Institute of Geography and Statistics (SIDRA database)
Roads	ROADS_09	Length of paved roads in kilometers in year 2009	Author's elaboration based on maps by The Brazilian Ministry of Transport
INSTRUMENTAL VARIABLES			
Lagged GDP per capita	GDP_PC_08	GDP per capita in dollars in year 2008	The Brazilian Institute of Geography and Statistics (SIDRA database)
GDP	GDP_09	GDP in dollars in year 2009	The Brazilian Institute of Geography and Statistics (SIDRA database)

Population in 1991	POP_91	Number of inhabitants in year 1991	The Brazilian Institute of Geography and Statistics (SIDRA database)
Population in 2000	POP_00	Number of inhabitants in year 2000	The Brazilian Institute of Geography and Statistics (SIDRA database)

Additional data involves geographical distances that were used in application of geographically weighted regressions (further in the text as GWR). Three distinct sets of distances were obtained: Haversine, road and time distances.

However, in the first place geographical coordinates (in decimal degrees) of municipality centers were obtained from The Brazilian Institute of Geography and Statistics. To convert them into Haversine distances in kilometers, the following formulas were used (here lat_i and lon_i are latitudinal and longitudinal coordinates in decimal degrees respectively, $radlat_i$ and $radlon_i$ are latitudinal and longitudinal coordinates in radians respectively and d^H is the distance between the two locations in kilometers):

$$radlat_i = \frac{lat_i \times \pi}{180}, radlon_i = \frac{lon_i \times \pi}{180}$$

$$a = \sin^2\left(\frac{\Delta radlat}{2}\right) + \cos\left(radlat_1\right) \times \cos\left(radlat_2\right) \times \sin^2\left(\frac{\Delta radlon}{2}\right)$$

$$d^H = 25484 \times \arctan\left(\frac{1-\sqrt{1-a}}{\sqrt{a}}\right)$$

Such approach has some obvious limitations. It does not take into account features of terrain. No doubt, some areas can be covered by mountains making the areas much harder to access. Natural obstacles make roads quite convoluted, thus increasing the distance one should travel from one city to another. Despite this, almost all studies that follow GWR approach limit themselves with Euclidian distances.

To account for road network this study considers distance by roads in kilometers between any two municipality centers. The data was extracted from Google interactive maps. However, road distances in kilometers do not reflect road quality or time, spent waiting for a ferry to transport traveler's vehicle across the river.

To take into consideration the differences in road quality and other circumstances that affect time of travelling, time distances were also used. The data was also extracted from Google interactive maps.

However, Google maps has a limitation in a way that not all roads are present on the map, that is, unpaved and natural landscape roads escape inclusion. Therefore, some missing values were inevitable. To fill in the missing data, distances from Haversine matrix were used. In case of time distances, speed of 80 km/h was selected (this is the average speed obtained by assessing available time and road distance data extracted from Google).

To compare different distance matrices, the author calculated Mantel statistic, given by the following formula (here d represents the number of elements in the lower or upper part of distance matrix, $\overline{x}$ and $\overline{y}$ stand for average distances between two different cities in the two matrices under comparison, s_x and s_y are standard deviations of all the distances between any two different cities for the two matrices):

$$r_M = \frac{1}{d-1} \sum_{i=1}^{n-1} \sum_{j=i+1}^{n} \left(\frac{x_{ij} - \overline{x}}{s_x} \right) \left(\frac{y_{ij} - \overline{y}}{s_y} \right)$$

The figures provided with strong anticipation that the final results of GWR method will be very different under the three approaches, since Mantel correlation between Haversine distance matrix and road distance matrix was 0.878, while Mantel correlation between the former and time distance matrix proved to be only 0.366.

Nevertheless, even time distance data is not perfect, as it measures time spent to access from one municipality center to another. Some municipality centers are situated close to the borders, one next to another. However, if the logging site is in the middle of such municipality, the results of time spent to access that area can be quite different, especially if the territory of that municipality is large.

Methodology

This study delivers results under both global and local approaches. In global approach the coefficients are assumed to be constant across different geographical regions, whereas local approach allows for variability in local coefficients. However, the complexity in nature of ties among various economic, geographical, demographical and political phenomena on the one side and deforestation on the other leads to endogeneity problem in almost any equation that aims at identifying determinants of deforestation. To account for it, this study will use two-stage least squares regression. The method provides with consistent estimates, given that instruments chosen are not flawed.

As for notation, x represents a set of exogenous variables, q is a set of endogenous variables, z includes instruments and y stands for deforestation. All capital letters represent vectors and matrices. Symbol ' means transposed matrix and symbol ^ represents predictions.

$$x = \begin{Bmatrix} EC\_GR\_00base, CATTLE\_09, AGR\_09, TIMBER\_09, CR\_AGR\_09, TENURE\_06, \\ POP\_GR\_91base \end{Bmatrix}$$

$$q = \{GDP\_PC\_09, POP\_10, ROADS\_09\}$$

$$z = \{GDP\_PC\_08, GDP\_09, POP\_91, POP\_00\}$$

$$y = \{DEF\_09\}$$

First, this study will assume no geographical differences across regions. The global model includes 7 exogenous variables (m=7), 3 endogenous variables (s=3) and 4 instruments (p=4). In the first stage of two-stage least squares approach s equations are computed to obtain predicted values of all endogenous variables. The procedure works as follows: each endogenous variable is regressed against all exogenous variables and all instruments. The obtained coefficients are substituted into equations to obtain predicted values of each endogenous variable. Formally:

$$q_r^i = \alpha_{0,r} + \sum_{j=1}^{m} \alpha_{j,r} x_j^i + \sum_{l=1}^{p} \alpha_{l,r} z_l^i + \varepsilon_r^i \text{ , where } r = \overline{1,s} \text{ and } i = \overline{1,n}$$

$$\hat{q}_r^i = a_{0,r} + \sum_{j=1}^{m} a_{j,r} x_j^i + \sum_{l=1}^{p} a_{l,r} z_l^i \text{ , where } r = \overline{1,s} \text{ and } i = \overline{1,n}$$

Second stage regression uses predicted values from the first stage. Dependent variable is regressed against all exogenous variables and predicted values of endogenous variables.

$$y^i = \beta_0 + \sum_{j=1}^{m} \beta_j x_j^i + \sum_{r=1}^{s} \beta_r \hat{q}_r^i + \varepsilon^i \text{ , where } i = \overline{1,n}$$

Further, let's denote matrix of all values of exogenous and endogenous variables as X, matrix of all values of exogenous and instrumental variables as Z and vector of values of dependent variable as Y. That is:

$$X = \begin{pmatrix} 1 & x_{11} & \dots & x_{1m} & q_{11} & \dots & q_{1s} \\ \dots & \dots & \dots & \dots & \dots & \dots & \dots \\ 1 & x_{n1} & \dots & x_{nm} & q_{n1} & \dots & q_{ns} \end{pmatrix}$$

$$Z = \begin{pmatrix} 1 & z_{11} & \dots & z_{1p} & x_{11} & \dots & x_{1m} \\ \dots & \dots & \dots & \dots & \dots & \dots & \dots \\ 1 & z_{n1} & \dots & z_{np} & x_{n1} & \dots & x_{nm} \end{pmatrix} \quad and \quad Y = \begin{pmatrix} y_1 \\ \dots \\ y_n \end{pmatrix}$$

Given the notation, the vector of coefficients for global model is calculated as follows:

$$B = \left[X'Z(Z'Z)^{-1} Z'X \right]^{-1} X'Z(Z'Z)^{-1} Z'Y$$

To obtain standard errors and t statistics for the two-stage least squares, one should apply the following formulas:

$$mse = \frac{\sum_{i=1}^{n}\left(y^i - \left(b_0 + \sum_{j=1}^{m} b_j x_j^i + \sum_{r=1}^{s} b_r q_r^i\right)\right)^2}{n-m-s-1}, \text{ where } i = \overline{1,n}$$

$$VCV = \left[X'Z\left(Z'Z\right)^{-1}Z'X\right]^{-1} \times mse$$

Next, the procedure follows usual patters – standard errors are diagonal elements of covariance matrix (VCV) and t statistics are computed by diving coefficient value by standard error for that coefficient.

However, the procedure of selecting instruments is not straightforward. Questions like which instruments to use, how many of them to use and are those instruments valid are inevitable. To reach the answers this study considered three main econometrical tests. First of all, the model must justify the use of instrumental variables at all. To assess this point the study interprets endogeneity test reported by Stata 11.0 together with results output. The test runs under the null hypothesis that the specified endogenous variables can be treated as exogenous. The test statistic follows chi-square distribution with number of degrees of freedom equal to number of endogenous variables used.

As for number of instruments, more instruments improve efficiency of two-stage least squares estimator. However, too many instruments can cause bias, especially if instruments used are only weakly correlated with endogenous variables (Söderbom, 2009). To take into consideration those two aspects (number and validity of instruments) the study uses Sargan-Hansen test of over identifying restrictions and "weak identification" test.

The former tests if instrumental variables are orthogonal to the error process and if they correlate with endogenous variables included in the model (Mileva, 2007). The test has a joint null hypothesis that the instruments are uncorrelated with the error term and that the excluded instruments are excluded correctly from the model. Sargan test statistics approximately follows chi square distribution with degree of freedom equal to the difference of the number of instrumental variables and the number of endogenous variables (Baum, 2003).

Weak instruments can cause biased estimators of instrumental variables and can provoke severe distortions in hypothesis testing. To check the strength of instruments the study uses "weak identification" test provided by Stata 11.0. It assumes i.i.d. errors. The test reported is an F version of the Cragg-Donald Wald statistics.

The procedure of selecting the best set of instruments has been done in four steps. Firstly, instruments were selected based on whether they can explain endogenous variables and cannot influence deforestation. For instance, population in year 2000 may explain population today, but have negligible effect on deforestation. Secondly, various combinations of instrumental variables were tested. Afterwards, those combinations, which passed all three tests described above, were saved. Finally, the one which was "the strongest" in a sense that it passed the test with the greatest margin, was chosen as the right one.

The next step in research is to assess possible spatial heterogeneity of factors affecting deforestation. To do this, the study employs GWR approach. The idea resembles the one from global regression, however, this time coefficients across municipalities are allowed to vary. The complexity of the research in a sense that GWR is carried out taking into account endogenous nature of some determinants and different distance measurement methods requires programming. The author used Gauss 10 programming package (the code of the program is available at the appendix section).

Firstly, let's define method of two-stage least squares with spatial heterogeneity formally:

1st stage

$$q_r^i = \alpha_{0,r}\left(lat^i, lon^i\right) + \sum_{j=1}^{m} \alpha_{j,r}\left(lat^i, lon^i\right) x_j^i + \sum_{l=1}^{p} \alpha_{l,r} z_l^i + \varepsilon_r^i, \text{ where } r = \overline{1,s} \text{ and } i = \overline{1,n}$$

$$\hat{q}_r^i = a_{0,r}^i + \sum_{j=1}^{m} a_{j,r}^i x_j^i + \sum_{l=1}^{p} a_{l,r} z_l^i, \text{ where } r = \overline{1,s} \text{ and } i = \overline{1,n}$$

2nd stage

$$y^i = \beta_0\left(lat^i, lon^i\right) + \sum_{j=1}^{m} \beta_j\left(lat^i, lon^i\right) x_j^i + \sum_{r=1}^{s} \beta_r\left(lat^i, lon^i\right) \hat{q}_r^i + \varepsilon^i, \text{ where } i = \overline{1,n}$$

Notice that coefficients for instrumental variables are not allowed to vary across space. The difference from global approach is that now each observation has its unique set of coefficients. Mathematically, to achieve this one must include weighting matrices in the next expression:

$$B_i = \left[X'W_iZ\left(Z'Z\right)^{-1}Z'X\right]^{-1} X'W_iZ\left(Z'Z\right)^{-1}Z'Y$$

The distance data was organized in square matrices.

$$D = \begin{pmatrix} 0 & d_{12} & \ldots & d_{1n} \\ d_{21} & 0 & \ldots & d_{2n} \\ \ldots & \ldots & \ldots & \ldots \\ d_{n1} & d_{n2} & \ldots & 0 \end{pmatrix}, \text{ where } d_{ij} = d_{ji} \text{ and } d_{ii} = 0$$

The weighting matrices were created by calculating weighs using chosen kernel function and then placing the data on weighted distances from one city to all the other cities into the diagonal of the weighting matrix, thus creating one weighting matrix per observation for each type of distance data.

$$W_i = \begin{pmatrix} k_{i1} & 0 & \dots & 0 \\ 0 & k_{i2} & \dots & 0 \\ \dots & \dots & \dots & \dots \\ 0 & 0 & \dots & k_{in} \end{pmatrix}, \text{ where } k_{ii} = 1$$

As far as kernel function is concerned, the choice of kernel function does not affect the results significantly (Brunsdon et al, 1998). Therefore, the chosen kernel function is the simplest one in the class of gradually decaying kernel functions and is defined as follows:

$$k_{ij} = \exp\left(\frac{-d_{ij}}{h}\right),\ k_{ij} \in (0,1]$$

Here h represents bandwidth, but let's focus on some features of the kernel function first. Notice that the figures of kernel are bounded between zero and one. The larger the distance, the lower the weight. Putting in a different way, those activities affecting deforestation, which take place at a distance, have lower weights than same activities, which occur at this place. For instance, coefficients for cattle ranching at this place would be higher than coefficient for cattle ranching some kilometers away. In global regression the distance does not matter and all such activities affecting deforestation have the same influence regardless the distances. That is, in case of global regression all weights would be equal to one.

Bandwidth is a key indicator in the whole GWR approach, since the choice of its size determines the results. First, notice that as size of bandwidth increases, kernel function approaches one:

$$\lim_{h\to\infty} k_{ij}\left(d_{ij}, h\right) = 1$$

That is, very large bandwidth overshadows the distance data. In all cases the weights become almost equal to one. In such a way weighting matrix approaches identity matrix:

$$\lim_{h\to\infty} W_i\left(k_{ij}\left(d_{ij}, h\right)\right) = I$$

As this happens, all sets of coefficients converge to a result generated by global approach. Thus, very high bandwidth values imply no spatial variation.

The computation of optimal bandwidth, as developed by Brunsdon et al (1998), commences by excluding i^{th} observation from the model to avoid minimization at h=0 (zero bandwidth leads to all zero weighting matrix, which in turn implies equality between real and predicted values of deforestation). In computational terms it means replacing the value of k_{ii}^{th} element with zero. The weighting matrix becomes as follows:

$$W_i = \begin{pmatrix} k_{i1} & 0 & \ldots & 0 \\ 0 & k_{i2} & \ldots & 0 \\ \ldots & \ldots & \ldots & \ldots \\ 0 & 0 & \ldots & k_{in} \end{pmatrix}, \text{ where } k_{ii} = 0$$

The variable to minimize is called cross validated sum of squares and is denoted as follows:

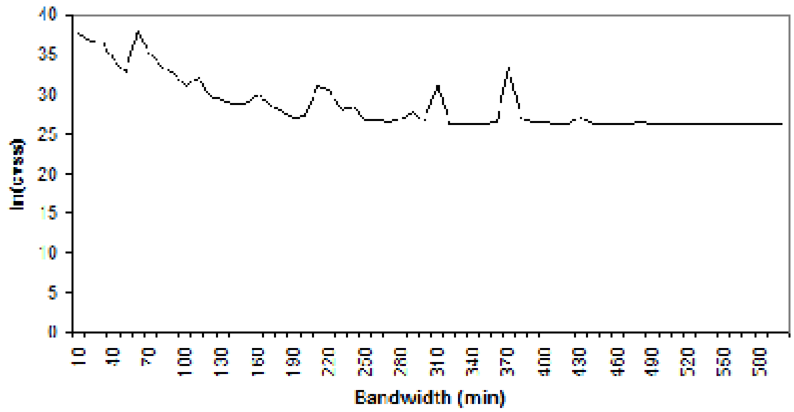
$$CVS(h) = \sum_{i=1}^{n} \left(Y - XB_i(h)\right)$$

$$cvss(h) = \sum_{i=1}^{n} cvs_i(h)$$

Surely, to solve this problem in practice is much more complicated. The approach this study pursues is to calculate cvss's for different values of bandwidth (see graph below). The values of cvss were calculated for bandwidths starting from 10 and ending at 600 with increments of 10. There is no need to select bandwidth values exceeding 600, since the graph becomes "flat" (surely, the graph does not become flat strictly speaking, some negligible fluctuations continue). This shape is dictated by the kernel function. Notice that kernel function decay quite rapidly and bandwidth equal to around 400 minutes already generates quite high weighs. This implies that weighting matrices resembles identity matrix, no matter if the size of bandwidth is 400 or few millions. However, the most importantly, those weighting matrices are very similar. Therefore, they provide similar estimates and similar cvss's. The cvss value at the "flat" part of graph is cvss for global regression.

However, the graph has one imminent challenge – it does not have a unique "minimum". At this point, the research has a loose end. Econometrical procedures cannot help to select one unique size of bandwidth in this case. Therefore, this is an issue of judgment by the experts. This study cannot answer the question whether there is sufficient spatial variation of local coefficients explaining deforestation in Legal Amazon. However, if there is, the best choice of the bandwidth would be at the point where "flat" section of the graph begins. The motivation for this idea lies on the previous discussion concerning relation between bandwidth size and degree of spatial variation. The bandwidth at the start of "flat" curve results in much higher degree of spatial variation than very high bandwidths. In this way it is easier to show spatial differences across municipalities, if they indeed exist. The other motivation was that bandwidth of 450 minutes (case with time distances) resulted in no significant spatial variation for agricultural activities, which seemed reasonable to the author. Indeed, the most variability in weights is achieved at bandwidth of 1150 minutes (standard deviation is 0.245) as opposed to chosen value of 450 (standard deviation equal to 0.206). Under the former bandwidth, local coefficients for agriculture vary significantly across space. However, if composition of cultivated plants is very different across regions and each plant depletes fertile lands at very different rate, the geographical differences may exist. If experts in agricultural field would argue in favor of agricultural differences across regions, the bandwidth in this study should be increased. This increase would have enormous effect of local coefficients for other variables too. For instance, correlation coefficient between local coefficients generated using bandwidth sizes of 450 and 1150 (the same distance measurement technique – time distances – was applied) was only 0.04. Therefore, the reader shall interpret the results within the framework this study develops.

Figure 1. Cross validated sum of squares for various sizes of bandwidth (case with time distances)



Source: author's elaboration

The lowest value of natural logarithm of cvss rounded to one tenth is 26.1. The lowest bandwidth for which cvss is equal to 26.1 is 450 minutes (in case of time distances). In case of Haversine distances, bandwidth size of 180 kilometers was selected. In case of road distances the chosen size of bandwidth was 340 kilometers.

Using selected bandwidths and the methodology described above the author obtained arrays of local coefficients. The coefficient estimates generated using different ways of distance measurement methods were compared by calculating correlation coefficients between vectors of the coefficients for each variable. This result is of a particular importance, as it enables to judge, whether distance measurement technique can significantly influence the final result of GWR method. The formula used is as follows:

$$r_{B^1,B^2} = \frac{\sum_{i=1}^{n}\left(b_i^1 - \overline{b^1}\right)\left(b_i^2 - \overline{b^2}\right)}{\sqrt{\sum_{i=1}^{n}\left(b_i^1 - \overline{b^1}\right)^2 \sum_{i=1}^{n}\left(b_i^2 - \overline{b^2}\right)^2}}$$

The last task is to check which variables, if any, vary sufficiently across space. Brunsdon et al (1998) suggest several methods. The most common is based on Monte Carlo approach. However, it has some limitations. Apart from the issue that the technique requires large amount of computational time, in case of large samples it detects even slightest deviations in local coefficients. Therefore, in cases with large samples, the test is highly likely to yield a highly statistically significant result for trivial fluctuations. An alternative Brunsdon et al (1998) offer is to compare variability in local coefficients with standard errors of the coefficients from the global model. Variability in GWR estimates of coefficients is calculated as follows:

$$\sqrt{v_j} = \sqrt{\frac{\left(b_{ij} - \overline{b_j}\right)^2}{n}} > ste\left(b_j\right)$$

The inequality above also is a rule to confirm or reject sufficient spatial variation across local coefficients.

Results

The second table presents the results for global approach computed by applying two-stage least squares approach. Simple OLS yielded very unstable results, omission or inclusion of a single variable, even not statistically significant, used to flip the signs and statistical significances of some coefficients. Therefore, controlling for endogeneity is crucial. The three statistical tests provided by Stata 11.0 (endogeneity, Sargan-Hansen and weak identification) helped in challenging endogenous nature of three variables – GDP per capita in year 2009, population in year 2010 and lengths of paved roads as of year 2009. As can be seen in the table below, endogeneity test concludes that those three variables selected are indeed endogenous (the null hypothesis stating that all variables treated as endogenous in the model could have been treated as exogenous was rejected). Sargan test statistics indicates that at least under 5% statistical significance level we cannot reject the null hypothesis that instruments are uncorrelated with error term and are excluded from the model correctly. Finally, Cragg-Donald Wald F statistics provided with the benchmark value in evaluating strength of instruments. The results indicated that instruments selected are very strong instruments for population variable and GDP per capita variable. However, these instruments are much weaker instruments to solve endogeneity in road variable. Nevertheless, the F value for roads exceeds benchmark value of Cragg-Donald Wald F test.

Coefficient for economic growth variable proved to be statistically insignificant. The explanation could be that higher incomes may be used also as prevention mechanism from deforestation rather than just alleviating access to it. Local governments may invest more to save forests by investing money in environmental police, creating new protected areas and monitoring them etc., which may absorb the money driven increase in pressure on forests created by particular business entities. However, the sign was negative, which contradicts Angelsen et al's (1999) argument that higher income is likely to increase pressure of forest resources. On the one hand, logging activities are quite costly, as they necessitate covering transportation and storage costs of cut down woods, as well as renting or buying machinery to cut trees and paying wages for lumberjacks and other personnel involved in the process. Therefore, higher incomes can open gates to deforestation. On the other hand, economic growth drives up prices of meat and agricultural production, thus reducing Brazil's competitiveness in global market. That is, importers of Brazilian production may switch to other trade partners, which can offer lower prices. Therefore, the demand of cattle and crops would fall, as well as the pressure on forests.

Table 2. Results of the global regression model

Method – two-stage least squares
Number of observations – 503
Dependent variable – def_09

GLOBAL	Coefficient	Standard error	T statistics
constant	26.68523	317.6389	0.084011
ec_gr_00base	-0.20867	0.17869	-1.167768
cattle_09	0.0065***	0.00035	18.55166
agr_ 09	0.021256***	0.002463	8.630052
timber_09	0.020385***	0.003068	6.644594
cr_agr_09	-8.18E-06***	1.47E-06	-5.549524
tenure_06	1.559983	3.205493	0.486659
pop_gr_91base	8.57E-05	0.000249	0.343896
gdp_pc_09	-0.01277	0.008263	-1.544962
pop_10	0.000323	0.000767	0.421899
roads _09	7.266289*	3.91347	1.856739

Endogeneity test	Chi-sq(3) 34.714	p value 0.000		
Sargan statistic	Chi-sq(1) 2.744	p value 0.098		
Weak identification test	Cragg-Donald Wald F statistics 9.1		F (2,491)	
			gdp_pc_09	1747.96
			pop_10	21772.75
			roads_09	18.16

* Statistically significant at 10% level
*** Statistically significant at 1% level

The study concludes that cattle ranching is major driver of deforestation in Legal Amazon. This result is reported in almost any study concerning this region. The reason of such find lies in the fact that cattle ranching activities are relatively more land intensive. This implies that land is the most important input in the business. As land resources are scarce, farmers must choose how to divide the land between forested area and land for cattle grazing. New areas of land are required not only for the expansion of cattle herd, but also because of depletion of land in which cattle graze. However, open economy and trade turns cattle ranching into profitable business and creates even more pressure on tropical rainforests. As economies in foreign countries grow, individuals tend to spend more on meat products, thus increasing the demand for beef and other meat products. As a consequence, such countries increase imports of meat production. Therefore, as demand for cattle activities in Brazil increase, more forests have to be cleared to supply cattle with necessary area for grazing. This effect in literature is known as "Hamburger" effect. Kaimowitz et al (2004) concluded that "Hamburger" effect is present in Legal Amazon.

Similar analysis would hold for agricultural activities, which, not surprisingly, were found to contribute to deforestation in Legal Amazon as table 2 indicates. Agriculture, like cattle ranching, requires land, which can only be gained at the expense of forests. The lesser effect on deforestation compared to cattle ranching can be explained in a way that plant cultivation is not as profitable as cattle business. Most other studies included only soybean as a proxy for agriculture and found that soybean cultivation drives deforestation.

Coefficient for timber values returned statistically significant result. The find is corroborated by most studies. The disagreement among researchers is in sign of the coefficient. This study finds that timber market stimulates deforestation and is in line with view expressed by Angelsen et al (1999). The economic explanation of why higher timber prices drives forest clearing is quite simple – higher price of wood means that economic agents receive more income for the same quantity of wood sold. Logging simply becomes more profitable. Opponents of this idea argue that higher timber prices would encourage managing forests more effectively. However, value of forests suffer from market failures, as it is zero priced in the market.

Coefficient for agricultural credit yielded statistically significant result with a negative sign, indicating that farmers invest in relatively more labour or capital intensive technologies. For instance, new additional worker may help to manage existing areas more efficiently, thus reducing the need to obtain new areas, currently occupied by forests. This may indicate that credit scheme is working as it increases effectiveness of management of current agricultural lands and takes away some pressure from forests at the same time. Although Oliveira el at (2011) found that credit is not a contributor to deforestation, it is worthy to notice that the variable they include is total credit (credit for both agricultural activities and cattle ranching).

As for tenure security, the results imply that it is not a driver of deforestation (the coefficient of tenure variable is statistically insignificant). Nevertheless, the coefficient has a positive sign, indicating that privatization of forests increases logging. This inference is in line with Angelsen et al (1999) conclusions. Conventional thesis of environmental economics states that property rights increase security of natural resources. This notion would be true in case of no marker failures. However, as forests are not included in market economy, the owners of forests undervalue their possessions and tend to believe that selling wood is more profitable. However, as tenure security does not seem to have any statistically significant effect on deforestation in Legal Amazon, the widely debated issue regarding improper management of property rights does not seem to be important when deforestation is considered.

This model included two variables for population – one describing the stock and the other the speed of growth of number of inhabitants. The results show that neither variable is a factor in deforestation. The find is in line with Angelsen et al (1999) observations. At first glance, it looks reasonable to assume that more individuals should increase the pressure on forests. However, it is likely that the highest share of population growth happens in cities (that is, in very concentrated places), where inhabitants engage in other activities than logging. Also, number of inhabitants in countries which import agricultural and meat products from Brazil may be much more important than inner population, especially if trade is very intensive and the foreign demand for such products is high.

Poverty, measured by GDP per capita, was found to be statistically insignificant. It is tempting to assume that deforestation is a problem of less developed country. This is because the tropical rainforests are situated in the regions, where poverty prevails. It is likely that not the poverty, but rather economic structure of the country determines the degree to which forests are exposed to deforestation. In areas where predominant sectors are land intensive (cattle ranching, agriculture etc.), deforestation is much bigger problem than in countries where more capital or labour intensive sectors prevail (tourism, manufacturing etc.). The finding is supported by Oliveira el at (2011), but contradicted by Hargrave (2009). However, as Angelsen et al (1999) stress, statistically significant result may be a consequence of failure to account for endogeneity.

Finally, the coefficient for road variable was found to be statistically significant at 10% significance level, although high standard error is a concern. The result implies that higher kilometrage of roads will increase deforestation. The result seems logical, as roads provide with better accessibility and reduce the costs of deforestation.

So far the coefficients were not allowed to vary across space. The next step is to relax this restriction by the means of GWR approach. However, before proceeding on analysis of the results, first consider the table below. It shows correlation coefficients between two vectors of local coefficients computed using the same GWR approach, however, with three different distance measurement techniques – straight line distances, evaluating Earth's curvature, in kilometers, distances by roads in kilometers and time distances using the roads in minutes. The result is of a particular importance, since array of local coefficients generated while using Haversine distances correlated with the array of local coefficients generated while using road distances only weakly or moderately. The difference is striking when local coefficients computed using time distances are compared to original Haversine approach – correlation coefficient is close to zero, indicating no correlation at all. Therefore, one of the key conclusions of this study is that researchers must be careful when interpreting the findings suggested by conventional approach, which incorporates only Euclidian distances. Given the opportunity exists, it is recommended to run GWR using time distance matrices.

Table 3. Correlation coefficients between vectors of local coefficients computed using different distance matrices (for cattle, road and credit variables)

	Haversine vs Road	Haversine vs Time	Road vs Time
Cattle ranching	0.498297	0.025739	0.109602
Road network	0.304993	0.000227	0.117381
Agricultural credit	0.422646	0.03793	0.177849

The further analysis will only consider GWR results generated using time distances and will be based on the assumption that "optimal" size of bandwidth is 450 minutes.

Coefficients for five variables – cattle ranching, agricultural credit, road network, population growth and stock – seem to vary across space sufficiently enough (see table 4). However, only the first three of them have influence on deforestation. Two important drivers of forest clearing – agricultural activities and timber values – does not seem to vary significantly. The results are quite different from Oliveira el at (2011), who find sufficient spatial variation in coefficients for GDP per capita, cattle ranching, soybean and sugar cane cultivation and non-timber extraction and insufficient spatial variation in coefficients for timber extraction, forestry, population density and credit. This empirical illustration demonstrates that the results must be treated with caution. The differences in results mostly depend on the extent to which endogeneity is addressed, on distance measurement technique and bandwidth choice.

However, the difference in results concerning agricultural activities can be explained by the notion that cultivated plants may have heterogeneous share in total agricultural area across different municipalities. However, when all these plants are described by a single variable, the differences are not distinguishable. At least, agricultural activities as a whole, in author's opinion, should not influence deforestation at different extent across different regions.

As for timber values, it is reasonable to observe little spatial differences in coefficients, since economic motivation to clear the land for timber does not vary much. The incentives may depend on perceptions of individuals on how much the forest is worth. However, the differences may be palpable across different cultures, though not likely across individuals living in the same country. After all, market fails to set a price on natural resources.

Cattle ranching seems to create pressure on forests to a different degree across Legal Amazon. The explanation may be that land fertility is not that important in cattle grazing, thus implying that pastures may exist in most kinds of terrain. Some municipalities may have a lot of territories previously occupied by agricultural plants, but no longer suitable for this kind of activity. On the other hand, cattle ranching may be present in those lands, thus taking away the need to clear forests further. In such municipalities, cattle business would create less pressure on forests.

As for agricultural credit, the money received as a credit may both push forward and restrain deforestation. Angelsen et al (1999) stress that increased investment in agriculture creates more jobs in the sector, thus creating more pressure on forests. On the other hand, farmers may invest in effective management of existing fields of plantations. Therefore, not just the variability is local coefficients is very likely, but also both signs of coefficients are expected.

Table 4. Results of GWR (time distances)

LOCAL	$\sqrt{v_j}$	$ste(b_j)$	$\frac{\sqrt{v_j}}{ste(b_j)}$
constant	215.241	317.6389	0.677628
ec_gr_00base	0.105103	0.17869	0.588184
cattle_09	0.000957	0.00035	2.731554*
agr_ 09	0.001766	0.002463	0.717066
timber_09	0.001946	0.003068	0.634182
cr_agr_09	2.72E-06	1.47E-06	1.843463*
tenure_06	0.622735	3.205493	0.194271
pop_gr_91base	0.000456	0.000249	1.831653*
gdp_pc_09	0.008149	0.008263	0.986227
pop_10	0.004705	0.000767	6.137468*
roads _09	14.5206	3.91347	3.710415*

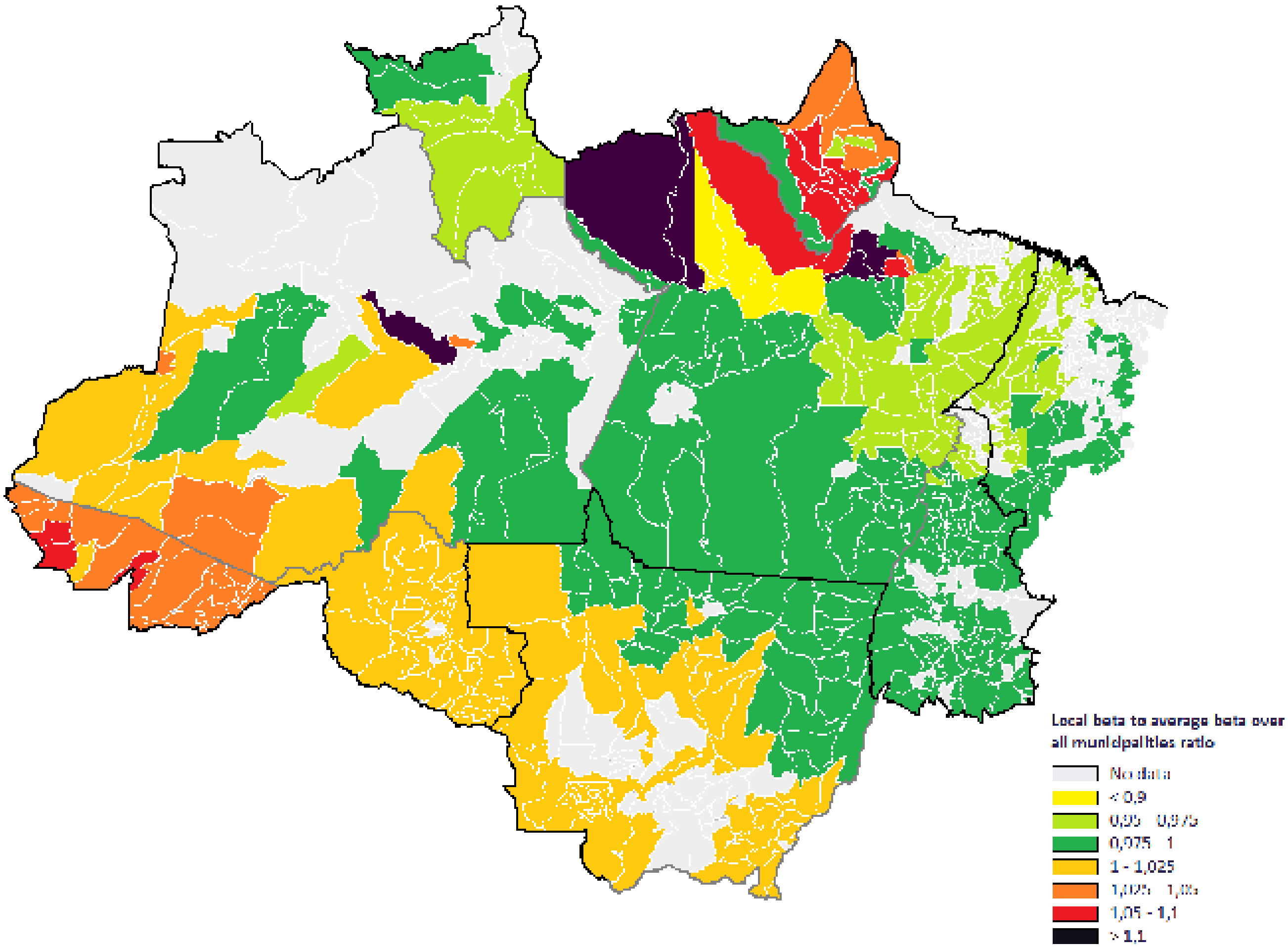
* Significant spatial variation

As far as road network is concerned, probably in most cases better accessibility will increase deforestation. However, the opposite also may happen, since roads also improve accessibility among different cities and connect remote farms with urban areas. As a result, some farmers may migrate to cities, as in most cases salaries in urban areas are higher than in rural ones. In such case road network would reduce deforestation. Therefore, variability in local coefficients for road variable could be fully justified.

As can be seen in figure 2, cattle ranching activities contribute relatively less to deforestation in Maranhão, Tocantins, Roraima, Northern part of Mato Grosso and in most parts of Pará, except the North. Cattle business becomes relatively stronger driver of forest clearing in the Southern part of Legal Amazon, which include almost all areas in Mato Gross, but the North, Rondônia and at least the South of Amazonas. Even stronger effect is found in Acre and Amapá, the two states, situated in the opposite sides of Legal Amazon. However, the highest relative impact of cattle ranching is in some municipalities, situated in the Northern and

North-West part of Pará. It is interesting to compare the results with Oliveira el at (2011), who find that cattle ranching contributes most heavily in the state of Pará, close to the border with Maranhão. However, this study concludes the opposite – the area mentioned is covered with light green colour, indicating relatively low contribution to deforestation. Oliveira el at (2011) also stress that some local coefficients of municipalities situated in the state of Roraima have negative sign, which contradicts common sense. This study finds relatively lower impact of cattle ranching on logging in Roraima, though it is always positive.

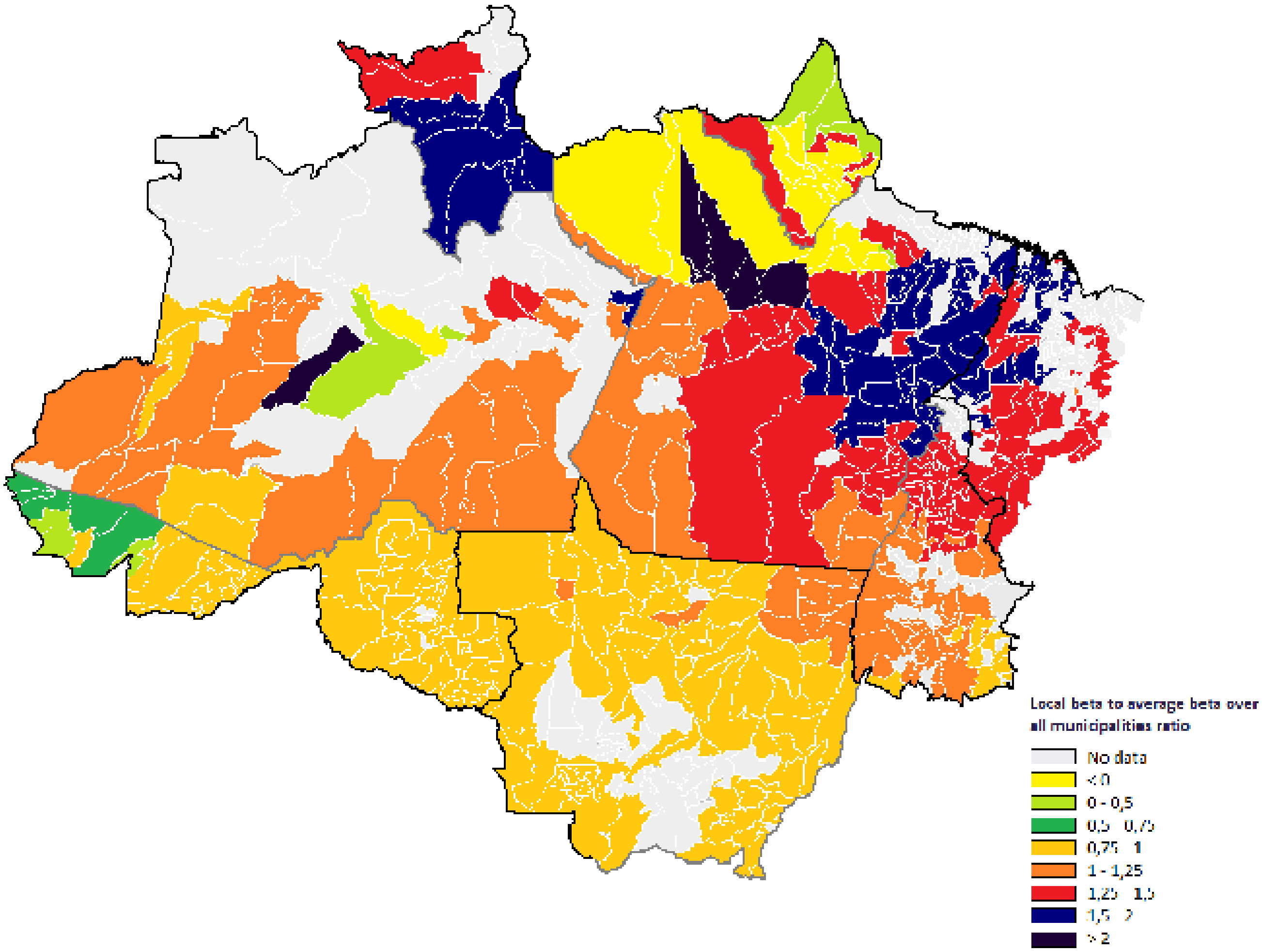
Figure 2. Spatial distribution of local coefficients for cattle ranching



Source: author's elaboration

Figure 3 reflects the distribution of local coefficients for road variable. The lowest relative impact is found in most parts of Amapá and in Northern and North-West part of Pará. In these areas the coefficients are negative, indicating that better road network reduces forest clearings. This result is reasonable as discussed earlier in the text. Amapá is the region where service and industrial sectors prevail. Potentially, the wages in such sectors are higher than in agriculture and cattle ranching. Therefore, better road connection with main cities may encourage some farmers to exchange their agricultural activities into occupation in services or industry. Nevertheless, in the majority of areas the contribution of road variable to logging is positive, though to very different extent. In the Southern part of Legal Amazon, more precisely, in states of Acre, Rondônia and Mato Grosso, the contribution of roads on forest clearings is lower than the average. In the mid part of Legal Amazon the importance of road impact generally increase in the direction from West to East. In the Northern part of Amazonas roads contribute moderately, the noticeable changes occur in the state of Pará, were relative pressure of roads on forests are much higher in the Eastern part of the state when compared to the West. Road network contributes heavily to deforestation in Maranhão, Northern part of Tocantins, East of Pará and Roraima. In some municipalities in the Northern part of Pará roads contribute to logging quite extremely, exceeding twice the average contribution.

Figure 3. Spatial distribution of local coefficients for roads

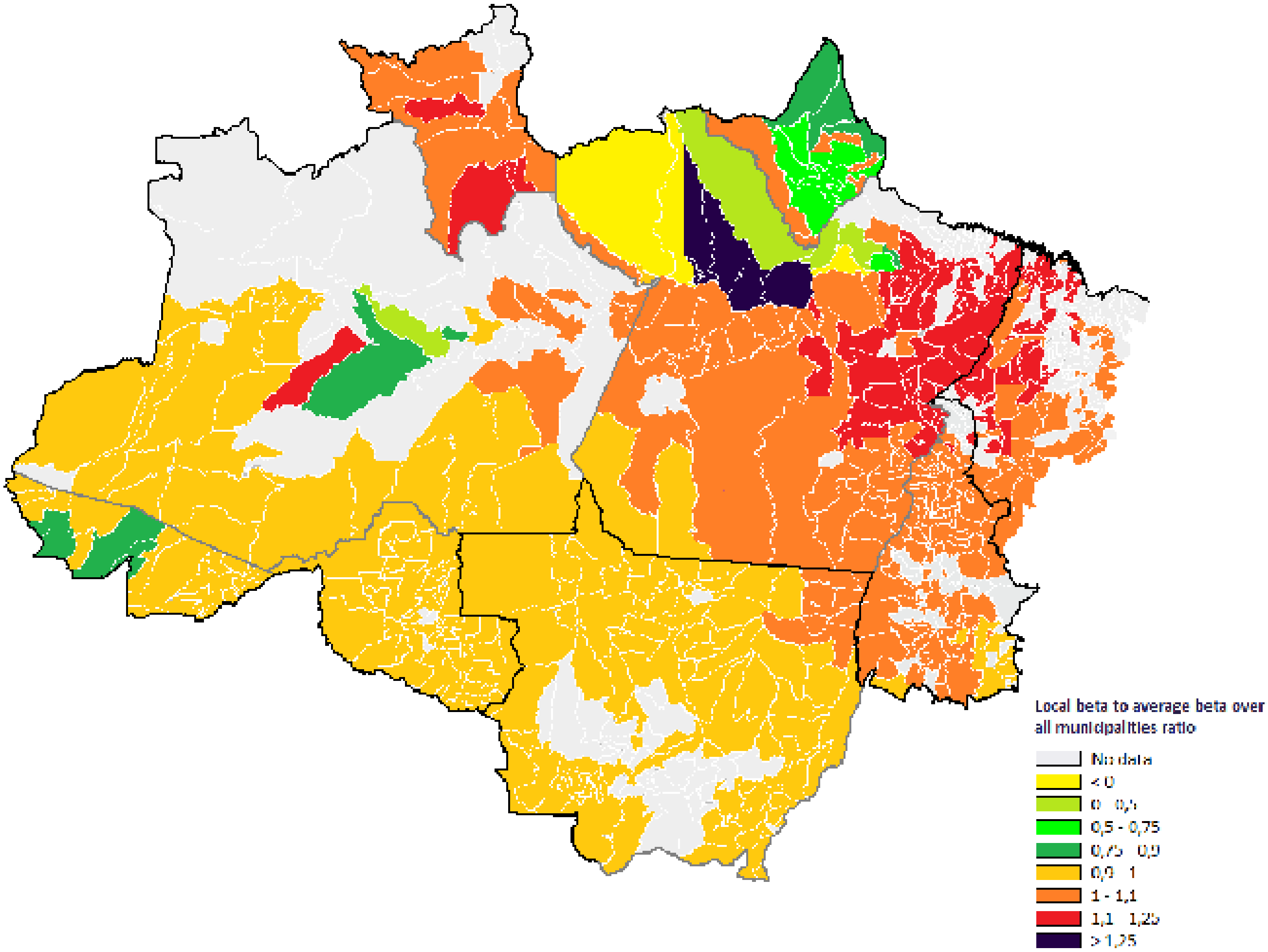


Source: author’s elaboration

Figure 4 represents spatial distribution of local coefficients for agricultural credit. In few municipalities, situated in North-East part of Pará, the coefficients are positive, indicating that money granted as agricultural credit increases deforestation. The result implies that to reduce the impact of deforestation, the governing bodies should reconsider the distribution of credit among municipalities, stop providing credit for those municipalities where it worsens the problem of logging and supply with more money those municipalities, where it reduces deforestation to the greatest extent. However, in the vast majority of areas agricultural credit has a negative impact, indicating more effective use of currently cultivated lands. In Amapá and mid Amazonas the money granted has relatively low effect on logging. As for the rest of

Legal Amazon, spatial differences are not so obvious, with few exceptions. Parts of Roraima, Eastern Pará and Western Maranhão seem to use the credit in a way that it reduces deforestation to the greater extent than in major parts of Legal Amazon. In few municipalities in the North of Pará agricultural credit contributes to reduced deforestation the most.

Figure 4. Spatial distribution of local coefficients for agricultural credit



Source: author’s elaboration

It is more convenient to analyze all three maps at once. The three graphs are like mirror reflection of each other – in areas, where cattle ranching has relatively lower impact on deforestation, road network and agricultural credit has a greater impact, and vice versa. However, the darker colours in the map for cattle ranching do not imply that deforestation is those areas is high, the message here could be that those areas may have poor accessibility by roads and do not invest much on labour or credit intensive technologies, thus making cattle ranching a relatively stronger contributor to logging. This would explain why the map reflecting cattle ranching is coloured in darker colours in areas with hardly accessible jungle.

The results may indicate that road network, squeezing from the North and the East inside the jungle is causing deforestation to a greater extent than in most areas in Legal Amazon. Also, the same areas use credit more effectively than the others in terms of saving forests.

Notice that changes in local coefficients occur slowly and smoothly in most parts of Legal Amazon. However, there are two regions, where the differences among neighboring municipalities are striking. One region is in the North-West part of Pará and the other – in the midst of Amazonas, though the patterns in the latter cannot be clearly distinguished due to missing data. Such difference could be observed in two neighboring regions, if one has a well developed road network with leading sectors of service or manufacturing, while the other is covered with mostly untouched forests and has poor accessibility. If this is a case, then roads may influence the reduction in deforestation in the former areas, where farmers may take advantage of better connection with urban areas and substitute agricultural activities to better paid jobs in services or manufacturing. The opposite would happen in forested regions – as new road would provide accessibility to vast areas of forest previously untouched forests. Also, as forested regions has fewer roads, the other logging driver – cattle ranching – has relatively more influence on forest clearings. Additionally, in urban regions with better road network forests are easier to reach. Therefore, money received as credit may encourage to clear lands rather than investing in effective management of currently cultivated lands. The opposite may hold in regions with dense forests. Here logging may be more expensive than in urban areas, thus forcing farmers to invest into quality of lands rather than quantity. Indeed, those municipalities to the North-West corner of Pará have only one major road and are host to many protected areas, whereas their Eastern neighbors have much denser road network.

Conclusions and suggestions

The study revealed that endogeneity is an important issue to consider in models explaining deforestation in Legal Amazon, since simple OLS and two-stage least squares approaches provide with different answers.

The results of global regression in general corroborates with observations made by Angelsen et al (1999): coefficients for agricultural acreage and timber market were found to be positive and statistically significant, coefficient for agricultural credit proved to be negative and statistically significant and coefficients for GDP per capita and both population growth and stock were statistically insignificant. The difference from what Angelsen et al (1999) stressed reflects in statistical insignificant coefficients for economic growth and tenure. However, the sign for tenure variable was in line with Angelsen et al's (1999) observations – privatization of lands has a positive effect on deforestation. Though this study disagrees with the sign for economic growth – the results indicate that economic growth has a negative effect on logging.

The results of GWR approach are open. This study was unable to answer the question whether spatial differences are present in Legal Amazon. The reason was the absence of one single "optimal" bandwidth. Therefore, the lowest "optimal" bandwidth was selected to contrast spatial differences if they indeed exist. The findings revealed significant spatial variation in cattle, agricultural credit, road and both population variables. However, only the first three have statistically significant affect on deforestation.

In state of Amapá cattle ranching plays relatively stronger role in driving deforestation, especially in Southern part of the state. On the other hand, contribution of roads to deforestation is relatively low, in Southern part roads even decrease logging. Agricultural credit too has relatively minor influence, especially in the South of Amapá. Roraima exhibit the opposite patterns – here cattle ranching has relatively lower influence on logging and road network as well as agricultural credit contribute relatively more, especially in the Southern part of the state. The forests in the region, encompassing East of Pará, West of Maranhão and North of Tocantins, are suffering relatively less from cattle ranching and relatively more from road network, agricultural credit reduces logging relatively more than in other areas of Legal Amazon. Similarly can be described the region encompassing the rest of Pará, except the North, the rest of Tocantins and the rest of Maranhão. However, here the amplitude of differences is much lower. States of Mato Grosso and Roraima seem not to have notable spatial variation. In Acme the influence of cattle ranching is slightly higher than the average and the influence of road network and agricultural credit – slightly lower. As for Amazonas, the patterns cannot be seen clearly, though it is very "colourful" region in a sense that different conclusions can be obtained for different municipalities in the state. One remarkable area is in the Northern and North-West part of Pará, where neighboring municipalities have completely opposite patterns regarding deforestation, ranging from positive influence of agricultural credit to highly negative, from negative influence of roads to highly positive, and from very low relative contribution of cattle ranching to very high.

However, the most important finding of this study lies in distance measurement technique. The results proved that the correlation between local coefficients computed using Haversine and road distances is quite weak, while local coefficients generated using Haversine and time distances do not correlate at all. This conclusion is of high importance.

Finally, it would be interesting to incorporate trade into the model. If data on exports concerning meat and agricultural products would be available on municipality level, it may have a significant effect on the results, since deforestation is a complex phenomenon and the reasons describing it are not bounded in the territory of Legal Amazon. Potentially, the demand of meat in such remote countries as Russia also shapes deforestation patterns in Brazil through trade mechanism.

Acknowledgements

I would like to thank my PhD supervisor William Nilsson for providing me with many helpful comments, suggestions and insights and programmer Ernestas Zabarauskas for extracting all the road and time distance data from Google interactive maps.

References

Angelsen, A.; Kaimowitz, D. (1999). Rethinking the Causes of Deforestation: Lessons from Economic Models. *The World Bank Research Observer, vol. 14, no. 1, pp. 73–98.*

Araujo, C.; Araujo Bonjean, C.; Combes, J. L. (2010). Does Land Tenure Insecurity Drive Deforestation in the Brazilian Amazon? Retrieved 10, 2012, from http://www.cerdi.org/uploads/pagePerso/ 28/2010.13.pdf

Baum, F.; Schaffer, M.; Stillman, S. (2003). Instrumental Variables and GMM: Estimation and Testing. *The Stata Journal, vol. 3, no. 1, pp. 1–31.* Retrieved 11, 2012, from http://www.stata-journal.com/sjpdf.html?articlenum=st0030

Brunsdon, C.; Fotheringham, S.; Chariton, M. (1998). Geographically weighet regression – modelling spatial non-stationarity. *Journal of the Royal Statistical Society. Series D (The Statistician), vol. 47, no. 3, pp. 431-443.*

Central Bank of Brazil (2013). Annual Statistics of Rural Credit. Retrieved 10, 2012, from http://www.bcb.gov.br/?RELRURAL

Fisher, H. (2010). Introduction to Stata - Handout 4: Instrumental variables. Retrieved 12, 2012, from http://www.econ.cam.ac.uk/postgrad/hcf26/teaching10/lecture41011.pdf

Google Maps (2013). Interactive map of Brazil. Retrieved 11, 2012, from https://maps.google.com/

Hargrave, J. (2009). Economic Causes of Deforestation in the Brazilian Amazon: An Empirical Analysis of the 2000s. Retrieved 10, 2012, from http://www.anpec.org.br/encontro2009/inscricao.on/arquivos/421-1d4f5be8e654ae48e805fd3d62afd01d.pdf

Hargrave, J.; Kis-Katos, K. (2011). Economic Causes of Deforestation in the Brazilian Amazon: A Panel Data Analysis for the 2000s. Retrieved 10, 2012, from http://www.vwl.uni-freiburg.de/iwipol/discussion_papers/DP17_Hargrave_Kis-Katos%20-%20Economic%20Causes%20of%20Deforestation%20in%20the%20Brazilian%20Amazon.pdf

Hwang, S. (2001). GAUSS Programming for Econometricians. Retrieved 12, 2012, from http://venus.unive.it/pasinigi/teaching/GaussIntro.pdf

Kaimowitz, D.; Angelsen, A. (1998). Economic Models of Tropical Deforestation: A Review. Retrieved 10, 2012, from http://www.cifor.org/publications/pdf_files/Books/Model.pdf

Kaimowitz D.; Mertens B.; Wunder S.; Pacheco P. (2004). Hamburger Connection Fuels Amazon Destruction: Cattle ranching and deforestation in Brazil's Amazon. Retrieved 12, 2012, from http://www.cifor.org/publications/pdf_files/media/Amazon.pdf

Mapsof.net (2013). Brazil maps. Retrieved 11, 2012, from http://mapsof.net/map/brazil-municipalities

Mileva, E. (2007). Using Arellano – Bond Dynamic Panel GMM Estimators in Stata. Tutorial with Examples Using Stata 9.0 (xtabond and xtabond2). Retrieved 11, 2012, from http://www.fordham.edu/economics/mcleod/Elitz-UsingArellano%E2%80%93BondGMMEstimators.pdf

Movable Type Scripts (2013). Calculate distance, bearing and more between Latitude/Longitude points. Retrieved 11, 2012, from http://www.movable-type.co.uk/scripts/latlong.html

Nerlove, M. (1998). Notes on Gauss. Retrieved 12, 2012, from http://www.arec.umd.edu/ mnerlove/ mnerlove.htm

Obersteiner, M.; Kindermann, G.; Rametsteiner E.; Sohngen, B. (2006). Economics of Avoiding Deforestation. Retrieved 10, 2012, from http://citeseerx.ist.psu.edu/viewdoc/summary?doi =10.1.1.168.1653

Oliveira, R. C.; Almeida, E. (2011). Deforestation in the Brazilian Amazonia and Spatial Heterogeneity: a Local Environmental Kuznets Curve Approach. Retrieved 10, 2012, from http://www.ufjf.br/poseconomia/files/2011/06/TD-005-2011-Eduardo-Simoes-de-Almeida.pdf

Pfaff, A.; Robalino, J. and others (2007). Road Investments, Spatial Spillovers, and Deforestation in the Brazilian Amazon. *Journal of Regional Science, vol. 47, no. 1, pp. 109–123.*

Rossi, E. (2006). Introduction to GAUSS Programming Language. Retrieved 12, 2012, from http://economia.unipv.it/pagp/pagine_personali/erossi/gaussnotes_06.pdf

Söderbom, M. (2009). Instrumental Variables, 2SLS and GMM. Lecture in Applied Econometrics. Retrieved 11, 2012, from http://www.soderbom.net/lec2n_final.pdf

The Brazilian Institute of Geography and Statistics, IBGE (2012). Cadastre of localities. Retrieved 11, 2012, from ftp://geoftp.ibge.gov.br/organizacao_territorial/localidades

The Brazilian Institute of Geography and Statistics, IBGE (2012). SIDRA database. Retrieved 11, 2012, from http://www.sidra.ibge.gov.br/bda/default.asp?z=t&o=1&i=P

The Brazilian Ministry of Transport, National Department of Transport Infrastructure, DNIT (2012). Multimodal maps. Retrieved 11, 2012, from http://www2.transportes.gov.br/bit/01-inicial/index.html

The Brazilian National Institute for Space Research, INPE (2012). PRODES database. Retrieved 10, 2012, from http://www.dpi.inpe.br/prodesdigital/prodes.php

Appendix

General part of Gauss 10 code

```
» load data[782,15]=source1;
» ma=data;
» m=packr(ma);
» mr=rows(m);
» load data[503,252]= source2;
» d1=data;
» load data[503,251]= source3;
» d2=data;
```

```
» d=d1~d2;
» xnoc=m[.,2]~m[.,3]~m[.,4]~m[.,5]~m[.,6]~m[.,7]~m[.,8]~m[.,9]~m[.,10]~m[.,11];
» c=ones(mr,1);
» x=c~xnoc;
» z=m[.,12]~m[.,13]~m[.,14]~m[.,15];
» y=m[.,1];
» i=z~c~m[.,2]~m[.,3]~m[.,4]~m[.,5]~m[.,6]~m[.,7]~m[.,8];
» ii=inv(i'*i);
» id=eye(503);
```

Part of Gauss 10 code for computing global regression with endogenous variables and instruments

```
» beta=inv(x'*i*ii*i'*x)*x'*i*ii*i'*y;
» yp=x*beta;
» e2=(y-yp)^2;
» mse=sumc(e2)/492;
» vcv=(inv(x'*i*ii*i'*x))*mse;
» v=diag(vcv);
» ste=sqrt(v);
» tstat=beta./ste;
» print beta;
```

Part of Gauss 10 code for computing optimal bandwidth

```
» h=10;
» do while (h<=600);
loop=1;
sum=0;
do while (loop<=503); restart: if(loop.<=503); k=exp(-d/h); k[loop,loop]=0;
betaw=inv(x'*(k[.,loop].*id)*i*ii*i'*x)*x'*(k[.,loop].*id)*i*ii*i'*y;
stew=sqrt(diag((inv(x'*(k[.,loop].*id)*i*ii*i'*x))*(sumc((y-
(x*(inv(x'*(km[.,loop].*id)*i*ii*i'*x)*x'*(k[.,loop].*id)*i*ii*i'*y)))^2)/492)));
tstatw=betaw./stew; cvss=sumc((y-(x*(inv(x'*(k[.,loop].*id)*i*ii*i'*x)*x'*(k[.,loop].*id)*i*ii*i'*y)))^2);
sum=sum+cvss; loop=loop+1; goto restart; endif; endo; print sum; h=h+10; endo;
```

Part of Gauss 10 code for computing GWR with endogenous variables and instruments

```
» k=exp(-d/h);
» h=450;
» loop=1;
» do while (loop<=503); betaw=inv(x'*(k[.,loop].*id)*i*ii*i'*x)*x'*(k[.,loop].*id)*i*ii*i'*y; print betaw;
stew=sqrt(diag((inv(x'*(k[.,loop].*id)*i*ii*i'*x))*(sumc((y-
(x*(inv(x'*(k[.,loop].*id)*i*ii*i'*x)*x'*(k[.,loop].*id)*i*ii*i'*y)))^2)/492)));
tstatw=betaw./stew; loop=loop+1; endo;
```