Causes of forest crime in Legal Amazon and underreporting: Markov Chain Monte Carlo approach

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

Jusys.tomas@gmail.com

January, 2014

The results are not updated and are subject to changes after all relevant comments will be gathered

Abstract. Deforestation in Legal Amazon remains among extensively researched topics in environmental economics. However, to this date no study eyed this problem from crime point of view. Illegal logging and forest crime in general are prevalent in Amazon forest. Recent database on embargoed areas in Legal Amazon is a perfect source for modeling deforestation from crime perspective. This study considers both cross section and panel models and follows Bayesian approach to model underreporting. This is done through Markov Chain Monte Carlo method, which is implemented in winBUGS software. Amazon jungle is classified into the Arc of Fire and Inner Amazon. The two territories are assumed to have different rates of underreporting. Seven models are presented: cross section model without taking into account underreporting, three cross section models, which consider underreporting and differ only in prior distributions of recovery rates, one panel data model without underreporting, one panel data model with underreporting assuming time-fixed recovery rate and, finally, one panel data model with and underreporting assuming time varying recovery rates. The last model is considered to be the most accurate. The results of that model reveal that forest stock, possibility of access by boat, value of extracted timber products, number of cattle animals and overall population influence number of forest crime cases positively, whereas coefficients on openness to trade, agricultural activities and rural credit are not statistically significant. The results also confirm complex relation between GDP and number of forest crime cases – the relation is positive for relatively poor and relatively rich municipalities and negative – for municipalities with moderate level of GDP. Cross section models reveal that road network increases the amount of illegal logging and other forest crime activities. Also, distinction between rural and urban population is important, since only the former affects the scale of illegal forest activities.

Keywords. Forest crime, embargoed areas, Legal Amazon, negative binomial, random effects, underreporting, Markov Chain Monte Carlo, winBUGS.

Introduction

Academic literature on deforestation in Legal Amazon is abundant. However, none of the previous studies eyed the problem from crime point of view. Illegal logging and other forest crimes are prevalent and uncontrollable threat to the whole environmental system not only in Brazil, but also globally. The Brazilian government quite recently launched a new project, under which audits of privately owned lands are carried out and infringers are punished by taking away their lands. This is reflected in publicly accessible database of embargoed areas. This information opens new gates to study deforestation in Legal Amazon from a different angle – to model the prevalence of forest crimes.

This study has three important goals. Most importantly, it approaches deforestation from crime perspective and identifies key determinants, affecting the amount of forest crime activities. Secondly, the research considers underreporting problem and classifies Amazon jungle into two zones, which differ in the percentage of potentially undiscovered forest crime sites. Underreporting is particularly high in dense jungle areas and can hardly be detected. Therefore, only small part of illegal crimes is reflected in official database. Finally, the third goal is to compare findings of models without underreporting with the findings of models which take into account underreporting.

The paper will contribute to existing literature in a way that it is the first attempt to model illegal forest clearings from crime point of view, treating the amount of cases of illegal logging and related forest crimes as a dependent variable. Also, the study will use winBUGS software and Markov Chain Monte Carlo (further in the text as MCMC) method to take into account underreporting issues. Usually the whole area under consideration is assigned only one distribution of recovery rate (Cariniceanu et al, 2003; McMillan et al, 2009). Recovery rate is the ratio between observed count and unobserved actual count of forest crime cases. This study identifies two areas with distinct level of underreporting and also evaluates the dynamics of the recovery rates in panel data models.

The author expects that this research will pave a way for new investigations in illegal logging and forest crimes in the Amazon jungle. The results of this paper should be valid for the discussions regarding what measures to apply in order to mitigate forest crime problem. Additionally, the methodology of underreporting should be useful for any type of scholar, who is interested in MCMC approach.

The article will be organized as follows. Firstly, the problem of forest crime will be presented. In the following section data and variables will be introduced. The proceeding section will explain methodology in detail, which will be followed by the results section. The last part will present concluding remarks and recommendations for future research.

The problem

The problem of illegal deforestation is still pervasive in Legal Amazon. Satellite imagery revealed that in the state of Pará 78 per cent of deforestation documented between August 2011 and July 2012 was illegal. It is estimated that more than 100 thousands of hectares were logged illegally during that period (Butler, 2013).

To combat illegal deforestation and other crimes related to forests and the environment, the Brazilian government launched a new project back in 2005, under which audits of privately owned areas are carried out and, under presence of any crime related to forests and environment, the lands are taken away. The database of embargoed areas records all embargoes due to environmental crimes, the majority of which are related directly or indirectly with forests. Several of the most common reasons of embargoes include:

- Destroying or damaging forests, cutting trees or other forms of natural vegetation in permanent preservation areas and areas with specially protected species without the permission of the court.

- Destroying, deforesting forests or damaging any type of native vegetation or native plant species subject to special protection in the legal reserve or forest easements without authorization or license from the competent environmental authority, or at odds with the approval granted, including plans for sustainable forest management.

- Clearing forests or other forms of vegetation, at odds with authorization provided by the Brazilian Institute of Environment and Renewable Natural Resources.

- Selling, exposing for sale, having in storage, transporting or storing timber, firewood, charcoal or other products of plant origin, without a valid license for the entire time of travel or storage, granted by the competent authority.

- Implementing projects for allotments without the competent environmental license.

- Preventing or hindering the natural regeneration of forests or other forms of native vegetation.

- Using fire in forests and other forms of vegetation without taking adequate precautions to ensure controllable burning process.

However, the extent to which the problem of illegal deforestation is faced differs between so called Arc of Fire and the rest of Legal Amazon. One of the actions to monitor and control illegal deforestation established by the federal government of Brazil was the Priority List of Municipalities. Almost all of those municipalities are situated in area dubbed as Arc of Fire, on the Eastern and Southern frontier of the Amazon forest. The list is edited annually by the Brazilian Ministry of Environment and can be accessed on its website. Municipalities are included in the list following criteria, which consider total area of forest cleared within municipality, total area cleared in the last three years and dynamics of the rate of deforestation in the last five years. Once in the list, municipalities receive support from the federal government in the implementation of actions aimed at reducing deforestation rates and also face strict auditing and control. Whenever municipalities, included in the Priority list, pass the criteria to be excluded from this list, they enter the list of municipalities under surveillance.

In this study municipalities, included in priority list in year 2009 and earlier, and all municipalities in the surveillance list were assigned to the Arc of Fire. In this way 43 municipalities belong to the Arc of Fire and 621 municipalities – to Inner Amazon. The average count of embargoes in the former varies from 13.4 to 38.1 in different years, and the average count in the latter zone varies from 2.1 to 4 embargoed areas. The lowest average counts for both regions were reported in 2006. Interestingly, the highest average count for the Arc of Fire was observed in 2012, whereas for Inner Amazon – in year 2008. The differences in average counts are clear. However, it is highly likely that the true unobserved difference is much smaller due to differences in underreporting between the two regions.

It is important to understand the implications of underreporting (extent to which forest crime sites are not disclosed) both from econometric and economic point of view.

The following example illustrates the importance of considering underreporting. The contribution of gross domestic product in determining the count of embargoed areas may be highly overstated in models, which do not consider underreporting, because lower GDP is linked with lower observed count (because of cut-offs in financing the process of tracking illegal logging activities). That is, when less financial and human resources are devoted to track infringers, less cases are identified and documented, and vice versa. However, actual unobserved count may not depend on GDP or depend to a lesser extent.

Therefore, ignoring underreporting (especially, when is it substantial) can lead to incorrect signs of the coefficients on the determinants and incorrect standard errors.

Earlier attempts to model underreporting in winBUGS environment, among others, can be found in Cariniceanu et al (2003) and McMillan et al (2009).

One of the reasons why the extent of detecting forest crimes is different between the two zones is that competent governmental institutions devote more financial and human resources to fight illegal logging in the Arc of Fire, that is, where deforestation is alarmingly high. However, there are other reasons. For instance, forest crime is easier and cheaper to track on the forest frontier than deep in the jungle due to the fact that the former is in close proximity to main cities and it is covered with dense road network. Secondly, territories in Inner Amazon are more humid and, therefore, may be covered by clouds for a very long period of time. This reduces the possibility to track any illegal deforestation via Satellite imagery.

Data and variables

The study uses data of 664 municipalities from Legal Amazon. Cross sectional analysis uses data of year 2010 with the exception of road variable, for which the data is available only for year 2009. Panel structure is formed from observations from 2006-2011 year period. The total number of observations for panel, therefore, is $664 \cdot 6 = 3984$.

Even though it is the first attempt to model causes of forest crimes in Legal Amazon, the topic is closely related to general problem of deforestation, which is extensively researched among scholars. Therefore, the covariates were selected based on previous works on deforestation in Legal Amazon, including Angelsen et al (1999), Assunção et al (2013), Faria et al (2013) and Kaimowitz et al (1998). The set of regressors mostly consists of economic, geographical, accessibility and land use variables. Unfortunately, data on political variables are not available. Some studies, like Mendes et al (2012), use reports from *Controladoria-Geral da União* (CGU) to create indices of corruption. However, these reports vary extremely in kind of information reported and are available only for quite small part of municipalities. Further, evidence based on indices derived from such unstructured information may be anecdotal. Another important missing variable is related to enforcement of laws related to deforestation. However, under panel models the absence of such data should not be such a drawback, since usually it has little temporal variation. The implication of this notion is that missing institutional covariates would not affect the performance of the model significantly (most importantly, the coefficients and standard errors).

Gross domestic product per capita variables can have a non-linear relation with dependent variable. This is in line with Environmental Kuznets Curve theory. However, GDP data is available at current prices only. It was deflated using consumer price index of Brazil (setting year 2006 as a base year). Unfortunately, data on consumer price index on municipality level is not being collected, thus preventing to consider the differences in inflation across municipalities in the research. Therefore, the data used in this study is pseudo-real GDP. The analysis provides reliable results in case inflation rates are similar all over Legal Amazon.

Also, notice that both square and cube of GDP per capita are scaled. This is important, because most statistical packages are not capable of dealing with large numbers (winBUGS does not load large numbers, *Stata* may fail to deliver some information in the output etc.).

The rural credit information is used in per capita terms to mitigate Intercorrelation between rural credit variables and agriculture and cattle variables.

The study splits population variable into rural and urban only under cross section models. Population for each year under panel structure is computed as a ratio between GDP and GDP per capita. Unfortunately, breakdown of GDP data by rural and urban areas in not available.

For detailed explanation of variables see the table below. The last column shows in which models a particular variable is used.

Table 1. Detailed description of the data

Variable	Code name	Explanation	Source	Models
DEPENDENT VARIABLE				
Embargoed areas	EMB	Number of embargoed areas	The Brazilian Institute of Environment and Renewable Natural Resources	1-7
INDEPENDENT VARIABLES				
Forest stock	FST	Area covered with forests, in square kilometers	The Brazilian National Institute for Space	1-7

			Research	
Openness to trade	OPT	Index, calculated by dividing the sum of values of imported and exported goods by GDP: $OPT_{it} = \frac{IMP_{it} + EXP_{it}}{GDP_{it}} \times 1000$	The Brazilian Ministry of Development, Industry and Foreign Trade and The Brazilian Institute of Geography and Statistics (SIDRA database)	1-7
Roads	RT	Length of paved, unpaved and natural roads in kilometers (year 2009)	Author's elaboration based on maps by The Brazilian Ministry of Transport	1-4
Boat	BOAT	Dummy variable on presence of access by boat (1 if access is possible)	The Brazilian Institute of Geography and Statistics (The Profile of Municipalities)	1-7
GDP per capita	GDPpc	GDP per capita in dollars	The Brazilian Institute of Geography and Statistics (SIDRA database)	1-7
Square of GDP per capita	GDPpc2sc	Square of GDP per capita in dollars, divided by 1000	Author's calculations	1-7
Cube of GDP per capita	GDPpc3sc	Cube of GDP per capita in dollars, divided by 1 million	Author's calculations	1-7
Cattle ranching	CATTLE	Number of cattle animals (heads)	The Brazilian Institute of Geography and Statistics (SIDRA database)	1-7
Agricultural activities	AGR	Area in square kilometers covered by temporal (yearly) agricultural plants	The Brazilian Institute of Geography and Statistics (SIDRA database)	1-7
Timber	TIMBER	Value of extracted timber products (round wood production, firewood and charcoal) in thousands of dollars	The Brazilian Institute of Geography and Statistics (SIDRA database)	1-7
Rural credit for agriculture	RCApc	The ratio of the amount of money in dollars, granted for agricultural activities, and the area used for cultivation of agricultural plants	Central Bank of Brazil and author's calculations	1-7
Rural credit for cattle ranching	RCCpc	The ratio of the amount of money in dollars, granted for cattle ranching, and the number of cattle animals	Central Bank of Brazil and author's calculations	1-7
Population	POP	Total number of inhabitants, calculated as:	The Brazilian Institute of Geography and Statistics (SIDRA database) and	5-7

		$POP_{it} = \frac{GDP_{it}}{GDPpc_{it}} \times 1000$	author's calculations	
Rural population	POPrur	Number of inhabitants in rural areas	The Brazilian Institute of Geography and Statistics (SIDRA database)	1-4
Urban population	POPurb	Number of inhabitants in urban areas	The Brazilian Institute of Geography and Statistics (SIDRA database)	1-4
Altitude	ALT	Altitude in meters of municipality's major city	The Brazilian Institute of Geography and Statistics	1-4
Time dummies	T06-T10	Year 2011 is chosen as a base year	Author	5-7

However, one problem related to the dependent variable remains unsolved. The database of embargoed areas reflects the date when a crime was documented. The problem is that the time lag between this date and time when the crime actually occurred may be substantial and the length of time lag itself may differ across municipalities. Unfortunately, this lag is unobserved and cannot be reliably inferred from the existing sources of information. The study assumes no time lag between the time when the crime is committed and the time when it is documented, but the note on this problem is worth mentioning.

Methodology

The study conducts both cross sectional and panel data analyses. Cross section analysis is comprised of four models: one presents simple approach without taking into account underreporting and the rest employs MCMC method to evaluate it, the only difference among the three models being prior distributions of recovery rates. The article also presents three models of panel data: one model does not take into account underreporting and the remaining two models consider underreporting: one of those models assumes constant recovery rate in the region dubbed as the Arc of Fire and the other relaxes this restriction. All three models additionally include time dummies to take into account time specific effects.

Potential endogeneity issues in cross sectional analysis are addressed by applying two stage residual inclusion method (further in the text as 2SRI).

All models are fitted using negative binomial distribution, since about half of counts are zeros, thus ensuring that the variance of the counts of embargoed areas exceeds the mean.

Cross sectional analysis

Since negative binomial distribution can be viewed as a gamma mixture of Poisson random variables, its conditional likelihood can be expressed as follows (StataCorp, 2009):

$$f\left(emb_i \mid u_i\right)=\frac{\left(u_i \mu_i\right)^{emb_i} e^{-u_i \mu_i}}{\Gamma\left(emb_i+1\right)}, \text{ where } \mu_i=\exp\left(\sum_{k=1}^{n} x_{ik} \beta_k\right)$$

In the equation above u_i are unobserved parameters, μ_i are expected counts of embargoed areas, x_{ik} stand for k^{th} regressor and β_k represent coefficients of the determinants. Parameters u_i have a $gamma(1/r,r)$ density, which can be written as follows (here r represents over dispersion parameter):

$$g(u)=\frac{u^{(1-r)/r} e^{-u/r}}{r^{1/r}\Gamma(1/r)}$$

This type of parameterization describes mean-dispersion negative binomial model and is applied to calculate Model 1 in this research.

The variables used are tested for potential endogeneity. Even though the study does not model deforestation directly, forest crime modeling is closely related topic. Therefore, arguments concerning deforestation and endogenous nature of particular variables may be valid. The literature often considers poverty and population variables as endogenous (Angelsen et al, 1999). Some studies suggest including road variables into the list of endogenous covariates (see Kaimowitz et al, 1998 and Andersen et al, 2002). However, only lagged data on road network is available and, therefore, this variable must be addressed as exogenous. Frankel et al (2005) treat trade as endogenous. Based on existing literature, three variables will be assumed to be exhibit endogenous relations with dependent variable. They include gross domestic product per capita, openness to trade and rural population. It is assumed that increased loggings may attract more inhabitants from rural areas. On the other hand, it does not attract city citizens due to the fact that individuals, who live in urban areas, hold job positions unrelated to deforestation. As opposed, occupations of inhabitants in rural areas are usually related to the environment.

Potential endogenous nature of three selected variables will be evaluated based on 2SRI approach, which provides unbiased and consistent estimates (Terza et al, 2008). In the first stage each potentially endogenous variable is regressed against all exogenous variables and selected instruments. The instrument set is comprised of one year lags of each potentially endogenous covariate and gross domestic products of years 2009 and 2010 to strengthen instrumental set. Note that gross domestic per capita has a non-linear relationship with dependent variable. Therefore, the variable, its square and its cube are treated as separate covariates. Notice that first stage regressions are linear. Therefore, the resulting coefficients can be directly used to predict potentially endogenous variables. The errors are computed by subtracting predicted and observed values of potentially endogenous covariates. Mathematically, it can be written as (here z_j are j^{th} instrumental or exogenous variables, $q^{(l)}$ represents endogenous covariate, $\varepsilon_i^{(l)}$ are the error terms, $\hat{q}_i^{(l)}$ stand for predicted values of l^{th} endogenous regressor, $b_j^{(l)}$ are estimated values of the coefficients on the regressors, m is total number of instrumental and exogenous variables and l is a number of equations to be estimated, or, equally, a number of endogenous covariates):

$$q_i^{(l)} = \sum_{j=1}^{m} z_{ij}\beta_j^{(l)} + \varepsilon_i^{(l)}$$

$$\hat{q}_i^{(l)} = \sum_{j=1}^{m} z_{ij} b_j^{(l)}$$

$$error_i^{(l)} = \hat{q}_i^{(l)} - q_i^{(l)}$$

Second stage is estimated using the negative binomial method described above. Expected count is modeled to depend on exogenous and endogenous covariates (not the predictions) and error terms from the first stage, which serve as unobserved confounders (Terza et al, 2008). That is:

$$\mu_i^{2sri} = \exp\left(\sum_{k=1}^{n} x_{ik}\beta_k + \sum_{l=1}^{p} error_{ip}\eta_p\right)$$

One advantage of 2SRI method is that by looking at statistical significance of the coefficients on error terms (η_p) one can decide whether potentially endogenous covariate should be treated as exogenous. If error terms are statistically insignificant, one should conclude that all regressors are exogenous. However, one important note is that standard errors of coefficients, computed using the two step procedure above, are incorrect. To correct for standard errors bootstrapping in *Stata* was employed. The whole two step method was bootstrapped using 500 repetitions to obtain consistent estimates of standard errors of the coefficients. Then Z statistics were computed as a ratio between a coefficient and bootstrapped standard error for that coefficient.

However, it is crucial that selected instrumental variables are strong instruments for potentially endogenous variables. Otherwise, the resulting standard errors can be too small, thus misleading the researcher to conclude that the variable tested is exogenous when it is not. Given that the first stage is linear, the two methods – 2SRI and 2SLS – provide with identical results (Terza et al, 2008). Therefore, command *ffirst* from *Stata's* two stage least squares method can be applied. One test statistic reported by *Stata* is Angrist-Pinchke first stage F statistic, which is used as a diagnostic for whether a particular set of instruments is strong for particular endogenous regressor. Critical values for that test are not available, but the test statistic is compared with Stock-Yogo weak identification test critical values for single endogenous regressor.

The procedure of testing for endogeneity was conducted in three steps. In the first step openness to trade, gross domestic product per capita, its square, its cube and rural population were treated as endogenous. After performing the procedure described above, the most statistically insignificant coefficient of error term was identified and the variable, to which corresponds that error term, was included into the list of exogenous covariates (all three gross domestic product per capita variables are treated as one in a sense that all three are excluded or retained in the list). In the second step the whole algorithm of 2SRI and bootstrapping was repeated once more, the only difference is that now one variable less was treated as endogenous and fewer instruments were used. After the procedure, the most statistically insignificant coefficient of error term again suggested which variable to exclude from potentially endogenous variable list. Finally, in the third step, the last remaining potentially endogenous covariate was tested.

The results revealed that all the variables can be treated as exogenous (see table below). These results are presented under methodology section because they have implications on the methodology explained further in the text.

However, endogeneity may be difficult to track due to potential time lag (the period between occurrence of forest crime and disclosure of that forest crime) in the dependent variable. If this lag is substantial, the dependent variable cannot influence any of potentially endogenous regressors simply because future events cannot influence present events.

To evaluate the scope of this problem, the 2SRI method was applied in the model, where independent variables are lagged by one year. This model rests on the notion that on average it takes one year to discover average illegal logging site. The results of endogeneity analysis revealed that coefficients on the error terms were even more statistically insignificant than under original case. However, this finding does not rule out the possibility of endogenous nature of particular variables, because the true lag of dependent variable is still unknown.

Table 2. Summary of endogeneity analysis

Error term	Coefficient	Bootstrapped SE	*Z* statistics	AP F test
STEP 1				
opt10error	-0.00928	0.007785	-1.19197	88.98*
gdp10pcerror	-0.000127	0.000179	-0.71069	119.83*
gdp10pc2scerror	9.90e-06	7.89e-06	1.254753	16.32**
gdp10pc3scerror	-9.52e-08	9.46e-08	-1.00634	42.16*
pop10rurerror	4.17e-06	2.58e-05	0.161628	672.64*
STEP 2				
opt10error	-0.012061	0.006968	-1.73084	93.01*
gdp10pcerror	-0.000066	0.000191	-0.34519	112.76*
gdp10pc2scerror	8.69e-06	8.97e-06	0.968785	14.39**
gdp10pc3scerror	-8.60e-08	1.01e-07	-0.85149	34.35*
STEP 3				
opt10error	-0.007735	0.011301	-0.68447	278.82***

* Passed 5% relative IV size critical value of Stock-Yogo test
** Passed 10% relative IV size critical value of Stock-Yogo test
*** Passed 10% maximal IV size critical value of Stock-Yogo test

The first model (Model 1) just introduced above does not consider underreporting. This is a severe limitation, since underreporting may be extremely high in dense jungle and it may be very different in the latter zone and in the Arc of Fire, where environmental inspectors can much more easily track the infringers. To take into account underreporting, winBUGS package was used, which relies on Gibbs sampling and is powered by MCMC algorithm. Short mathematical description of negative binomial model with underreporting is offered below:

$$emb_i \sim negbin(p_i, r)$$

$$p_i = \frac{r}{r + \mu_i}$$

$$\mu_i = C_i \times S_i$$

$$\ln C_i = \sum_{k=1}^{n} x_{ik} \beta_k$$

$$S_i = \begin{cases} R_1 \text{ for } i < 621.5 \\ R_2 \text{ for } i > 621.5 \end{cases}$$

$$\beta_k \sim norm(0, 0.001)$$

$$R_1 \sim beta(a_1, b_1)$$

$$R_2 \sim beta(a_2, b_2)$$

$$r \sim gamma(1, 0.1)$$

Popular parameterization of negative binomial model was used, which is applicable in many situations. For instance, Sartorius (2013) uses it in modeling age-specific mortality and Pardoe et al (2003) – in modeling consumer preferences. Here p_i are probabilities of a given amount of failures before r successes. Expected count (μ_i) is the product of unobserved actual count of embargoed areas (C_i) and the recovery rate (R_i). The data was sorted in a way that the first 621 observations correspond to Inner Amazon and the rest belongs to the Arc of Fire. Therefore, R_1 is a recovery rate for Inner Amazon and R_2 is a recovery rate for the Arc of Fire. The logarithm of unobserved actual count is regressed against the determinants of embargoed areas (x_{ik}). Coefficients (β_k) are fitted using normal distributions and are given weak priors. Recovery rates follow beta distributions with shape parameters α and β. Finally, the dispersion parameter r is assumed to follow gamma distribution with mean equal to 1 and variance of 0.1.

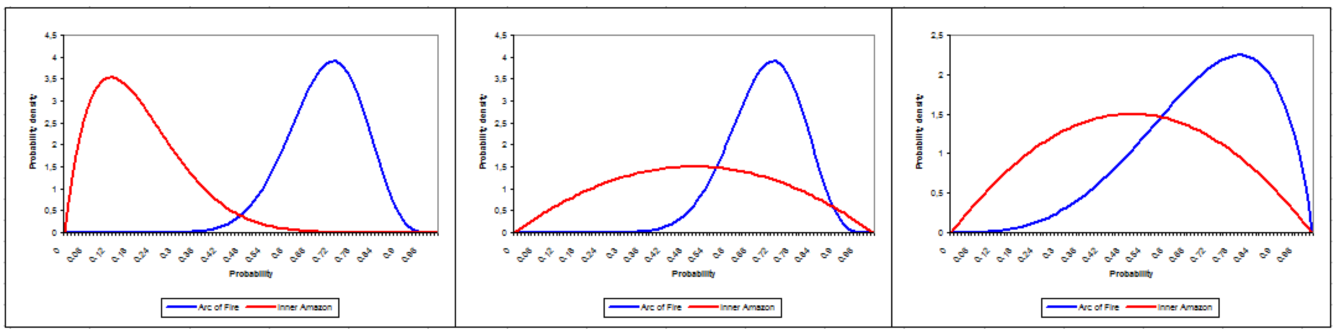
Since recovery rates are unobserved, three different models with different prior distributions of the recovery rates were run to check if the results are sensitive to the choice of R_1 and R_2 distributions. The mean of any beta distribution is given by the following formula (parameters α and β are shape parameters of beta distribution):

$$E_{beta}[R] = \frac{\alpha}{\alpha + \beta}$$

The first model with underreporting (or Model 2) assumes mean recovery rate equal to 0.2 in Inner Amazon and 0.7 – in the Arc of Fire. Both distributions are assigned low variance, accounting to the belief that deep Amazon almost certainly hides a lot of yet undiscovered forest crime sites and Arc of Fire region almost certainly has the majority of forest crime cases discovered and documented. Model 3 makes a different assumption concerning deep Amazon jungle – it allows almost for any recovery rate due to the fact that there is little information on illegal logging and other forest crime activities in the area. The mean recovery rate is selected to be 0.5. Finally, Model 4 retains the idea of previous model regarding Inner Amazon, but also gives more variance for recovery rate of the Arc of Fire to account for higher uncertainty. However, mean recovery rate remained unchanged and equal to 0.7.

The graphs below depict all three choices of prior distributions. The distributions of recovery rates for each graph are presented below figure 1.

Figure 1. Prior distributions of recovery rates (Model 2-Model 4 and Models 6a, 6b and 6c)



Model 2 (left graph): $R_1 \sim beta(2,8)$ and $R_2 \sim beta(14,6)$

Model 3 (middle graph): $R_1 \sim beta(2,2)$ and $R_2 \sim beta(14,6)$

Model 4 (right graph): $R_1 \sim beta(2,2)$ and $R_2 \sim beta(4.2,1.8)$

One of key parts of MCMC method is the choice of initial values. Zero values were selected as initials for coefficients, means of recovery rate distributions were selected as starting values for recovery rates and mean value of 1 - as an initial for dispersion parameter.

Starting 20000 iterations were used as a burn-in period and further 80000 iterations to fit the models. A burn-in period is a number of starting samples to be discarded. The purpose of this procedure is to allow the chain to stabilize and mitigate the effects of the initial values (Thompson el al, 2006).

Panel data analysis

Sometimes panel data may reveal aspects that cannot be seen under cross sectional analysis. The most obvious advantage is that panel structure has more observations and, therefore, can provide with more accurate results. Secondly, time aspect is also reflected. Finally, panel data models can be an efficient remedy against omitted variable bias if those missing variables exhibit little temporal variation.

The very first question to answer in panel models is what kind of approach – fixed or random effects – to follow. There is no consensus in the literature on how the type of model should be selected. The general rule suggests choosing random effects model whenever no correlation between individual specific effects and independent variables is anticipated. However, this correlation is unobserved. Clarck et al (2013) note that the choice between the two approaches is a tradeoff: fixed effects models will produce unbiased estimates, but are subject to high sample-to-sample variability, whereas random effects models almost always introduce some bias in estimates, but also greatly reduce the variance of those estimates. This notion implies that random effects model can still be preferred in the presence of some degree of correlation between individual effects and the regressors. Clarck et al (2013) also show that broadly used Hausman test fails to offer correct type of model. They argue that the test does not have sufficient statistical power to reliably detect departures from the null hypothesis. The way to simultaneously assess bias and variance of estimators is root mean squared error (RMSE). Therefore, the measure of both fixed and random effects approaches was computed. The formula is presented below (here λ_I represents expected count of embargoed areas, N is the total number of observations and subscript I marks observations of pooled data):

$$RMSE = \sqrt{\frac{1}{N}\sum_{i=1}^{N}\left(emb_I - \lambda_I\right)^2}$$

RMSE of fixed effects model (with time dummies) is calculated to be 11.24, whereas the same measure of random effects model (with time dummies) is equal to 11.68. The difference between the two figures is negligible. This was further confirmed by *Stata's ttest*. The author turned to random effects model due to following reasons: 1) random effects model allowed to include one accessibility variable which does not vary in time (altitude variable was dropped from the models because it exhibits little spatial variation and no temporal variation thus creating problems for winBUGS to fit the models), 2) fixed effects models drop all observations without within-group variance of the dependent variable, which constitutes quite substantial share in this research.

Further, let's introduce random effects negative binomial model for panel data. In the model described below the over dispersion parameter follows beta distribution and is allowed to vary randomly across groups. Then the joint probability of counts for i^{th} group can be expressed as follows (StataCorp, 2009):

$$\Pr\left(EMB_i \middle| \mathrm{X}_i\right) = \frac{\Gamma(\alpha+\beta)\Gamma\left(\alpha+\sum_{t=1}^{T}\lambda_{it}\right)\Gamma\left(\beta+\sum_{t=1}^{T}emb_{it}\right)}{\Gamma(\alpha)\Gamma(\beta)\Gamma\left(\alpha+\beta+\sum_{t=1}^{T}\lambda_{it}+\sum_{t=1}^{T}emb_{it}\right)}\prod_{t=1}^{T}\frac{\Gamma\left(\lambda_{it}+emb_{it}\right)}{\Gamma\left(\lambda_{it}\right)\Gamma\left(emb_{it}+1\right)},$$

$$\text{where } EMB_i = \left(emb_{i1},...,emb_{iT}\right) \text{ and } X_i = \left(X_{i1},...,X_{iT}\right)$$

Here α and β represent shape parameters of beta distribution and λ_{it} is expected count of embargoed areas for i^{th} observation in t^{th} time period. The capital X_i is i^{th} vector of T matrices of the values of independent regressors.

Time dummies were included in to the model (referred to as Model 5) to account for time specific effects. Therefore, the expected count is modeled as follows ($time_t$ are dummy variables for each year, but year 2011, and γ_t are coefficients on those time dummies):

$$\lambda_{it} = \exp\left(\sum_{k=1}^{n} x_{itk}\beta_k + \sum_{t=1}^{T-1} time_t\gamma_t\right)$$

From this point underreporting is introduced. To check how important is to consider temporal changes in recovery rates of the Arc of Fire, the study calculates three models with fixed recovery rate and three models, where varying recovery rates are assumed (the models will be referred to as Models 6a, 6b, 6c and Models 7a, 7b, 7c respectively later in the text).

In case of panel models with time-fixed recovery rate, the same prior distributions of recovery rates, applied under cross section analysis, were tested (see figure 1). Short description of Model 6 is presented below:

$$emb_{it} \sim negbin\left(p_{it}, r\right)$$

$$p_{it} = \frac{r}{r + \lambda_{it}}$$

$$\lambda_{it} = C_{it} \times S_{it}$$

$$\ln C_{it} = \sum_{k=1}^{n} x_{itk}\beta_k + \sum_{t=1}^{T-1} time_t\gamma_t + re_i$$

$$S_{it} = \begin{cases} R_1 \text{ for } i < 621.5 \\ R_2 \text{ for } i > 621.5 \end{cases}$$

$$re_i \sim norm\left(0, \tau\right)$$

$$\beta_k \sim norm\left(0, 0.001\right)$$

$$R_1 \sim beta\left(2, 8\right)$$

$$R_2 \sim beta\left(12, 8\right)$$

$$\tau \sim gamma\left(0.001, 0.001\right)$$

$$r \sim gamma\left(1, 0.1\right)$$

The notation is the same as under cross section case. The main difference in this model is that it includes random effects. Usually random effects are assumed to follow normal distribution. Indeed, McCulloch et al (2012) show that misspecification of the shape of random effects distribution does not matter, except for the intercept. According to McCulloch et al (2012), simulation studies indicate that most aspects of statistical inference are highly robust to normality assumption. Therefore, random effects (re_i) were assumed to follow normal distribution with zero mean and precision *tau* (τ), which, in turn, is assumed to follow gamma distribution.

The successful fit of the model depends a lot on initial values. Failure to provide the model with appropriate starting values of the parameters may result in problematic posterior distributions and, consequently, misleading results. The author follows the same reasoning as under cross section case. However, this time all individual specific effects must be initialized also. One simple way is to input zero values as initials. However, estimates of individual effects from pooled fixed effects model were used instead to improve the shape of posterior distributions. Since data is pooled, simple cross sectional negative binomial model, explained in the first paragraphs of this section, can be applied. The only difference is that equation of expected count additionally includes N-1 dummy variables for each observation (except the base observation). Mathematically:

$$\mu_I^{FE} = \exp\left(\sum_{k=1}^{n} x_{Ik}\beta_k + \sum_{t=1}^{T-1} time_t \gamma_t + \sum_{m=1}^{N-1} d_{Im}\delta_m \right)$$

Then coefficients on dummies of individual effects (δ_m) were used as initial values of random effects in winBUGS. Initial 20000 iterations were discarded (burn-in period) and 80000 subsequent iterations were used to fit the models.

Further, models with time varying recovery rates were implemented. The algorythm of models with time varying recovery rates resembles the algortyhm of Models 6a, 6b and 6c, already introduced in the text. To adapt the formulee,six different recovery rates for the Arc of Fire region should be introduced, thus reflecting the temporal changes. The code of the last model is available in the Appendix A.

The temporal dynamics of underreporting in the Arc of Fire was investigated under three different scenarios (for sensitivity analysis). Only Inner Amazon retained fixed recovery rate. The later notion relies on the assumption that underreporting in Inner Amazon does not change over time, because no substantial technological advancements were made in Satellite imagery during the period under consideration. Nor any other major illegal logging related initiatives were carried out in that region.

Back in 2005 (just before the launch of the program that takes away lands for crimes against forests) The Brazilian Institute of Environment and Renewable Natural Resources estimated that it can only identify 10 percent of all illegal logging activities in the Brazilian Amazon

Greenpeace, 2005). Based on this note and properties of beta distribution[1] the recovery rate for Inner Amazon was set to follow beta distribution with mean 0.2.

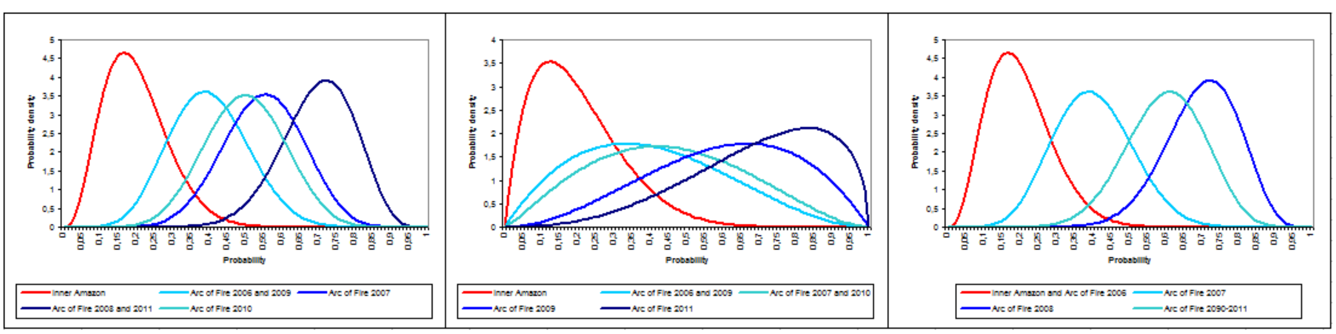
Conversely, the Arc of Fire region faces ever increasing attention from inspectors of the government. Therefore, the level of underreporting can change over the years accordingly. Three different sets of prior distributions were considered to check how robust the results are to the choice of the distributions of the recovery rates. The summary of the assumed means of recovery rates in the Arc of Fire over the years is presented below:

	2006	2007	2008	2009	2010	2011
Case 1	0.4	0.55	0.7	0.4	0.5	0.7
Case 2	0.4	0.45	0.6	0.4	0.45	0.7
Case 3	0.2	0.4	0.7	0.6	0.6	0.6

The first case assumes effective combat against illegal forest activities in the initial years of the program until year 2008. It also assumes a sharp drop of the recovery rate in the Arc of Fire due to the impact of global crisis and financial cut-offs in budget of auditing the area. Finally, rapid recovery is assumed. Case 2 models similar patterns as Case 1, but it assumes much slower rate of reduction of underreporting to account for the potentially long process of optimizing the system of land auditing. The last case assumes very low recovery rate initially, rapid decrease in undisclosed cases and stable recovery rate after the global crisis. The last scenario is least likely, but it is necessary to evaluate how the final result depends on the prior beliefs about recovery rates.

However, deciding on the means of prior distributions is only half the job done. Controlling for variance is equally important. As can be seen in figure 2, cases 1 and 3 are assigned pretty low variance, which reflects quite high degree of certainty about selected mean values. Under case 2, large variance is used to support the belief of uncertainly surrounding chosen means of recovery rates.

Figure 2. Prior distributions of time varying recovery rates (cases 1, 2 and 3 depicted on the left, middle and right graphs respectively)



[1] Beta distributions with mean equal to 0.1 and higher variance assign extremely high probablility that the true recovery rate is zero and sets very low variance for bell-shaped distributions, which is too restrictive given relatively high degree of uncertainty about the true recovery rate.

$$\text{Case 1: } R_1 \sim beta(4,16); R_2, R_5 \sim beta(8,12); R_3 \sim beta(11,9); R_4, R_7 \sim beta(14,6); R_6 \sim beta(10,10)$$
$$\text{Case 2: } R_1 \sim beta(2,8); R_2, R_5 \sim beta(2,3); R_3, R_6 \sim beta(2.25, 2.75); R_4 \sim beta(3,2); R_7 \sim beta(3.5,1.5)$$
$$\text{Case 3: } R_1, R_2 \sim beta(4,16); R_3 \sim beta(8,12); R_4 \sim beta(14,6); R_5, R_6, R_7 \sim beta(12,8)$$

Over reporting could also happen in coming years. As more and more financial and human resources are devoted to track illegal forest activities and technological aspect regarding Satellite imagery is ever improving, observed count in that specific year approaches the true count. Besides, illegal activities from the past years are also revealed. Therefore, the sum of the two may exceed the true count of forest crimes. Note that beta distribution is not suitable for modeling over reporting, since it is bounded between 0 and 1, thus automatically ruling out any possibility of over reporting. For purposes of this study this is not a problem, since undiscovered cases of illegal forest activities remains prevalent.

Measures of goodness of fit were reported to check how much predictive power each model has and to compare models of same sort. McFadden determinantion coefficient was used to measure the fit of maximum likelihood models, whereas deviance information criterion (DIC) was used to serve the same purpose in MCMC models. The DIC tool was initialized only after burn in periods were completed.

The results

Cross section results

Cross sectional analysis revealed that simple negative binomial regression without considering underreporting reflects the situation very accurately (see table 3). Only two differences can be observed: under models with underreporting the intercept is statistically insignificant and rural credit for cattle ranching influences the number of embargoed areas (all results are reported at 5% statistical significance level).

The first discrepancy may be down to the fact that the shape of random effects distribution may be misspecified (McCulloch et al, 2012) or due to autocorrelation between sequential draws of parameters in MCMC algorithm. The second difference may be explained by looking at posterior distributions of the parameters (see Appendix B, figure 3), since the distribution of the coefficient for rural credit does not fit normal distribution properly. Surely, it might be that the differences are observed due to underreporting issues.

However, in general MCMC models are well fitted, as the shape of posterior distributions of various parameters reveal, though the majority of the graphs are slightly skewed either to the right or to the left. The analysis of autocorrelation between sequential draws of parameters showed that most of the parameters are free of that problem, except the coefficients of the intercept and GDP.

It is also apparent that the results are not sensitive to the choice of prior distributions of the two recovery rates.

Not surprisingly, the results suggest that forest stock positively influences the number of embargoed areas. The larger the forest, the easier is to hide illegal logging and other forest crimes. Considering this, the number of illegal activities could increase more than proportionally with the size of the forest. However this notion is not backed by the academic literature.

As for openness to trade, the models suggest that it does not influence the number of embargoes. However, this does not imply that trade is not important. The data on openness to trade reflects international trade only and cannot capture internal movements of chopped woods. Those cities from which the goods are exported can be viewed as hubs. Products for trade originate in various places in Legal Amazon, then these products are either shipped or transported in trucks to main cities (usually, ports). Here the trade flows are documented and this is the information, which is available for the research. Since main cities are usually situated far from forests or in areas with very low vegetation, the scale of illegal logging there is low. This is why available data cannot capture relations between trade and illegal logging.

Table 3. Coefficients of cross sectional analysis (for Model 1 standard errors are reported in parenthesis, for Models 2 to 4 credible intervals are presented)

Methods – maximum likelihood (Model 1) and Markov Chain Monte Carlo (Models 2 to 4)
Number of observations – 664
Dependent variable – EMB10

	Model 1	Model 2	Model 3	Model 4
McFadden R^2				
DIC				
Mean R1	-	0.35	0.48	0.49
Mean R2	-	0.67	0.71	0.74
Over dispersion parameter				
cons	-1.555342* (0.2742948)	-0.5414 [-1.325,0.281]	-0.8075 [-1.565,1.04e-04]	-0.8429 [-1.64,0.003526]
FST10	0.0000305* (8.72e-06)	2.88E-05* [1.19e-05,4.75e-05]	2.97E-05* [1.29e-05,4.83e-05]	2.97E-05* [1.29e-05,4.84e-05]
OPT10	-0.0009794 (0.0006426)	-0.00107 [-0.002587,3.08e-04]	-0.001055 [-0.002558,3.16e-04]	-0.001063 [-0.002574,3.26e-04]
RT09	0.0032729* (0.0009339)	0.003181* [0.001343,0.005129]	0.003223* [0.001409,0.005152]	0.003194* [0.001368,0.00508]
BOAT	0.3287401* (0.1568322)	0.3436* [0.03012,0.6611]	0.3461* [0.03132,0.6607]	0.3385* [0.02384,0.6613]
GDPpc10	0.0001548* (0.0000348)	1.73e-04* [9.61e-05,2.56e-04]	1.63E-04* [8.98e-05,2.34e-04]	1.71E-04* [9.95e-05,2.48e-04]
GDPpc2sc10	-3.78e-06* (1.24e-06)	-4.57e-06* [-7,8e-06,-1.76e-06]	-4.20E-06* [-6.98e-06,-1.5e-06]	-4.54E-06* [-7.58e-06,-1.86e-06]
GDPpc3sc10	2.39e-08*	3.19e-08*	2.86E-08*	3.18E-08*

	(1.09e-08)	[6.87e-09,6.31e-08]	[5.02e-09,5.56e-08]	[7.43e-09,6.12e-08]
CATTLE10	1.96e-06* (5.88e-07)	1.36e-06* [1.75e-07,2.58e-06]	1.60E-06* [3.98e-07,2.87e-06]	1.55E-06* [3.28e-07,2.8e-06]
AGR10	-1.49e-06 (1.28e-06)	-1.57e-06 [-4.16e-06,1.18e-06]	-1.53E-06 [-4.16e-06,1.23e-06]	-1.43E-06 [-4.08e-06,1.39e-06]
TIMBER10	0.0000216* (7.77e-06)	1.66e-05* [2.88e-06,3.33e-05]	1.85E-05* [3.99e-06,3.56e-05]	1.83E-05* [3.94e-06,3.55e-05]
RCApc10	-0.0000732* (0.0000363)	-7.36e-05* [-1.5e-04,-2.44e-06]	-7.48E-05* [-1.49e-04,-3e-06]	-7.49E-05* [-1.5e-04,-4.3e-06]
RCCpc10	-0.0000829 (0.0001769)	-4.37e-04* [-0.00127,-4.45e-06]	-4.47E-04* [-0.001278,-9.51e-06]	-4.41E-04* [-0.001267,-8.16e-06]
POPrur10	0.0000386* (8.94e-06)	4.27e-05* [2.51e-05,6.2e-05]	4.11E-05* [2.33e-05,5.98e-05]	4.15E-05* [2.41e-05,6.02e-05]
POPurb10	9.44e-07 (1.10e-06)	1.45e-06 [-4.6e-07,4.06e-06]	1.51E-06 [-4.18e-07,4.18e-06]	1.49E-06 [-4.57e-07,4.11e-06]
ALT	-0.0001058 (0.0007457)	-8.78e-06 [-0.001482,0.001544]	-2.23E-05 [-0.001482,0.001499]	-5.00E-05 [-0.001565,0.001501]

*Statistically significant at 5% level (Model 1) / 95% credible interval suggests statistical significance (Models 2 to 4)

The coefficients for accessibility variables are positive and statistically significant. Better network of roads not only facilitates the access to dense forests, but also ensures cheaper and faster transport of chopped woods.

The richness of a particular region also has implications on the number of embargoes. Under poverty individuals tend to clear forests illegally simply to make a living. As the level of life increases, the population tends to obey the forest laws. However, when large sums of money become available, individuals can acquire machinery for cutting tress, rent trucks, hire and pay salaries to employees. Even though infringers face sanctions when disclosed, the profit earned from forest crimes may exceed the losses incurred.

The cattle market was also found to be a significant contributor to the amount of illegal forest activities. Since land is limited and cattle business is very land intensive, new areas need to be cleared to feed animals at the expense of forests. However, not always this can be done legally. If cattle ranching is very lucrative, the farmers may prefer expanding cattle grazing territories and risk being caught rather than obey to forest laws and loose money because of failure to meet the demand of cattle meat and other cattle products.

The first unexpected result is statistical insignificance of the coefficient on agricultural area. One of the differences between cattle ranching and agriculture is that cattle animals can graze in different places, whereas plants under cultivation cannot be transported to another area. If illegally logged areas are cultivated, under disclosure the farmer looses not only the land, but also plants that grow in that land. Another explanation may be that agriculture business is not so profitable and it is not worthy to pay fines for illegal deforestation. Yet another notion is that agricultural lands in some cases are previous lands of cattle grazing. In such a case new areas need not to be cleared, the only change is the change of activity in that land.

Timber market is an important driver of forest crime. Since stricter regulations on illegal logging are under way and fewer forests remain standing, the price of timber products will rise in the future. Therefore, the profitability of timber market will increase. Paradoxically, cutting, transporting and selling timber and timber products create a lot of positive implications in the economy. For instance, it increases employment in Legal Amazon, creates ties of trade with other countries, satisfies the tastes of consumers and raises country's GDP.

Rural credit for agriculture reduces the number of embargoed areas. Under cross section analysis that money seems to serve as a safety rig against forest crimes, since the more money is received, the lower the willingness to commit some sort of forest crime and, as a result, loose land, agricultural harvest and financing.

As for rural credit for cattle ranching, the results depend on the model. The three MCMC models with underreporting suggest statistical significance, whereas simple negative binomial regression concludes that money for cattle ranching does not influence the amount of forest crimes. However, it is unclear, whether this is due to econometrical problems or underreporting. Both conclusions seem reasonable. Money may discourage from illegal activities. On the other hand, cattle ranching may be so profitable that it is worthy keep expanding grazing areas.

The results also imply that the more inhabited rural areas are, the more forest crimes occur. Individuals, who live outside cities, are involved in occupations directly (loggers) or indirectly (farmers) linked with forests and their resources.

Conversely, the number of inhabitants in urban areas does not influence illegal activities in forests. This is intuitive, because citizens engage in occupations unrelated to forests and, therefore, do not create pressure on them. The author would recommend splitting the data on population in rural and urban whenever possible. Usually, the number of inhabitants in cities exceeds rural population several times. Therefore, considering the whole population in deforestation models may hide the potential importance of rural population.

Altitude does not seem to have an effect on the amount of forest crimes. This result is not intuitive, since areas in the mountains usually are harder to access and the terrain itself is not so favorable for trees to grow compared to the areas close to fresh water. On the other hand, most areas in Legal Amazon are situated in relatively low altitude. Therefore, the differences may not be substantial.

Panel results

The coefficients, recovery rates and other parameters of panel data models are relatively well fitted with the exception of rural credit coefficient (see figure 4 in Appendix B).

Recovery rates. Estimated means of recovery rates of Model 7 reveal that Amazon jungle (excluding the Arc of Fire) shelters a vast amount of illegal loggers and other forest-related infringers. More than 85% of them are not disclosed (see table 4). This findings is backed by the results of Model 6, were mean recovery rate in Inner Amazon is predicted to be 17.81%. The Arc of Fire exhibits much higher mean recovery rates, where roughly half of infringers were disclosed over the 5 year period (in comparison, Model 6 predicts fixed mean recovery rate equal to 60.77%). However, the rate of disclosure was notably different over the years. The first loop of recovery rates between years 2007 and 2008 may suggest larger investment in the program or potential increase in efficiency of catching forest-related infringers due to learning by doing process. The second loop could be attributed to global crisis and consequential cut-offs in the budget.

Table 4. Estimated mean recovery rates in percentage (Model 7)

Inner Amazon	Arc of Fire				
R1 (2006-2010)	R2 (2006)	R3 (2007)	R4 (2008)	R5 (2009)	R6 (2010)
14.3	45.29	46.37	62.64	62.96	39.36

The models. Panel data analysis uncovered similar patterns as cross section models. Negative binomial model and MCMC models with underreporting offer similar conclusions. The differences include: 1) in MCMC models the coefficient on intercept is statistically insignificant and positive 2) the coefficients on openness to trade turned from positive into negative when underreporting was considered, though remained statistically insignificant, 3) the coefficient on agricultural area became statistically insignificant in MCMC models (see table 5).

The only difference in the results between cross section and panel analyses is that coefficients on rural credit are statistically insignificant in the later case. This finding suggests that credit system is inefficient in reducing the number of illegal activities against forests. In other words, the potential to earn from cattle ranching and agriculture businesses is significant enough to ignore the financial supply from the Central Bank.

Another interesting aspect is the inclusion of time dummies into panel regressions to capture time specific effects. The results show that year is an important factor in explaining the amount of forest crimes under any model. Statistically significant coefficient in model, that disregards underreporting, may suggest that the differences among years arise from the variability in detected forest crime spots. However, statistically significant coefficients on time dummies in MCMC models imply that the process of illegal logging is dynamic. This dynamicity may stem for fluctuations in the demand of meat, agrarian, timber and other products, which shape the extent to which land owners are willing to initiate illegal timber extraction and other crimes of similar sort.

The fact that the results depend very little on the choice of the method (maximum likelihood or MCMC), the choice of prior distributions in MCMC models, the choice between cross sectional and panel analyses and the choice between models with or without controlling for underreporting, makes the results robust and reliable.

Table 5. Coefficients of panel data analysis (for Model 5 errors are reported in parenthesis, for Models 6 and 7 credible intervals are presented)

Methods – maximum likelihood (Model 5) and Markov Chain Monte Carlo (Models 6 and 7)
Number of observations per period – 664
Number of periods – 6
Type of panel – balanced
Dependent variable – EMB

	Model 5	Model 6	Model 7
McFadden R^2			
DIC			
Over dispersion parameter			
cons	-1.575505* (0.1348585)	0.0338 [-0.5486,0.7621]	0.2741 [-0.2194,0.8255]
FST	0.000012* (2.66e-06)	1.98E-05* [1.11e-05,2.88e-05]	1.97E-05* [1.04e-05,2.87e-05]
OPT	0.0002168 (0.0002593)	-2.59E-04 [-0.00107,5.39e-04]	-2.72E-04 [-0.001083,4.99e-04]
BOAT	0.3341389* (0.0812131)	0.5055* [0.2639,0,7526]	0.5061* [0.2687,0.7457]
GDPpc	0.0000939* (0.0000138)	1.42E-04* [1.05e-04,1.74e-04]	1.42E-04* [1.04e-04,1.78e-04]
GDPpc2sc	-2.35e-06* (4.13e-07)	-3.25E-06* [-4.18e-06,-2.1e-06]	-3.28E-06* [-4.35e-06,-2.18e-06]
GDPpc3sc	1.23e-08* (2.94e-09)	1.65E-08* [6.87e-09,2.35e-08]	1.67E-08* [8.31e-09,2.45e-08]
CATTLE	2.29e-06* (2.24e-07)	2.99E-06* [2.22e-06,3.75e-06]	2.98E-06* [2.21e-06,3.77e-06]
AGR	1.94e-06* (6.77e-07)	8.66E-07 [-1.21e-06,3.02e-06]	8.44E-07 [-1.27e-06,2.94e-06]
TIMBER	9.98e-06* (1.88e-06)	7.55E-06* [9.7e-07,1.44e-05]	7.30E-06* [6.27e-07,1.42e-05]
RCApc	-0.0000122 (6.72e-06)	-1.40E-05 [-3.54e-05,7.21e-06]	-1.40E-05 [-3.55e-05,7.38e-06]
RCCpc	-0.0000281 (0.0000389)	-8.85E-05 [-2.46e-04,6.56e-06]	-8.64E-05 [-2.43e-04,7.65e-06]
POP	2.46e-06* (4.7e-07)	2.88E-06* [1.69e-06,4.08e-06]	2.89E-06* [1.71e-06,4.13e-06]
T06	0.2037432* (0.0701315)	0.1868* [0.02219,0.3519]	0.1701 [-0.007313,0.3481]
T07	0.3094318* (0.0671115)	0.329* [0.1753,0.486]	0.3093* [0.142,0,4738]
T08	0.7239081*	0.742*	0.6865*

	(0.0599443)	[0.5967,0.8899]	[0.5302,0.8403]
T09	0.3540905* (0.0603945)	0.3288* [0.1801,0.4804]	0.273* [0.1133,0.4334]

*Statistically significant at 5% level (Model 5) / 95% credible interval suggests statistical significance (Models 6 and 7)

Finally, it is interesting to compare the results of the study with the findings in deforestation literature, which tries to identify the determinants of logging in Legal Amazon. The spectrum of results for some variables is diverse, but for some variables there is a consensus. These variables include cattle ranching, agricultural activities, timber market and road network. The researchers agree that all four listed variables drive deforestation and that the strongest contributor is the amount of cattle. The results of this study are in line with these findings except for agricultural business. It seems that agriculture is not responsible for illegal logging and other forest crime activities, implying that the deforestation caused by plant cultivation is mostly legal.

Conclusions and recommendations

This study could not identify endogenous nature of the regressors. However, this finding may be a consequence of potential time lag in the dependent variable.

Further, underreporting issues were considered. Under cross section case the only notable difference in the results between maximum likelihood (ML) model without underreporting and MCMC models with underreporting is that the later suggest statistically significant coefficient on the variable of rural credit for cattle ranching. Under panel case, the ML and MCMC models disagreed on statistical significance of the coefficient on agricultural area. The preferred model suggests that agricultural area does not influence the number of embargoes.

Posterior distributions of various parameters in MCMC models revealed that the models are relatively well fitted. Only the distribution of the coefficient on rural credit for cattle ranching does not follow normal distribution adequately. As for autocorrelation between sequential draws of parameters, the process of fitting coefficients on intercept and GDP suffered from this problem. Otherwise, the problem does not seem to be substantial to raise worries.

Some sensitivity analysis was conducted – that is, how sensitive the results are to the choice of the prior distributions of the recovery rates. The findings were identical regardless of prior distributions used. Also, the results of the time-fixed recovery rate type models were compared with the results of the time varying recovery rates type models. The results did not change except for time dummy of year 2006, whose coefficient is statistically insignificant in the later model.

Both cross sectional and panel analyses provided with very similar results, though one notable difference was observed. Both types of models (when underreporting was addressed) concluded that forest stock, length of roads, GDP per capita, timber market and cattle ranching contributes to the amount of illegal logging and other related forest crime activities. In both types of models the coefficients on openness to trade and agricultural area were found to be statistically insignificant. Panel analysis concluded that total population effects number of embargoes positively. Under cross section analysis the data was splitted into rural and urban population. Only the former was concluded to positively influence the dependent variable (the coefficient on urban population was found to be statistically insignificant). Cross sectional analysis additionally included accessibility by boat and altitude variables. The results showed that presence of connection by boat increases the number of embargoes, whereas altitude has no effect. Panel models included time dummies to capture time specific effects. The coefficients of time dummies are statistically significant. However, one difference in results stands out. Cross sectional analysis concluded that both rural credit for agriculture and cattle ranching reduces the amount of illegal forest activities, whereas panel models provide with statistically insignificant coefficients on rural credit variables.

Politics play a significant role in fighting against illegal logging. Reliable data on enforcement of forest laws and level of corruption would greatly enrich the analysis. Future research could also find a way to incorporate the network of protected areas into the models.

Another possible improvement of this paper lies in econometric method. Since large amount of zero counts may imply the problem of excessive zeros, conventional negative binomial model may not be the best choice (at least from theoretical perspective). Possibly, zero inflated negative binomial method would provide with superior fit. To justify zero inflated models, two different processes – one generating zero counts and one generating positive counts – must be present. For example, in some areas were forest stock is low (urban areas or depleted territories), it is extremely difficult to hide illegal logging. Therefore, in those areas it is reasonable to assume that no infringements will happen.

However, current statistical packages are unable to offer zero inflated models for panel data. Despite this, the idea of excessive zero counts may serve for future research.

Acknowledgements

I would like to thank my PhD supervisor William Nilsson for providing me with many helpful comments, suggestions and insights and the University of the Balearic Islands for providing me with scholarship for this research.

References

Andersen, L.; Granger, C.; Reis, E.; Weinhold, D.; Wunder, S. (2002). The Dynamics of Deforestation and Economic Growth in the Brazilian Amazon. *Cambridge University Press.* Retrieved 11, 2012, from http://catdir.loc.gov/catdir/samples/cam033/2002067617.pdf

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.*

Assunção, J.; Gandour, C.; Rocha, Romero; Rocha, Rudi (2013). Does Credit Affect Deforestation?

Evidence from a Rural Credit Policy in the Brazilian Amazon. *Climate Policy Initiative.* Retrieved 06, 2013, from http://climatepolicyinitiative.org/wp-content/uploads/2013/01/Does-Credit-Affect-Deforestation-Evidence-from-a-Rural-Credit-Policy-in-the-Brazilian-Amazon-Technical-Paper-English.pdf

Butler, R. (2013). Illegal logging remains rampant in Brazil. Retrieved 12, 2013, from http://news.Mongabay.com/2013/1023-illegal-logging-brazil.html

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

Clark, T.; Linzer, D. (2013). Should I Use Fixed or Random Effects? *Emory University*. Retrieved 09, 2013, from http://polmeth.wustl.edu/media/Paper/ClarkLinzerREFEMar2012.pdf

Crainiceanu, C. M., J. R. Stedinger, D. Ruppert, and C. T. Behr (2003). Modeling the U.S. national distribution of waterborne pathogen concentrations with application to Cryptosporidium parvum. *Journal of Water Resource Research, vol. 39, no. 9, 1235, doi:10.1029/2002WR001664.*

Faria, W.; Almeida, A. (2013). Relationship between Openness to Trade and Deforestation: Empirical Evidence from the Brazilian Amazon. *The University of São Paulo. Regional and Urban Economics Lab.* Retrieved 06, 2013, from http://www.usp.br/nereus/wp-content/uploads/TD_Nereus_03_20131.pdf

Frankel J.; Rose A. (2005). Is Trade Good or Bad for the Environment? Sorting Out the Causality. *Review of Economics and Statistics, 87, no. 1.*

Greenpeace Internationa (2005). Logging in the Amazon. Retrieved 12, 2013, from http://www.greenpeace.org/international/en/campaigns/forests/amazon/logging-in-the-amazon/

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

McCulloch, C.; Neuhaus, J. (2012). Misspecifying the Shape of a Random Effects Distribution: Why Getting It Wrong May Not Matter. *Journal of Statistical Science, vol. 26, no. 3, pp. 388–402.*

McMillan, G. P.; Bedrick E.; C'deBaca, J. (2009). A Bayesian Model for Estimating the Effects of Drug Use when Drug Use may be Under-reported. *Journal of Addiction, vol. 104, issue 11, pp. 1820-1826.*

Mendes, C.; Porto, S. (2012). Deforestation, economic growth and corruption: a nonparametric analysis on the case of Amazon forest. *Applied Economics Letters, vol. 19, issue 13, pp. 1285-1291.*

Pardoe, I.; Durham, C. (2003). Model choice applied to consumer preferences. *In Proceedings of the 2003 Joint Statistical Meetings, Alexandria, VA. American Statistical Association.*

Sartorius, B. (2013). Modelling determinants, impact, and spacetime risk of age-specific mortality in rural South Africa: integrating methods to enhance policy relevance. *Journal of Global Health Action 2013, 6: 19239.*

StataCorp. 2009. *Stata Statistical Software: Release 11*. College Station, TX: StataCorp LP.

Terza, J.; Basu, A.; Rathouz, P. (2008). Two-stage residual inclusion estimation: Addressing endogeneity in health econometric modeling. *Journal of Health Economics, vol. 27, issue 3, pp. 531-543.*

The Brazilian Institute of Environment and Renewable Natural Resources (2013). Retrieved 06, 2013, from https://servicos.ibama.gov.br/ctf/publico/areasembargadas/ConsultaPublicaAreasEmbargadas.php

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 Institute of Geography and Statistics, The Profile of Municipalities (2013). Retrieved 06, 2013, from http://www.ibge.gov.br/munic2009/

The Brazilian Ministry of Development, Industry and Foreign Trade (2013). Retrieved 06, 2013, from http://www.desenvolvimento.gov.br/sitio/sistema/balanca/

The Brazilian Ministry of Environment (2013). The list of priority municipalities in Brazil. Retrieved 06, 2013, from http://www.mma.gov.br/images/arquivo/80120/lista_prioritarios_doc14_10_2013.pdf and http://www.mma.gov.br/images/arquivo/80120/Lista_site_monit_ doc14_10_2013.pdf

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

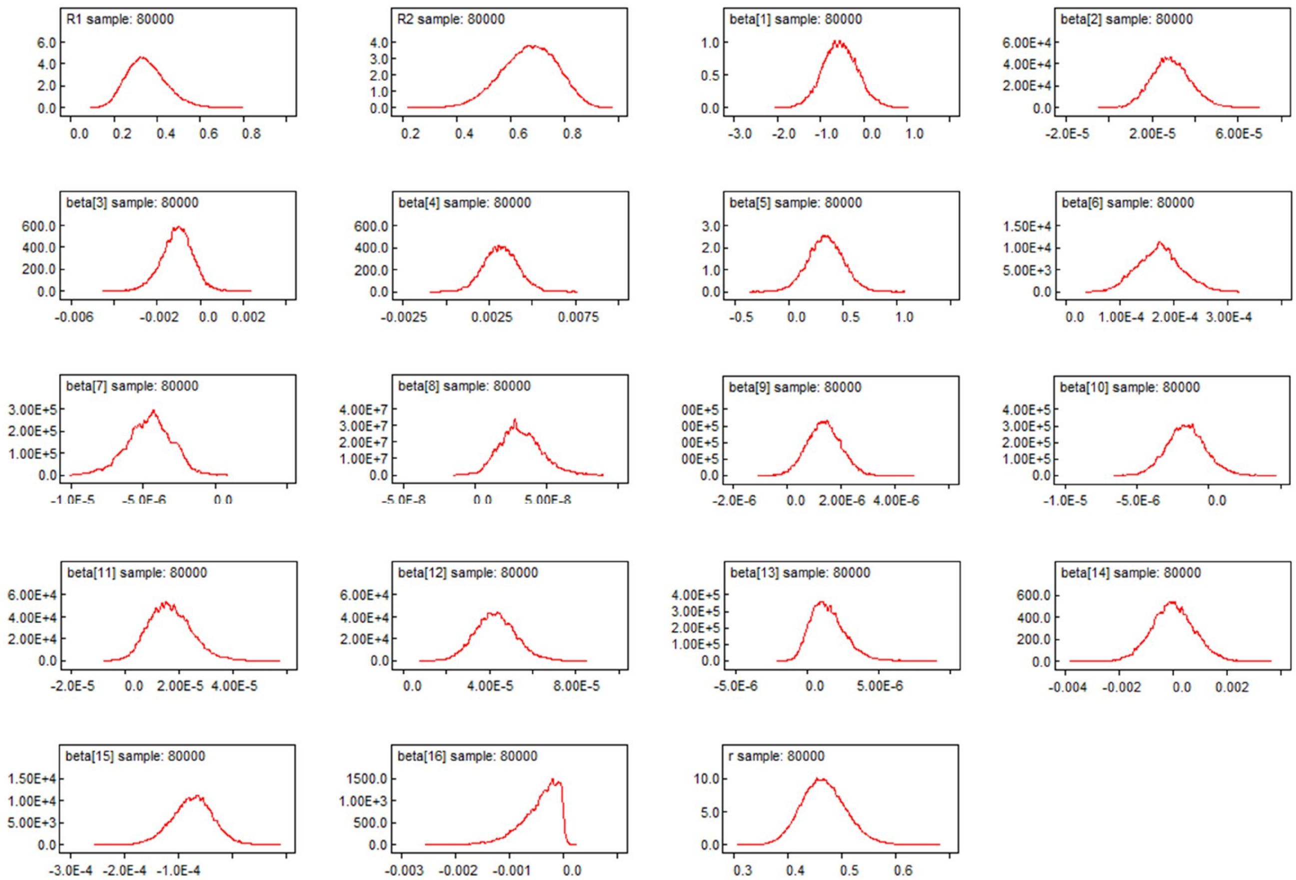
Thompson, J.; Palmer, T.; Moreno, S. (2006). Bayesian Analysis in Stata using WinBUGS. *The Stata Journal, vol. 6, no. 4, pp. 530–549.*

Appendix A: winBUGS code (Model 7)

```
model{
for (i in 1:664) {
for (j in 1:5)
{ v1[i,j] ~ dnegbin(p[i,j],r)
p[i,j]<-r / (r+lambda[i,j])
lambda[i,j]<-C[i,j]*S[i,j]
S[i,j]<-step(621.5-i)*R1+step(i-621.5)*step(1.5-j)*R2+step(i-621.5)*step(2.5-j)*step(j-
1.5)*R3+step(i-621.5)*step(3.5-j)*step(j-2.5)*R4+step(i-621.5)*step(4.5-j)*step(j-3.5)*R5+step(i-
621.5)*step(j-4.5)*R6
log(C[i,j]) <-
beta[1]+beta[2]*v2[i,j]+beta[3]*v3[i,j]+beta[4]*v4[i,j]+beta[5]*v5[i,j]+beta[6]*v6[i,j]+beta[7]*v7[i
,j]+beta[8]*v8[i,j]+beta[9]*v9[i,j]+beta[10]*v10[i,j]+beta[11]*v11[i,j]+beta[12]*v12[i,j]+beta[13]*v
13[i,j]+beta[14]*v14[i,j]+beta[15]*v15[i,j]+beta[16]*v16[i,j]+beta[17]*v17[i,j]+re[i]}
re[i]~dnorm(0,tau)}
for (k in 1:17)
{ beta[k]~dnorm( 0.0, 0.001 ) }
R1~dbeta(2,8)
R2~dbeta(8.4,11.6)
R3~dbeta(8,12)
R4~dbeta(10.6,9.4)
R5~dbeta(14.4,5.6)
R6~dbeta(8.2,11.8)
tau~dgamma(0.001,0.001)
r~dgamma(1,0.1)
}
```

Appendix B: posterior distributions

Figure 3. Posterior distributions of recovery rates, coefficients and other parameters (Model 2)



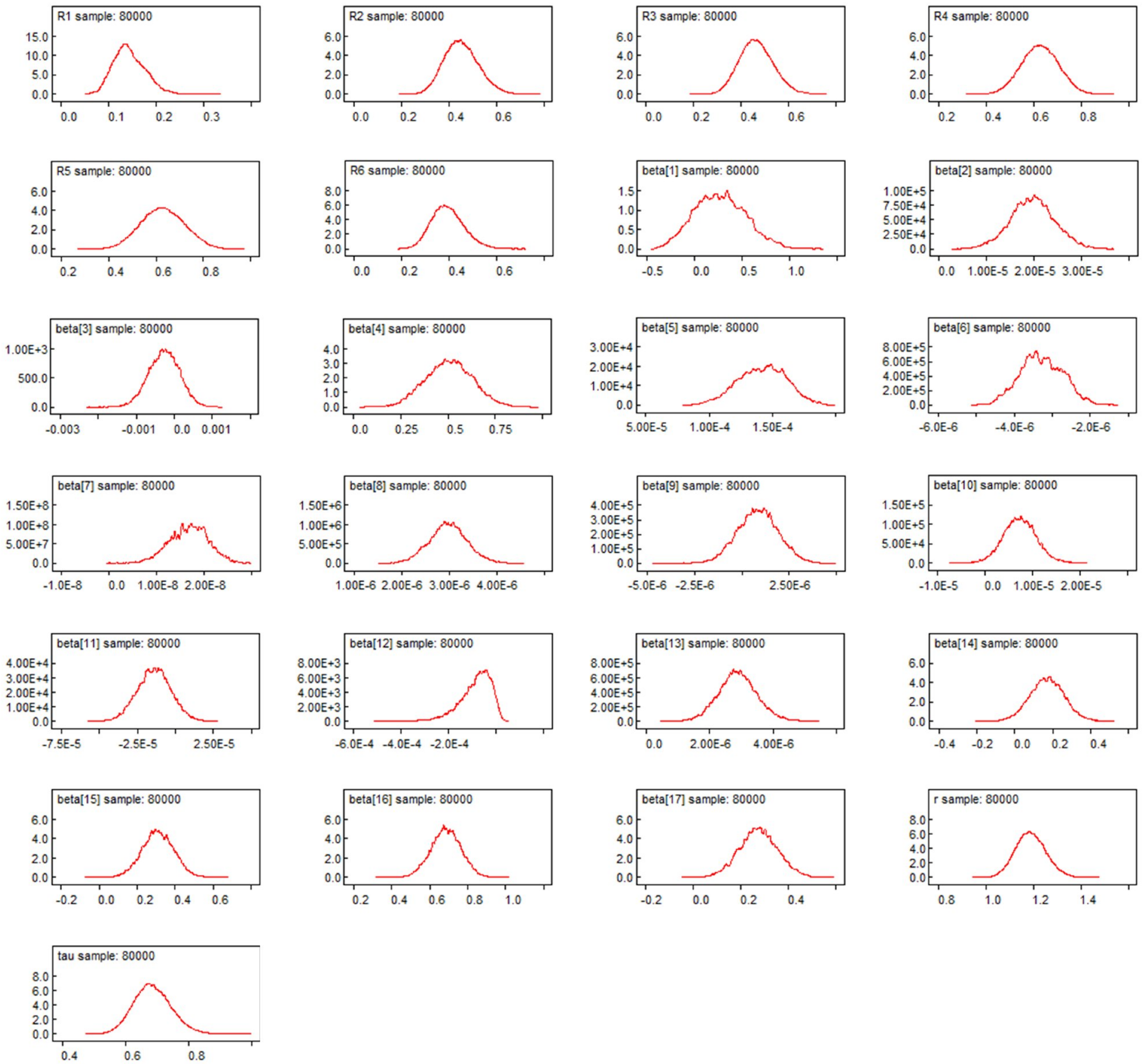
Recovery rates:
- R1 – Inner Amazon
- R2 – Arc of Fire

Coefficients:
- beta[1] – intercept
- beta[2] – forest stock
- beta[3] – openness to trade
- beta[4] – length of roads
- beta[5] – acessibility by boat
- beta[6] – GDP per capita
- beta[7] – GDP per capita squared

Coefficients:
- beta[8] – GDP per capita cubed
- beta[9] – cattle animals
- beta[10] – agricultural area
- beta[11] – value of timber products
- beta[12] – rural population
- beta[13] – urban population
- beta[14] – altitude
- beta[15] – rural credit for agriculture
- beta[16] – rural credit for cattle ranching

r – over dispersion parameter

Figure 4. Posterior distributions of recovery rates, coefficients and other parameters (Model 7)



Recovery rates:
- R1 – Inner Amazon
- R2-R6 – Arc of Fire 2006-2010

Coefficients:
- beta[1] – intercept
- beta[2] – forest stock
- beta[3] – openness to trade
- beta[4] – acessibility by boat
- beta[5] – GDP per capita
- beta[6] – GDP per capita squared
- beta[7] – GDP per capita cubed

Coefficients:
- beta[8] – cattle animals
- beta[9] – agricultural area
- beta[10] – value of timber products
- beta[11] – rural credit for agriculture
- beta[12] – rural credit for cattle ranching
- beta[13] – population
- beta[14]-beta[17] – time dummies 2006-2009

r – over dispersion parameter
tau – precision of random effects distribution

Appendix C: technical notes

The data. Some discrepancies in data were observed. Population variable for year 2010 was calculated as a ratio between GDP and GDP per capita for that year. The resulting figures were compared with population figures reported by the Brazilian Institute of Geography and Statistics. A couple of notes should be made. Firstly, in few cases two municipalities in two different states of Legal Amazon have exactly the same name. The problem arises, because different lists from different sources sort those names in different order. Secondly, data on population variable for couple of municipalities is incorrect. However, this can be easily observed and corrected manually.

winBUGS code. The logical sentence below represents the way recovery rates are assigned depending on geographical location of the municipality and year:

```
S[i,j]<-step(621.5-i)*R1+step(i-621.5)*step(1.5-j)*R2+step(i-621.5)*step(2.5-j)*step(j-
1.5)*R3+step(i-621.5)*step(3.5-j)*step(j-2.5)*R4+step(i-621.5)*step(4.5-j)*step(j-3.5)*R5+step(i-
621.5)*step(j-4.5)*R6
```

The *step* function in winBUGS assigns value 1 when the expression within the brackets is positive and value 0 – when it is negative. For instance, expression step(621.5-i) is assigned value 1 for the starting 621 observations, and value 0 – for the rest of observations. To clarify, consider municipality Itupiranga (i=635) in year 2008 (j=3). The code selects the correct recovery rate as follows:

$$S = 0 \cdot R_1 + 1 \cdot 0 \cdot R_2 + 1 \cdot 1 \cdot 1 \cdot R_3 + 1 \cdot 1 \cdot 0 \cdot R_4 + 1 \cdot 1 \cdot 0 \cdot R_5 + 1 \cdot 0 \cdot R_6 = R_3$$