Factors Affecting Forest Area Changes in Cambodia: An Econometric Approach

Tetsuya Michinaka¹, Motoe Miyamoto¹, Yasuhiro Yokota¹, Heng Sokh², Sethaphal Lao² & Vuthy Ma²

¹ Forestry and Forest Products Research Institute, Japan
² Forestry Administration, Ministry of Agriculture, Forestry and Fisheries, Kingdom of Cambodia

Correspondence: Tetsuya Michinaka, REDD Research and Development Center, Forestry and Forest Products Research Institute, 305-8687, Tsukuba, Ibaraki, Japan. Tel: 81-29-829-8635. E-mail: zhangyt@affrc.go.jp

Received: December 25, 2012   Accepted: April 5, 2013   Online Published: April 11, 2013
doi:10.5539/jsd.v6n5p12          URL: http://dx.doi.org/10.5539/jsd.v6n5p12

Abstract
Clarifying factors affecting forest area changes is critical to implementing REDD+ scheme properly. We analyzed some socio-economic factors and clarified their relationships with deforestation in Cambodia for the period of 2002 to 2010. A panel data analysis was conducted for 18 provinces, while six other provinces were deleted from the list because only a small amount of their land was forested. Time effects, cross-sectional dependence, serial correlation in idiosyncratic errors, and heteroskedasticity were tested, and robust variance matrix estimations were obtained to solve the problems of heteroskedasticity and serial correlation. The model estimation results showed that population, gross agricultural production and large-scale plantation development have negative impacts on forest area changes. On the other hand, the impacts of rice cultivation, gross industrial production, household income and house floor area by household were found not to be significant. Overall, however, the results indicated that forests in Cambodia still face pressure from the increases in population, agriculture production, and the enlargement of land development. As the increase in productivity of agriculture gives a better use of current agricultural land and lessens the pressure on forest, intensifying agriculture is important. It is also important to develop industry and other economic ventures to grow national economy while not imposing pressure on forest. This research reminds decision makers to use discretion when developing large-scale plantations.

Keywords: deforestation, Cambodia, REDD+, econometric analysis, large-scale plantation development

1. Introduction
Clarifying the affecting factors for deforestation is important for lessening, halting, and reversing the deforestation that, combined with forest degradation, contributes approximately 17% of the global greenhouse gas emissions (Intergovernmental Panel on Climate Change [IPCC], 2007). Deforestation, the conversion of forest to other land uses (Food and Agriculture Organization of the United Nations [FAO], 2001; Schoene, Killmann, von Lupke, & LoyocheWilkie, 2007), has been the topic of many studies and documents due to its impacts on climate change, national economic development and rural livelihoods (Angelsen, 2008; Angelsen et al., 2009; Springate-Baginski & Wollenberg, 2010). A new scheme, REDD (reducing emissions from deforestation and forest degradation in developing countries), was originally developed to reduce emissions that resulted from deforestation and forest degradation in developing countries. Since 2007, REDD has expanded to REDD+, as it now includes conservation of carbon stocks, sustainable management of forests, and enhancement of forest carbon stocks in reducing the adverse impacts from forest-related activities (United Nations Framework Convention on Climate Change [UNFCCC], 2007). Developed countries are to provide new and additional resources, including forestry and investments through international institutions (UNFCCC, 2009). As one of the developing countries, Cambodia attracts attention regarding its readiness of REDD+ (Forestry and Forest Products Research Institute [FFPRI], 2011).

The forest area in Cambodia has decreased at an annual rate of approximately 0.5% of the total land area from 2002 to 2010 (Forestry Administration [FA], 2011), with a total forest land area of 11.10 million ha, equal to 61.15% of the land area, in 2002, 10.71 million ha (59.09%) in 2006 and 10.36 million ha (57.07%) in 2010. This makes Cambodia a country with high forest cover and also high deforestation. In addressing the deforestation and related issues in Cambodia, the government of Cambodia has made efforts to prepare for the
implementation of the REDD scheme, and it entered the stage of implementation of the UN-REDD Programme in August 2011 (UN-REDD Programme [UN-REDD], 2011). In the National Programme Document of the UN-REDD Programme for Cambodia, direct and indirect factors both within the forest sector and outside of the forest sector for deforestation and forest degradation are summarized (Cambodia, Food and Agriculture Organization of the United Nations [FAO], United Nations Development Programme [UNDP], & United Nations Environment Programme [UNEP], 2010). Within the forest sector, direct factors include unsustainable logging, fire (Note 1) and unsustainable wood fuel collection, while indirect factors include slow speed of permanent forest demarcation activity, an inadequate enforcement of forest laws, a demand for wood energy, etc. Outside of the forest sector, direct factors include the clearance of land for agriculture, the expansion of settlements, and the development of infrastructure, while the indirect factors include population increases, poverty, increasing incomes, demands for resources, low agricultural yields, large-scale agro-industrial developments, etc.

Apart from the above-mentioned document, some other studies describe Cambodia forest resources and forest management. Wolf (1996) noted that logging for the export market is the primary cause of deforestation in Cambodia. Additionally, from the perspective of trade, Lang and Chan (2005) analyzed, in their working report series, the impact of China’s demand for forest products on Southeast Asian countries including Cambodia after China’s logging ban in 1998. Kim-Phat, Ouk, Uozumi and Ueki (1999) may have been the first to examine the causes of deforestation in Cambodia from a wider view in which the impacts of war, population growth and illegal logging were analyzed. Kim, PHAT, Koike and Hayashi (2005) further analyzed the forestry development in six different political regimes in Cambodia, i.e., before 1970, 1970 to 1975, 1975 to 1979, 1979 to 1989, 1989 to 1993 and 1993 to present. They concluded that deforestation in Cambodia was caused by “intensive bombardments during war time, forest clearing for resettlements, wood extraction for re-construction after war, agricultural cultivation, and indiscriminate logging in recent years”. De Lopez (2001) analyzed the deforestation in Cambodia using a stakeholder management approach and concluded that several primary stakeholders have been ignored, including local communities, the Royal Cambodian Armed Forces, the Ministry of Environment and Cambodia's neighbors. He further argued that the deforestation would not stop without a comprehensive stakeholder management framework. Kao, Iida, Inoue and Ohga (2005), using a forest concession company as an example, assessed the management of this company based on criteria and indicators of the Montreal Process and concluded that the management situation is still challenging. Poffenberger (2009) discussed the implementation of a pilot REDD project and factors of deforestation and forest degradation in Cambodia and noted that a large-scale land development program might frustrate the Forestry Administration’s efforts to achieve the national target of retaining forest cover at 60% in year 2015. He also noted the importance of the engagement of forest-dependent people to control local factors in the REDD project. Recently, not only deforestation but also forest biomass and the carbon stock have become topics of research as the technologies of monitoring and remote sensing progress (Top, Mizoue, & Kai, 2004; Sasaki, 2006; Kiyono et al., 2011; Toriyama et al., 2011).

Although many studies have analyzed the socio-economic factors related to deforestation that are important for implementing REDD+, analyses for quantitatively verifying these qualitative analysis results for Cambodia are scarce (Note 2). Dasgupta, Meisner and Wheeler (2005) analyzed the nexus between poverty and environment to the cases of Cambodia, Lao PDR, and Vietnam at the provincial and district levels using a regression analysis of deforestation and other environmental problems with population and poverty population. Top et al. (2009) analyzed the biodiversity issue in Kampong Thom Province, Cambodia, and found significant negative correlations between population density surrounding clusters and tree density, basal area, stand volume, aboveground biomass, and species richness and diversity. Shigeyama, Mizoue, Kajisa and Yoshida (2011) analyzed the four factors affecting deforestation in Cambodia, which include the distance from roadways, altitude, the distance from a community forest and the distance from the country border, by using a forest cover maps with a logistic regression approach, and they found that various factors played different roles during different periods (1992 to 1996, 1996 to 2002 and 2002 to 2005). In this research, the deforestation was analyzed from the perspective of geographical and spatial relationships, but socio-economic factors were not considered. The reasons for fewer quantitative analyses for Cambodia may include the short history of peaceful political regime (the civil war ended in 1998) combined with the problem of data comparability and availability. The objective of this research was to clarify the affecting factors or factors of deforestation in Cambodia by adopting an econometric approach. We undertook a panel data analysis of the forest area changes in Cambodia on a provincial level for the period 2002 to 2010.

In Section 2, we will provide a qualitative review of the factors of deforestation. In Section 3, the data, methods and empirical model for the econometric analysis will be discussed. Section 4 will implement a panel data
analysis that includes related tests and robust estimation results. The discussion section will conclude the analysis, followed by acknowledgements, notes and references.

2. Cause Map for Deforestation and Forest Degradation

Based on the literature review and field investigation, a cause map for deforestation and forest degradation is depicted in this section. An introduction to forest resources in Cambodia, however, is introduced first.

2.1 Forest Resources in Cambodia

Cambodia, bordered by Thailand, Laos and Vietnam, has a total land area of 18 million ha. With a GDP per capita below US$1000, Cambodia is one of the developing countries. According to the 2008 General Population Census of Cambodia (GPCC), the total population of Cambodia in 2008 was 13.4 million, of which 19.5% lived in urban areas and 80.5% lived in rural areas (National Institute of Statistics [NIS], 2009). The total population had increased by 2 million between 1998 and 2008, which is an annual population growth rate of 1.54% (NIS, 2009). In the past four decades, Cambodia’s forest area has been rapidly declining. During the regime of the communist leader Pol Pot (1975-1979), over 2 million Cambodians were forced out of their cities and towns into remote areas, and an estimated 1.3 million ha of dense forests and at least 300,000 ha of edaphic forests were cleared for settlement and fuel wood (Kim et al., 2005). Between 1 and 2 million people were killed by execution, starvation, etc. during this regime (Kim et al., 2005). When people returned to their home villages after the end of the regime, they found their houses destroyed; thus, the homes needed to be rebuilt. During the first decade of the 2000s, the trend of deforestation did not change, indicating that wood was important both during the war and after the war and that the forest remains an important natural resource in Cambodia.

During the past several decades, organizations, such as the FAO, UNDP, the Mekong Secretariat, the Japan Forest Technology Association [JAFTA], the Danish International Development Assistance [DANIDA], etc., and Cambodian government agencies implemented forest inventory in Cambodia in different ways. Figure 1 shows some of the forest inventory results sourced from FA (2011). Before the 1970s, the forest area remained relatively constant. Since the 1970s, however, the forest area began to decline due to after effects of the Vietnam War as Cambodia suffered from unstable political situations and unsustainable logging in the area. Remote sensing was introduced to assess the forest covers of 1992/93 and 1996/97, and a few ground truth point was undertaken. Since 2002, however, adequate ground truthing was conducted (Pak, Leng, & Leang, 2010), and the quality of mapping forest has been improved. In 2002, the system for inventorying forests was simplified in comparison to previous inventory methods. Forests were now classified into evergreen forest, semi-evergreen forest, deciduous forest and other forests. The categories of forests have also changed in that wood and shrub land is no longer considered as forest. Therefore, we will undertake the analysis to the forest area changes from 2002 to 2010.

![Figure 1. Forest cover change in Cambodia from 1965 to 2010](image)

Source: FA, 2011.
2.2 Cause Map for Deforestation and Forest Degradation

Based on the literature review and field survey of various forest regions in Cambodia, we created the cause map shown in Figure 2. Our field survey includes Provinces of Odder Meanchevy, Pursat, Kampong Chhnang, and Mondulkiri, which are located in the northwest, Tole Sap Lake and northeast regions respectively. Forest area decreased in all these provinces except Kampong Chhnang, in which forest area increased from 2006 to 2010. These provinces are also among the 18 provinces for panel data analysis.

The items in the large circle on the right-hand side are direct factors, both within and outside of the forest sector. These direct factors are actually caused by the demand for agricultural products and poverty alleviation, migration, and the demand for timber from international and domestic markets. The underlying reasons for these increased demands may well be due to population growth and the need for economic development. In the chain between the need for economic development and population growth, low agricultural productivity and the slow development of industry may also be contributing factors behind the direct factors. Additionally, as village or rural development is also important for improving the livelihood of the rural population, such development of infrastructure, social welfare, house construction and other industry developments are also important to impose direct and indirect impacts on forests. In one hand, these aspects reflect the general level of the development of the society and the economy. On the other hand, human beings increase their resource demands during the course of development. The square frames above the yellow circles show the background or indirect factors, including political factors, law enforcement, low institutional capacity, social customs and norms (claiming land through utilization), and a low awareness of environmental protection. These background factors affect the behaviors of stakeholders, including ordinary consumers, villagers, migrants, traders, large-scale plantation developers, etc.

Figure 2. Cause map for deforestation and forest degradation in Cambodia

3. Data and Method

3.1 Variables and Data

Based on the above cause map, we gathered data for variables related to forest area and socio-economic development. In Cambodia, due to the limited budget, little reforestation is implemented. Accordingly, in the research, only variables on the demand side for forest land and products are considered. With the international expertise and financial support from the United Nations Development Programs (UNDP), the Asian
Development Bank (ADB), the United Nations Child’s Fund (UNICEF), the International Labor Organization (ILO), the Sweden International Development Cooperation Agency (SIDA), the United Nation Population Fund (UNFPA), and the Japan International Cooperation Agency (JICA), Cambodia has made great progress in the field of statistics. Even so, it is too difficult to gather data suitable for reliable quantitative analysis for some of the factors previously described. Therefore, in addition to total forest area, which is the dependent variable, seven independent variables are considered in our econometric analysis. These independent variables include agricultural GVA (gross value added, million Riels), industrial GVA (million Riels), household income (million Riels), house floor area by household (m²), total population (1,000 persons), rice cultivation area (ha) and large-scale plantation development (mainly economic land concession, or ELC). All of the variables expressed in Riels are deflated by the GDP deflator and were changed to a constant value in 2000 (World Bank, 2012).

As previously stated, considering the data comparability, 2002 to 2010 serves as the period of analysis. However, this period is too short for an annual time-series analysis to get statistically significant estimations. We gather panel data at the provincial level and adopt a panel data analysis for our study. The Forestry Administration (FA) issued provincial forest area data at four-year intervals for 2002, 2006 and 2010. Accordingly, there are three time points, 2002, 2006 and 2010, in the panel data. There are 24 provinces in Cambodia including the municipal city of Phnom Penh. Eighteen provinces remained on the list, while 6 other provinces, including Phnom Penh, were deleted because of their small amount of forest area (Note 3). The 18 provinces considered in our research are Banteay Meanchey, Battambang, Koh Kong, Kratie, Kampong Cham, Kampong Chhnang, Kampong Speu, Kampong Thom, Kampot, Mondul Kiri, Preah Vihear, Pursat, Ratanak Kiri, Siem Reap, Sihanouk Ville, Pailin, Stung Treng and Oudor Meanchey (Otdar Meanchey).

Viewing the left side of the cause map (Figure 2), we understand that the economic development and the development of agriculture and industry are important for the entire society including the forest sector. Therefore, we discuss this issue first. Usually, the gross domestic product (GDP) reflects the size of the economy in a country expressed in the market value of the products in agriculture, industry, service, and others. When dealing with regions, such as provinces in a country, the GDP is sometimes referred to as the gross regional product, GRP (Viet, 2010). In Cambodia, however, neither the provincial GDP nor the provincial GRP is calculated, but rather, the provincial GVA (gross value added) is calculated. Oeur and Yim (2009) explained that, in Cambodia, the production approach and expenditure approach are used to calculate the GDP. In the former approach, the GDP is the sum of the gross value added (GVA) of the agriculture sector, industry sector, services sector, and tax on production less the subsidies and the FISIM (Financial Intermediation Services Indirectly Measured). In this research, two variables, agricultural GVA (million Riels) and industrial GVA (million Riels), are considered in the model to examine the impacts of the development of agriculture and industry on forest resources. The data are sourced from the NIS, the National Institute of Statistics of Cambodia.

The second factor is disposable income or GDP per capita, from the perspective of the demand by consumer. In Cambodia, there is no statistical annual income data by household or per capita or GDP per capita at the provincial level. As a proxy, we use household data obtained from the Socio-Economic Survey. The Socio-Economic Survey had been conducted in 1993-94, 1996, 1997, 1999, 2004, and every year since 2007. In Cambodia, most households have income from self-employment, as many of the households are involved in small-scale business and agriculture. As a result, the quality of the income data makes it difficult to be precise. Therefore, we add household income as an independent variable in an effort to capture the general situation. Another variable, floor area of house by household, was also sourced from the Socio-Economic Survey. As shown in the cause map in section 2, timber is consumed in countryside houses; therefore, we consider that the demand for timber in the construction industry may be a factor in the deforestation of the area. We verify this hypothesis herein. Still, as it is difficult to find suitable data for the construction industry, we take the average square meters of floor area of the house per household as a proxy for the development of the construction industry. For model estimations, we need the annual provincial data from 2002, 2006 and 2010. Therefore, we calculate the household income and average squared meters of floor area per household using an interpolation approach based on the Socio-Economic Survey data from NIS and assuming that the population changes smoothly.

Population is a factor that is considered in most socio-economic analyses (Michinaka & Miyamoto, 2013). We use the total population, which represents the size or magnitude of the population. The rural population is not included in the model estimation because it has a high correlation with the total population for the 18 provinces used in our study. The correlation coefficient between these two variables is 0.980 (Pearson’s correlation). The General Population Census of Cambodia was administered in 1998, which is more than three decades after the last census in 1962. The 1998 census provides a detailed population dataset. Since 1998, two additional
large-scale population surveys were published, the Cambodia Inter-Censal Population Survey 2004 and the General Population Census of Cambodia 2008. Apart from these three data sets, other population data for Cambodia are all projections. Because we need yearly provincial population data for 2002, 2006 and 2010, we calculate the data using an interpolation approach.

Next, the area of cultivated rice (ha) is added to the analysis. Rice is the most important agricultural product in Cambodia. With the increase in population, the demand for agricultural products also increases. Therefore, the pressure due to the population increases may also cause an increase in the area used for rice cultivation. The rice cultivation area increased from 2.2 million ha in 2002 to 2.8 million ha in 2010 (MAFF, 2011). The area of land used for rice cultivation is added into the model estimation because we believe that it may be one of the factors in the deforestation from the view of the land use. The rice cultivated area data are sourced from the MAFF annual reports.

The last variable, large-scale plantation development is a dummy variable in the estimation. The objectives of the large-scale plantation development are mainly to develop intensive agricultural and industrial-agricultural activities, to develop the land in an appropriate and perpetual manner, to increase employment and to generate state, provincial, or communal revenues (MAFF, 2012). However, these objectives may not be easy to fulfill (Kao et al., 2005). Based on the MAFF website profile of ELC companies, we summarized some basic facts about the implementation of large-scale plantation development in each province. As a dummy variable, it takes one as its value when large-scale plantation development is implemented in the province during that year; otherwise, it takes zero as its value. In this way, the impact of large-scale plantation development can be analyzed.

3.2 Method

In this research, a panel data analysis is applied to estimate the impacts of socio-economic factors on forest area changes because, first of all, panel data analysis overcomes the difficulties of data availability faced by time series data analysis. The panel data, or longitudinal data, comprise a dataset based on observations of the same individuals over multiple periods; thus, it is wider than time-series data and longer than cross-sectional data. Here, a panel dataset is composed of 18 provinces at three time points, totally 54 observations. Panel data analyses are increasingly used by applied researchers for the analysis of multiple country factors, particularly with respect to deforestation (Barbier & Burgess, 2001; Nguyen Van & Azomahou, 2007; Michinaka & Miyamoto, 2013). Hsiao (2003) describes the advantages of panel data analysis, positing that (a) it can solve problems that cannot be solved by either time-series data or cross-sectional data alone; (b) it can increase the degrees of freedom and (c) it can improve the efficiency of estimations. Greene (2000) argues that panel data analysis allows the researcher far greater flexibility in modeling differences in behavior across individuals.

The basic linear regression model for panel data analysis usually takes the following forms (Croissant & Millo, 2008):

\[ y_{it} = \alpha_i + x_{it} \beta + u_{it} \]  
\[ y_{it} = \alpha + x_{it} \beta + \mu_i + \lambda_t + \epsilon_{it} \]  

where \( i = 1, \ldots, N \) is the individual (firm, province, country, etc.) index, and \( t = 1, \ldots, T \) is the time index. In Formula 1, \( u_{it} \) is a disturbance term. In Formula 2, the disturbance term is expressed by three components: \( \mu_i \) represents unobserved cross-sectional (individual) effects for \( N \) cross sections, \( \lambda_t \) represents unobserved time effects for \( T \) time periods, and \( \epsilon_{it} \) represents random or idiosyncratic disturbances; \( \alpha_i \) becomes \( \alpha \), forming part of the intercept, while the rest of the intercept is derived from error components depending on their algorithm. In other words, individual effects and time effects can be reflected in the final model. In this research, for example, individual effects may result from the land area of the province, the distance to the national border, or the distance to Phnom Penh, the capital of Cambodia. While these factors may impose different impacts on different provinces, they do not change over time. There are seven factors discussed in the research, while other factors, including the land area, the distances to national borders and the capital, are treated as unobserved factors, and their impacts, if any, are included in the error components. As these factors do not change over time for an individual (province), they are considered individual effects. On the other hand, time effects allow researchers to control for aggregate common shocks that occurred at some point in time, and though there may be different impacts during different years, the factors remain the same for all the individuals (provinces), as they do not change with the cross-sectional dimension. For example, all the provinces simultaneously face the same international environment, such as the global economic crisis, and the same domestic environment, such as changes in national policy. These factors differentiate one year from another.
A number of assumptions are usually made with respect to panel data models. By assuming parameter homogeneity, i.e., no individual or time effects, the intercept $\alpha$ and slope parameter $\beta$ remain constant for all $i$ and $t$, and the disturbance term $u_{it}$ is treated as random with mean 0. Thus, the above model would be a pooled or pooling OLS model. There are two basic frameworks used to generalize panel data model when the above assumption is rejected. The fixed effects approach (FE) assumes different intercepts, i.e., recognizing the presence of individual effects, $\alpha_i$, but the same slope, $\beta$, across cross-sectional unit $i$. In a one-way FE model, only one type of effect, either individual effects or time effects, is included, even though time effects are usually not discussed. As for the pooled OLS model, the FE model is also estimated by an ordinary least squares regression (OLS). Because the individual effects and/or time effects are captured in the model as dummy variables, with one dummy variable for each individual or period in the model, the FE model is also called the least squares with dummy variables (LSDV), or within-group estimator. Of course, in this way, the degree of freedom is greatly lessened compared with the following approach when $N$ is large.

As an alternative, the random effects approach (RE) assumes that the individual-specific and time-specific component error terms ($\mu_i$ and $\lambda_t$, respectively) and the overall error $u_{it}$ are random variables drawn from a normal distribution and are independently and identically distributed, with the assumption that these error components are uncorrelated with the explanatory variables. The FE model, in contrast, assumes the individual effects are correlated with the independent variables. The RE model is estimated using a generalized least square estimate (GLS). The Hausman test statistic and the $p$-values for the difference between FE and RE estimates are calculated in this research.

In the presence of heteroskedasticity and serial correlation, robust estimations of the error covariance matrix are necessary. In this research, not only are the robust estimations applying to the FE and RE models introduced, but the robust estimation for the general feasible-generalized least square estimate (FGLS) model is also introduced. The general FGLS model is estimated using a two-step estimation process. First, an OLS or FE model is estimated, and then its residuals are used to estimate an error covariance matrix using an FGLS analysis (Wooldridge, 2002; Croissant & Millo, 2008).

### 4. Tests and Model Results

A panel data analysis allows for the estimation of three types of models, the pooled OLS model, the FE model and the RE model. To determine the preferred model, the F-test and the Hausman test, which test parameter heterogeneity and the correlation relationship between individual effects and the independent variables, should be administered. In this research, a series of tests and robust error covariance estimations are conducted. R Packages were used in the analysis (R Core Team, 2012; plm (Croissant & Millo, 2008), sandwich (Zeileis, 2004), and lmtest (Zeileis & Hothorn, 2002)).

#### 4.1 Test Results

In this research, before obtaining final panel data model estimations, we conduct a series of tests that assess not only parameter heterogeneity and the correlation as stated above but also the time (fixed) effects, cross-sectional dependence or contemporaneous correlation, serial correlation and heteroskedasticity.

The test results are shown in Table 1. First, for testing parameter heterogeneity, an F-test for individual effects based on variance analysis is conducted with an alternative hypothesis of significant effects. The large F-test statistics suggest that the individual effects are significant; thus, the FE model is preferred over the pooled OLS model.

The second test is the Hausman test, which compares the FE model with the RE model (Hausman, 1978). As stated herein, the individual effects (including time effects) can be treated as a fixed but unknown constant differing among individuals, or they can be treated as random effects assuming that they are drawn from an iid distribution. The Hausman test examines the difference between the two estimators in the above treatments. If the RE model specification is correct, the two estimates should be near each other rather than far apart. The Hausman test set the null hypothesis as the RE model, which is correct. The alternative hypothesis is the presence of the correlation between the individual effects and the independent variables, i.e., not a random relationship. The result shows that the $p$-value is very small, thus indicating that the FE model is preferred over the RE model.

The third diagnostic tool is the Breusch-Pagan Lagrange multiplier test (Breusch & Pagan, 1980), which is used to compare the RE model with the pooled OLS model. The Breusch-Pagan statistic is a Chi-squared ($\chi^2$) Lagrange multiplier statistic. The result shows that significant panel effects are detected, implying that the RE
model is better than the pooled OLS model, i.e., the GLS estimation should be used in place of the OLS. For this research, the FE model is the best among the three models when considering the individual effects.

Table 1. Test results

<table>
<thead>
<tr>
<th>Test items</th>
<th>Methods</th>
<th>Results</th>
<th>Remarks</th>
</tr>
</thead>
<tbody>
<tr>
<td>(1) Comparing models: FE vs pooled</td>
<td>F test for individual effects</td>
<td>F = 715.917, df₁ = 17, df₂ = 29, p-value = 0.000</td>
<td>FE model is better than pooled model</td>
</tr>
<tr>
<td>(2) Comparing models: FE vs RE</td>
<td>Hausman test</td>
<td>χ² = 31.882, df = 7, p-value = 0.000</td>
<td>FE model is better than RE model</td>
</tr>
<tr>
<td>(3) Comparing models: pooled vs RE</td>
<td>Lagrange Multiplier Test (Breusch-Pagan test)</td>
<td>χ² = 26.064, df = 1, p-value = 0.000</td>
<td>RE model is better than pooled model</td>
</tr>
<tr>
<td>(4) Testing for time effects</td>
<td>a) F test</td>
<td>F = 1.668, df₁ = 2, df₂ = 27, p-value = 0.208</td>
<td>No significant effects</td>
</tr>
<tr>
<td></td>
<td>b) Lagrange Multiplier Test (Breusch-Pagan)</td>
<td>χ² = 0.687, df = 1, p-value = 0.407</td>
<td></td>
</tr>
<tr>
<td>(5) Testing for cross-sectional dependence</td>
<td>Pasaran CD test</td>
<td>FE: z = -0.808, p-value = 0.419</td>
<td>No cross-sectional dependence</td>
</tr>
<tr>
<td></td>
<td></td>
<td>RE: z = -0.794, p-value = 0.427</td>
<td></td>
</tr>
<tr>
<td>(6) Testing for serial correlation</td>
<td>Breusch-Godfrey/Wooldridge test</td>
<td>FE: χ² = 19.099, df = 3, p-value = 0.000</td>
<td>FE model at 1% significant level; not in RE at 5%</td>
</tr>
<tr>
<td></td>
<td></td>
<td>RE: χ² = 7.507, df = 3, p-value = 0.057</td>
<td></td>
</tr>
<tr>
<td>(7) Testing for heteroskedasticity</td>
<td>Studentized Breusch-Pagan test</td>
<td>BP = 39.172, df = 24, p-value = 0.026</td>
<td>Heteroskedastic</td>
</tr>
</tbody>
</table>

Note: RE: fixed effects model; RE: random effects model.

The fourth test is for time effects. For the FE (fixed effects) model, time (fixed) effects are usually ignored because the individual (cross-sectional) effects explain a large part of the variances. If two effects are considered together, there will be more dummy variables, thus further lessening the degree of freedom. In this research, time effects are calculated by the F-test and the Breusch-Pagan Lagrange multiplier test (Breusch & Pagan, 1980). The results of both tests indicate that the effects are not significant. This may be due to the data structure, which includes 18 provinces and 3 time points, which is short and wide. Therefore, it is not necessary to add time effects into the model estimation in the following analysis.

Next, we test for cross-sectional dependence or contemporaneous correlation. Cross-sectional dependence can arise due to spatial or spill-over effects, or it may be due to unobserved or unobservable factors (Baltagi & Pesaran, 2007). As a result, cross-sectional dependence can lead to bias in tests results. The null hypothesis in the Pesaran CD test of independence concludes that residuals across individuals are not correlated (Pesaran, 2004). The test is conducted for both the FE model and the RE model, and the results indicate that there is no cross-sectional dependence. Therefore, it is not necessary to perform corrections for cross-sectional dependence. Serial correlation in errors or disturbances across periods is also tested for both the FE model and the RE model. No serial correlation is an important assumption for classical multiple linear regression estimations. Serial non-correlation does not imply that observations \( y_i \) and \( y_j \) are uncorrelated but implies that deviations of observations from their expected values are uncorrelated (Greene, 2000). Greene (2000) describes the assumptions concerning the variance and covariance of the disturbance in the following formulas:
\[ \text{Var}[\varepsilon_i | X] = \sigma^2, \text{ for all } i = 1, \ldots, n, \text{ and Cov}[\varepsilon_i, \varepsilon_j | X] = 0, \text{ for all } i \neq j \] (3)

In the former equation, the constant variance is defined as homoskedasticity; in the latter equation, the non-correlatedness across observations is identified as non-autocorrelation or serial non-correlation. A first-order serial correlation process can be written as \( \varepsilon_t = \rho \varepsilon_{t-1} + v_t \), where rho (\( \rho \)) is the correlation coefficient between the two errors. Serial correlation may result from misspecification, such as due to the omitted variable. Therefore, efforts are made to add possible variables to avoid misspecification in this research. Apart from misspecification, serial correlation may result from the variables that are serial-correlated across observations. If errors or residuals suffer from serial correlation or autocorrelation, the estimations are not necessarily biased, but the statistical inference results are incorrect. Serial correlation is detected in this research using the Breusch-Godfrey/Wooldridge test (Breusch, 1978; Godfrey, 1978; Wooldridge, 2002). For the fixed effects model, the null hypothesis of no serial correlation is rejected at the 1% level, while the null hypothesis for the random effects model is rejected at the 10% level but not rejected at the 5% level. Therefore, computing the robust variance matrix for the estimators becomes necessary.

The last test is for heteroskedasticity. When the assumption of homoskedasticity is not met, errors are heteroskedastic. Breusch and Pagan (1979) noted that in the presence of heteroskedastic disturbances, the loss in efficiency may be substantial, the estimated standard errors are biased, and inference may be invalid. Therefore, computing the robust variance matrix for the estimators becomes necessary, as in the situation of serial correlation. In this research, heteroskedasticity is tested using the Studentized Breusch-Pagan test (Breusch & Pagan, 1979). Breusch and Pagan (1979) devised the hypothesis on variance that \( \sigma^2_t = h(z_t' \alpha) \), where \( z_t \) is a vector of independent variables and \( \alpha \) is a vector of unrestricted parameters functionally uncorrelated to the \( \beta \) coefficients; if \( \alpha = 0 \), the model is homoskedastic. A Lagrangian multiplier statistic for \( \alpha = 0 \) was described. The above test results are displayed in Table 1.

4.2 Model Results

As stated in the previous section, in the presence of a serial correlation of errors and heteroskedasticity in the model, robust covariance matrix estimation should be conducted to obtain a valid statistical inference. Wooldridge (2002) argues that nothing will be lost in using robust standard errors and test statistics even if the classical assumption holds. Here, this method is used to devise estimators for the appropriate asymptotic covariance matrix and to obtain consistent estimations. Croissant and Millo (2008) note that in the context of panel data, the most general version also proves consistent against serial correlation and not only under heteroskedasticity. For the FE model in this research, the “arellano” method is used, and consistent estimations are obtained versus both heteroskedasticity and serial correlation (Croissant & Millo, 2008; Arellano, 1987). Apart from the FE and RE models, a general feasible generalized least square (general FGLS) approach is also introduced to address the situation in which the idiosyncratic errors are generally heteroskedastic and serially correlated across t. Because serial correlation is also detected in the FE model, the general FGLS is undertaken by adopting the FE estimations in the first step. In the second step, the residuals of the FE estimations are used to estimate an error covariance matrix by a general FGLS analysis (Wooldridge, 2002). The estimation results are shown in Table 2.

<table>
<thead>
<tr>
<th>Variables</th>
<th>Fixed Effects</th>
<th>Random Effects</th>
<th>General FGLS (fixed)</th>
</tr>
</thead>
<tbody>
<tr>
<td>Population (1,000 persons)</td>
<td>-316.01**</td>
<td>-317.92**</td>
<td>-274.19**</td>
</tr>
<tr>
<td>Household Income (million Riel)</td>
<td>-1.99</td>
<td>-2.26</td>
<td>-2.54**</td>
</tr>
<tr>
<td>Rice cultivation area</td>
<td>-0.28</td>
<td>-0.31</td>
<td>-0.28</td>
</tr>
<tr>
<td>Agricultural GVA (million Riel)</td>
<td>-0.10***</td>
<td>-0.09***</td>
<td>-0.08**</td>
</tr>
<tr>
<td>Industrial GVA (million Riel)</td>
<td>0.05</td>
<td>0.04</td>
<td>0.01</td>
</tr>
<tr>
<td>Floor area per household (m²)</td>
<td>790.93</td>
<td>754.33</td>
<td>597.04</td>
</tr>
<tr>
<td>Large-scale plantation development (dummy variable)</td>
<td>-22,426.10*</td>
<td>-21,037.00***</td>
<td>-21,305.00**</td>
</tr>
<tr>
<td>AIC</td>
<td>19.318</td>
<td>19.799</td>
<td>19.352</td>
</tr>
<tr>
<td>SC or BIC</td>
<td>19.576</td>
<td>20.057</td>
<td>19.610</td>
</tr>
</tbody>
</table>

Note: (1) ***: significant at 1% level; **: significant at 5% level; *: significant at 10% level.

(2) All the variables in Riel are using constant values based on the GDP deflator from the World Bank (2012).
These models show similar results. First, population, agricultural gross value added (GVA) and the dummy variable large-scale plantation development all have significant coefficients. The evidence shows that the increase in population and agricultural production yields negative impacts on the changes in the forest area in Cambodia, and the link between large-scale plantation development and decreases in forest area has been clarified. The rice cultivation area, the industrial GVA and the house floor area per household (a proxy for construction industry) do not have significant effects on deforestation. According to the parameter values, an increase of every 1,000 persons may cause a forest area decrease of approximately 300 ha. Increases in agricultural production by a million Riel result in a 0.10 ha decrease in forest area. The results of the dummy variable indicate that the forest area decreased by 21,000 ha to 22,400 ha between 2002 and 2010 in a province that implemented large-scale plantation development compared with a province that did not implement the development. As for the income, general FGLS estimations yielded a different result than the other two estimations in which the coefficient is significant. This result implies that the increase in income may put pressure on the changes in forest area or impose negative impacts.

Akaike information criterion (AIC) and Schwarz criterion (SC or BIC) are used to measure the goodness of fit for models. For AIC or SC values, the smallest one gives the best goodness of fit. AIC and SC values can be calculated by (Greene, 2000):

\[ AIC = \log(\text{RSS}/n) + 2K/n \quad (4) \]
\[ SC = \log(\text{RSS}/n) + K\log(n) \quad (5) \]

Where RSS is the sum of squared residuals, n is the number of observations, and K is the number of parameters.

5. Discussion and Conclusions

Population is an important factor with respect to its effect on the changes in forest area in Cambodia. As shown in the cause map in Section 2.2, as the total population increases, the demands for land, agricultural products, timber products and new job opportunities also increase. Low agricultural productivity leads more people to focus on agriculture, and the increase in agricultural production relies on the extension of agricultural land areas. This observation explains the significant impacts of agriculture production on forest area changes. However, the evidence does not show much impact of rice cultivation area on forest area changes as expected. First, it means the conversion of forestland is small for rice cultivation and large for the cultivation of other crops, such as rubber, cassava, sugar cane, etc. Second, in the past decade, rice production by volume in Cambodia has increased faster than the rice cultivation area, indicating that rice productivity has been increasing. Average yield per hectare of paddy rice increased from 1.9 ton in 2002 to 3.0 ton in 2010 (NIS, 2012a). When rice productivity increases, the pressure from rice production on the forest is reduced.

Since 1993, when the general election was held under the guidance of the United Nations Transitional Authority in Cambodia (UN-TAC), Cambodia embraced the rapid development of society and the economy. Entering the 2000s, Cambodia became one of the countries with the fastest economic growth in the world. From 2002 to 2007, the average annual growth rate was over 10% (at constant value), although it was less than 7% in 2008 due to the global financial crisis (NIS, 2012b). The increase in household income results in an increase in various demands. Because annual provincial per capita income data are not available, income per household is used. But the data for household income are also difficult to accurately assess considering the large number of people living in rural areas who receive income from many aspects. Additionally, the size of the household has also changed from 1998 to 2008. The census data show that the average household size has decreased from 5.2 to 4.7. Specifically, for urban areas, the average household size has decreased from 5.5 to 4.9, while for rural areas, this value has decreased from 5.1 to 4.6 (NIS, 2009). Therefore, the household income data might not reflect the real level of per capita GDP or disposable income due to the issue of data accuracy and changes in the size of household. The increase in per capita income might not be well reflected by household income. AIC and SC results show that the FE model is the best in three models and in the FE model, the household income is not significant, which is not a result as expected.

By verifying the impacts of population and agricultural production, it seems difficult to stop deforestation immediately, as each of the aforementioned factors continues to increase. If the productivity of agricultural production can be further increased, the pressure on the forest from these factors can be alleviated. Furthermore, the contents of agriculture production must be examined to determine if there are ways to intensify agriculture other than by relying on the conversion of land use, which results in deforestation.
Industrial gross value added (GVA) and house floor area per household do not significantly affect changes in forest area. This result suggests that it is important to develop the manufacturing and processing industries and other industries while intensifying agriculture. In fact, the share of agricultural value added has decreased from 31.47% in 2002 to 26.45% in 2008, while the share of industrial value added has increased from 24.68% in 2002 to 27.45% in 2008, as calculated by the 2000 constant price (NIS, 2012b).

The United Nations Development Programme (UNDP) has issued the annual human development index (HDI) since 1990. According to the report, the HDI for Cambodia in 2011 was 0.523, an increase from 0.494 in 2010. However, the HDI is still very low in comparison to those of other countries (UNDP, 2012). From a global perspective, the impact of the socio-economic factors of a country may begin to change from negative to positive as the country’s HDI increases (Michinaka & Miyamoto, 2013). Therefore, for Cambodia, these negative signs may be expected to shift and become positive as the society and economy continue to advance to a higher level.

Overall, the results indicated that forests in Cambodia still face pressure from the increases in population and agriculture production in which the enlargement of land development is still important. Intensifying agriculture helps to increase the productivity of agriculture and gives a better use of current agricultural land, therefore, lessens the pressure on forest. It is also important to develop industry and other economic ventures to grow national economy while not imposing pressure on forest. The last factor in the analysis, large-scale plantation development, may be good for a growing economy with respect to income growth and employment, but it must be properly implemented.

This research analyzed the affecting factors for deforestation in Cambodia by adopting panel data analysis. Due to data availability, there are some factors discussed in literature review and cause map, such as stakeholders, exports, road construction, poverty, migration, and employment in different industries, whose relationships with deforestation have not been clarified quantitatively; when data availability get improved, it is worth implementing further research.

Acknowledgements
This collaborative research conducted by the Forestry and Forest Product Research Institute (FFPRI), Japan, and Forestry Administration, Cambodia, was supported by the emergency project to develop the structure for promoting the REDD+ action funded by the Forestry Agency, Japan. The authors would like to thank Mr. Chivin Leng, Forestry Administration, Cambodia, for his valuable comments and suggestions to improve the quality of the paper before submission. The authors thank anonymous reviewers for their constructive comments. The authors also thank Mr. La Pengly and other staff members from Conservation International Cambodia for their contributions to our field trips.

References


Cambodia, Food and Agriculture Organization of the United Nations [FAO], United Nations Development


**Notes**

Note 1. While the Cambodia UN-REDD Program roadmap has described fire as a direct factor of deforestation and forest degradation in Cambodia, the program has explained that the role fire plays is disputed and unclear. Considering the use of fire and natural forest fires throughout the long history, fire does not seem to be the factor for deforestation and forest degradation in recent decades.

Note 2. Both quantitative and qualitative studies have strengths and weaknesses.

Note 3. Those provinces with less than 20,000 ha of forest area are considered as outliers and are deleted from the list. This criterion is somewhat arbitrary, but it should not be a problem when considering the fact that the total forest area in these 18 provinces shares over 99% of the total forest area in Cambodia.