JARQ 59 (2), 175-186 (2025) https://www.jircas.go.jp

Socioeconomic Factors Influencing Forest Area Changes in China: Regional Panel Autoregressive Distributed Lag Modeling Approach

Jiaze TAN[1]* and Satoshi TACHIBANA[2]

[1] Fukushima Prefectural Centre for Environmental Creation, Miharu, Fukushima, Japan
[2] Laboratory of Forest Resources and Society, Division of Forest and Biomaterials Science, Graduate School of Agriculture, Kyoto University, Kyoto, Japan

Abstract

While the global forest area is decreasing overall, it is experiencing growth in China. This study uses econometric methods to analyze the causal factors behind this increase, focusing particularly on socioeconomic factors. Previous research has given scant attention to stationarity, and few studies have considered regional characteristics such as China's forest resource distribution and the economic disparity between Inland and Coastal areas. We used panel data collected between 1993 to 2018 from four regions in China. Panel unit root and panel cointegration tests were employed, and panel ARDL models were specified. The results reveal that the long-term relationship between forest area and gross regional product per capita (GRPPC) varies by region. A linear relationship with a positive slope was realized between forest area and GRPPC in the Northeast and Central regions. However, an inverted U-shaped relationship is found between forest area and GRPPC in the East and West regions. Throughout the analysis, the Northeast and Central regions appeared to be in the recovery phase of forest area under a positive relationship with economic growth. In contrast, the East and West regions reached a stable phase with sufficient recovered forests and a constant level of economic development.

Discipline: Forestry
Additional key words: economic analysis, forest transition hypothesis, panel cointegration test, panel unit root test, regional characteristics

Introduction

Forests and their products have attracted attention as measures for mitigating global warming, which is a major international issue. According to the Food and Agriculture Organization of the United Nations (FAO) Global Forest Resource Assessment 2020, while global forest area is decreasing, China ranked first in the world in terms of net increase in forest area from 2010 to 2020 (FAO 2020). Various hypotheses have been proposed regarding the mechanism of forest area increase, including the forest transition (FT) hypothesis by Mather (1992b) and the U-shaped hypothesis of forest resources (U-shaped hypothesis) by Nagata et al. (1994).

This study focused on the impact of socioeconomic factors on forest area transformation in four Chinese regions from 1993 to 2018. When using time-series data for quantitative analysis, it is essential to use tests corresponding to "spurious regressions" and analytical methods corresponding to the data with cointegration relationships. However, these methods have rarely been used in this field. One solution is to use the panel autoregressive distributed lag (ARDL) model described (Nishiyama et al. 2019). Moreover, previous studies of the U-shaped hypothesis have generally used national-level analyses. However, in countries such as China, large disparities exist between Coastal and Inland areas, and the stages of development vary among regions. Therefore, the relationship between forest area and socioeconomic factors is not uniform among countries. However, regional-level time-series data often have insufficient sample sizes than national-level time-series data.

*Corresponding author: tan_jiaze_01@pref.fukushima.lg.jp
Received 17 May 2024; accepted 2 September 2024; J-STAGE Advanced Epub 30 January 2025.
https://doi.org/10.6090/jarq.24J05

Therefore, the insufficiency of sample size can be avoided using panel data that combines time-series and cross-sectional data.

This study aimed to clarify whether the relationship between forest area and socioeconomic factors varies regionally in the same country in the short and long term. Considering the possibility of a stable phase in some regions, we conducted an empirical analysis using a panel ARDL model with China as a case study. The novelty of this method is confirming the stationarity of the variables in regional panel data for China and further elaborating the analysis from short- and long-term perspectives.

Literature review

The FT hypothesis (Mather 1992b) states that if time is plotted on the horizontal axis to represent the historical development of land use and forest area is plotted on the vertical axis to represent land use change, forest area shifts from decreasing to increasing over time. Sakamoto et al. (2014) reported that the FT factors vary according to age and region. Another hypothesis related to the FT is the U-shaped hypothesis proposed by Nagata et al. (1994). This hypothesis states that when gross national product per capita is considered a measure of the stage of economic development on the horizontal axis, and forest area, forest area ratio and growing stock of forest are considered measures of the amount of forest resources on the vertical axis, the amount of forest resources can be drawn as U-shaped relationship from the perspective of economic development over time. In the U-shaped hypothesis, two patterns are considered the final stable phase: the first is a reversal after the decline phase, moving from the recovery phase to the stable phase. The second is an "L"-shaped pattern, moving directly from the decline phase to the stable phase. In European countries, Mather (1992b) distinguished three stages of forest use according to the combination of forest ownership system and purpose of use. Inoue (1992) distinguished four phases in the historical development of forest-use patterns. In Japan, the long-term use of forests and forest areas is divided into four phases, with the main determinants being population (POP) growth and the developmental stages of productivity (Yorimitsu 1984). Several empirical analyses have been conducted on the U-shaped hypothesis. Zhang et al. (2006) analyzed the impact of socioeconomic factors, such as GDP per capita, on forest areas in China from 1990 to 2001. After applying panel data analysis at the national and regional levels and regression analysis at the provincial level, they found that China was in the later stages of forest changes. Wang et al. (2007) criticized this study for using dummy variables that divided China into North and South and argued that variables such as temperature should be included. Then, panel data analysis was conducted on the relationship between forests and socioeconomic factors in Cambodia from 2002 to 2010 (Michinaka et al. 2013). Michinaka & Miyamoto (2013) used cluster analysis to group countries from 1995 to 2005, focusing on the relationship between forest area and the human development index for each group. Tan et al. (2020) highlighted the possibility of a second reversal in each Chinese province by conducting an empirical analysis using the model that introduced square and cubic terms. They found the possibility of a second reversal in several provinces.

However, these quantitative analyses did not include these three perspectives—first, stationarity and cointegration relationship tests. If regression analysis is performed as a nonstationary process, there is the risk of "spurious regression," whereby a false relationship is found among variables that are not related. Engle & Granger (1987) highlighted the concepts of unit roots and cointegration in these problems. Since then, researchers have used the ARDL model, which corresponds to the stationarity argument (Pesaran & Shin 1999). Recently, research has been conducted on the panel ARDL model, an analytical method that extends the ARDL model to panel data. Caravaggio (2020b) divided 114 countries into three groups based on income and analyzed the relationship between deforestation rates and economic growth. The study found that the recovery of forest areas is slowing in high-income countries. Second, the impact of socioeconomic factors on forest areas varies by region. China exhibits significant regional disparities, and its developmental stages may differ among regions. However, few studies have analyzed such relationships at the regional level. Third, the pattern was verified when approaching the stable phase. An important question posed by the U-shaped hypothesis was whether a high level of stability with abundant forests (U-shaped) or a stable state with depleted forests (L-shaped) could be achieved. However, few studies have examined the transition to a stable phase. Tan et al. (2021) tested such unit roots and cointegration in an empirical study of the U-shaped hypothesis. Their study used unit root and bound tests for cointegration in China from 1973 to 2018, and it analyzed the relationship between forest area and socioeconomic factors from both short and long-term perspectives. They found that, although China was already in the latter part of the U-shape, further economic growth could slow the rate of forest area increase.

Methods and data

1. Methods

We used panel data to group several provinces into four regions in China. Panel data are rich in information because they combine time-series and cross-sectional data (Kitamura 2005). However, the dynamic effect in the model, the error term, and the lag variable of the dependent variable must be correlated (Greene 2019). However, Pesaran & Smith (1995) obtained a consistent estimator by averaging the OLS estimator regressed on each individual (Chigira et al. 2011, Pesaran & Smith 1995). They proposed a procedure called the PMG estimator, which constrains the long-run coefficients to be identical but allows the short-run coefficients to differ among groups (Pesaran & Smith 1995). We used the PMG estimator in our panel ARDL model. Our analysis method first performed panel unit root and panel cointegration tests, followed by estimating a panel ARDL model. When a cointegration relationship was detected among nonstationary variables in the panel ARDL model, the short-run impact of changes in the independent variable could be analyzed by estimating the error correction model (ECM). The ECM represents a short-run relationship equation in which errors from long-run equilibrium are corrected. The Akaike Information Criterion (AIC) was used to determine the length of the lag terms for each independent variable, aimed at achieving an optimal model.

First, a panel unit root test was performed to examine whether the time-series data was stationary. This test involved analyzing several individuals and examining whether they exhibited a unit root (Chigira et al. 2011). One representative test utilized was the Levin, Lin, and Chu (LLC) test proposed by Levin et al. (2002). The basic idea was to extend the augmented Dickey-Fuller (ADF) test, a representative unit root test for time-series and panel data, and perform a t-test on the pooled ordinary least squares regression (OLS) estimator. However, the LLC test assumes that dynamic parameters are identical among individuals. The IPS test further assumes that dynamic parameters differ among individuals (Chigira et al. 2011). The Fisher ADF test further allows for heterogeneity in the constant and trend terms and for the size of the time series data to differ among individuals (Maddala & Wu 1999, Choi 2001). If such a Fisher-type panel unit root test uses the PP test instead of the ADF test, it is called the Fisher PP test (Dritsaki & Dritsaki 2013). Overall, a comprehensive decision, considering the results of several unit root tests, was necessary due to discrepancies in the outcomes of these tests (Murao 2019). The maximum number of lags was set at four, and lag selection was used SIC (Schwarz Information Criterion). Constant terms were included, but trend terms were not included in panel unit root tests. A panel cointegration test was performed after the panel unit root test. The Pedroni test was utilized for robustness, allowing for heterogeneity among individuals (Chigira et al. 2011). The maximum number of lags was set at four, and lag selection was used SIC. Constant terms were included, but trend terms were not included in this test. If no unit roots were found in the residuals, this confirmed the existence of a long-run equilibrium relationship. However, deviation from the long-run equilibrium occurred in the short-run. An error correction term was then developed as an ECM (E-Views 12 2021). This model facilitated the calculation of OLS estimators for each subject in both the short and long runs to estimate the average magnitude of the coefficients (Chigira et al. 2011). The panel ARDL model allowed the development of the ECM equation (5), expressing the short-run relationship by rewriting the basic equation (1). Equation (1) represents the panel ARDL model with different parameters for each individual, *i*.

$$Y_{i,t} = a_{i,0} + \sum_{n=1}^{p} \varphi_{i,n} Y_{i,t-n} + \sum_{j=1}^{k} \sum_{l_j=0}^{q_j} \omega_{i,j,l_j} X_{i,j,t-l_j} + \varepsilon_{i,t} \quad (1)$$

where $a_{i,0}$ is the constant term; $\varphi_{i,n}$ is the parameter of the lag term of the dependent variable; $\omega_{i,j}$ is the coefficient of the lag term of the *k* independent variables $X_{i,j}(j = 1 \ldots k)$; and $\varepsilon_{i,t}$ is the disturbance term. *p* is the lag length of the lag term of the dependent variable $Y_{i,t}$, q is the lag length of the independent variable $X_{i,j}$. If a long-term equilibrium relationship existed, its equation in the ARDL model follows from equation (1):

$$Y_{i,t} = b_{i,1} X_{i,1,t} \ldots + b_{i,k} X_{i,k,t} \quad (2)$$

The long-term equilibrium relationship parameters $b_{i,j}$ can be derived from equation (1).

$$b_{i,j} = \sum_{l_j=0}^{q_j} \omega_{i,j,l_j} / (1 - \sum_{n=1}^{p} \varphi_{i,n}) \quad (3)$$

The error in the long-run equilibrium can be expressed as follows:

$$EC_{i,t} = Y_{i,t} - (b_{i,1} X_{i,1,t} + \cdots + b_{i,k} X_{i,k,t}) \quad (4)$$

Furthermore, the ECM expressed equation (1) as follows:

$$\Delta Y_{i,t} = a_{i,0} + \sum_{n=1}^{p-1} \varphi_{i,n} \Delta Y_{i,t-n} + \sum_{j=1}^{k} \sum_{l_j=0}^{q_j-1} \beta_{i,j} \Delta X_{i,j,t-l_j} + \theta_i EC_{i,t-1} + \delta_{i,t} \quad (5)$$

Errors in the long-run equilibrium relationship were adjusted using the error correction term. $EC_{i,t-1}$ is the time-series calculated in Equation (5), and error-correction term θ_i is the adjustment coefficient, representing the speed at which errors from the long-run equilibrium are adjusted. $\beta_{i,j}$ is a parameter of independent variables, and $\delta_{i,t}$ is the error term. The maximum number of lags for the panel ARDL model analysis was set at four. The panel ARDL model equation was selected using the AIC. We used R-4.1.2 to check serial correlation, but E-views 12 was used for all other analyses.

2. Data

Four regional panel datasets were created from provincial data (Fig. 1). The divisions of Northeast, East, Central, and West were based on the 11th Five-Year National Economic and Social Development Plan of the People's Republic of China (adopted in March 2006). Chapter 19 of this plan describes the implementation of the provincial development strategy (China State Council Gazette 2006). Our study focused on assessing the impact of socioeconomic factors on forest areas. Therefore, classifying regions according to their economic growth was considered reasonable for analyzing the relationship between socioeconomic factors and forest areas.

Based on previous studies (Nagata et al. 1994, Sakamoto et al. 2014), forest area was designated as the dependent variable, while the variables shown in Table 1 were used as independent variables. First, at a low economic level, an increase in gross regional product per capita (GRPPC) was found to have a negative impact on forest areas because of the increase in logging associated with an increase in demand for forest products. However, once the GRPPC reached a certain level, afforestation for wood production increased, driven by heightened environmental awareness and higher forest product prices stemming from increased demand. This, in turn, was considered to have a positive impact on forest area. By including a squared term in the GRPPC, the U-shaped or inverse U-shaped relationship between the GRPPC and forest area could be inferred. Additionally, POP is thought to act in decreasing and increasing forest areas. Forests are converted into agricultural land to meet the increased demand for food due to population growth, leading to a decrease in forest area. During industrialization, as agricultural productivity increased and people became more affluent, the need for forest transformation and conservation became more important. In China, forest areas may increase consistent with population growth under the state's mass-mobilization afforestation policy, such as compulsory tree-planting during the "Tree-planting festival." The percentage of participants eligible for compulsory tree planting increased from approximately 34% between 1982 and 1988 to 85% in 1993 (Wunshiou 1996). Regarding trade openness (TO), recent studies on the U-shaped hypothesis indicated that the impact of trade has become important (Sakamoto et al. 2014). As China joined the World Trade Organization in 2001 and has been actively promoting trade liberalization, this factor was included as a variable to account for its effects. Previous studies of the environmental Kuznets curve (EKC) hypothesis have indicated that increased trade has a negative impact on the environment owing to increased economic activity. However, technological innovation in developed countries positively impacts environmentally friendly technologies through trade (Churchill et al. 2018). The increase of the primary industry ratio (IS1) in the social structure is expected to have a negative impact on the forest area because of an increase in deforestation to secure land for agriculture (Nagata et al. 1994). The increase of the secondary industry ratio (IS2) in the social structure leads to industrial development and an increase in deforestation to secure fuelwood, building materials, roads, and industrial use (Mather 1992a).

Fig. 1. Map of four regions of China

*Northeast: 3 provinces; East: 10 provinces; Central: 6 provinces; West: 11 provinces

For this study, panel data for 26 years (1993-2018) are used. This period ensured the continuity of annual data. We used forest area data from the China Forest Resources Report from 2014 to 2018 (China National Forestry and Grassland Administration 2019). In the 5th Forest Resource Survey, the definition of forest area was changed, and the data were corrected in the same rules as in Tan et al. (2020, 2021). For the independent variables, data from the annual editions of the Statistical Yearbook of China (China National Bureau of Statistics 1999-2022) were used for GRPPC, POP, and TO. For GRPPC, provincial GRPs were divided by the provincial population. As these figures (in RMB) are nominal for each year, they were calculated using the World Bank's GDP deflator (2015 = 100) (World Bank 2021). Then, we used total population data. Additionally, TO was calculated using the value of exports and imports by province, deflating the sum of these values in the same way as GRPPC, and dividing it by the provincial GRP. Furthermore, IS1 and IS2 were calculated using the deflated provincial gross primary and secondary industry output values divided by the provincial GRP. The descriptive statistics are showed in Table 2.

Results

Results of the panel unit root tests are shown in Tables 3 to 6. Panel cointegration tests were used to determine whether the variables had a cointegration relationship. We assumed 13 models as listed in Table 7. The significant level of the Pedroni test was used at the 10% level and the results are shown in Table 8. Models that met most of the 11 statistics presented in the Pedroni test were determined to be cointegrated. In the Northeast, the unit root test showed that *ln*POP and *ln*IS2 were

Table 1. Variables and sign conditions

Variable	Sign	References
Forest area (FA)		China Forest Resources Report 2014-2018
Per capita Gross Regional Product (GRPPC)	+ / −	China Statistical Yearbook (Annual)
Population (POP)	+ / −	China Statistical Yearbook (Annual)
Trade Openness (TO)	+ / −	China Statistical Yearbook (Annual)
Primary Industry ratio (IS1)	−	China Statistical Yearbook (Annual)
Secondary Industry ratio (IS2)	−	China Statistical Yearbook (Annual)

Table 2. Descriptive statistics

	Mean	Median	Maximum	Minimum	SD
NorthEast : *ln*FA	11.411	11.225	12.210	10.716	0.554
NorthEast : *ln*PGDP	10.069	10.074	11.089	8.877	0.620
NorthEast : *ln*POP	8.165	8.246	8.387	7.846	0.197
NorthEast : *ln*TO	–4.610	–4.940	–1.754	–6.736	1.418
NorthEast : *ln*IS1	2.609	2.565	3.336	1.988	0.337
NorthEast : *ln*IS2	3.857	3.896	4.086	3.203	0.167
East : *ln*FA	9.472	9.861	11.466	5.242	1.722
East : *ln*PGDP	10.456	10.496	11.765	8.825	0.729
East : *ln*POP	8.140	8.319	9.337	6.553	0.869
East : *ln*TO	–2.305	–2.244	1.908	–6.659	2.079
East : *ln*IS1	1.867	2.086	3.635	–1.204	1.124
East : *ln*IS2	3.766	3.894	4.096	1.792	0.314
Central : *ln*FA	10.801	10.701	11.595	9.769	0.572
Central : *ln*PGDP	9.710	9.709	11.016	8.439	0.715
Central : *ln*POP	8.635	8.687	9.182	8.010	0.325
Central : *ln*TO	–5.688	–5.847	–3.235	–7.669	1.221
Central : *ln*IS1	2.695	2.700	3.500	1.482	0.468
Central : *ln*IS2	3.849	3.853	4.119	3.555	0.123
West : *ln*FA	11.127	11.461	12.474	6.582	1.229
West : *ln*PGDP	9.597	9.543	11.171	8.043	0.779
West : *ln*POP	7.663	7.905	9.361	5.447	1.061
West : *ln*TO	–6.393	–6.630	–2.139	–9.293	1.407
West : *ln*IS1	2.819	2.836	3.829	1.988	0.404
West : *ln*IS2	3.737	3.753	4.067	2.851	0.196

non-stationary with first differences. Therefore, we conducted cointegration tests using a combination of variables that did not include these two variables. The results showed that most of the statistics in Model 02, 03, and 08 were cointegrated, indicating the long-run relationship. In addition, model selection using AIC led to the selection of Model 08. The results of the same cointegration tests for the models selected in the East, Central, and West directions are listed in Table 8. These results demonstrate the long-run relationship among the four regions of China. Short-run error corrections for long-run equilibrium relationships were developed using a panel ARDL model based on the results described above. Furthermore, regression coefficients for the short and long-run models were estimated. Independent variables were selected using the AIC, with AIC = 4 used as the maximum value.

Table 9 shows the results of the long-run model estimation for the Northeast region. The coefficient of GRPPC was positive at the 0.1% significance level. This indicates that GRPPC has positive effect on forest area in the long run. TO and IS1 were negative at the 0.1% significance level. Table 10 shows the estimation results for the short run. The ECT was not significant. Table 11 shows the results of the long run estimations for the East. The coefficient of GRPPC was positive and its squared term was negative, both were considerable at the 0.1% significance level. This indicates that GRPPC has an inverse U-shaped effect on forest area in the long run. Table 12 shows the estimation results for the short run. The ECT was significant at the 1% level. This indicates the existence of the short-run relationship, such that any error in the long-run equilibrium from the coefficient values would be corrected by approximately 2.87 years. The sign of the lagged term for forest area was positive and significant at the 5% level. The lagged TO term was positive and significant at the 5% level. Table 13 shows the results over a long run for the Central. The coefficient of GRPPC was positive and significant at the 0.1% level. This indicates that GRPPC has a positive long-run effect on forests. Table 14 shows the results for the short run. The ECT was not significant. Table 15 shows the results of the long-run estimations for the West. The coefficient of GRPPC was positive, and its squared term was negative; both were significant at the 0.1% level. This indicates that GRPPC has an inverse U-shaped effect on the forest area in the long run. The sign of the coefficient for the population was positive and significant at the 0.1% level. This indicates that the increasing of population has a positive long-run effect on forest areas. Table 16 shows the estimated results for the short run. The ECT was significant at the 1% level. This indicates the existence of the short-run relationship, such that any error in the long-run equilibrium from the coefficient values would be corrected by approximately 2.08 years.

We estimated the turning points for the East and West using the estimated coefficients. These two regions showed an inverted U-shaped relationship between forest area and GRPPC. The turning points in the East and West were approximately 100,000 and 60,000 Yuan, respectively. In the East, four provinces reached the turning point during the advanced period of analysis

Table 3. Panel unit root test: Northeast

Variable	LLC	IPS	ADF	PP
*ln*FA	3.435	3.954	0.240	0.499
*ln*GRPPC	–2.594**	–0.397	6.027	4.775
*ln*GRPPC^2	–3.589***	–0.882	8.414	3.546
*ln*POP	–1.005	–0.269	5.262	27.247**
*ln*TO	–0.541	0.489	2.822	3.101
*ln*IS1	0.716	1.383	2.977	1.164
*ln*IS2	3.839	2.582	2.645	1.643
Δ*ln*FA	–57.582***	–38.262***	53.991***	53.418***
Δ*ln*GRPPC	–2.231**	–2.081**	16.6**	16.51**
Δ*ln*GRPPC^2	–2.143*	–2.071*	16.6*	16.341*
Δ*ln*POP	–1.71*	–0.080	4.818	4.076
Δ*ln*TO	–6.729***	–5.58***	37.325***	37.279***
Δ*ln*IS1	–4.939***	–5.494***	37.8765***	55.989***
Δ*ln*IS2	–0.290	–0.547	8.745	36.453***

***$P < 0.01$, **$P < 0.05$, *$P < 0.1$

Table 4. Panel unit root test: East

Variable	LLC	IPS	ADF	PP
*ln*FA	0.999	1.000	0.999	0.998
*ln*GRPPC	–2.594**	–0.397	6.027	4.775
*ln*GRPPC^2	–3.589***	–0.882	8.414	3.546
*ln*POP	–1.010	–0.269	5.262	27.246**
*ln*TO	–0.541	0.489	2.822	3.101
*ln*IS1	0.716	1.383	2.977	1.164
*ln*IS2	3.839	2.582	2.645	1.643
Δ*ln*FA	–57.582***	–38.262***	53.991***	53.418***
Δ*ln*GRPPC	–2.231*	–2.081*	16.599*	16.507*
Δ*ln*GRPPC^2	–2.142*	–2.071*	16.598*	16.341*
Δ*ln*POP	–1.712*	–0.081	4.818	4.076
Δ*ln*TO	–6.729***	–5.58***	37.325***	37.279***
Δ*ln*IS1	–4.939***	–5.495***	37.877***	55.989***
Δ*ln*IS2	–0.290	–0.547	8.745	36.453***

***$P < 0.01$, **$P < 0.05$, *$P < 0.1$

Table 5. Panel unit root test: Central

Variable	LLC	IPS	ADF	PP
*ln*FA	–10.358***	–9.502***	312.453***	34.281*
*ln*GRPPC	–4.394***	–0.592	26.326	31.335
*ln*GRPPC^2	–3.494***	0.346	20.203	19.499
*ln*POP	–3.379***	0.697	18.439	24.862
*ln*TO	–3.602***	–0.286	17.141	16.512
*ln*IS1	–0.074	4.276	4.105	3.816
*ln*IS2	1.365	2.537	21.057	19.037
Δ*ln*FA	–9.159***	–12.623***	151.456**	85.018***
Δ*ln*GRPPC	–4.65***	–5.263***	75.581***	73.589***
Δ*ln*GRPPC^2	–4.646***	–5.285***	75.384***	73.584***
Δ*ln*POP	–10.959***	9.989***	123.611***	126.955**
Δ*ln*TO	–10.191***	–9.07***	110.005***	109.7***
Δ*ln*IS1	–13.016***	–12.329***	154.541***	170.351***
Δ*ln*IS2	–9.579***	–9.027***	111.618***	108.849***

***$P < 0.01$, **$P < 0.05$, *$P < 0.1$

Table 6. Panel unit root test: West

Variable	LLC	IPS	ADF	PP
*ln*FA	–11.952***	–18.882***	336.33***	66.37***
*ln*GRPPC	–3.34***	1.170	18.545	3.107
*ln*GRPPC^2	–2.748**	2.028	16.508	1.967
*ln*POP	–3.52***	0.823	35.063*	95.571***
*ln*TO	0.771	1.769	19.865	20.996
*ln*IS1	–2.417**	0.579	41.696**	2.985
*ln*IS2	–0.264	0.799	13.766	10.843
Δ*ln*FA	–10.055***	–9.153***	116.423***	104.521***
Δ*ln*GRPPC	–5.745***	–6.345***	90.254***	102.376***
Δ*ln*GRPPC^2	–5.782***	–6.307***	89.687***	101.171***
Δ*ln*POP	–4.204***	–5.781***	79.387***	82.162***
Δ*ln*TO	–12.558***	–13.0211***	169.851***	192.84***
Δ*ln*IS1	–12.874***	–12.771***	166.751***	187.463***
Δ*ln*IS2	–12.068***	–12.52***	163.669***	166.474***

***$P < 0.01$, **$P < 0.05$, *$P < 0.1$

(Beijing 2015, Tianjin 2015, Shanghai 2016, Jiangsu 2018). In the West, Inner Mongolia reached its turning point in 2011. In the Northeast and Central, the relationship between the forest area and GRPPC showed a monotonic increase. These results indicated that the turning point varies among the regions. The analysis period of this study was in the stable phase of the U-shaped hypothesis. We tested serial correlation in the error terms of the model residuals for each province. The ACF function was used to check whether there was a serial correlation, based on whether the value of the serial correlation coefficient exceeded the 95% confidence interval. Although the results for the Northeast are noteworthy, they indicated that the four regions covered by this analysis were generally not serially correlated. Model fitting is shown in Figure 2. This indicated that the fitted values fit the actual values for most of the provinces. In addition, the residual values were scattered at zero, and there was no explicit trend. Therefore, the model was

Table 7. List of models assumed for different combinations of variables

	*ln*GRPPC	*ln*POP	*ln*TO	*ln*IS1	*ln*IS2	*ln*GRPPC^2
Model 01	○	○				
Model 02	○		○			
Model 03	○			○		
Model 04	○				○	
Model 05	○	○	○			
Model 06	○	○		○		
Model 07	○	○			○	
Model 08	○		○	○		
Model 09	○		○		○	
Model 10	○	○				○
Model 11	○		○			○
Model 12	○			○		○
Model 13	○				○	○

Table 8. Pedroni test

	NE: model 08	East: model 11	Central: model 11	West: model 10
panel v	2.172*	2.012*	1.424	1.229
panel rho	–1.806*	0.382	–1.239	0.472
panel pp	–2.886**	–0.670.	–3.538***	–0.516
panel adf	–2.676**	–2.939**	–4.139***	–4.795***
w-panel v	2.21*	2.204*	1.33	2.225*
w-panel rho	–1.950*	–0.709	–1.279	0.286
w-panel pp	–3.098**	–2.428**	–3.398***	–2.161*
w-panel adf	–2.531**	–4.316***	–3.553***	–3.563***
group rho	–1.377	0.401	–0.549	0.394
group pp	–3.392***	–2.399**	–3.996***	–2.200*
group adf	0.165	–4.980***	–3.822***	–5.931***

***$P < 0.01$, **$P < 0.05$, *$P < 0.1$

Table 9. Panel ARDL model—long run: Northeast

Dependent variable: *ln*FA	Coefficient	SE	t value	p value
*ln*GRPPC	0.056***	0.001	39.610	0.000
*ln*TO	–0.007***	0.001	–10.737	0.000
*ln*IS1	–0.015***	0.002	–8.983	0.000

***$P < 0.01$, **$P < 0.05$, *$P < 0.1$

Table 10. Panel ARDL model—short run: Northeast

Dependent variable: *ln*FA	Coefficient	SE	t value	p value
ECT	0.401	0.530	0.756	0.457
Δ*ln*FA (–1)	0.429	0.247	1.739	0.095
Δ*ln*FA (–2)	0.414	0.341	1.217	0.235
Δ*ln*FA (–3)	0.079	0.043	1.796	0.085
Δ*ln*GRPPC	0.027	0.042	0.646	0.524
Δ*ln*GRPPC (–1)	–0.005	0.024	–0.194	0.848
Δ*ln*GRPPC (–2)	–0.029	0.026	–1.129	0.270
Δ*ln*GRPPC (–3)	0.011	0.034	0.330	0.744
Δ*ln*TO	–0.002	0.002	–0.952	0.351
Δ*ln*TO (–1)	0.002*	0.001	2.136	0.043
Δ*ln*TO (–2)	0.007	0.001	1.908	0.068
Δ*ln*TO (–3)	0.003***	0.004	21.990	0.000
Δ*ln*IS1	–0.002	0.000	–0.171	0.866
Δ*ln*IS1 (–1)	0.000	0.001	–0.038	0.970
Δ*ln*IS1 (–2)	0.028*	0.013	2.070	0.049
Δ*ln*IS1 (–3)	0.013	0.007	1.776	0.088
C	–4.216	5.690	–0.741	0.466

***$P < 0.01$, **$P < 0.05$, *$P < 0.1$

Table 11. Panel ARDL model—long run: East

Dependent variable: *ln*FA	Coefficient	SE	t value	p value
*ln*GRPPC	2.097***	0.305	6.884	0.000
*ln*GRPPC^2	–0.091***	0.015	–6.019	0.000
*ln*TO	0.009	0.009	1.086	0.280

***$P < 0.01$, **$P < 0.05$, *$P < 0.1$

Table 12. Panel ARDL model—short run: East

Dependent variable: *ln*FA	Coefficient	SE	t value	p value
ECT	–0.349***	0.110	–3.183	0.002
Δ*ln*FA (–1)	0.319*	0.151	2.106	0.038
Δ*ln*FA (–2)	0.131	0.103	1.262	0.210
Δ*ln*GRPPC	–8.343	7.361	–1.133	0.260
Δ*ln*GRPPC (–1)	0.396	1.773	0.223	0.824
Δ*ln*GRPPC (–2)	0.657	3.486	0.188	0.851
Δ*ln*GRPPC (–3)	1.569	1.131	1.387	0.167
Δ*ln*GRPPC^2	0.362	0.325	1.114	0.268
Δ*ln*GRPPC^2 (–1)	–0.032	0.081	–0.401	0.689
Δ*ln*GRPPC^2 (–2)	–0.027	0.157	–0.170	0.866
Δ*ln*GRPPC^2 (–3)	–0.076	0.577	–1.311	0.193
Δ*ln*TO	0.012	0.010	1.195	0.235
Δ*ln*TO (–1)	–0.016*	0.008	–2.050	0.043
Δ*ln*TO (–2)	–0.004	0.016	–0.270	0.790
Δ*ln*TO (–3)	–0.005	0.008	–0.615	0.540
C	–0.755**	0.212	–3.569	0.001

***$P < 0.01$, **$P < 0.05$, *$P < 0.1$

Table 13. Panel ARDL model—long run: Central

Dependent variable: *ln*FA	Coefficient	SE	t value	p value
*ln*GRPPC	0.238***	0.038	6.282	0.000
*ln*TO	0.004	0.026	0.138	0.890

***$P < 0.01$, **$P < 0.05$, *$P < 0.1$

Table 14. Panel ARDL model—short run: Central

Dependent variable: *ln*FA	Coefficient	SE	t value	p value
ECT	–0.087	0.054	–1.610	0.111
Δ*ln*FA (–1)	0.496***	0.122	4.067	0.000
Δ*ln*GRPPC	0.032	0.022	1.438	0.154
Δ*ln*GRPPC (–1)	–0.018	0.025	–0.744	0.459
Δ*ln*GRPPC (–2)	–0.039	0.022	–1.794	0.076
Δ*ln*TO	0.004	0.003	1.295	0.199
Δ*ln*TO (–1)	0.005	0.004	1.044	0.299
Δ*ln*TO (–2)	0.006*	0.003	2.017	0.046
C	0.732	0.461	1.586	0.116

***$P < 0.01$, **$P < 0.05$, *$P < 0.1$

Table 15. Panel ARDL model—long run: West

Dependent variable: *ln*FA	Coefficient	SE	t value	p value
*ln*GRPPC	1.517***	0.027	56.289	0.000
*ln*GRPPC^2	–0.069***	0.001	–47.344	0.000
*ln*POP	0.717***	0.038	19.016	0.000

***$P < 0.01$, **$P < 0.05$, *$P < 0.1$

Table 16. Panel ARDL model—short run: West

Dependent variable: *ln*FA	Coefficient	SE	t value	p value
ECT	–0.481***	0.164	–2.931	0.004
Δ*ln*FA (–1)	0.206	0.125	1.654	0.101
Δ*ln*FA (–2)	–0.062	0.102	–0.612	0.542
Δ*ln*GRPPC	–2.130	2.032	–1.049	0.297
Δ*ln*GRPPC (–1)	8.089	8.043	1.006	0.317
Δ*ln*GRPPC (–2)	6.319	7.108	0.889	0.376
Δ*ln*GRPPC (–3)	–8.521	10.620	–0.804	0.424
Δ*ln*GRPPC^2	0.089	0.092	0.970	0.334
Δ*ln*GRPPC^2 (–1)	–0.409	0.401	–1.021	0.310
Δ*ln*GRPPC^2 (–2)	–0.311	0.339	–0.917	0.361
Δ*ln*GRPPC^2 (–3)	0.444	0.551	0.805	0.423
Δ*ln*POP	1.645	2.403	0.684	0.495
Δ*ln*POP (–1)	–4.488	2.269	–1,978	0.051
Δ*ln*POP (–2)	3.482	4.732	0.736	0.463
Δ*ln*POP (–3)	–5.528	7.079	0.781	0.437
C	–0.957**	0.313	–3.058	0.003

***$P < 0.01$, **$P < 0.05$, *$P < 0.1$

suitable for the data. Notably, owing to the presence of lag terms, there were no fitted values for the first four years.

Discussion and conclusions

We applied a panel ARDL model and conducted regional-level time-series analysis while addressing the issues of insufficient sample size and stationarity. We estimated the short- and long-run models separately and found that the short-run model was almost insignificant. Tables 17 and 18 present summaries of the results of the analysis for the short and long runs.

The Northeast and Central demonstrated short-run relationships, as previously described. Fast-growing trees were actively established in the East (Forestry and Forest Products Research Institute 2010), whereas Great Western Development may have encouraged rapid afforestation in the West, including the return of the Grain for Green Project, causing short-run relationships between the East and West regions (Seki et al. 2009). The long-run relationship between forest areas and GRPPC varied among regions. This relationship was positive in the Northeast and Central, but was an inverted U-shape in the East and West. In this study, the Northeast and Central China were in the reforestation stage of the U-shaped hypothesis. However, East and West China may be in the next stage of the U-shaped stagnation phase. In terms of variables other than GRPPC, TO and IS1 were negative and significant in the Northeast. Caravaggio (2020b) pointed out that trade openness has positive but small impact on deforestation, and our study also found similar results. Shi et al. (2017) analyzed data from 1977-2013 in Heilongjiang province, a province in Northeast China, and found that agriculture had a strong and negative impact on the forests. This long-term negative relationship exists with forest areas, consistent with sign conditions. In the East and Central regions, factors other than GRPPC

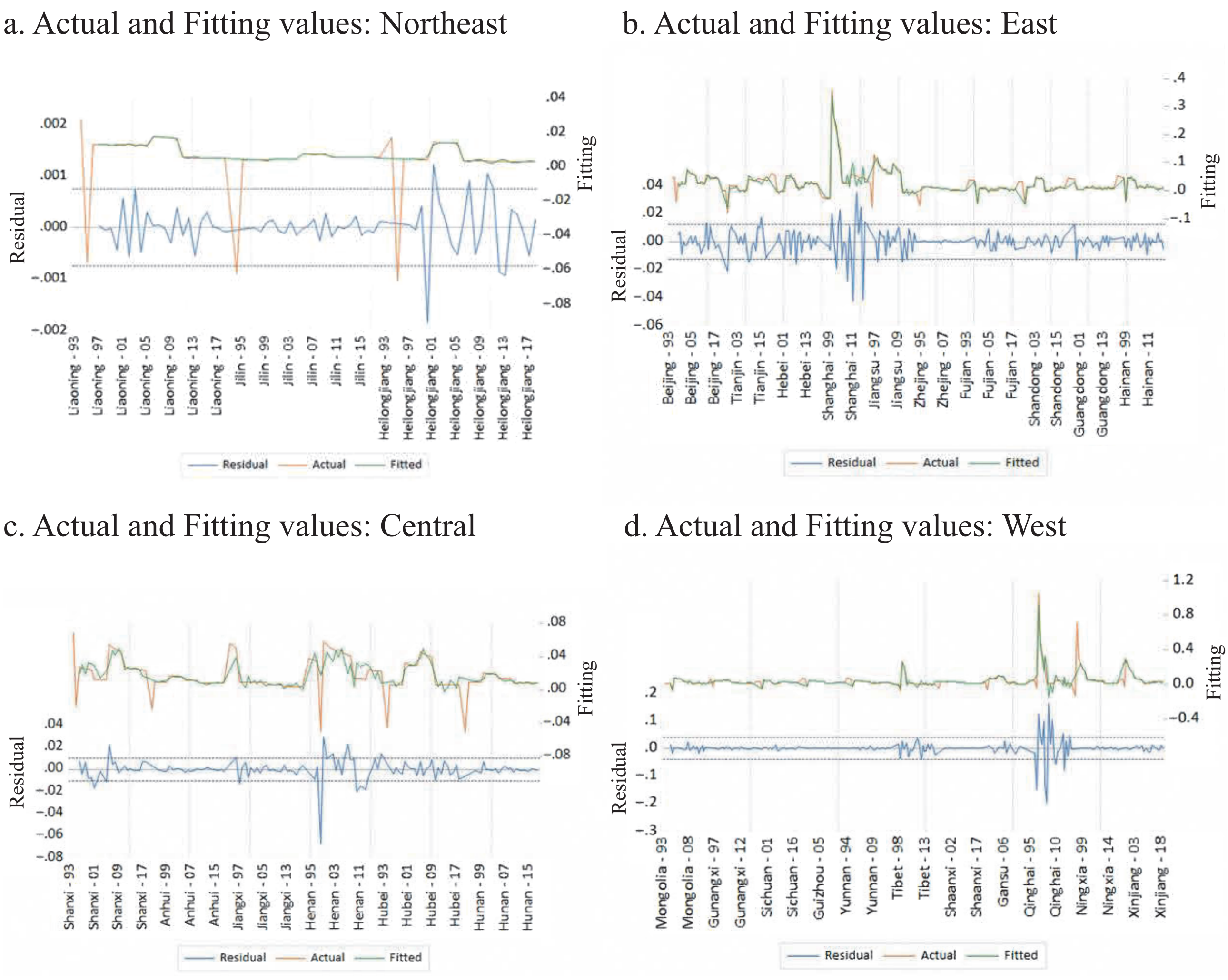

Fig. 2. Actual and Fitting values of four regions

Table 17. Panel ARDL model analysis results matrix table (long run)

	Northeast	East	Central	West
GRPPC	+	+	+	+
$GRPPC^2$		−		−
POP				+
TO	−	×	×	
IS1	−			

*+: positive and significant
−: negative and significant
×: not significant

Table 18. Panel ARDL model analysis results matrix table (short run)

	Northeast	East	Central	West
GRPPC	×	×	×	×
$GRPPC^2$		×		×
POP				×
TO	+	−	+	
IS1	+			
ECT	×	−	×	−

*+: positive and significant
−: negative and significant
×: not significant

had no effect on forest area. In the West, population, as a variable other than GRPPC, was positive and significant, and a long-term positive relationship existed with forest areas, which is consistent with the results of Tan et al. (2020). This suggests that in Western China, which has many arid areas, POP growth may have a relatively small effect on forest area expansion because of the influence of policies such as the Tree-planting Festival.

In summary, although China is in the reforestation stage of the U-shaped hypothesis (Tan et al. 2020, 2021, Zhang et al. 2006), the analysis revealed regional differences. It has been suggested that some regions are in the phase in which forest areas are recovering as the economy develops (Northeast and Central), whereas others are in the phase in which forest area expansion stagnates despite further economic development (East and West). The panel ARDL model can be used to discuss the results of short and long-run model analyses. Most of the variables were not significant in the short run. However, they become significant in the long run. The ECT was negative and significant in the East and West, but not significant in the Northeast and Central regions. There are 10 and 11 provinces in the East and West, 3 and 6 provinces in the Northeast and Central regions, with the latter two regions with an insignificant ECT, having fewer provinces. Caravaggio (2020a) notes that 25 years of FRA data are insufficient to incorporate many adjustments for long-lived resources like forests. So, it may be possible that the error correction from the long-term equilibrium was not adequately represented in these regions. Meanwhile, Caravaggio (2020b) indicated that because forest transformations are slow, deviations from equilibrium are not rapidly adjusted. This may indicate that the deviations from the long-run equilibrium were gradually corrected. Compared to Caravaggio (2020b), the speed of adjustment to a long-run equilibrium in East and West China in this analysis was found to be a little faster. In the East and West, where the relationship between forest areas and GRPPC had an inverted U-shape, the turning point was in the latter half of the analysis period, when it stopped at approximately 100,000 and 60,000 Yuan, respectively. Wang et al. (2019) noted that the turning point of the U-shape was at approximately 27,000 Yuan, whereas Zhang et al. (2006) found that China was already in the latter half of the U-shape. Tan et al. (2021) conducted a national-level analysis targeting China using the ARDL model and concluded that GRPPC is the most important variable, and once the economy reaches a certain level of development, its further impact may slow down. The reason for the decline in forest area expansion may be that high-income regions are attempting sustainable forestry by focusing on the management of existing afforested land rather than creating new afforestation (Seki et al. 2009). Hao (2019) divided China into three regions and analyzed the relationship between GRPPC and timber production using the generalized method of moments. The Eastern region reached its turning point at approximately 47,000 Yuan, whereas the Western region reached its turning point at approximately 9,000 Yuan. An empirical analysis by Caravaggio (2020b) based on the EKC hypothesis showed that after the second reversal, forest area expansion tapered off in high-income countries. Although Caravaggio (2022) notes that there are two turning points, only the second turning point was observed in this study. It may be effected, however, by the shorter analysis period.

From the results of this study, we concluded as follows.

1) The results of the analysis using the panel ARDL model showed that GRPPC had the largest impact on the forest area.

2) In China, regional differences exist in the relationship between forest areas and socioeconomic factors in the same country.

3) China was already in the latter stage of the U-shape curve; however, as the economy grew, the

increase in forest area slowed and may have reached the stable phase in the East and West.

Funding details

This study did not receive any specific grants from funding agencies in the public, commercial, or non-profit sectors.

Acknowledgements

We thank the reviewers for their valuable comments.

References

Caravaggio, N. (2020a) Economic growth and the forest development path: A theoretical re-assessment of the environmental Kuznets curve for deforestation. *Forest Policy and Economics*, **118**, 1-11.

Caravaggio, N. (2020b) A global empirical re-assessment of the Environmental Kuznets curve for deforestation. *Forest Policy and Economics*, **119**, 1-12.

Caravaggio, N. (2022) Economic growth and forest transition in Latin America. *Forest Policy and Economics*, **135**, 1-11.

Chigira, H. et al. (2011) *Dynamic Panel Data Analysis*, Chisen Shokan, Tokyo [In Japanese].

China National Bureau of Statistics (1999-2022) China Statistical Yearbook. http://www.stats.gov.cn/tjsj/ndsj/. Accessed on 20 January 2022 [In Chinese].

China National Forestry and Grassland Administration (2019) *China Forest Resources Report 2014-2018*. China Forestry Publishing House, China [In Chinese].

China State Council Gazette (2006) 11th Five-year Plan of China. http://www.gov.cn/gongbao/content/2006/content_268766.htm. Accessed on 10 March 2022 [In Chinese].

Choi, I. (2001) Unit root test for panel data. *Journal of International Money and Finance*, **20**, 249-272.

Churchill, S. A. et al. (2018) The environmental Kuznets curve in the OECD: 1870-2014. *Energy Economics*, **75**, 389-399.

Dritsaki, C. & Dritsaki, M. (2013) Hysteresis in unemployment: an empirical research for three member states of the European Union. *Theoretical and Applied Economics*, **20**, 35-46.

Engle, R. F. & Granger, C. W. J. (1987) Cointegration and error correction: Representation, estimation and testing. *Econometrica*, **55**, 251-276.

E-Views 12 (2021) User's Guide. https://www.eviews.com/help/helpintro.html#page/content%2Fpanel-Estimating_a_Panel_Equation.html%23ww193253. Accessed on 20 March 2022.

FAO (2020) Global forest resources assessment 2020 Main report. https://www.fao.org/3/ca9825en/ca9825en.pdf. Accessed on 10 February 2022.

Forestry and Forest Products Research Institute (2010) *China's forests, forestry and timber industry – Current status and prospects*, J-FIC, Japan [In Japanese].

Greene, W. H. (2019) *Econometric analysis, 8th ed.* Pearson Education Limited, UK.

Hao, Y. et al. (2019) Relationship between forest and economic growth: Empirical evidence from China. *Journal of Cleaner Production*, **214**, 848-859.

Inoue, M. (1992) Basic policy for tropical forest conservation based on the characteristics of forest-use patterns. *Review of Forestry Culture*, **13**, 27-32 [In Japanese].

Kitamura, Y. (2005) *Panel data analysis*. Iwanami Shoten, Japan [In Japanese].

Levin, A. et al. (2002) Unit root test in panel data: asymptotic and finite-sample properties. *J. Econometrics*, **108**, 1-24.

Maddala, G. S. & Wu, S. (1999) A Comparative Study of Unit Root Tests with Panel Data and a New Simple Test. *Oxford Bulletin of Economics and Statistics*, **61**, 631-652.

Mather, A. S. (1992a) *Global Forest Resource* (translated by Minoru Kumazaki). Tsukiji Shokan, Japan, pp.78-102 [In Japanese].

Mather, A. S. (1992b). *The forest transition*. Area, **24**, 367-379.

Michinaka, T. & Miyamoto, M. (2013) Forest and human development: An analysis of the socio-economic factors affecting global forest area changes. *J. Society of Forest Planning*, **18**, 141-150.

Michinaka, T. et al. (2013) Factors affecting forest area changes in Cambodia: An econometric approach. *J. Sustainable Development*, **6**, 12-25.

Murao, H. (2019) *VAR empirical analysis in R—from the basics of time series analysis to forecasting.* Ohmsha, Japan [In Japanese].

Nagata, S. et al. (1994) *Forest resource use and restoration—economic logic and natural logic*. Rural Culture Association, Japan [In Japanese].

Nishiyama, Y. et al. (2019) *Econometrics: Statical data analysis for empirical economics.* Yuhikaku Publishing, Japan [In Japanese].

Pesaran, M. H. & Shin, Y. (1999) An autoregressive distributed-lag modeling approach to cointegration analysis. *In* Econometrics and Economic Theory in the 20th Century, Cambridge University Press, pp. 371-413.

Pesaran, M. H. & Smith, R. (1995) Estimating long-run relationships from dynamic heterogeneous panels. *J. of Econometrics*, **68**, 79-113.

Sakamoto, M. et al. (2014) Trends in research on the transition of forest areas: Focus on the forest transition theory. *Forest Economy* (*Ringyo Keizai*), **67**, 1-16 [In Japanese].

Seki, Y. et al. (2009) *Reforestation in China: Beyond socialism and marketism*. Ochanomizu shobo, pp. 241-254 [In Japanese].

Shi, M. et al. (2017) An empirical analysis of the driving forces of forest cover change in northeast China: *Forest Policy and Economics*, **78**, 78-87.

Tan, J. et al. (2020) An empirical study on the U-shaped hypothesis of forest resources using a non-linear model: A case of China. *Forest Economy* (*Ringyo Keizai*), **73**, 1-14 [In Japanese with English summary].

Tan, J. et al. (2021) Short- and long-run effects of socio-economic factors on forest dynamics in China. *Journal of Forest Economics* (*Ringyo Keizai Kenkyu*), **104**, 74-81 [In Japanese with English summary].

Wang, J. et al. (2019) Economic growth, government policies, and forest transition in China. *Regional Environmental Change*, **19**, 1023-1033.

Wang, S. et al. (2007) Is China in a later stage of a U-shaped

forest resource curve? A re-examination of empirical evidence. *Forest Policy and Economics*, **10**, 1-6.
World Bank (2021) World Bank Open Data. https://data.worldbank.org/indicator/NY.GDP.DEFL.ZS?locations=US. Accessed on 20 January 2022.
Wunshiou, Y. (1996) Mandatory tree planting scheme in China. *Forest Economy*, **49**, 15-23.
Yorimitsu, R. (1984) *Forest and green resources in Japan*. Toyo Keizai Inc., pp. 8-18 [In Japanese].
Zhang, Y. et al. (2006) Impact of socio-economic factors on the changes in forest areas in China. *Forest Policy and Economics*, **9**, 63-76.