International Journal of Statistical Sciences *ISSN 1683-5603*
Vol. 22(1), 2022, pp 197-212
© 2022 Dept. of Statistics, Univ. of Rajshahi, Bangladesh

The Latent-Space Adjusted Approach: An Alternative form in Modeling Economic Development and Forest Cover

Dipak K. Dey[*] and Brisilda Ndreka

Department of Statistics, University of Connecticut, Storrs, CT 06269

*****Correspondence should be addressed to Dipak K. Dey
(Email: dipak.dey@uconn.edu)

[Received September 12, 2021; Accepted March 30, 2022]

Abstact

Peer effect, sometimes referred to as contagion effect has become an essential part of social network analysis. This paper proposes a novel data-driven framework that borrow the "contagious effect" concept to analyze the influence of the economical income status in the forest resources of a country. Specifically, considering that called “contagious economic effect”as a regressor in a dynamic linear model we can “capture” if the "attitude" of the countries within the same economical level is similar according to the proportion of forest changes. For the interence part, we first apply latent-space adjusted approach to obtain unobserved trait, which are extracted from social network structure. Then, the latent variables are used in the model to correct biases when identifying contagion effect. Firstly, the analysis suggests that countries in the same economic level does not "behave" the same toward the forest change. Secondly, the model comes up with indentifying economical and social factors playing a significant role in the variable of interest.

Keywords: Deforestation, latent trait, contagious effect, economy level, social factor.

AMS Subject Classification: 91B76.

1. Introduction

According to the Center for International Forestry Research (CIFOR), since 1990, more than 300 million hectares of forests have been destroyed. Where the causes of deforestation and forest degradion ca be socioeconomic, political and natural factors. Even the collection of deforestation causes is very large, studies have suggested that environmental degradation maybe a consequence of expansions in economic activities. Economic growth is consistently viewed as having effects on environmental quality, leading by the idea that predicting economic growth, can project environmental consequences as well. These conclusions followed the interest of researchers in exploring the link between the economy and the environment.

It is well known in the environmental economics literature the inverted-U-shaped curve (known as the environmental Kuznets curve, Kuznets (1955)), describing relationship between environmental degradation and economic growth, Grossman and Krueger (1991). Which suggests that environmental degradation initially rises with per capita income. However, with economic growth comes an increased demand for environmental quality, leading to environmental aspect improved. This theory has been empirically confirmed by several other researchers referring to Culas (2012); Ahmed et al. (2016) or Duan et al. (2016) etc. However, other studies negated this existence, suggesting that environmental degradation will start to rise again beyond a certain income level, Bhattarai et al. (2009); Álvarez-Herranz and Balsalobre Lorente (2016). Furthermore, is pretended that the relation between economic growth and environmental issues in countries going through the process of transition from a centrally planned to a free-market economy has never been analyzed, Bernaciak (2013).

Under these pros and cons of income level in the environmental degradation, we aim to answer the following research questions: Do the countries in the same income level behave similarly toward environmental degradation? In other words, can we pretend that lower middle-income countries have the same attitude on forest change (the same question raised for other economic level as well)? Analyzing these questions, will provide us in a better argument why relationship curves between environment and per capita income has different approach in different studies.

This paper contributes to the existing literature that when comes to define the pattern between income and environmental, we should consider being in the scenario that even within the same income level countries may not behave the same.

Since forest change is not just affected by economic development, we also include variables to control for population growth and agriculture development.

This study starts by going down the line of longitudinal social network analysis. We perceive the income level groups as a social network, in which the states are considered as nodes of a graph that are connected if they share the same income level, consider those the graph edges. To define the income status GNI per capita, Atlas method is used.

We build the analysis under latent space model by Hoff et al. (2002), where inference is performed under Bayesian frameworks, Markov Chain Monte Carlo procedures used to estimate latent positions. First, we estimate the latent unobserved variable using latent-space adjusted approach, which according Xu (2018), has the potential to correctly identify social influence effects when there is an unobserved variable that co-determines the influence and the selection process. Then, we use the dynamic linear model Friedkin and Johnson. (1990) to categorize the significance of exposure term (contagious effect).

Our approach not only allows us to test if the countries on the same economical stage share the same behaviour towards forest change but also to identify socioeconomic determinants in forest cover across countries.

The format of this paper is as follows: Section 2 describes the data used and inclusion of participants variables in the study. Section 3 outlines the theory of the model fitting. Section 4 describes implementation in dynamic linear model. Furthermore, considering longitudinal data coming from a non- normal distribution, approaches based on generalized linear equations such as quasi-binomial and beta regression are applied. Finally, Section 5 contains concluding remarks and directions for future research and related implementation. Appendix provides extended explanation as needed.

2. The involved study factors of Forest Cover

The data set is accessed by the:

https://databank.worldbank.org/reports.aspx?source=world-development-indicators from the World Bank website. The inference is done on 219 countries worldwide. This compiled data set is pulled from four data sets linked by time (year) during the past 30 years based on 1990, 2000, 2010, 2018 data and place to inspire the forest cover change analysis. The study consists of several socioeconomic factors in forest cover across countries. These factors are used not only when looking at the direct impact of the interested variable that is proportion of forest area defined as:

$$FA_i = \frac{forest\ area\ of\ i\text{–}th\ country}{country's\ total\ land\ area} \quad (1)$$

for $i = 1,2, \ldots ,219$.

This is also used in creating new predictors for our fitting model. The proportion of the forest area in country i (FA_i) is assumed to be impacted by its level of GDP per capita (GDP_i), the growth rate of GDP per capita (ΔGDP_i) population density (P_i), rural population growth (ΔR_i) and the agriculture raw material export ($ARM_{exp}i$). GNI per capita (Atlas method); was used to categorised the income level in each country. According to:

https://documents1.worldbank.org/curated/en/408581467988942234/pdf/WPS7528.pdf

the World Bank's Classification of Countries by Income, the atlas method threshold is updated as result of taking into consideration inflation and other economic factors. We referred to Table 1 to categorize the income status of the countries, approximately to the closest year.

Table 1: Decline in thresholds relative to world GNI per capita (current US, Atlas method).

source: https://documents1.worldbank.org/curated/en/408581467988942234/pdf/WPS7528.pdf

Bank fiscal year	1978	1984	1989	1999	2009	2016
Calendar year of data	1976	1982	1987	1997	2007	2014
Upper bound thresholds for income groupings, GNI per capita						
Low	250	410	480	785	935	1045
Lower middle	...	1670	1940	3125	3705	4125
Upper middle	...	...	6000	9655	11455	12735
World GNI per capita	1623	2567	3290	5491	8394	10779

Missing data: In the data studied, more that than 5% of them are considered missing values. While being under these conditions multivariate imputation is performed to replace them. We are assuming that data is missing at random (MAR) meaning that the probability that a value is missing depends only on observed value and can be predicted using them and processing with MICE (Multivariate Imputation via Chained Equations) package by R using CART (Classification and Regression Tree) predictive algorithm. *Why CART?* CART is a nonparametric approach so we do not have to depend on any specific distributional assumptions, CART is also not significantly impacted by outliers and also deals with considerable missing data.

3. Socioeconomic determinants on forest area and fitting model

3.1. Theoretical Model Description

In this study the analysis is built under the idea of the network influence model Manski (1993), where the countries represent nodes of the graph and edges connects states that belong to the same income level at a given time $\boldsymbol{t}$ *(represented by year).*

We are considering that the proportion of forest area of $i-th$ country at time t (FA_{it}) to be a function of the past forest cover proportion area, to the GDP per capita at time t GDP_{it}, to the annual percentage of growth rate ΔGDP_{it}, population density P_{it}, rural population grow R_{it}, their agriculture raw material export $ARM_{exp}it$ and some latent covariaties unidentified C_i, for example can be the level of corruption that we are not involving in the model. We start with a dynamic linear model Friedkin and Johnsen (1990):

$$FA_{it} = \beta_0 + \beta_1 FA_{it-1} + \beta_2 \frac{\Sigma Z_{ijt-1} FA_{jt-1}}{\Sigma Z_{ijt-1}} + \beta_3 GDP_{it} + \beta_4 \Delta GDP_{it} + \beta_5 P_{it} + \beta_6 R_{it} + ARM_{exp}it + C_i + \epsilon_{it} \quad (2)$$

for $i = 1,2,\ldots,219$, and for notation purpose let muve $t = 1,2,3,4$ where; $t = 1$ refers to data in 1990; $t = 2$ year 2000; $t = 3$ year 2010; $t = 4$ year 2018,

where referring to Xu (2018), FA_{it-1} lagged dependent variable is the previous "behaviour" (in the study the proportion of forest area at time $t-1$) of country i, Z_{ijt-1} is a dummy variable, with values of 1 if the countries correspond to the same income levels and 0 otherwise, considering now $\frac{Z_{ijt-1}Y_{jt-1}}{Z_{ijt-1}}$ as the exposure term. The rest are the covariates that are considered that might impact the variable of interest. The estimation of β_2, referred to as the contagious effect has an important role in our analysis. Since we provide an answer to the question: Do the countries in the same income levels have the same "behavior" on forest change?

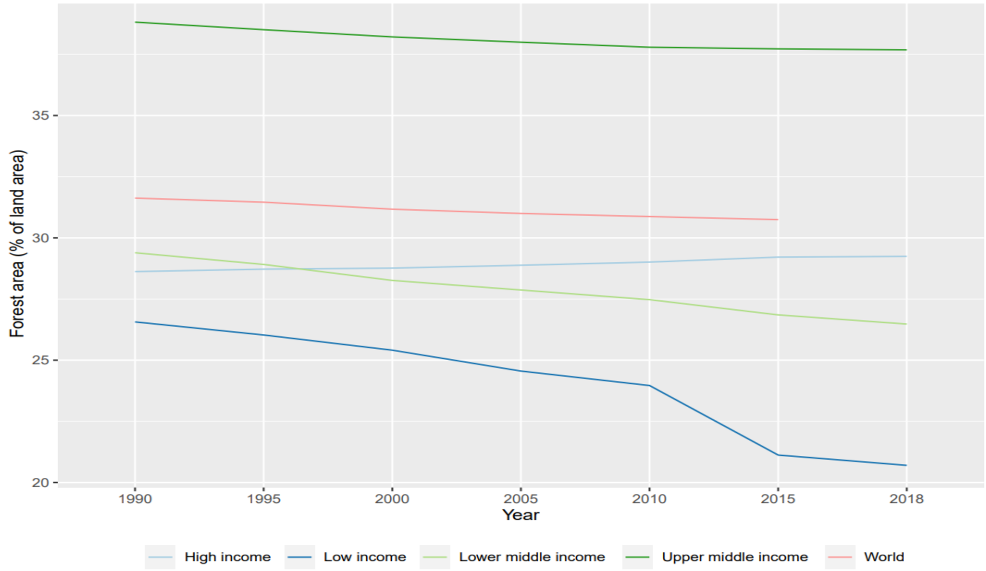

Figure 1: Percentage of forest area change in the last 30 years, controlling for income level.

Having a a close view at the graph above, the proportion of forest area during these last thirty years are visualized and are given as percentages for four categories of income level and worldwide as well. Clearly it is noted that the world is losing forest area, the red curve decreases in time but after 2005 deforestation becomes smoother. Focusing on the curves that represent the income level category, we see that they act differently. The high income countries with a stable economy are gaining forest area over time. The opposite tabloid is presented by the dark blue curve representing low income countries where the deforestation is continuing with a drastic decrease after 2015. In a overall view during these past thirty years for the two other categories low middle income and upper middle income again is noted by a loss in forest area. However after 2010 deforestation is faltering for the upper middle income countries . For the lower middle income where economy is in developing process this phenomena is increasing. Since even within the same income level categories the curves act differently for different periods of time it doesn’t make it easy to conclude the impact of economy level on forest area .

Going back to the fitting model Eq.2, the latent variable is estimated using the latent space model Hoff (2002), concretely used as proxies of the unobserved trait the "latent space position". This last one defined by

$$log\ \ odds(Z_{ij} = 1|c_i, c_j, x_{ij}, \alpha, \beta) = \alpha + \beta' x_{ij} - |c_i - c_j|, \tag{3}$$

where Z_{ij} indicates an interaction between i and j, x_{ij} the vector of covariates, c indicates the latent social position of i and j, and $|c_i - c_j|$ which is the Euclidean distance between them. The parameters α and β are estimated using either Maximum-Likelihood Estimation (MLE) or Markov Chain Monte Carlo (MCMC) methods and the latent position c can be estimated by Minimum Kullback-Leibler (MKL) estimates Shortreed (2006). In other words, if two countries in the same income level that are close to each other in terms of GDP per capita growth, population density, rural population growth and proportion of forest cover, would also be close also under unobserved trait. Working with longitudinal data, the estimated value of latent positions for all available time point are used as proxies of the unobserved trait and process with the use of OLS to estimate parameters Xu (2018).

4. Model fitting

Under this development, FC_t consists of a vector of proportion of forest area per each state in three time point 2000, 2010, 2018. Ensuring comparability with the lagged effect, FC_{t-1} is a vector that lies in with the forest area proportion at 1990, 2000, 2010 across countries (lagged dependent variable in 2000 are consider data at 1990, for 2010 are considered values at 2000 and for 2018 forest area are considered 2010 data set). The value of the rest of explanatory variables GDP_t, ΔGDP, P_t, R_t, ARM_{exp} and contagious terms is data coming from 2000, 2010, 2018.

Consider the forest data described above our interests lie in modeling the forest area proportion (FA) to time t as a function of the forest cover proportion at $t-1$ (FA_{t-1}) to the GDP and growth rate of GDP (ΔGDP), to the population density (P), rural population growth (R), to the export of the raw material (ARM) and some unobserved covariate defined as latent variable (C).

The latent space model 3 was used to estimate "latent social position" (C_i) for each time point (recalling that in our analysis data we studied at four different time periods i= 1,2,3,4) and used it as proxies for the unobserved trait. Estimating the latent variable C_i the next step consists in fitting the influence model.

The first attempt was considered dynamic linear regression of the response on the covariates Eq.2. Even while not having a multicollinearity issue (vif values less than 2) the estimated regression displayed evidence of heteroskedasticity; the p-value for Koenker's (1981) homoskedasticity test was < 0.0001. Asymmetric of the residuals, even if we consider instead that transformation of the regression on the covariates was permanent; estimating of H_0 : the regression error is normally distributed, the test statistic equals to 0.44691 and the corresponding p-value is less than 0.001.

Handling asymmetric distributions of the continuous proportions data and having that quasimodel has more relaxed assumption, started up a quasi-binomial model. The influential models Eq. 2 comes in the form:

$$logit(FA_{it}) = \beta_0 + \beta_1 FA_{it-1} + \beta_2 \frac{\Sigma Z_{ijt-1} FA_{jt-1}}{\Sigma Z_{ijt-1}} + \beta_3 GDP_{it} + \beta_4 \Delta GDP_{it} + \beta_5 P_{it} + \beta_6 R_{it} + ARM_{exp} it + C_i \quad (4)$$

where $logit(FA_{it}) = log \frac{FA_{it}}{1-FA_{it}}$.

From the fitting model we have that from the covariates for an $\alpha = 10\%$ intended factors that impact forest area proportion to time t are; lagged effect (proportion area at time $t-1$); annual percentage grow rate of GDP per capita; population density and rural population growth, see Table 3 in Apendix. The exposure term coefficients corresponding with the influence that country j may have in the forest area proportion of the $i-th$ one, for $i,j = 1,\ldots,219$ are not considered significant. Hosmer-Lemeshow goodness of fit test concluded that the logistic response function (quasi-binomial) is appropriate (null hypothesis is not rejected $p-value$: 0.678)

Consider the nature of the data: continuous variables that assume values in the unit interval; trouble with non-constance variance then beta regression was the next attempt at modeling the proportion of forest area Ferrari and Cribari-Neto (2004). Having that FC (dependent variable) takes the extremes 0, a transformation of the data is performed as $(FC * (n-1) + 0.5)/n$ where n is the sample size, see Smiths and Verkuilen (2006).

Starting out with beta regression with logit link to the mean and identity for specific function, Eq. 4 now under beta distribution, after likelihood-ratio test on variable selection was applied we have reached a fitting-model similar to the quasi-binomial one, reffer to Table 4.

While measuring the model performance a repeated k-fold coss-validation is applied, with 10-fold and 3 repeats. The process consists of splitting the data into 10-folds and repeating this 3 times. Cross-validation was used for beta regression in order to select the best link for the fitted model. The performance test stated for loglog link function was considered a better fit to data even with not significant differences, Table 5. The influence model would come as follow:

$$loglog(FA_{it}) = \beta_0 + \beta_1 FA_{it-1} + \beta_2 \frac{\Sigma Z_{ijt-1} FA_{jt-1}}{\Sigma Z_{ijt-1}} + \beta_3 GDP_{it} + \beta_4 \Delta GDP_{it} + \beta_5 P_{it} + \beta_6 R_{it} + ARM_{exp} it + C_i \quad (5)$$

where $loglog(FA_{it}) = log(-log(1 - FA_{it}))$.

The likelihood ratio test states for a slightly differences in the significance of exploratory variables as well as those from previous two models, see Table 5. In terms of significance at level 10%, the model point out that anly FC_{t-1}; growth rate pf GDP per capita; population density has a significant impact in predicting proportion area.

As shown in the Table 2, the beta regression model with loglog link performs slightly better in terms of RMSE. The difference between the logit and the loglog link is tiny and we would prefer to move on with the logit model here because it is more convenient in interpretation.

Table 2: Model performance

Indices of model performance	AIC	BIC	R^2	RMSE	Sigma	Log_loss	PCP
Quasi-binomial model	-	-	-	0.08	0.22	0.05	0.02
Beta regression model (logit link)	-1544.794	-1517.923	0.735	0.083	0.852	-	-
Beta regression model (loglog link)	-1697.360	-1674.968	0.859	0.079	0.960	-	-

Ending up in the beta regression with only the significant covariates:

$$logit(FA_{it}) = \beta_0 + \beta_1 FA_{it-1} + \beta_2 \Delta GDP_{it} + \beta_3 P_{it} + \beta_4 R_{it} \quad (6)$$

where $logit(FA_{it}) = log\frac{FA_{it}}{1-FA_{it}}$.

5. Conclusion

Modeling the change of the proportion of forest area involving contagious effect as covariate explains the possibility of spread of deforestation or reforestation phenomena across counties or regions. Another important concept underlying this scenario is that latent variables are used as a regressor. Estimating a covariate that may have an impact in the response means that we increase that chances in a more precise predicted model.

Through the analysis above is ended up in identifying economical and social variables that need to be taken in consideration to evaluate the future "destiny" of forest change. The model considered did not "testify“ that countries within the same economical status behave the same on how the proportion in forest area changes within states over time. We conclude an insignificant contagious effect under a 10% significance threshold that we worked. The proportion of forest area decreases with a population density increase, a social-economical factor significant for predicting change in the proportion forested area of a country. Rural population growing in this analysis so far is not considered a potential factor that affects deforestation. The forest area condition 10 years before gives an important information in predicting how this would change further years. The annual percentage growth rate of GDP per capita "is pretending" to have a positive impact in a forest cover of a country.

References

[1] Arnau, J., Bono, R., Bendayan, R., and Blanca, M. J. (2016). Analyzing longitudinal data and use of the generalized linear model in health and social sciences. Quality Quantity.

[2] Chavance, M., Escolano, S., Romon, M., Basdevant, A., De Lauzonguillain, B., Charles, and Aline, M. (2010). Latent variables and structural equation models for longitudinal relationships: an illustration in nutritional epidemiology. BMC Medical Research Methodology.

[3] Chien, L.-C. (2011). Diagnostic plots in beta-regression models. Journal of Applied Statistics 38, 1607–1622.

[4] Cribari-Neto, F., and Zeileis, A. (2010). Beta regression in r. Journal of Statistical Software, Articles 34, 2, 1–24.

[5] CUaresma, C., Jesus, D. O., Steffen, F., Ian, M., Michael, O., Linda, S., and Brian, W. (2017). Economic development, and forest cover: Evidence from satellite data. Scientific Report.

[6] Espinheira, P. L., Ferrari, S. L., And Cribari-Neto, F. (2008). Influence diagnostics in beta regression. Computational Statistics Data Analysis 52, 9, 4417–4431.

[7] Ferrari, S., and Cribari-Neto, F. (2004). Beta regression for modelling rates and proportions. Journal of Applied Statistics 31, 7, 799–815.

[8] Friedkin, N. E., and Johnsen, E. (1990). Social influence, and opinions. Journal of Mathematical Sociology, 193–206.

[9] Hoff, P., Raftery, A., and Handcock, M. (2002). Latent space approaches to social network analysis. Journal of the American Statistical Association 97, 1090–1098.

[10] JACK. Beta regression (betareg) with caret and train.

[11] Joshi, P., and Beck, K. (2016). Environmental kuznets curve for deforestation: evidence using gmm estimation for oecd and non-oecd regions. iForest - Biogeosciences and Forestry 10, e1–e8.

[12] Koenker, R. (1981). A note on studentizing a test for heteroscedasticity. Journal of Econometrics 17, 1, 107–112.

[13] Liang, K.-Y., and Zeger, S. L. (1986). Longitudinal data analysis using generalized linear models. Biometrika 73, 1, 13–22. 9

[14] L. S., Z., Liang, K.-Y., and Albert, P. S. (1988). Models for longitudinal data: A generalized estimating equation approach. Biometrics 44, 4, 1049–1060.

[15] Pereira, G. H. (2019). On quantile residuals in beta regression. Communications in Statistics - Simulation and Computation 48, 302 – 316.

[16] Ritchie, H., and Roser, M. (2021). Forests and deforestation. Our World in Data.

[17] Shortreed, S., Handcock, M. S., and Hoff, P. (2006). Positional estimation within a latent space model for networks. Methodology 2, 1, 24–33.

[18] Smithson, M., and Verkuilen, J. (2006). A better lemon squeezer? maximum-likelihood regression with beta-distributed dependent variables. Psychological methods 11 1, 54–71.

[19] Van Buuren, S., And Groothuis-Oudshoorn, K. (2011). mice: Multivariate imputation by chained equations in r. Journal of Statistical Software, Articles, 1–67.

[20] Wang, M., And Wang, W. (2015). Bias-corrected maximum likelihood estimation of the parameters of the weighted lindley distribution. Communication in Statistics- Simulation and Computation 46.

[21] XU, R. (2018). Alternative estimation methods for identifying contagion effects in dynamic social networks: A latent-space adjusted approach. Social Networks, 101–117.

[22] YEO, I., J. R. (2000). A new family of power transformations to improve normality or symmetry. Biometrika, 954–959

[23] Contreras-Hermosilla, A. (2000). The Underlying Causes of Forest Decline. CIFOR Occasional paper no 30. Center for International Forestry Research, Bogor, Indonesia.

[24] Kuznets, S. (1955). Economic Growth and Income Inequality. The American Economic Review, 45, 1-28.

[25] Grossman, G. M., Krueger, A. B. (1991). Environmental impacts of the North American Free Trade Agreement. NBER. Working paper 3914.

[26] Culas R. J. (2012). REDD and forest transition: tunneling through the environmental Kuznets curve. Ecol Econ 79:44–51

[27] Ahmed A, Uddin G. S, Sohag K. (2016). Biomass energy, technological process and the environmental Kuznets curve: evidence from selected European countries. Biomass Bioenergy 90:202–208

[28] Duan L, Guo Y, Zhu H, Yu K. (2016). The effects of FDI, economic growth and energy consumption on carbon emissions in ASEAN-5: evidence from panel quantile regression. Econ Model 58:237–248

[29] Bhattarai K, Paudel K. P, Poudel B. N. (2009). Searching for an environmental Kuznets curve in carbon dioxide pollutant in Latin American countries. J Agric Appl Econ 41(1):13–27

[30] Álvarez-Herranz A, Balsalobre Lorente D. (2016). Economic growth and energy regulation in the environmental Kuznets curve. Environ Sci Pollut Res 23(16):16478–16494

[31] Bernaciak, A. (2013). The environmental Kuznet curve in transition countries on the example of Poland. Economic and Environmental Studies (2013) (E&ES) [ISSN:] 2081-8319 [Volume:] 13 [Year:] 2013 [Issue:] 3 [Pages:] 279-293

Appendix

Table 2: Parameter estimate of quai-binomial model. Indicating significant predictor; lagged effect, growth rate of GDP, population density a rural population growth.

Coefficients	Estimate	Std. Error	t value	$\Pr(>\lvert t \rvert)$
(Intercept)	-1.785e+00	6.839e-01	-2.610	0.00926 **
lagg_FC	5.004e+00	1.149e-01	43.569	<2e-16 ***
exposure	-2.547e+00	2.054e+00	-1.240	0.21546
ARM_exp	2.149e-03	3.160e-03	0.680	0.49670
GDP	1.174e-06	1.082e-06	1.085	0.27842
grow_GDP	1.973e-02	6.422e-03	3.072	0.00221 **
P	-3.761e-05	1.449e-05	-2.596	0.00966 **
grow_rural	-2.250e-02	1.353e-02	-1.663	0.09677.
latent	3.521e-04	5.746e-04	0.613	0.54025

Table 3: Parameter estimate of beta regression model, with logit link function to mean and identity to precision function.

Coefficients (mean model with logit link):	Estimate	Std. Error	z value	Pr(> |z|)
(Intercept)	-2.685e+e00	5.190 e-0.2	-51.735	<2e-16***
lagg_FC	5.239e+00	1.080e-01	48.531	<2e-16 ***
grow_GDP	1.361e-02	6.225e-03	2.186	0.0288 *
P	-7.194e-05	1.457e-05	-4.937	7.95 e-07 ***
grow_rural	-2.355e-02	1.294e-02	-1.820	0.0688 .

Table 4: Parameter estimate of beta regression model, with loglog link function to mean and identity to precision function.

	Estimate	Std.Error	z value	Pr(> |z|
(Intercept)	-1.318e+00	3.240e-01	-4.067	4.76 e-05**
lagg_FC	3.183e+00	5.697e-02	55.872	<2e-16 ***
exposure	3.531e-01	9.749e-01	0.362	0.717
ARM_exp	1.902e-03	1.629e-03	1.167	0.243
GDP	2.511e-07	5.015e-07	0.501	0.617
grow_GDP	6.743e-03	2.988e-03	2.257	0.024 *
P	-2.519e-05	5.645e-06	-4.461	8.15e-06***
grow_rural	-4.060e-03	6.157e-03	-0.659	0.510
latent	1.969e-04	2.732e-04	0.721	0.471

Table 5: Selected variable detailed description

Variable name	**Code name**	**Short descriptions/ Notes**
Forest area (% of land area)	FA	Forest area is land under natural or planted stands of trees of at least 5 meters in situ, whether productive or not, and excludes tree stands in agricultural production systems and trees in urban parks and gardens. The indictor is derived by dividing total area under forest of a country by country’s total land area and multiplying by 100. https://data.worldbank.org/indicator/AG.LND.F

		RST.ZS?locations=1W&name_desc=false
GDP per capita (current US$)	GDP	GDP per capita is gross domestic product divided by midyear population. https://www.worldbank.org/en/home
GDP per capita growth (annual %)	ΔGDP	Annual percentage growth rate of GDP per capita based on constant local currency. https://www.worldbank.org/en/home
Population density	P	Population density is midyear population divided by land area in square kilometers. Population is based on the de facto definition of population, which counts all residents regardless of legal status or citizenship–except for refugees not permanently settled in the country of asylum, who are generally considered part of the population of their country of origin. Land area is a country's total area, excluding area under inland water bodies, national claims to continental shelf, and exclusive economic zones. In most cases the definition of inland water bodies includes major rivers and lakes. https://www.worldbank.org/en/home
Rural population growth (annual %)	ΔR	Rural population refers to people living in rural areas as defined by national statistical offices. It is calculated as the difference between total population and urban population. https://data.worldbank.org/indicator/SP.RUR.TOTL.ZG
GNI per capita, Atlas method	GNI	GNI per capita (formerly GNP per capita) is the gross national income, converted to U.S. dollars using the World Bank Atlas method https://datahelpdesk.worldbank.org/knowledgebase/articles/378832-what-is-the-world-bank-atlas-method, divided

		by the midyear population https://data.worldbank.org/indicator/AG.LND.FRST.ZS?locations=1W&name_desc=false. GNI per capita was used in analysis to categorized the states income level where: \$1,035 or less a year before; lower middle-income economies are those with a GNI per capita between \$1,036 and \$4,045; upper middle-income economies are those with a GNI per capita between \$4,046 and \$12,535; high income economies are those with a GNI per capita of \$12,536 or more. https://datahelpdesk.worldbank.org/knowledgebase/articles/ 378834-how-does-the-world-bank-classify-countries.
Agricultural raw materials exports (% of merchandise exports)	ARM_{exp}	Agricultural raw materials comprise SITC section 2 (crude materials except fuels) excluding divisions 22, 27 (crude fertilizers and minerals excluding coal, petroleum, and precious stones), and 28 (metalliferous ores and scrap) https://data.worldbank.org/indicator/TX.VAL.AGRI.ZS.UN