ARTICLE IN PRESS

Journal of Management Science and Engineering xxx (xxxx) xxx

Science Press

Contents lists available at ScienceDirect

Journal of Management Science and Engineering

journal homepage: www.keaipublishing.com/en/journals/journal-of-management-science-and-engineering/

Determinants in the forecast of the gross national income of China and India from 1952 to 2015

Angel Colmenares [a], Xu Yan [b, c, *], Weiguo Wang [c, **]

[a] School of Statistics, Dongbei University of Finance and Economics, Dalian, China
[b] School of Mathematics, Ocean University of China, Qing Dao, China
[c] School of Data Science and Artificial Intellegence, Dongbei University of Finance and Economics, Dalian, China

A R T I C L E I N F O

Article history:
Available online xxx

Keywords:
Gross National Income (GNI)
Stepwise
Ridge
Lasso
Elastic net
AutoRegressive Distributed Lag (ARDL)

A B S T R A C T

In 2015, China and India's population represented approximately 35.74% of the total number of people living in the world. Due to the historical context and behavior of the most relevant indicators, this study proposes to utilize a wide variety of demographic, economic, and production indicators from 1952 to 2015 to assess their impact on the GNI in China and India. A comprehensive and new fangled modeling process with stepwise, regularization and distributed lag regression approaches is presented. Accordingly, theoretical results were corroborated through extensive diagnostic tests and an empirical check of the models' predictive capacity. The findings show that GNI in China is most influenced by variables such as reserves in foreign currency and the dependency ratio; whereas, variables of energy production and birth rate were generated for India. Therefore, it's the timing for China to relax the universal two-child policy. Due to the current value below the substitution rate, a gloomy outlook for China's future population and economy is predicted. Conversely, a positive outlook is forecasted for India, given the low price in the future of oil- India's primary raw material.

1. Introduction

In recent decades, China's dramatic economic growth has been studied extensively (Yueh, 2013; Wu, 2003; Knight and Ding, 2012; Zheng, 2018). The changing demographics of the country have also received scholarly attention (Mo and Wei, 2018; Li, 1992; Ansley, 1984), due to the unique strict birth control policies that have led to a rapid increase in the country's aging population rate. On the contrary, India has not seen such accelerated economic and demographic changes. However, the Indian economy experienced a sustained increase, contributing to the steady growth in its population (AllaouaZoubida, 1997; Jati, 2009; Eichengreen et al., 2010). In 2018, the United Nations reported that the estimated 1,352,642,280 of India's population was approaching the size of the estimated Chinese population of 1,427,647,7860.

Although India and China are different, they have undergone similar transformations since the second half of the twentieth century. The funding of the People's Republic of China (PRC) occurred in 1949. In 1950, one year after the PRC's

* Corresponding author. School of Mathematics, Ocean University of China, Qing Dao, China.
** Corresponding author. School of Data Science and Artificial Intellegence, Dongbei University of Finance and Economics, Dalian, China.
E-mail addresses: yan_xu@dufe.edu.cn (X. Yan), wwguo@dufe.edu.cn (W. Wang).

https://doi.org/10.1016/j.jmse.2021.06.001

Please cite this article as: A. Colmenares, X. Yan and W. Wang, Determinants in the forecast of the gross national income of China and India from 1952 to 2015, Journal of Management Science and Engineering, https://doi.org/10.1016/j.jmse.2021.06.001

formation, India gained its independence, becoming the Republic of India. With the use of strategic five-year plans, both countries have planned their economy's trajectory, even though the form of government is different. Furthermore, they formally joined the World Trade Organization (WHO) six years apart, with India joining in 1995 and China in 2001. As previously outlined, China and India have similarities in areas of their historical events, the size of their population, and the planning of their economies. These similarities have sufficient latitude to conduct a comparative study between the two countries, taking the Gross National Income (GNI) as the primary dependent variable.

In this study, we analyze the impact of the following explanatory variables on the GNI between 1952 and 2015: foreign exchange reserves, trade balance, national government expenditure, energy production, production of fish, total population, literacy rate, median population age, death rate, dependency ratio, fertility rate, life expectancy, and birth rate.

To assess the relationship between the economic variables mentioned, this study employs the use of regression models, since traditionally they have been used as a starting point. Therefore, when there is a large number of explanatory variables, one of the primary methods for selecting the variables of interest is stepwise selection. The use of stepwise analysis for selecting variables implemented nine different criterias (*Akaike, Corrected form of Akaike, Bayesian, Hannan and Quinn, R-square, adjusted R-square, Mallows Cp, Significant Levels)* for the three types of paths that can be taken (*forward selection, bac*kward elimination, or mixed regression). We obtained 27 selection possibilities and chose the best model with the highest frequency. In the second step of the modeling process, a regularization regression analysis is applied. We implement the novel method which is called 'induced smoothed lasso' proposed by Cilluffo (Cilluffo et al., 2019) to obtain the residuals.

The novel method enabled us to compare the regularized regression results with the traditional stepwise approach to teste our hypotheses. After selecting the variables in the first two steps, we implement the lag models with the most relevant variables in the third stage of the modeling process. In the third step, a temporal dimension in our study to understand the long-term effects of the independent variables between the GNI are included. After applying these three-step modeling processes, we have selected the most relevant variables for each transformation and verified our hypotheses using all the available data. The different types of experiments are conducted to verify the predictive capacity of our models. Finally, we analyze the future scenarios based on the most recent values of each country's selected variables.

An extensive range of tests can be conducted for different models. Including traditional autocorrelation tests such as: the Box-Pierce and the Ljung-Box tests, the Augmented Dickey-Fuller, Phillips-Perron, the Kwiatkowski-Phillips-Schmidt-Shin stationarity test, and the multicollinearity and Granger Causality tests. We follow the methodology proposed in Peña and Slate (2006), which the Global Validation of Linear Models Assumptions was implemented. This approach involves four tests; Skewness, Kurtosis, Linearity, and Heteroscedasticity. These tests are complemented with nine additional tests to substantiate the autocorrelation and correlation of the residuals with the dependent variable. We also performed additional tests for heteroscedasticity and linearity.

Due to the rigorous analysis engaged in this paper, it is not only to propose an extensive statistical procedures in assessing a wide variety of economics and demographics indicators of China and India, but also to identify the indicators which have the most significant impacts on the GNI for China and India during the period in the study. The regression methods in the paper will verify the hypotheses and predictions at different time horizons. It is a new perspective and novel methodologies to compare the structures and increasing potentials of the GNI of China and India.

The outline of this paper is as follows. Section 2 presents the theoretical frameworks and specific methodologies. Section 3 briefly introduces the definitions, contexts, historical evolution and descriptive analysis of all the indicators and methodologies. Section 4 provides a further application to corroborate each of the models' predictive capacity in different scenarios. The results of the experiments are also presented in this section. Section 5 summarizes the main findings.

2. Theoretical frameworks

In this section, a brief overview of the most relevant technical elements and methodologies for each type of analysis and test are introduced.

2.1. Criterions for model selection

To compare the different models adjusted to the same sample, besides of goodness-of-fit criteria, we also consider the variable selection strategies that automatically or semi-automatically construct subsets of all possible models of the best fit. Information criteria indicators measure the complexity of the model used as a linear or approximately linear function of the number of its parameters. Criteria such as Mallows Cp, AIC, Corrected form of Akaike, Bayesian, and Schwarz are based on this principle. If only the R-square is considered, it can lead to over-parameterized models. The adjusted R squared tells us what percentage of the dependent variable's variation is collectively explained by all the independent variables. In Table 1,a summary of the proposed information criterion to use and the principal reference for this in-depth review is presented.

It is proposed to consider all the criterias seen above in the three different selection strategies of stepwise regression analysis.

Table 1
Information criteria to use.

Information Criterion	Notation	Reference
Akaike	AIC	Akaike (1974)
Corrected form of Akaike	AICc	Hurvich and Tsai (1993)
Bayesian	BIC	Schwarz (1978)
Hannan and Quinn	HQ	Hannan and Quinn (1979)
R-square	Rsq	Hocking (1976)
adjusted R-square	adjRsq	Hocking (1976)
Mallows Cp	CP	Mallows (1973)
Significant Levels	SL	Elliott et al. (2006)

2.2. Multicollinearity

When there are highly correlated variables, collinearity or multicollinearity can occur. Basically, this is where some of the regression models' assumptions may not be fulfilled, in some cases leading to spurious regressions despite being the “best” among all the options. Therefore, the verification of this type of model's hypotheses plays an essential role before reaching any definitive conclusion, especially starting with the review of the correlation between the variables to be used.

The basics of overall multicollinearity diagnostics measures proposed in Imdadullah et al. (2016); Farrar and Glauber (1967); Kovács et al. (2005); Chatterjee and Price (1977); Theil (1950) are presented in Table 2.

A specific measure in the case of regression is the variance-inflation factors proposed by Fox et al. in 1992 (Fox and Monette, 1992). The variance-inflation is a useful diagnostic because it directly specifies the damage imposed by collinearity on the estimation accuracy.

2.3. Assumptions of linear regression

In linear modeling, there are several proposed assumptions including *linearity, stationaries variables*, *heteroskedasticity, no correlation with residuals, autocorrelation*, *normality of the typified residual with zero mean.*

Table 3 summarizes the tests for the validation of each of the assumptions or dimensions mentioned above.

As an alternative for verification in a summarized manner in Farrar and Glauber (1967), the Global Validation of Linear Model Assumptions are proposed, where the dimensions of Skewness, Kurtosis, Linearity, Heteroscedasticity are established. It is also proposed to use this type of test to corroborate some of the models' results.

2.4. Granger Causality

According to Granger, if a lagged variable is correlated with another variable's future values, one variable is said to cause the other. However, the previously mentioned reason is incorrect to claim causality. It is possible that a lagged variable spuriously correlates with another variable only because it is a leading indicator and not because there is genuinely causality (especially if they are non-stationary time series). However, this is a limitation that must be supplied with reason and the literature. In any case, what can be said is the opposite; if there is no such correlation, then the lagged variable “does not cause” the other.

2.5. Stepwise regression analysis

Traditionally, regression models have been used to study the relationship between economic variables. When there is a large number of explanatory variables, one of the primary methods for variable selection is the stepwise selection method. The best model can be selected using three different methods: forward selection, backward elimination, or mixed regression. We present a provisional regression equation for each method, which includes some variables (regressors included) and others not included (regressors absent).

Table 2
Overall multicollinearity diagnostics measures.

Test	Reference
Determinant of normalized correlation matrix |X'X |:	Imdadullah et al. (2016)
Farrar Chi-Square	Farrar and Glauber (1967)
Red Indicator	Kovács et al. (2005)
Sum of Lambda Inverse	Chatterjee and Price (1977)
Theil's Method	Theil (1950)
Condition Number	Imdadullah et al. (2016)

Table 3
Tests for assumptions testing of the models by dimension.

Dimension	Test	Notation	Reference
Autocorrelation	Box – Pierce	Bx-P	Box and Pierce (1970)
	Ljung – Box	L-B	Ljung and Box (1978)
	Breusch - Godfrey	B-G	Breusch (1978); Godfrey (1978)
	Durbin - Watson	D-W	Durbin and Watson (1950)
Correlation	Pearson's product-moment correlation coefficient	Cor	Soper et al. (1917)
Heteroscedasticity	Breusch - Pagan	B-P	Breusch and Pagan (1979)
	Goldfeld - Quandt	G-Q	Goldfeld and Quandt (1965)
	Harrison - McCabe	H-M	Harrison and McCabe (1979)
Linearity	Harvey - Collier	H-C	Harvey and Collier (1977)
	Rainbow	Rai	Utts (1982)
	Ramsey Regression Equation Specification Error Test	Ram	Ramsey (1969)
Stationarity	Augmented Dickey-Fuller		Fuller (1976)
	Phillips - Perron		Phillips and Perron (1988)
	Kwiatkowski-Phillips-Schmidt-Shin		Kwiatkowski et al. (1992)

For the forward selection procedure, the regression equation does not include a regressor. The variables are then introduced one by one to decrease the criterion. The backward regression procedure proceeds analogously, but it begins with an equation that includes all the regressors. They are excluded one by one, as long as the increase in the criterions caused by the exclusion is not excessive. In the mixed procedure, the inclusion and exclusion of variables in the regression line are alternated. Thus, allowing a variable that is initially included to be later discarded when the presence of another variable renders its contribution to reducing the criterion insignificant.

2.6. *Regularization regression analysis*

To overcome the problem of collinearity, regularization methods have been suggested for model selection to avoid overfitting in predictive techniques. Within the regularized regression methods, we have three mains options: Ridge, Lasso, and Elastic Net.

2.6.1. *Ridge*

This technique was initially recommended by Hoerl and Kennard (Hoerl, 1970) in 1970 as a method to avoid the adverse effects of the collinearity problem in a linear model estimated by least squares. It is very similar to least squares, except that the coefficients are estimated by minimizing a different amount by including a penalty term in the objective function and contracting the regression coefficients. The higher the penalty term, the greater the contraction of the coefficients.

Unlike the least square method, Ridge regression produces a different set for each value of the penalty and not a single vector of estimated coefficients; if the penalty becomes zero, ordinary least squares are formed.

One of the disadvantages of this method is that it contracts all the coefficients towards zero, but does not nullify any of them. Therefore, there is no selection of variables, remaining all the variables in the model. This fact is inconvenient in those studies that have a high number of explanatory or predictive variables. The latter problem is avoided by proposing Lasso regression.

2.6.2. *Lasso*

Motivated by the objective of finding a linear regression technique that, by contracting the coefficients carry out a variable selection, Tibshirani (1996) in 1996, proposed the Lasso technique (least absolute shrinkage and selection operator). It is a regularized linear regression technique, like Ridge, but with a slight difference in a penalty that has significant consequences.

In particular, starting from a specific value of the complexity parameter, the Lasso estimator produces null estimates for some coefficients and non-null for others, with which Lasso performs a kind of selection of variables continuously. With this, Lasso reduces the variability of the estimates by reducing the coefficients and, at the same time, produces interpretable models by reducing some coefficients to zero.

2.6.3. *Elastic net*

Zou and Hastie (2005) in 2005 suggested a variable selection and regularization technique acknowledged as Elastic Net. The Elastic Net retains the advantages of Lasso, automatically makes the variable selection and continuous contraction, and at the same time overcomes some of its limitations. With this method, groups of correlated variables can be selected. To accomplish this, two penalty terms are included.

2.7. Distributed lag regressions analysis

2.7.1. Distributed lag

Dynamic economic models typically characterize the economy's path in terms of its long-run equilibrium (steady-state) (Hashem, 2015) and involve modeling expectations, learning, and adjustment costs. There is a variety of dynamic specifications used in applied time series econometrics. Most linear distributed lag models discussed in the literature belong to the rational distributed lag models category. Early examples of these types of models include the polynomial and geometric distributed lag models. The rational distributed lag model is also called the autoregressive distributed lag (ARDL) model.

2.7.2. Cointegration and error correction model

Two or more variables are said to be cointegrated if they are individually integrated (or have a random walk component) (Hashem, 2015), but there are linear combinations of stationary combinations. An error correction model (ECM) is a type of multiple time series model most commonly used for data where the variables have a long-run stochastic trend (Engle and Granger, 1987; Banerjee et al., 1998), also known as cointegration. ECMs are a theoretically-driven method, convenient for approximating both short-term and long-term properties of one set of time series data on another. The term error-correction relays to the fact that the last period's deviation from a long-run equilibrium, the error, influences its short-run dynamics. Thus, ECMs estimate the speed at which a dependent variable returns to equilibrium after a change in other variables.

2.7.3. AutoRegressive Distributed Lag (ARDL)

In general, the regression equation only gives us a short-term relationship between the variables. It does not provide evidence about the long-term behavior of the parameters in the model. A problem is created since researchers are most concerned about long-term relationships between the variables of interest. With the specification of the ECM, we can incorporate both long-term and short-term information. 2003 Nobel Prize-winning economists Engle and Granger recommended a two-step procedure. The first step consists of estimating the long-term multipliers in a static regression equation (the cointegration equation) and testing for cointegration. The dynamics are specified in the second step, where an error-correction model is formulated and estimated using the residuals from the first step as equilibrium errors.

The ECM is based upon the ordinary least square coefficient of the lagged dependent variable in an autoregressive distributed lag model augmented with the regressors' leads. This test was expanded by Pesaran, Shin, and Smith (Pesaran et al., 2001) by introducing a different way of testing the relationship between a dependent variable and a set of regressors when it is not known with certainty whether the underlying regressors are trend- or first-difference stationary. They also used the traditional ARDL approach to analyze the long-term relationships when the underlying variables are cointegrated in the first order. Following Pesaran and Shin (1997) (Pesaran, 1997), the study provides a simple framework for testing the existence of a single level relationship between the regressor and regressors. The framework for testing is used when it is not known with certainty whether the regressors are purely integrated with the original values or order zero I(0), integrated into this first lag or first order I(1), or mutually cointegrated.

3. Variables analysis and selection process

In this section, a descriptive, stepwise, penalization, and distributed lag regression analysis will be carried out sequentially to find the models, checking if the models meet all the theoretical elements seen in the previous section.

3.1. Variables definition

The primary source for India's variables is the Directorate General of Commercial Intelligence & Statistics (DGCI&S). In the case of China, the National Bureau of Statistics. Table 4 presents the variables, its denomination, and a brief conceptualization of each.

3.2. Correlations and visual inspection of the variables

The alignment of the variables with appropriate definitions in Table 4 provide a basis to present a visual comparison of the fourteen (14) variables assessed in China and India. The visual comparisons are offered according to the similar chronological arrangement used in Table 4.

Fig. 1 shows that the GNI of the two countries are quite similar to each other from 1950 to 2020.

From Fig. 2, there is almost no imports and exports before the reform and opening-up policy of China, so the India's imports and exports are much hinger at that moment. China's total imports and exports climbed from $18 million in 1980 to $431.5 billion in 2019, an average annual growth rate of 26.1 percent, contributing to an all-around opening up and especially exceeding India in the recent decade.

In the case of India, inflows by foreign institutional investors and a weak dollar have pushed up forex reserves. At the mean time, China's foreign currency reserves have been already the world's largest and over 7 times than India. China's reserves have ballooned as the central bank buys up dollars generated from its huge trade and influx of foreign investment, as we can see in Fig. 3.

Table 4
Notation and definition of variables to use.

Indicator or Variable	Variable Name	Variable Definition
Foreign Exchange Reserves	RES	The money or other assets held by a central bank or other monetary authorities to pay its liabilities if needed. E.g., the currency issued by the central bank, as well as the various bank reserves deposited with the central bank by the government and other financial institutions (Eichengreen, 2011).
Trade Balance	TRAD	It is the difference between the monetary value of a nation's exports and imports over a certain period. China, a high-growth economy, has tended to trade surpluses. (O'Sullivan and Sheffrin, 2003)
National Government Expenditure	GV.EX	The expenditure of all government consumption, investment, and transfer payments (Barro and Grilli, 1994). This spending can be financed by government borrowing or taxes. Changes in government spending is a significant component of fiscal policy used to stabilize the macroeconomic business cycle.
Gross National Income	GNI	The total domestic and foreign output claimed by residents of a country, consisting of gross domestic product (GDP), plus factor incomes earned by foreign residents, minus income earned by nonresidents' domestic economy. (Todaro and Smith, 2011).
Energy Production	EN.PR	The total energy generated every year by all the sources in kilowatt-hours (kWh), which is a 1-kW amount of power delivered in 1 h. One of the most common usages is billing the amount of energy delivered by electric utilities. The abbreviation is “kWh. “Megawatt-hours is one million watts amount of power delivered in 1 h. Energy is one particular input to the production function since it is non-substitutable. It is an essential element for economic production and, therefore, economic growth (Stern, 2010).
Production of Fish	FISH	The total of the output of Seawater Aquatic Products, in 10000 tons. Given its extensive coastlines, ranking China as the tenth and India as the fifteenth country with the most extensive coastlines globally (Coastline of the world fa), fishing is an essential element of production and economic growth in both countries.
Total Population	POPU	All nationals present in, or temporarily absent from a country, and aliens permanently settled in a country. This indicator shows the number of people that usually live in an area. Growth rates are the annual changes in population resulting from births, deaths, and net migration.
Literacy Rate	LIT	The total number of literate persons in a given age group expressed as a percentage of the total population in that age group. The Adult Literacy Rate measures literacy among persons aged 15 years and above, and the Youth Literacy Rate measures literacy among persons aged 15 to 24 years. Previous studies show how the investment in human capital, that is, in education and skills training, is three times as important to economic growth over the long run as an investment in physical capital, such as machinery and equipment (Burnett, 2006)
Median age	MD.AG	Median age divides a population into two numerically equal groups; that is, half the people are younger than this age, and half are older.
Death Rate	DEAT	A measure of the number of deaths (in general, or due to a specific cause) in a particular population, scaled to that population's size, per unit of time. The Mortality Rate is typically expressed in units of deaths per 1,000 individuals per year; thus, a Mortality Rate of 9.5 (out of 1,000) in a population of 1,000 would mean 9.5 deaths per year in that entire population or 0.95% out of the total. (Crude death rate (per 1,0, 2008)
Dependency Ratio	DP.RA	An age-population ratio of those typically not in the labor force (the dependent part ages 0 to 14 and 65+) and those typically in the labor force (the productive part ages 15 to 64). It is used to measure the pressure on the productive population
Fertility rate	FERT	A measure of the fertility of an imaginary woman who passes through her reproductive life subject to all the age-specific fertility rates for ages 15–49 that were recorded for a given population in a given year. This rate is the number of children a woman would have if she was subject to prevailing fertility rates at all ages from a single given year and survives throughout all her childbearing years.
Life expectancy	LIF.EX	The average number of years a person born in a given country is expected to live if Mortality Rates at each age were to remain steady in the future.
Birth Rate	BIRTH	The total number of live births per 1,000 in a population in a year or period (technically, births/population rate).

Government Spending refers to public expenditure on goods and services and is a major component of the GDP. Government spending policies like setting up budget targets, adjusting taxation, increasing public expenditure and public works are very effective tools in influencing economic growth. Fig. 4 shows that the national government expenditure are quite similar of the two countries during the period, but China has a rapid growth in recently and exceeded India.

China and India are aggressively investing in energy production. As we can see in Fig. 5, since 1950s to 2015, China's energy productivity is always greater than India and the difference is enlarged dramatically in recent years because of the rapid economic growth. China and India together account for almost two-thirds of the world's coal consumption.

Together, China and India account for 37 percent of the world's population. Both of the countries total fertility rate (TFR) has been declining and got the impressive changes. But China has lower TFP and higher Literacy rate while India has the lower levels of education and women's status, less effective governance, and more entrenched traditional values and customs, this is shown in Fig. 6.

The seawater aquatic production can reflect the consumption. The increasing population of middle class in China and India impact the demand of fish and other seawater aquatic. Fig. 7 shows that India could be the world's largest consumer market, surpassing China in the future because of the stable increasing population of the middle class.

The fertility rate both China and India are gradually declined. But China has a dramatically decrease after the one-child policy at the end of 1970s. One of the important theory of the neoclassical macroeconomics is Solow model which emphasizes that the economic growth of any of the country can be achieved with the help of these three input factors labor, capital, and technological growth. Because of the decreasing fertility rate in Chian, it is obviously that the economic growth in China in the next 20 years will significant influenced by the low fertility rate now. In the historical data, we can see on Fig. 8 that the lower

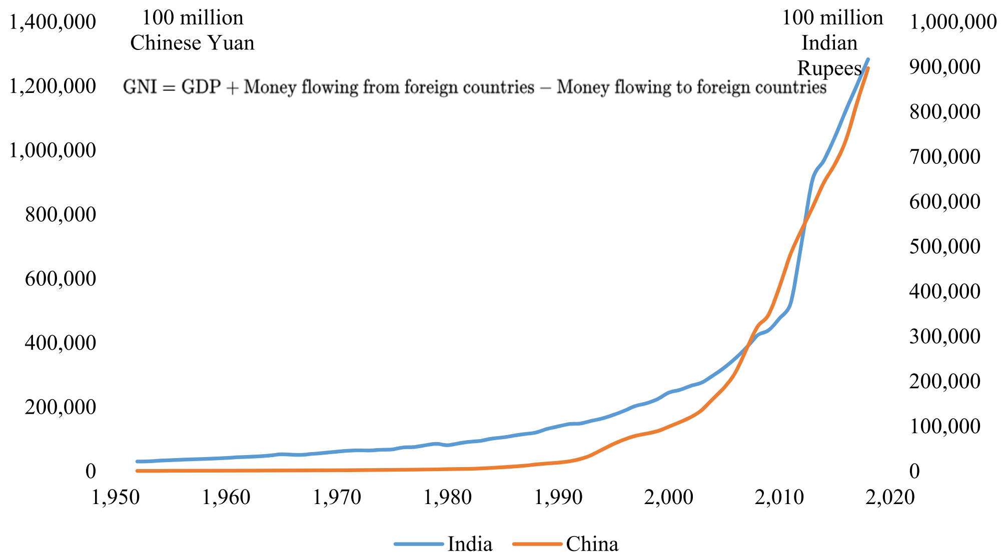

Fig. 1. Gross National Income (100 million yuan).

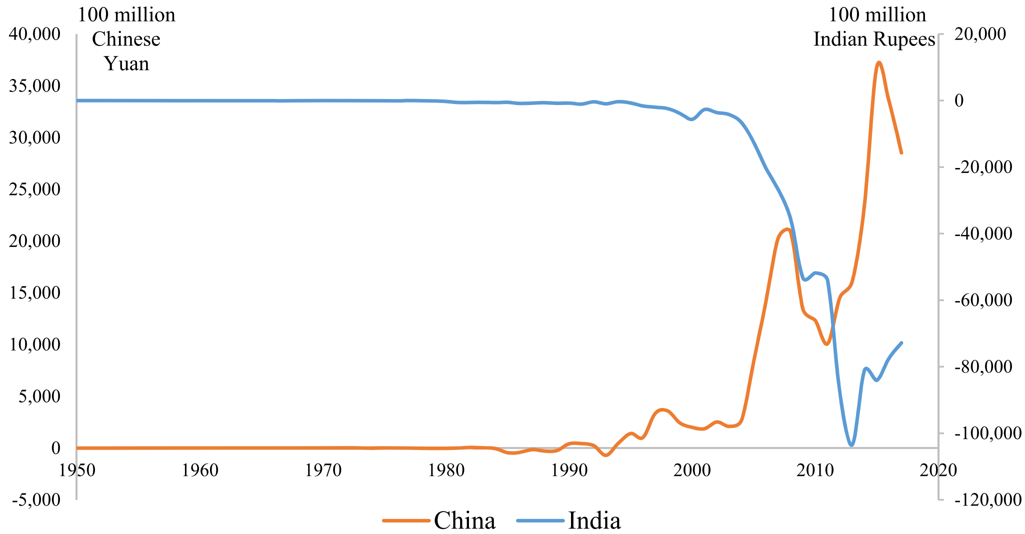

Fig. 2. Balance of Total Imports and Exports, China vs. India 1950–2017.

fertility rate in 1980s to 1990s already influenced the labor supply in China. On the contrary, India also has decreasing fertility rate and only a little bit higher than China now but the decreasing are much stable comparing to China. We are comparing the demographic structures to evaluate the effections of the one-child policy in China.

The life expectancy of China in Fig. 9 is always a little bit higher than in India. Only in 1960 to 1970, because of the great chinese famine as we can see on Fig. 11, but the Chinese life expectancy was quite close to Indian.

The birth rate of China has been changing sharply influence by Population policy intervention, especially during the 1950s to 1970s, as we can see in Fig. 10. After the one-child policy, China's birth rate keep on decreasing while India's birth rate always stable and influence much by the economic growth.

China and India are the two most populous countries in the world, with China home to about 1.44 billion people and India to 1.38 billion in 2020. China and India together account for about 36% of total world population and 67% of Asia population. Fig. 12 shows that India will exceed China become the most populous country in the future.

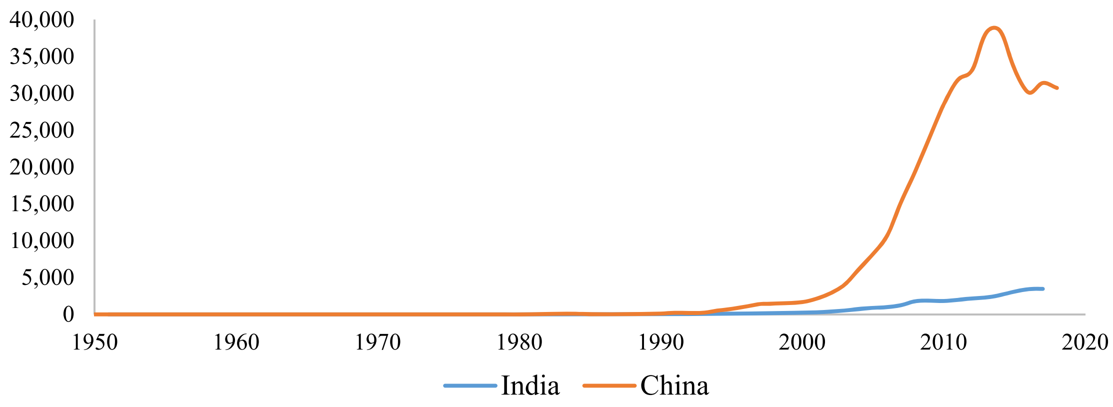

Fig. 3. Foreign Exchange Reserves India and China in USD (100 million).

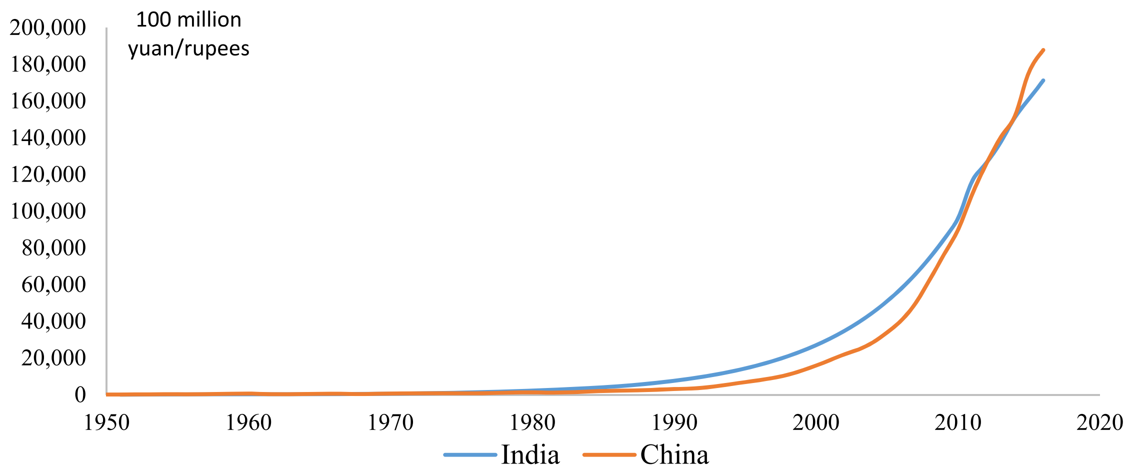

Fig. 4. National Government Expenditure (100 million yuan/rupees).

As the impacts of China's one-child policy are felt, and India's population continues to grow, the two nations' demographics are diverging. We can see in Fig. 13 how the median age of Chinese is the same as with the USA and arrives at 38.4 years old while the Indian are only 28.4 years old. The demographics structure are quite influence the economic growth by the theory of Solow in the neoclassical economics.

The child dependency ratio is expected to remain stable in China but the old-age dependency ratio is rising steadily after the one-child policy, as we can see in Fig. 14.

While it's clear that the world's two most populous countries have some key similarities, they are both on very different demographic paths at the moment. China's population has plateaued, and will eventually decline over the remainder of the 21st century. There is plenty of room to grow economically, but the weight of an aging population will create additional social and economic pressures, in Fig. 16 we can see the change by big-range of ages in China, decreasing the youngest groups rapidly. On the other hand, India is following a more traditional demographic path, as long as it is uninterrupted by drastic policy decisions, we can see this soft change by big-range of ages in Fig. 15. The country will likely top out at 1.6–1.7 billion people and will have enough supply of labor, before it begins to experience the typical demographic transition already experienced by more developed economies in North America, Europe, and Japan. As general reference Fig. 17 shows the trend by big-range of ages for the rest of the World, in general, India has an evolution more similar to the rest of the world. Concluding this descriptive analysis, and checking the cross-correlation values, on Fig. 18 we can see the how the variables closest to the GNI are positive related to economic factors, such as reserves and government spending, then we have variables

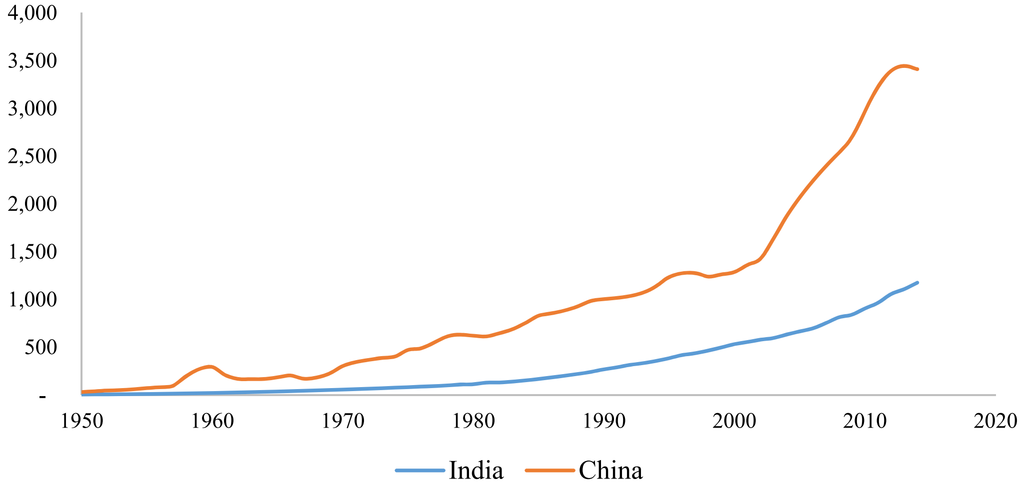

Fig. 5. Energy generated (Billion KWH).

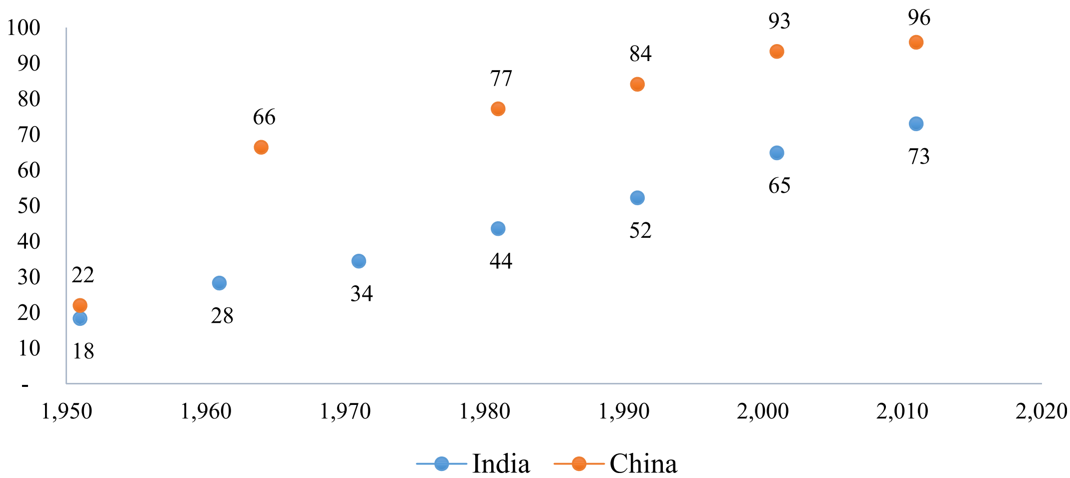

Fig. 6. Literacy rate, national population census (%).

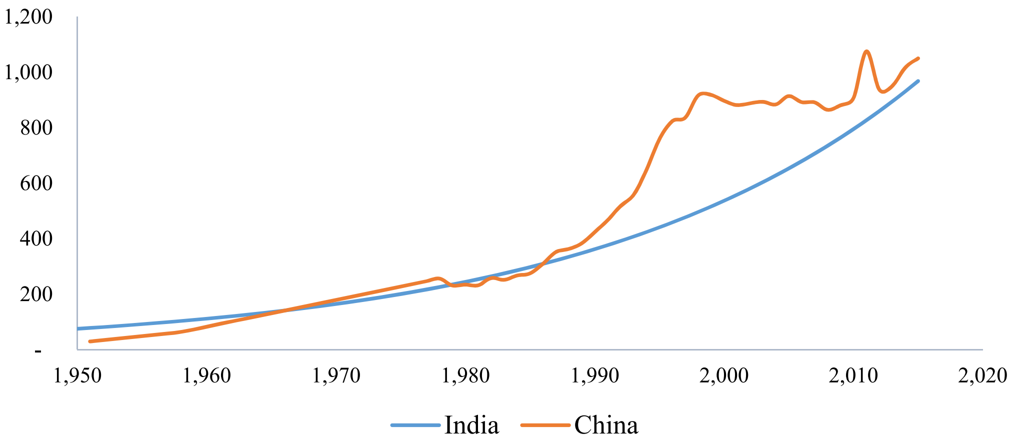

Fig. 7. The output of Seawater Aquatic Products, Fish (10000 tons).

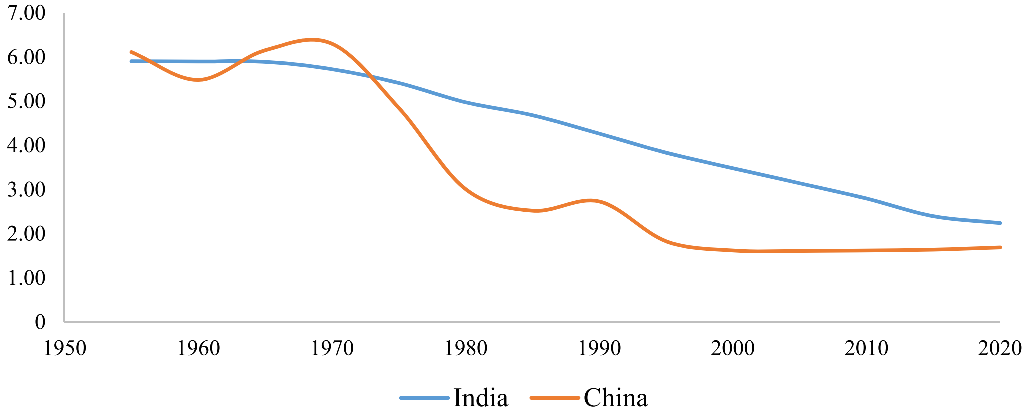

Fig. 8. Total fertility (live births per woman).

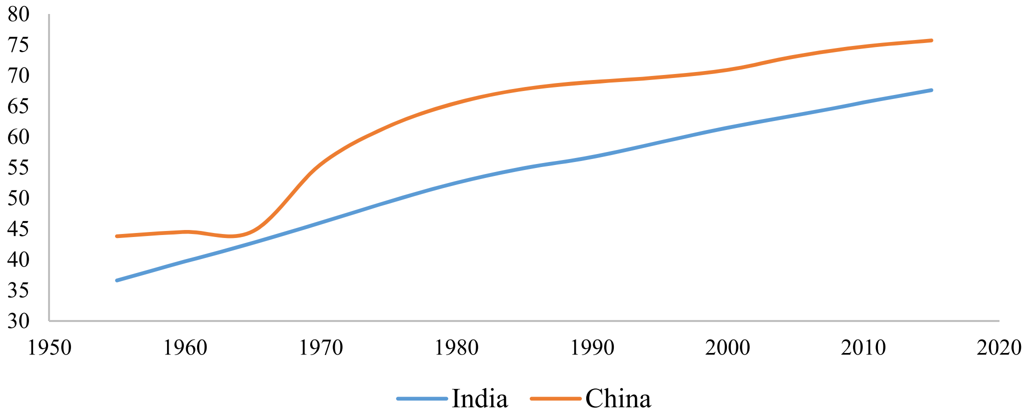

Fig. 9. Life expectancy in years.

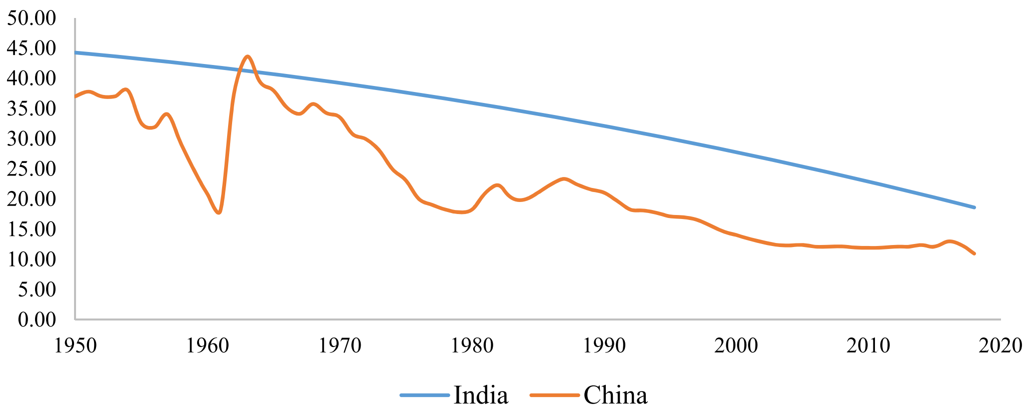

Fig. 10. Birth rate.

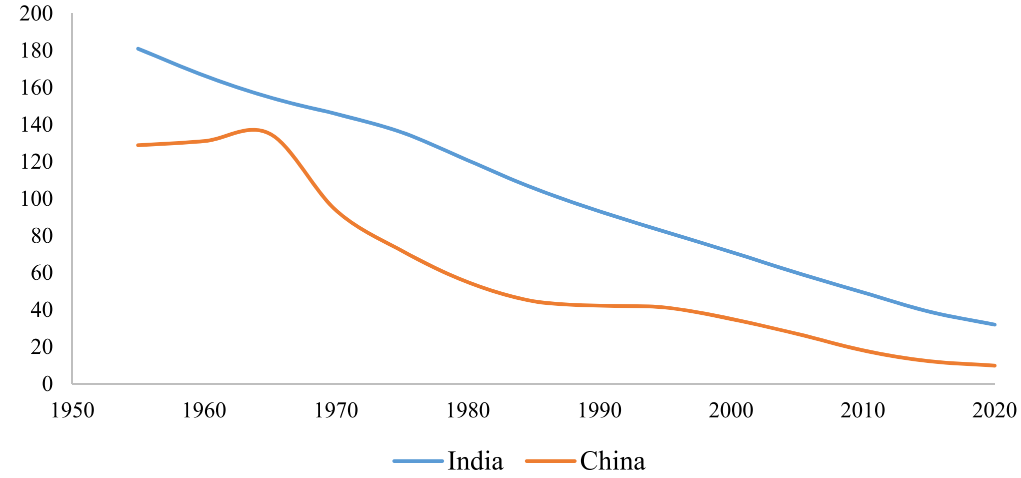

Fig. 11. Death rate.

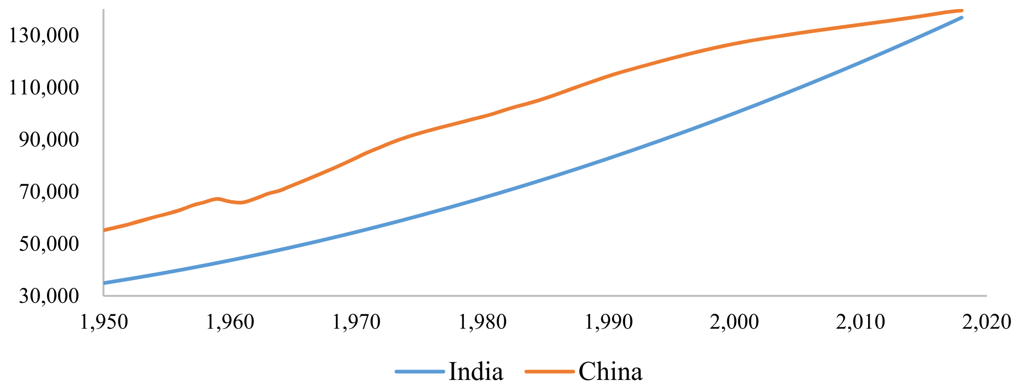

Fig. 12. Total Population (year-end) (10000 persons).

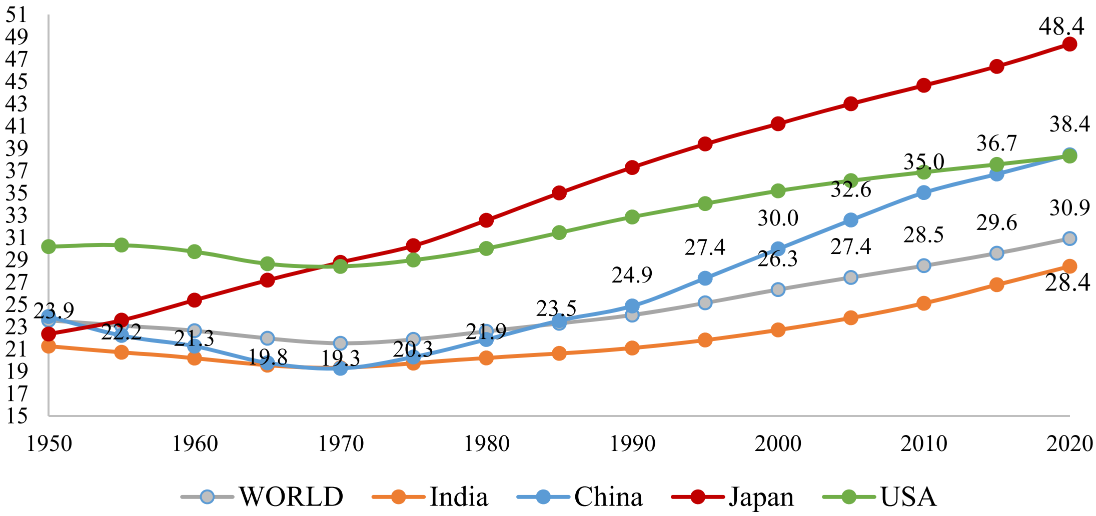

Fig. 13. Median age.

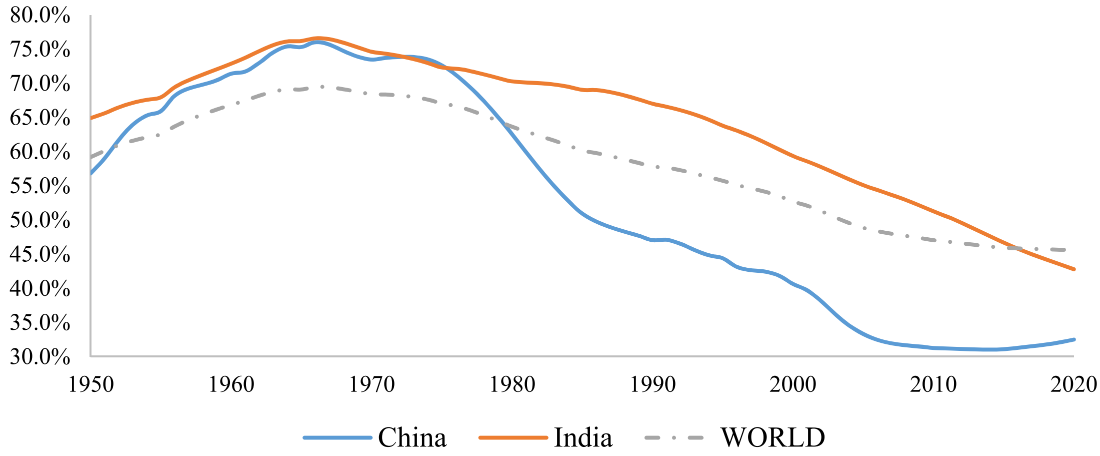

Fig. 14. Total dependency ratio (under 15 + after 70)/(between 15 and 70).

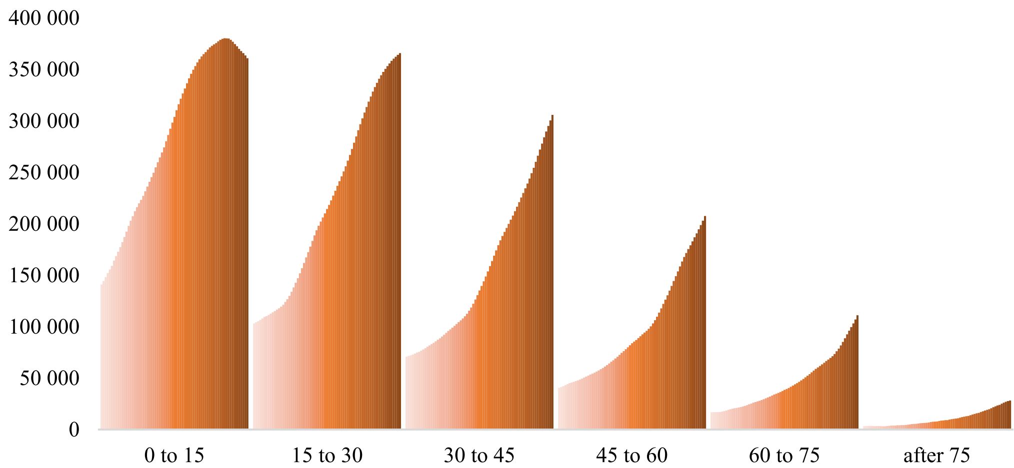

Fig. 15. Evolution in the time of population by a big-range of ages, India.

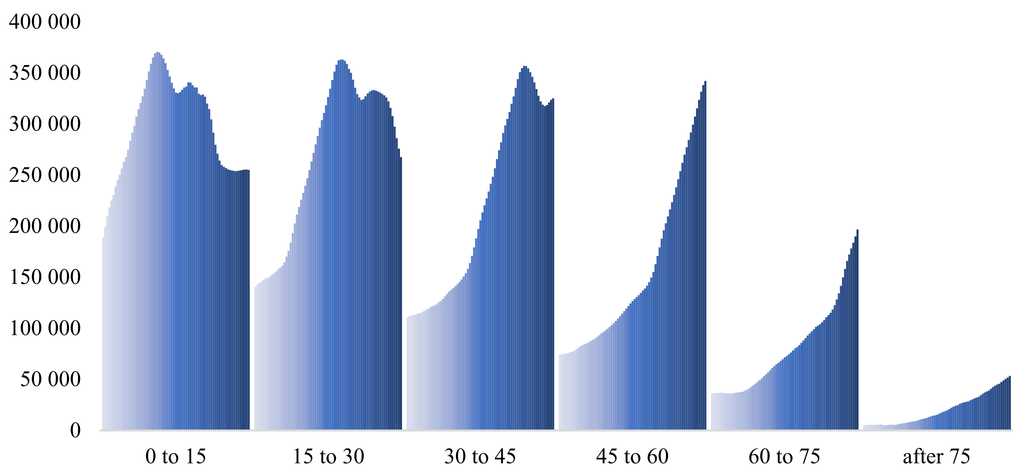

Fig. 16. Evolution in the time of population by a big-range of ages, China.

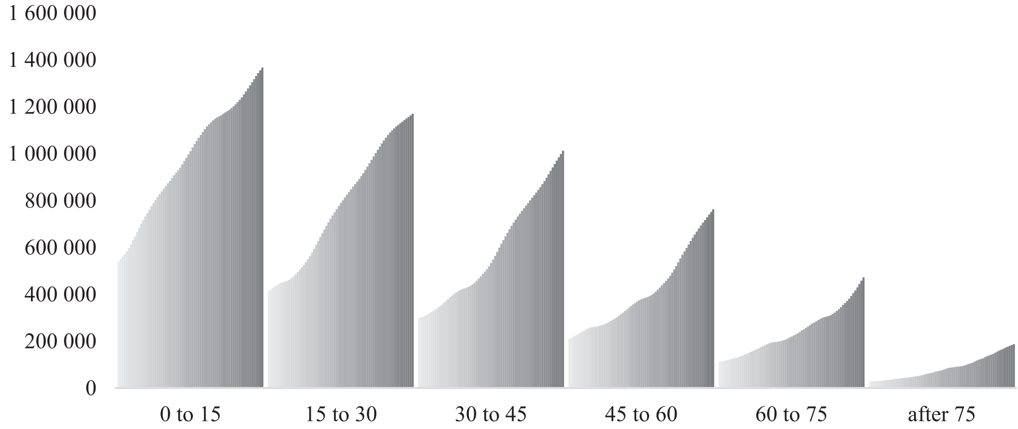

Fig. 17. Evolution in the time of population by a big-range of ages, rest of the world.

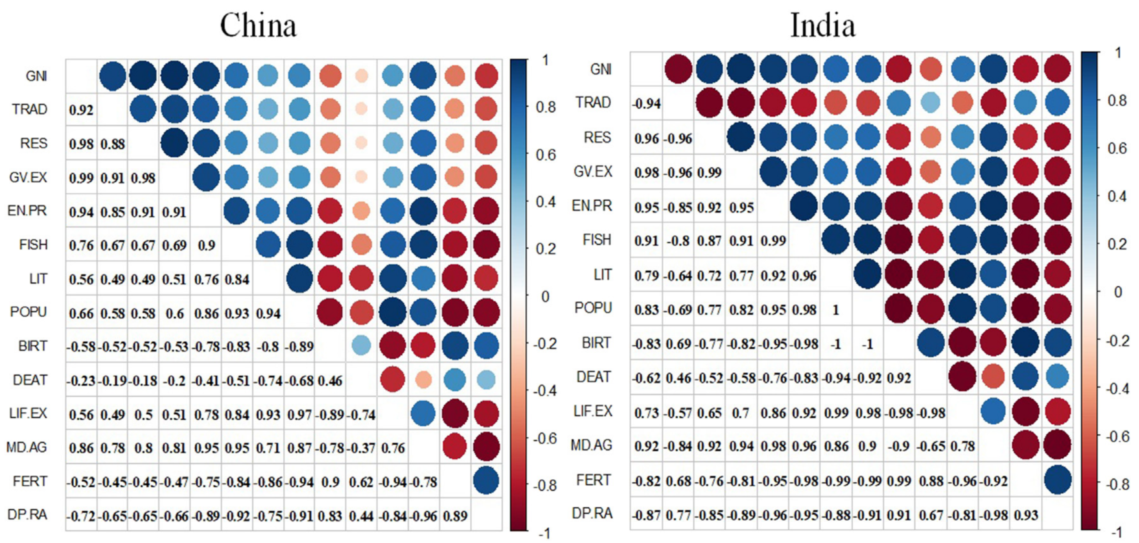

Fig. 18. Cross-correlation for both countries and all the variables.

associated with production such as energy production and fishing, finally all variables associated with demographic aspects based on the literacy rate, usually with more negative relation with the GNI.

3.2.1. Context, historical evolution, and descriptive analysis of indicators

After the founding of the current socialist government, the PRC established five-year plans as the characteristic of the socialist economies. The change in demographic indicators was an extreme and initiating abrupt increase in the birth rate during the first two plans. Subsequently, there was an increase in the mortality rate due to food supply problems(MacFarquhar, 1987), causing birth control policies to be established (Potts, 2006). The study period (1952–2015) for this paper covers the first twelve five-year plans; beginning, with the five-year plan created by Mao Zedong, the founder of the republic, to the former president Hu Jintao.

The 3rd Plenary Session of the 11th Central Committee of the Communist Party of China in 1978 marked the country's economy's opening and modernization. This new direction for China first emerged in the ninth five-year plan implemented from 1996 to 2000. The first medium-term plan was carried out under a socialist market economy and consisted of a cross-century strategy. In 2001 there was a substantial increase in Chinese energy production, which coincided with China's formal membership to the World Trade Organization.[1] The government spending, reserves, and energy stand out as the variables

[1] http://www.china.org.cn/english/49058.htm.

Table 5
Results of the overall multicollinearity diagnostics measures by country.

Country	China		India	
Test	Value	Result	Value	Result
Determinant |X'X|:	0	T	0	T
Farrar Chi-Square:	2.10E+03	T	5.36E+03	T
Red Indicator:	7.45E-01	T	8.73E-01	T
Sum of Lambda Inverse:	2.28E+03	T	2.36E+07	T
Theil's Method:	5.32E-01	T	1.03E+00	T
Condition Number:	5.32E+02	T	1.01E+05	T

with the highest correlation to the GNI. Their coefficients are 0.99, 0.98, and 0.94, respectively, showing the Chinese government's strong economic development and growth efforts.

The most populated democratic country in the world is India. However, after the triumph of independence in 1950 came unprecedented tragedies such as civil violence (Metcalf and Metcalf, 2006). In the early 1950s, India adopted a mixed economic growth model to balance the role of the market with that of the state, and like China, India established five-year plans. The first Prime Minister to establish five-year plans was Jawaharlal Nehru. Subsequently, incumbent Prime Ministers such as Lal Bahadur, Indira Gandhi, and Rajiv Gandhi also generated five-year plans during their tenure. The Nehru–Gandhi family and the Indian National Congress party ruled India for close to forty years, since its foundation until 1989.

In the 1990s, India experienced a period of political destabilization. In the 90s, India had five different presidents, with some presidents serving one year in office. For example, Vishwanath Pratap Singh was in office from 1989 to 1990 and resigned due to religious and caste issues. However, under President Pamulaparthi Venkata Narasimha Rao, the economy was formally liberalized in 1991 by opening the country to foreign investment. With this impulse, India became an official member of the World Trade Organization in 1995. Winning the elections in 1996, Atal Bihari Vajpayee, during his presidency, deepened India's economic reforms by increasing taxes privatization[2]with the ninth five-year plan. Unlike China, the evolution of India's indicators is smoother and without significant changes throughout the study period. An almost linear trend is noted in the case of demographic indicators, which can be attributed to the lack of state. Economic growth becomes exponential after the reforms of the ninth five-year plan. However, the balance of total imports and exports is the opposite of China. Whereas India deepens its deficits, China experiences a large trade surplus.

3.2.2. Multicollinearity results

If the overall multicollinearity diagnostics were applied to the GNI of each country as the variable of response and the rest of the variables as regressors, the results in Table 5 would be achieved. The column “Results” indicates with “T,”which test gives true multicollinearity.

In all the measures for both countries, a high degree of collinearity was detected. High colinearity can be a problem in the modeling process; therefore, the initial way to solve this situation is to transform the variables.

3.2.3. Variables transformation

From the time-series graphs, an exponential type of behavior such as high collinearity, was seen in many of the variables, especially from the variable under study (GNI). Therefore, it is proposed to transform the variables using First difference, logarithmic and Growth Rate. For the application of the First difference, there are two approaches. The use of the approaches is due to negative values and differences between the variables. Hence, a re-escalation is made of all the variables between 1 and 100 in the time-series plots in Figs. 19 and 20. There it is observed that as the transformations are applied, the series are more stabilized, and their trend is eliminated.

3.2.4. Autocorrelation and stationarity test

The results to verify the autocorrelation will be done through the Box-Pierce and Ljung-Box tests. Regarding stationarity, the Augmented Dickey-Fuller, Phillips-Perron, and Kwiatkowski-Phillips-Schmidt-Shin tests are implemented. All of this in three different ways, with drift and trend, only with train or without any of them. Since the behavior of the evolution of the tests turned out to be similar between the variables, only the results of the p-values corresponding to the tests for the GNI will be shown. The results are presented in Table 6.

3.2.5. Variance-inflation factors

For calculating the variance-inflation factors, an initial linear regression of the GNI as a function of all the other variables for the original series, and each type of transformation was applied. The VIF coefficients can be interpreted as the percentage

[2] https://niti.gov.in/planningcommission.gov.in/docs/plans/planrel/fiveyr/9th/vol1/v1c1-2.htm.

Table 6
Results of the autocorrelation and stationarity test for original and transformed series by country.

Country/Series		China					India				
Test		Orig	Log	Log Diff	Gro	Gro Diff	Orig	Log	Log Diff	Gro	Gro Diff
Box-Pierce		0	0	0	0	0.05	0	0	0.46	0	0.07
Ljung-Box		0	0	0	0	0.04	0	0	0.44	0	0.06
Augmented Dickey-Fuller	**L**	0.99	0.99	0.21	0.21	0.01	0.99	0.99	0.01	0.01	0.01
	D	0.99	0.99	0.28	0.28	0.01	0.99	0.99	0.01	0.01	0.01
	D-T	0.99	0.85	0.24	0.24	0.01	0.99	0.99	0.01	0.01	0.01
Phillips-Perron	**L**	0.99	0.99	0.3	0.3	0.01	0.99	0.98	0.01	0.01	0.01
	D	0.99	0.99	0.25	0.25	0.01	0.99	0.98	0.01	0.01	0.01
	D-T	0.99	0.98	0.08	0.08	0.01	0.99	0.99	0.01	0.01	0.01
Kwiatkowski-Phillips-Schmidt-Shin	**L**	0.1	0.02	0.1	0.1	0.1	0.1	0.02	0.02	0.08	0.1
	D	0.1	0.05	0.1	0.1	0.1	0.1	0.1	0.1	0.03	0.1
	D-T	0.09	0.04	0.1	0.1	0.1	0.1	0.01	0.02	0.1	0.1

Table 7
Variance-inflation factors of all the variables by country.

Country	China					India				
Variable	Orig	Log	Log Diff	Gro	Gro Diff	Orig	Log	Log Diff	Gro	Gro Diff
POPU	9E+02	4E+02	14.1	14.1	1.9	1E+07	2E+05	3E+03	1E+03	2.3
BIRT	15.4	31.5	1.4	1.4	1.4	6E+06	1E+04	5E+02	3E+02	1.7
LIF.EX	2E+02	70.6	16.9	16.9	1.1	5E+06	1E+04	8E+02	5E+02	2.5
MD.AG	3E+02	32.4	2.3	2.3	1.3	6E+05	1E+02	3.9	1E+03	1.3
DEAT	5.6	10.5	1.2	1.2	1.8	6E+05	6E+04	3E+02	147.1	1.4
FISH	83.2	91.5	2.4	2.4	1.2	2E+05	2E+05	9E+02	1.1	1.3
FERT	34.8	51.6	1.3	1.3	1.0	2E+05	4E+04	1E+03	86.9	1.6
LIT	1E+02	30.1	22.1	22.1	1.3	3E+03	6E+03	5E+02	8.9	1.2
EN.PR	3E+02	86.6	1.7	1.7	1.5	1E+03	5E+03	1.6	3.6	1.2
GV.EX	42.7	2E+02	2.1	2.1	1.8	1E+03	5E+03	10.8	1.7	1.4
DP.RA	2E+02	89.5	1.5	1.5	1.1	1E+02	7E+02	37.3	2.9	1.4
RES	1E+02	3E+02	2.1	2.1	1.2	1E+02	235.1	1.9	1.4	1.3
TRAD	7.9	18.7	1.3	1.3	1.3	30.5	8.0	4.9	1.3	1.3

Table 8
Variables selected with the proposed stepwise regression analysis by transformed series and country.

Country	Transformation	Variables Selected						Frequency
		BIRT	DEAT	DP.RA	EN.PR	GV.EX	TRAD	
India	Log Diff		X	X	X		X	9
	Gro	X			X	X		12
	Gro Diff				X	X		11
China	Log Diff			X		X	X	6
	Gro							6
	Gro Diff				X	X		14

of the variance inflated for each coefficient; with a value of 1, the variable is not correlated, between 1 and 5 moderately correlated, and greater than 5, there is a highly correlated variable. With this, an idea of the improvement that results in a decrease of multicollinearity when transforming the variables is seen. Table 7 offers the results.

Table 9
Multicollinearity test for variables selected with the proposed stepwise regression analysis by transformed series and country.

Country	China			India		
Test	Log Diff	Gro	Gro Diff	Log Diff	Gro	Gro Diff
Determinant |X'X |:	F	F	F	F	F	F
Farrar Chi-Square:	T	T	T	T	T	T
Red Indicator:	F	F	T	F	F	F
Sum of Lambda Inverse:	F	F	F	F	F	F
Theil's Method:	F	F	F	F	F	F
Condition Number:	F	F	F	F	F	F

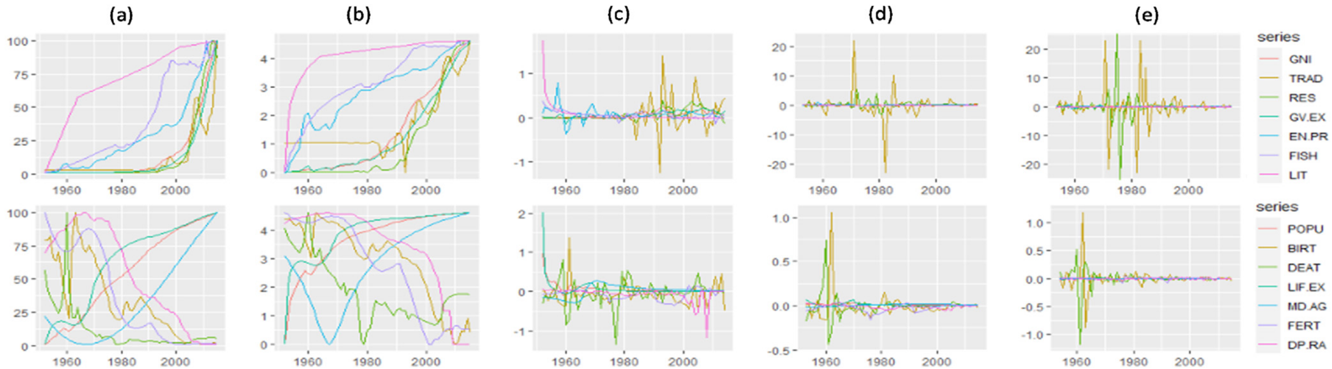

Fig. 19. China time series plot per variable and type of transformation. (a) Variables rescaled between 1 and 100, (b) Log of the rescaled variables, (c) Difference of the Log, (d) Growth rate, (d) Difference of the Growth rate.

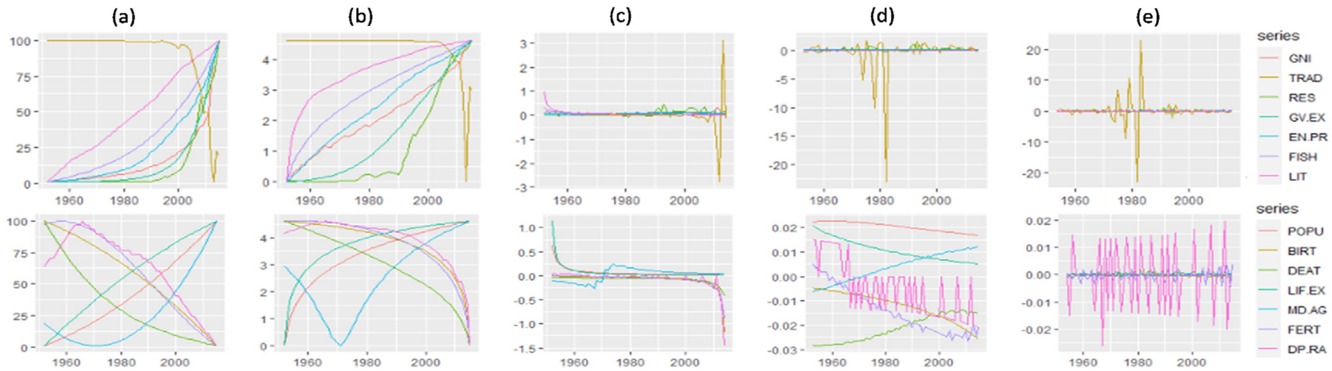

Fig. 20. India time series plot per variable and type of transformation. (a) Variables rescaled between 1 and 100, (b) Log of the rescaled variables, (c) Difference of the Log, (d) Growth rate, (d) Difference of the Growth rate.

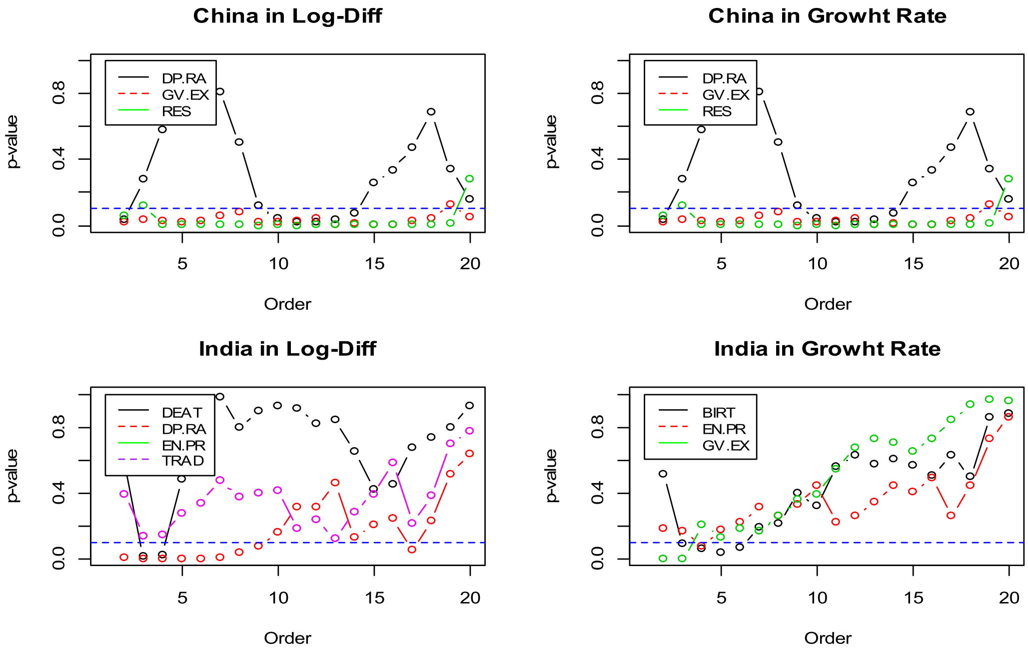

Fig. 21. P-values of the Granger Causality test by series and variables selected with the proposed stepwise regression analysis.

3.3. *Stepwise regression analysis*

The stepwise regression analysis is performed as the first approach to make a selection of the variables for each transformation. The three types of model selection methods applied are; these are; forward, backward, and bi-direction. For the three methods, the nine criteria seen in the theoretical section is used to determine the order in which effects enter or leave at each step. The data used is all the history (1950–2015) to get the maximum advantage of the information.

Table 10
Result of the Global Validation of Linear Model Assumptions applied to the models selected with the proposed stepwise regression analysis.

Country	China			India		
Directional Test	Log Diff	Gro	Gro Diff	Log Diff	Gro	Gro Diff
Global	0.00	0.00	**0.65**	0.00	0.00	0.00
Skewness	0.00	0.00	**0.16**	**0.14**	0.00	**0.02**
Kurtosis	**0.02**	**0.02**	**0.74**	**0.50**	0.00	0.00
Linearity	**0.38**	**0.38**	**0.89**	**0.09**	**0.45**	0.01
Heteroscedasticity	**0.18**	**0.18**	**0.56**	0.00	0.00	**0.02**

Table 11
Result of assumptions diagnostic tests to the models selected with the proposed stepwise regression analysis.

Country/Transformation		China			India		
Dimension	Test	Log Diff	Gro	Gro Diff	Log Diff	Gro	Gro Diff
Autocorrelation	**B-G**	0.00	0.00	**0.64**	**0.13**	0.05	**0.26**
	D-W	0.00	0.00	**0.49**	**0.92**	0.01	**0.88**
Correlation	**Cor**	0.00	0.00	0.00	0.00	0.00	0.00
Heteroscedasticity	**B-P**	**0.15**	**0.15**	**0.14**	**0.61**	0.00	**0.32**
	G-Q	0.00	0.00	**0.47**	**1.00**	0.00	**0.14**
	H-M	0.00	0.00	**0.34**	**1.00**	0.01	**0.25**
Linearity	**H-C**	**0.46**	**0.46**	**0.62**	**0.19**	**0.71**	**0.93**
	Rai	**0.99**	**0.99**	**0.40**	**0.16**	0.00	0.06
	Ram	**0.69**	**0.69**	**0.33**	**0.12**	**0.05**	**0.02**

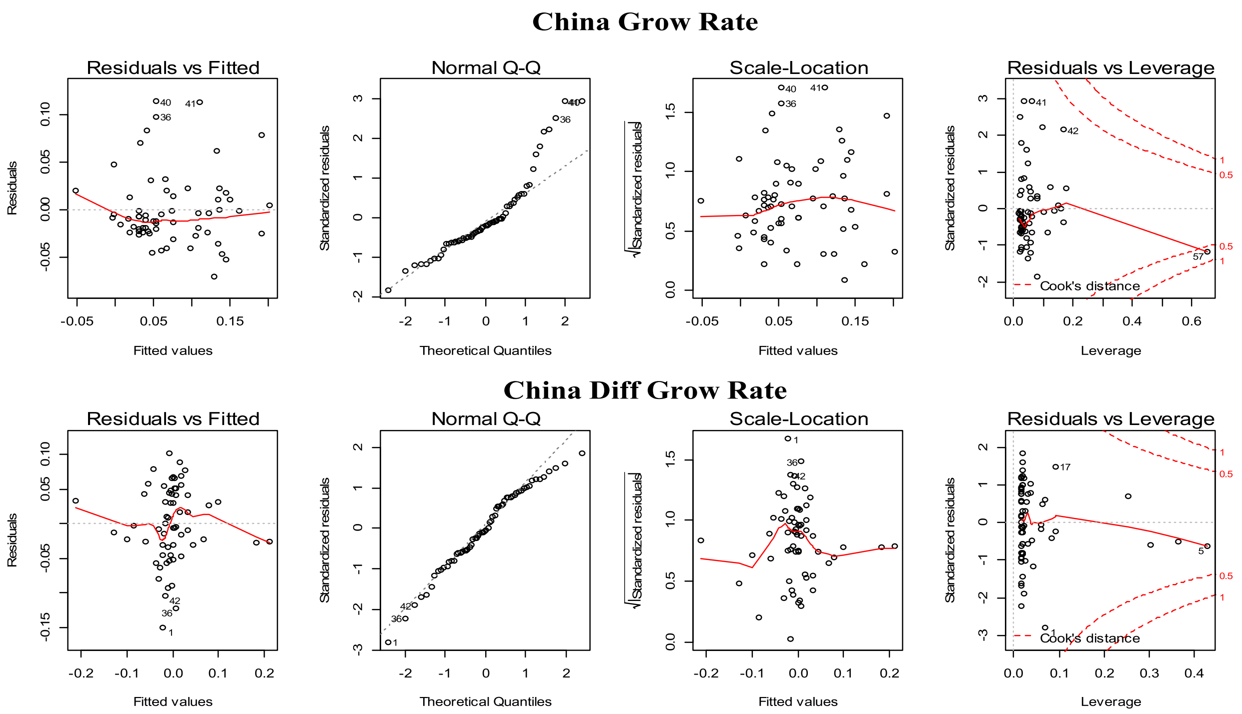

Fig. 22. Residual diagnostic for the models selected with the proposed stepwise regression analysis for the China Growth Rate and his first difference.

3.3.1. Model selection

Based on these 27 selection processes, the combinations with the highest frequency are shown in Table 8. Only the variables selected by country and transformation are shown.

For China's case, the frequency column shows that within the 27 selection criteria used though matter the variables that were selected, the series in logarithms and growth rate gave the same variables. This is because both transformations usually generate approximate results.

3.3.2. Multicollinearity and Granger Causality of variables selected

With these six selections, we again apply the comprehensive multicollinearity diagnostics tests, like before, with the GNI of each country as the variable of response. In Table 9, the results show whether the test gives an affirmative answer to the multicollinearity (T) or not (F).

Now, the existence or not of a correlation between two variables does not necessarily imply causality. That is, one variable is correlated with another does not always imply that one of them is the cause of the changes in the values of another. In the next figure, the graphs of the different p-values for different lags of the variables in the Log-Diff and Gro transformations are presented. The results show us that there is a sense of causality in the selected variables, especially in the initials lags. As shown in Fig. 21, the blue line represents the 10% significance margin of the test.

3.3.3. Tests for the models selected

Applying the Global Validation of Linear Models Assumptions to the models selected by the different selection processes of the stepwise regression analysis, the results of the p-values are presented in Table 10.

Those values that pass the test for a degree of significance more significant than 1% are highlighted in bold. The difference in the growth rate for China was the only model that passed all the tests. This is the model with the variables energy production and government spending. In India's case, the difference in logarithm passes three of the tests, presenting a problem of heteroscedasticity. As a complement to the previous tests, another group of tests is shown in Table 11 due to the different dimensions seen in the theory section.

These tests confirm the previous results, with the addition that no model presents non-correlated residuals with the GNI. It seems that the models for the series of the growth rate for both countries are the series that meets the most theoretical requirements. As an example, in Fig. 22, the residues' changes will be shown on China's growth rate model and his first

Table 12
Result of assumptions diagnostic tests to the Ridge regression by country and transformed series.

Country/Transformation		China			India		
Dimension	Test	Log	Gro	Gro Diff	Log	Gro	Gro Diff
Autocorrelation	B-G	0.00	0.00	**0.90**	**0.15**	**0.12**	**0.71**
	D-W	0.00	0.00	**0.53**	0.00	0.01	**0.82**
Correlation	Cor	0.00	0.00	0.00	0.00	0.00	0.00
Heteroscedasticity	B-P	**0.06**	**0.24**	**0.99**	**0.29**	0.00	**0.21**
	G-Q	0.00	0.00	**0.38**	**0.98**	0.01	**0.02**
	H-M	**0.96**	0.00	**0.14**	**0.99**	0.00	**0.09**
Linearity	H-C	**0.02**	**0.97**	**0.89**	**0.82**	**0.90**	**0.67**
	Rai	0.00	**0.22**	**0.57**	**0.30**	0.01	**0.06**
	Ram	0.00	**0.48**	**0.22**	**0.18**	0.00	0.00

Table 13
Result of assumptions diagnostic tests to the Lasso regression by country and transformed series.

Country/Transformation		China			India		
Dimension	Test	Log	Gro	Gro Diff	Log	Gro	Gro Diff
Autocorrelation	B-G	0.00	0.00	**0.90**	**0.66**	**0.11**	**0.25**
	D-W	0.00	0.00	**0.45**	**0.08**	**0.04**	**0.87**
Correlation	Cor	**0.73**	0.00	0.00	**0.79**	0.00	0.00
Heteroscedastic	B-P	**0.07**	**0.07**	**0.83**	**0.03**	0.01	**0.14**
	G-Q	0.00	0.00	**0.30**	**0.53**	0.00	**0.25**
	H-M	**0.83**	0.00	**0.19**	**0.84**	0.00	**0.38**
Linearity	H-C	**0.08**	**0.20**	**0.65**	**0.89**	**0.38**	**0.87**
	Rai	0.00	**0.99**	**0.55**	**0.02**	0.00	**0.06**
	Ram	0.00	**0.92**	**0.39**	0.00	**0.13**	**0.02**

Table 14
Variables selected by Lasso and Elastic Net regression by country and transformed series.

Country	China						India					
Serie	Log		Gro		Gro Diff		Log		Gro		Gro Diff	
Variable/Model	L	E	L	E	L	E	L	E	L	E	L	E
BIRT	X							X	X	X		
DEAT	X	X					X	X				X
DP.RA	X					X						
EN.PR	X				X	X	X	X				X
FERT		X						X				
FISH	X	X					X	X				
GV.EX	X	X	X	X	X	X	X	X		X	X	X
LIF.EX	X	X				X						
LIT	X						X					
MD.AG	X	X										
POPU	X							X				
RES	X	X	X	X		X	X	X				
TRAD		X					X	X				

Table 15
Diagnostic test for the distributed lag regression of China, using the variables selected by Stepwise"S" and Elastic Net "E" regression, with two lags "I" or optimized lags "II."

Serie	Model	Correlation	Heteroscedastic			Linearity		
			B-P	G-Q	H-M	H-C	Rai	Ram
Log Diff	**S-I**	0.00	0.12	0.00	0.00	0.83	1.00	0.65
	S-II	0.00	0.29	0.00	0.00	0.84	0.95	0.86
	E-I	0.00	0.21	0.00	0.00	0.51	0.99	0.90
	E-II	0.00	0.41	0.00	0.00	0.53	0.91	0.23
Gro Diff	**S-I**	0.47	0.07	0.37	0.28	0.64	0.71	0.39
	S-II	0.49	0.88	0.13	0.00	0.76	0.31	0.61
	E-I	0.32	0.94	0.15	0.03	0.48	0.59	0.15
	E-II	0.56	0.01	0.93	0.00	0.38	0.54	0.58

Table 16
Diagnostic test for the autoregressive distributed lag regression of China, using an autoregressive distributed lag.

Serie	Dimension/Test						
	Correlation	Heteroscedastic			Linearity		
		B-P	G-Q	H-M	H-C	Rai	Ram
Log Diff	0.1	0	0.55	0	1	0.31	0
Gro Diff	0.69	0.21	0.61	0.05	0.68	0.35	0.81

difference. There we can see how the residuals have a relatively close Gaussian behavior, especially in the difference in the growth rate.

Once the initial Stepwise regression analysis has been carried out, we already have an initial selection of variables for each transformation with their corresponding tests. The next procedure will be to carry out a different approach to the traditional OLS when introducing penalties with the Regularization regression Analysis.

3.4. *Regularization regression analysis*

In this section, the three types of regression seen in the theoretical framework will be applied. Once more, the entire history with all the information available will be considered to make the selection of the variables.

3.4.1. *Ridge regression*

The Ridge regression will be applied for the logarithm series, the growth rates, and the difference in the growth rates for all the variables. Then the test results if the standardized residuals generated by the methodology proposed by Cilluffo et al. (2019) are taken into account for the test. In Table 12, the result of these tests is presented.

The results obtained in this case bear some resemblance to those seen for the models selected in the stepwise regression analysis; however, a more significant number of variables are used here. The penalty parameter selection was carried out in two stages, with first modeling and calculating the mean of the absolute error (MAE). Secondly, the metric using cross-validation taking 20 data in each group, per the parameter that minimizes said metric.

3.4.2. *Lasso regression*

The Lasso regression will be performed in two steps equally, initially with all the variables and using k-fold cross-validation taking $k = 20$ data points to find the lambda parameter that minimizes the MAE. The model will then be applied again with the selected variables, and the final lambda and the standardized residuals generated by the methodology proposed by Cilluffo et al. (2019) will be calculated. Table 13 shows the results.

In this case, it was possible to break the correlation between the errors and the GNI for the logarithms models. In the case of India, almost all the tests were passed.

3.4.3. *Elastic net regression*

Since the Elastic Net can be seen as a midpoint between Ridge and Lasso, the same procedure applied in those cases will be followed. However, this time we will look for the two parameters required in this model. As seen in the theoretical section, the metric selected in this case is the MAE. Next, we will see the results of the variables selected by the Lasso model and the Elastic Net in each transformation for comparison purposes. The column with the letter “L” stands for Lasso model variables and “E” for Elastic Net variables.

In China's case, government spending variables are confirmed, but it emerges as a variable reserve when considering the elastic net model. Government spending is also shown as a determining factor for the GNI for India. Like mortality rate and energy production, other variables were selected in the stepwise models and two of the transformations.

3.5. *Distributed lag regressions analysis*

In the distributed lag regression analysis, we use the results obtained previously, specifically the variables selected by the Stepwise regression analysis denoted by “S” and the Elastic Net regression denoted by “E.” All available data will also be used.

3.5.1. *Distributed lag regression*

The classical distributed lag regression is performed in two different ways. Initially, with just two lags, grade one denoted by “I,” and alternatively with an optimized model based on the metric MAE denoted “II.” With which we will have the models based on the variables obtained with Stepwise and low grade denoted by “S-I” and with an optimized degree of lags “S-II.” In this case, to summarize results, only the p-values of the tests will be shown in China's case in Table 15.

Table 17
F-statistic of the Pesaran, Shin, and Smith cointegration test.

Serie/Indicator	Log Diff		Gro Diff	
Country	F-statistic	Lags	F-statistic	Lags
China	3.9	(4,3,1)	10.73	(1,2,1,1)
India	15.40	(1,1,1,1,1)	26.44	(1,2)

Table 18
Variables selected by Stepwise "S," Lasso "L" and Elastic Net "E" regressions by country and series.

Country	China								India							
Serie	Gro				Gro Diff				Gro				Gro Diff			
Variable/Model	S	L	E	T	S	L	E	T	S	L	E	T	S	L	E	T
BIRT									X	X	X	3				
DEAT															X	1
DP.RA	X			1			X	1								
EN.PR					X	X	X	3	X			1	X		X	2
GV.EX	**X**	**X**	**X**	**3**	**X**	**X**	**X**	**3**	**X**		**X**	**2**	**X**	**X**	**X**	**3**
LIF.EX							X	1								
RES		X	X	2			X	1								
TRAD	X			1												

Table 19
Explanation of the proposed experiments for the case of 15 years of forecast.

Train-Verify ratio/ Years	Sliding Window 55-69	70-84	85-99	00-14	Train-Verify ratio/ Years	Cumulative Window 55-69	70-84	85-99	00-14
	15	15			1:1	15	15		
1:1		15	15		2:1	20		15	
			15	15	3:1	30			15

Table 20
Parameters of the proposed experiments.

Type of training window		Sliding		Cumulative	
Concept/Experiment		I	II	I	II
Number of Years for:	Training	15	20	15,30,45	20,40
	Verifying			15	20
Repetition of the experiment		3	2	3	2
Train to verify the ratio		1:1		1:1, 2:1,3:1	1:1, 2:1

Table 21
Experimental results by country and transformed series.

Country	China						Global		India								General	
Serie	Log		Gro		Gro Diff				Log		Gro		Gro Diff		Global			
Metric/Model	M	R	M	R	M	R	M	R	M	R	M	R	M	R	M	R	M	R
STW	2.9	1	2.9	1	5.5	4	3.8	1	5.1	1	2.6	2	6.5	3	4.7	2	4.3	1
LAS	3.2	2	3.2	2	5.4	2	3.9	3	5.1	2	2.6	1	6.4	2	4.7	1	4.3	2
DSL1	3.4	3	3.4	4	5.4	3	4.1	4	5.9	3	4.6	4	6.8	5	5.8	3	4.9	3
ADL1	4.2	7	4.2	6	5.5	5	4.6	6	6.1	4	4.6	3	6.8	6	5.8	4	5.2	4
RID	4.2	6	4.6	7	5.8	8	4.9	7	6.4	5	5.0	5	6.2	1	5.9	5	5.4	5
DSL2	3.9	5	3.9	5	5.6	6	4.5	5	6.6	6	5.4	7	7.0	7	6.3	6	5.4	6
ELN	3.4	4	3.2	3	5.1	1	3.9	2	7.3	8	8.0	8	6.6	4	7.3	8	5.6	7
ADL2	4.8	8	4.8	8	5.6	7	5.0	8	6.8	7	5.2	6	7.2	8	6.4	7	5.7	8

Table 22
Variables selected by country and transformed series with the Stepwise and Lasso regression.

Country	China						India					
Serie	Log		Gro		Gro Diff		Log		Gro		Gro Diff	
Variable/Model	LAS	STW	LAS	STW	LAS	STW	LAS	STW	LAS	STW	LAS	STW
BIRT							**X**	**X**	**X**		**X**	**X**
DEAT										X		
DP.RA		X		X						X		
EN.PR					X	X		**X**		**X**		**X**
GV.EX	**X**	**X**	**X**	**X**	**X**	**X**		X				X
RES	X		X									
TRAD		X		X						X		

All the tests were finally overcome for the series of the growth rate on the first difference with the variables of the elastic net regression. The lags selected for the best models were varied. In the case of China 3(three) and a single lag was selected for the differentiated series in logarithm or growth rates, respectively. In the case of India, one lag and 4 (four), respectively, were selected. The results show that, in general, it is not necessary to take into account ancient information to make the GNI prediction.

3.5.2. Autoregressive distributed lag regression

For autoregressive models, optimization will be accomplished by calculating the optimal orders (lag structure) for short-term relationships and the autoregressive part of the ARDL model. The limit test is applied (ARDL) with the approach seen in the theoretical section (Pesaran, 1997); the criterion used will be BIC ***(see** Table 16**, for the result of the tests)***. In China's case, the results will be shown when taking the variables of the stepwise regression for the series in Log-diff (Table 8. TRAD, GV. EX, DP. RA), while for the series in Growth-diff, the selected ones by the Lasso regression (Table 14. GV. EX, EN. PR). The same models are chosen for the case of India, Stepwise for Log-diff (DEAT, DP. RA, EN. PR, TRAD) and Lasso for Growth-diff (GV.EX).

In this case, we again have positive results in practically all tests. The test results of the Pesaran, Shin, and Smith cointegration tests (Pesaran et al., 2001) are given in Table 17. The results show that there is a long-term or cointegrating relationship between the variables. The optimal lag is just one step back. Likewise, the results are only shown in China's case with the optimized models to the difference of the Growth rate. The notation of the lags contains the autoregressive order as the first parameter. The rest are the lags of the variables chosen for each series.

In all cases, the test is passed with a 10% significance in India with 1%. The results indicate that the selected variables have a cointegration relationship or a long-term relationship.

3.6. Summary of variables selected

Now in Table 18, the variables selected by each of the Stepwise "S," Lasso "L," and Elastic Net "E" approaches are reviewed in the following table.

In column T, we have the total repetition of the variable between the models, being in both cases, government spending the variable that is most repeated for the two series shown. As mentioned earlier, despite having different forms of governments, both countries have planned their economy through five-year plans. Therefore, it makes sense that government spending is selected with a high frequency in all models and transformations. Also, the production of energy shows a high impact in the GNI. In previous research, theoretical and empirical evidence indicated that energy use and production play a key role in enabling growth because production is a function of capital, labor, and energy (Stern, 2010).

For India, the birth rate emerges like a variable to consider. Previous studies show how birth rate declines have a strong medium-term positive impact on per capita income growth through labor supply or "dependency" effects (Brander and Dowrick, 1994). For China, the reserves emerge like a critical variable to consider. Here is a necessary highlight that these foreign reserves have a crucial role in economic growth and stability as it helps the economy perform effectively in global markets. In both series but with less frequency, the dependency ratio for china is selected. Such a result shows the impact of the birth control policies in the boosting of economic growth.

4. Experiments and discussion

Based on the previous results, we propose different experiments in this section to corroborate each of the models' predictive capacity in different scenarios.

4.1. Proposed experiments

The proposed experiments are used to test the models when the size of the data or time window used for its construction and the prediction is changed. In this sense, it is proposed to divide the study period into three or four periods of equal length.

In this case, we are going to start the history from the year 1955 since some models require lagged information. Thus, having 60 years of history, there would be temporal windows of 15 years if it is divided into four parts and 20 years if it is divided into three parts. The first scheme is denoted by the type "I" experiment and the second by "II."

In considering the data for the construction of the models, a "Sliding" scheme is defined. The last 15 or 20 years of information, as the case may be, is considered. In another "Cumulative" scheme, the information from the beginning of the history (1955) is cumulatively taken into account. Table 19 shows a serial construction of the different experiments for the 15-year prediction case. The blue period is the historical one taken into consideration, and the green period is the prediction. Additionally, Table 20 shows the general configuration of all the experiments.

The models to be applied will be those previously seen. They will be denoted as Stepwise "STW," Ridge "RID," Lasso "LAS," Elastic Net "ELN," Distributed Lag "DSL" and Autoregressive Distributed Lag regression "ADL."

4.2. Experimental results

Now we will see the results of the experiments in aggregate form for the Symmetric mean absolute percentage error (sMAPE) metric (Makridakis, 1993) denoted by "M." This metric will be taken into account since it is an asymmetric relative measure, it allows a direct comparison between the models to facilitate the choice of the best model (this is the model with the lowest sMAPE). A column with the ranking "R" is included; this value will simply take the values from 1 to 8, with 1 being the model with the lowest sMAPE and 8 the one with the worst result. The results for both countries with the series in logarithms, growth rate, and the difference in the growth rate are in Table 21.

In the last column, the general average for both countries by the model is presented with the best result obtained by the proposed method with the Stepwise regression. The Lasso type regression occupies the second place, and both models occupy the first or second positions in all series and both countries. The variables chosen for each series and model are presented in Table 22.

The repeated variables in all the series are highlighted in bold. In the case of China, government spending is also repeated in both models. This difference arises because, unlike China, the Indian government has not implemented birth control policies. The correlation between the GNI and the birth rate was also verified. China being only -0.58, while the case of India is -0.86, according to Fig. 18.

5. Conclusion

In this paper, a novel methodology is adopted to explain trends of GNI in China and India between 1952 and 2015. The rescaling of variables between 1 and 100 and subsequent transformation was necessary for the logarithm and percentage of change calculation. The data were analyzed using three different approaches: stepwise, regularization, and distributed lag. An extensive analysis of both descriptive and statistical perspectives was perforrned. Tests were conducted to calculate: multicollinearity, autocorrelation, stationarity, and variance-inflation factors. The results showed that the logarithm's proposed transformation and the growth rate are necessary to meet the assumptions of linear regression models.

The first analysis was stepwise regression. Twenty-seven (27) different combinations of methods and criteria were considered for each type of transformation, and the models were selected with the highest frequency of selection for each data series. The chosen models were subjected to a multitude of tests to verify assumptions. Then, the Global Validation of Linear Models Assumptions was implemented. Next, nine additional tests to corroborate the autocorrelation and correlation of the residuals with the response variable, the GNI, were conducted. Additional tests for heteroscedasticity and linearity were performed.

The second analysis was a regularization regression that considered three techniques: Ridge, Lasso, and Elastic Net. Based on the methodology proposed by Cilluffo (Cilluffo et al., 2019), we conducted tests on the residuals in the three cases. The parameters were obtained with cross-validation using 20 data points and the MAE metric. In the ridge regression, all the variables were taken into account, and we obtained improved test results for the Stepwise regressions. However, taking into account all the variables, the model did not allow a direct analysis of the most relevant variables. In the case of the Lasso and elastic net regressions, some improvement was also achieved. With the last model, it was possible to avoid the correlation between the residuals and the GNI. The chosen variables in this second analysis confirmed those from the stepwise analysis, with the addition of the Foreign Exchange Reserves for China and the Death Rate for India.

The third analysis, with distributed lag models, was conducted to obtain the long-term relationship between the GNI and the independent variables selected during the two previous analyses. The methodology proposed by Pesaran and Shin (Pesaran, 1997) were used. In both cases, it was determined that outdated information was not necessary to determine the GNI. All models were constructed under a sliding prediction scheme, dividing the history into three or four equal time windows of 15 and 20 years.

In the case of China, National Government Expenditure was a variable repeated in all models, this due to China's five-year economic planning policies. In a socialist state like China, the government determines the economy's direction through influence in government spending levels. Two other variables were selected; the balance of total imports and exports and the dependency ratio. All these variables were consistent with events throughout China's economic history. China is often referred to as the "World Factory." In 2005, the country produced 70 % of the world's toys, 60% of its bicycles, 50% of its shoes, and 33% of its luggage. For the same year, China also manufactured half of the world's microwave ovens, one-third of its TV sets and air conditioners, a quarter of its washers, and one-fifth of its refrigerators (Zhang, 2006). Accordingly, the Trade

Balance is also an essential variable for representing economic growth. When looking at the dependency ratio, we see that the chinese government's birth control policies have yielded positive economic results. However, China's current population become of aging. Indicators like the dependency ratio are showing a trend contrary to the one that is desired. As the dependency ratio is starting to rise gradually, some estimate it to reach 42.2 in 2020.[3]

To control population growth and boost economic growth, China's total fertility rate has declined dramatically since 1970s, mainly as a consequence of the national population policy. Nevertheless, China's birth control policy currently continues, although it has been relaxed to allow families to have two children. Therefore, any restrictions on birth rate no longer seem to make sense in the present day, as it would be difficult for the Chinese population to return to the fertility levels of 30 years ago.

Our results analyzed the long-term relationship between the dependency ratio and the GNI. They indicate that maintaining birth control policies could have adverse effects in the long-term. By 2020, China is expected to have an estimated birth rate of 11.6[4] and a level below the substitution rate of 1.6[5], which amounts to fewer than two children per couple. According to this projection, China could face a situation similar to Japan, where the estimated replacement rate is 1.43, and the dependency ratio is 69[6].

China's population policy has undergone a series of changes, which have affected the nation's economy and social development. A measure that is likely to become more costly as the most prolonged current birth control policies are kept in place. In such a case, based on the theory of Solow model, some Chinese researchers proposed the advices that China should more reliant on technology progress rather than mass labor. But the economic growth at the moment by the technology progress is uncertain. On the other hand, China's miracle in the past 40 years shows that reforms in China has profound influenced the economic growth. For example, China's reform and opening-up policy underwent vast changes to its economic system and boost the economic growth over the past 40 years. Therefore, the upholding and improving the current population policy to optimize the population structure will be striking important to the GNI. The future reforms of population and other resources allocation is vital for economic growth in China. With more investment in specialized labor, energy production of China is also vital for predicting the GNI.

In India's case, energy production and birth rate were the variables selected for the most accurate predictions. Contrary to China, the Indian government has not implemented birth control policies. For this reason, these indicators have maintained a more linear trend. In 2020, the birth rate has declined and is estimated to be 18.2[7]. Furthermore, to stay above the substitution rate of 2.35[8]. The Indian economy is heavily dependent on oil imports.[9] The pronounced increase in oil price is one of the main reasons for explaining the trade balance trend. However, the medium- and long-term outlook seem to be favorable for India[10]. While India is not a socialist state like China, the Indian government has historically planned its economy. For this reason, the government spending variable was selected in two of the proposed transformations.

Therefore, the paper provides a meaningful contribution to the literature on variable interactions with GNI. Using economic data from China and India, the study explained how the policies applied by each country regarding the variables selected could positively or negatively affect the GNI. In India's case, government spending, energy production, and birth rate were the variables with the best performance. Therefore, their planning economy and energy policy focused on securing adequate energy resources to meet its economy's growing demands. The first two variables controlled by the government lead the economy's route, together with the steady decrease of birth rate, contribute to economic growth. In China's case, we find three principal variables; government spending, the balance of total imports and exports, and the dependency ratio. Three principal characteristics explain these variables: 1. The socialist government that leads the way of the economy. 2. The trade balance is a critical element in the growth of the economy of China, and finally. 3. The birth control policies that show how they boosted the economic growth through the decrease of the dependency ratio indicator. The results from this paper urge both countries to consider the variables assessed in this paper when developing and implementing five-year plans in the future. By doing this, they will be able to benefit economically in the long term and curtail undesirable policy-related outcomes that may seek to destabilize their economy.

Declaration of competing interest

The authors declare that there is no conflict of interest regarding the publication of this paper.

Acknowledgments

3 https://www.cia.gov/library/publications/the-world-factbook/geos/ch.html.
4 https://www.cia.gov/library/publications/the-world-factbook/geos/ch.html.
5 https://www.cia.gov/library/publications/the-world-factbook/fields/356.html.
6 https://www.cia.gov/library/publications/the-world-factbook/geos/ja.html.
7 https://www.cia.gov/library/publications/the-world-factbook/geos/in.html.
8 https://www.cia.gov/library/publications/the-world-factbook/geos/in.html.
9 https://www.files.ethz.ch/isn/89787/73.pdf.
10 http://pubdocs.worldbank.org/en/339801451407117632/PRN01Mar2015OilPrices.pdf.

This work was supported by the National Natural Science Foundation of China (no. 71773012, 12026239), Natural Science Foundation of Liaoning Province LN2020J35 and Research Project of Dongbei University of Finance and Economics (DUFE2020Y22).

References

Akaike, H. (1974). A new look at the statistical model identification. *IEEE Trans. Automat. Contr., 19*(6), 716–723. https://doi.org/10.1109/TAC.1974.1100705. MR 0423716.

Allaoua, Kherous, & Zoubida. (1997). India, Sustaining Rapid Economic Growth. In *World Bank Country Study*. Washington, USA: World Bank Group.

Ansley, J. (1984). *Coale. Rapid Population Change in China, 1952-1982*. Washington, USA: National Academies.

Banerjee, A., Dolado, J., & Mestre, R. (1998). Error-correction mechanism tests for cointegration in a single-equation framework. *J. Time Anal., 19*(3), 267–283. https://doi.org/10.1111/1467-9892.00091

Barro, Robert, & Grilli, Vittorio (1994). *European Macroeconomics*. Ch. 15–16. Macmillan. ISBN 0-333-57764-7.

Box, G. E. P., & Pierce, D. A. (1970). Distribution of residual autocorrelations in autoregressive-integrated moving average time series models. *J. Am. Stat. Assoc., 65*(332), 1509–1526. https://doi.org/10.1080/01621459.1970.10481180

Brander, J. A., & Dowrick, S. (1994). The role of fertility and population in economic growth. *J. Popul. Econ., 7*(1), 1–25. https://doi.org/10.1007/bf00160435

Breusch, T. S. (1978). Testing for autocorrelation in dynamic linear models. *Aust. Econ. Pap., 17*, 334–355. https://doi.org/10.1111/j.1467-8454.1978.tb00635.x

Breusch, T. S., & Pagan, A. R. (1979). A simple test for heteroskedasticity and random coefficient variation. *Econometrica, 47*(5), 1287–1294. https://doi.org/10.2307/1911963.JSTOR1911963. MR 0545960.

Burnett, Nicholas, & EFA Global Monitoring Report team at UNESCO. (2006). *Global Monitoring Report 2006: Literacy for Life*. Paris, France: United Nations Educational, Scientific and Cultural Organization. ISBN 92-3-104008-1.

Chatterjee, S., & Price, B. (1977). *Regression Analysis by Example*. New York: John Wiley & Sons.

Cilluffo, G., Sottile, G., La Grutta, S., & Muggeo, V. M. (2019). The Induced Smoothed lasso: a practical framework for hypothesis testing in high dimensional regression. *Stat. Methods Med. Res.*. , Article 096228021984289. https://doi.org/10.1177/0962280219842890

Coastline of the world factbook.Central intelligence agency. https://www.cia.gov/library/publications/resources/the-world-factbook/fields/282.html.

Crude Death Rate (Per 1,000 Population) Based on World Population Prospects the 2008 Revision, United Nations. Retrieved 22 June 2010.

Durbin, J., & Watson, G. S. (1950). Testing for serial correlation in least squares regression. *I. Biometrika, 37*(3–4), 409–428. https://doi.org/10.1093/biomet/37.3-4.409

Eichengreen, Barry (2011). *Exorbitant Privilege: the Rise and Fall of the Dollar and the Future of the International Monetary System*. USA: Oxford University Press.

Eichengreen, Barry, Gupta, Poonam, & Kumar, Rajiv (2010). *Emerging Giants, China, and India in the World Economy*. New York, USA: Oxford University Press.

Elliott, G., Granger, C. W. J., & Timmermann, A. G. (2006). *Handbook of Economic Forecasting*. North Holland.

Engle, R. F., & Granger, C. W. J. (1987). Co-integration and error correction: representation, estimation, and testing. *Econometrica, 55*(2), 251. https://doi.org/10.2307/1913236

Farrar, D. E., & Glauber, R. R. (1967). Multicollinearity in regression analysis: the problem revisited. *Rev. Econ. Stat., 49*(1), 92. https://doi.org/10.2307/1937887

Fox, J., & Monette, G. (1992). Generalized collinearity diagnostics. *J. Am. Stat. Assoc., 87*(417), 178–183. https://doi.org/10.1080/01621459.1992.10475190

Fuller, Wayne A. (1976). *Introduction to Statistical Time Series*. New York: Wiley. http://www.loc.gov/catdir/enhancements/fy0607/76006954-t.html.

Godfrey, L. G. (1978). Testing against general autoregressive and moving average error models when the regressors include lagged dependent variables. *Econometrica, 46*, 1293–1301. JSTOR 1913829.

Goldfeld, S. M., & Quandt, R. E. (1965). Some tests for homoscedasticity. *J. Am. Stat. Assoc., 60*(310), 539–547. https://doi.org/10.1080/01621459.1965.10480811

Hannan, E. J., & Quinn, B. G. (1979). The determination of the order of an autoregression. *J. Roy. Stat. Soc. B, 41*(2), 190–195.

Harrison, M. J., & McCabe, B. P. M. (1979). A test for heteroscedasticity based on ordinary least squares residuals. *J. Am. Stat. Assoc., 74*(366), 494. https://doi.org/10.2307/2286361

Harvey, A. C., & Collier, P. (1977). Testing for functional misspecification in regression analysis. *J. Econom., 6*(1), 103–119. https://doi.org/10.1016/0304-4076(77)90057-4

Hashem, M. (2015). *Pesaran - Time Series and Panel Data Econometrics*. Oxford University Press.

Hocking, R. R. (1976). A biometrics invited paper. The analysis and selection of variables in linear regression. *Biometrics, 32*(1), 1. https://doi.org/10.2307/2529336

Hoerl, A. E., & Kennard, R. W. (1970). Ridge regression: biased estimation for non-orthogonal problems. *Technometrics, 12*(1), 55–67.

Hurvich, C. M., & Tsai, C.-L. (1993). A corrected Akaike information criterion for vector autoregressive model selection. *J. Time Anal., 14*(3), 271–279. https://doi.org/10.1111/j.1467-9892.1993.tb00144.x

Imdadullah, M., Aslam, M., Altaf, S., & mctest. (2016). An R package for detection of collinearity among regressors. *R J., 8*. https://doi.org/10.32614/RJ-2016-0628495-505, 2.

Jati, K. (2009). *Sengupta, Chiranjib Neogi. India's New Economy, Industry Efficiency, and Growth*. New York, USA: Palgrave Macmillan.

Knight, John, & Ding, Sai (2012). *China's Remarkable Economic Growth*. London, United Kingdom: Oxford University Press.

Kovács, Péter, et al. (2005). A new measure of multicollinearity in linear regression models. *Int. Stat. Rev./Rev. Int. Stat., 73*(3), 405–412. JSTOR, www.jstor.org/stable/25472683.

Kwiatkowski, D., Phillips, P. C. B., Schmidt, P., & Shin, Y. (1992). Testing the null hypothesis of stationarity against the alternative of a unit root. *J. Econom., 54*(1–3), 159–178. doi:10.1016.

Li, Muzhen (1992). *The Population of Modern China*. New York, USA: Springer.

Ljung, G. M., & Box, G. E. P. (1978). On a measure of lack of fit in time series models. *Biometrika, 65*(2), 297–303. https://doi.org/10.1093/biomet/65.2.297

MacFarquhar, R. (1987). The Cambridge history of China. In *The People's Republic. Part 1, the Emergence of Revolutionary China, 1949-1965* (vol. 14). Cambridge: Cambridge University Press.

Makridakis, S. (1993). Accuracy measures: theoretical and practical concerns. *Int. J. Forecast., 9*(4), 527–529. https://doi.org/10.1016/0169-2070(93)90079-3

Mallows, C. L. (1973). Some comments on C P. *Technometrics, 15*(4), 661. https://doi.org/10.2307/1267380

Metcalf, B., & Metcalf, T. R. (9 October 2006). *A Concise History of Modern India* (second ed.). Cambridge University Press, 978-0-521-68225-1.

Mo, Long, & Wei, Yuhong (2018). *China's Demographic Dilemma of Population Aging and Population Control*. Singapore: Springer.

O'Sullivan, Arthur, & Sheffrin, Steven M. (2003). *Economics: Principles in Action* (p. 462). Upper Saddle River, New Jersey 07458: Pearson Prentice Hall. ISBN 0-13-063085-3.

Peña, E. A., & Slate, E. H. (2006). Global validation of linear model assumptions. *J. Am. Stat. Assoc., 101*(473), 341–354. https://doi.org/10.1198/016214505000000637

Pesaran.Shin-An Autoregressive Distributed Lag Modelling Approach to Cointegration Analysis.

Pesaran, M. H., Shin, Y., & Smith, R. J. (2001). Bounds testing approaches to the analysis of level relationships. *J. Appl. Econom., 16*(3), 289–326. https://doi.org/10.1002/jae.616

Phillips, P. C. B., & Perron, P. (1988). Testing for a unit root in time series regression. *Biometrika, 75*(2), 335–346. https://doi.org/10.1093/biomet/75.2.335

Potts, M. (2006). China's one-child policy. *BMJ, 333*(7564), 361–362. https://doi.org/10.1136/bmj.38938.412593.80

Ramsey, J. B. (1969). Tests for specification errors in classical linear least squares regression analysis. *J. Roy. Stat. Soc. B, 31*(2), 350–371. JSTOR 2984219.

Schwarz, Gideon E. (1978). Estimating the dimension of a model. *Ann. Stat., 6*(2), 461–464. https://doi.org/10.1214/aos/1176344136. MR 0468014.

Soper, H. E., Young, A. W., Cave, B. M., Lee, A., & Pearson, K. (1917). On the distribution of the correlation coefficient in small samples. Appendix II to the papers of "Student" and R. A. Fisher. A co-operative study. *Biometrika, 11*(4), 328–413. https://doi.org/10.1093/biomet/11.4.328

Stern, David (2010). *The Role of Energy in Economic Growth, CCEP Working Papers*. Centre for Climate & Energy Policy, Crawford School of Public Policy, The Australian National University. https://EconPapers.repec.org/RePEc:een:ccepwp:0310.

Theil, H. (1950). A rank invariant method of linear and polynomial regression analysis, i, ii, iii. In , *53*. *Proceedings of the Koninklijke Nederlandse Akademie Wetenschappen, Series A Mathematical Sciences* (pp. 386–392), 521-525, 1397-1412.

Tibshirani, R. (1996). Regression shrinkage and selection via the Lasso. *J. Roy. Stat. Soc. B, 58*(1), 267–288.

Todaro, M. P., & Smith, S. C. (2011). *Economic Development 11* (vol. 10). Addison-Wesley, Pearson. ISBN, 0-13.

Utts, J. M. (1982). The rainbow test for lack of fit in regression. *Commun. Stat. Theor. Methods, 11*(24), 2801–2815. https://doi.org/10.1080/03610928208828423

Wu, Yanrui (2003). *China's Economic Growth, A Miracle with Chinese Characteristics*. London, United Kingdom: Routledge.

Yueh, Linda (2013). *China's Growth, the Making of an Economic Superpower.* London, United Kingdom: Oxford University Press.

Zhang, Kevin (2006). China as a world factory. In *Routledge Studies in the Growth Economies of Asia*. New York. USA.

Zheng, Xinli (2018). *China's 40 Years of Economic Reform and Development, How the Miracle Was Created*. Beijing, China: Springer Singapore.

Zou, H. y, & Hastie, T. (2005). Regularization and variable selection via the elastic net. *J. Roy. Stat. Soc. B, 67*(2), 301–320.