Regression Based on Examining Population Forecast Accuracy

A Poongodai, R Suhasini, R Muthukumar

Abstract:*India is the World's second largest populated country after China. The population of India is around 15% of the world while its geographical area is just 2.4 percent of the world. Slowing down population growth andimproving population quality are the key factor in improving India's economic development.For this, population forecasting plays an important role. Population forecasting is widely used for planning and budgeting. Hence, there is need of accurate forecasting. This study investigates the population forecast accuracy by using regression models. It also examines the effect of growth rate and population size on the forecast errors.*

Keywords:*Population Forecast, Forecast accuracy,Population size, Growth Rate, Regression, Forecast error, precision and bias*

I. INTRODUCTION

Population forecast is a projection based on certain assumptionsthat yields a realistic picture of the probable future population.

If projections are estimated for the whole country, they are called total projections. When projections are made for a region/ state, they are called regional or sectoral projections. Total projections are easy to make as compared with regional projections. [2]

In fact, all projections are based on past data for the future population. They are called forward projections. However, if we want to check the accuracy of past population data, projections are made about the past population. These are called backward projections. Backward projections are also made if population census was not done for a period. [2]

In fact, there are many methods are followed to do population forecasting. Whatever method we follow, population forecastinghas become very important in the field of the economy.

State level, population forecasting is widely used for planning and budgeting. Population forecasting helps to identify the trends in the growth of the population between two censuses.

Accordingly, it will mostly be used for the following purposes:

- Economic Development planning
- Social development planning
- Business planning
- Investment Planning
- Infrastructure planning

Hence the forecasting should be accurate enough to support the above-said activities. Such important purposes make it necessary to examine the precision and bias of forecasting.

Countless studies were done in overseas to find out the effect of growth rate and population size on forecast accuracy by analysing forecast errors. These studies measured the population size in the launch year and growth rate over the base period. Some studies found that increase in population improves precision but an increase in the absolute growth rate deteriorates it.[7][14] In the case of bias, some studies found that it showednull/meagre relationship with population size while it showed a positive relationship with the growth rate. [4][9][13] These are generally accepted characteristics of population forecast errors. [5]

In India there are just few studies found on population forecasting methods, but no studies were conducted to evaluate forecast error and forecast accuracy. This paper evaluates population forecast accuracy. Using a data set from 31 states in India for census yearsfrom 1901 to 2011, state wise population forecasts are constructed and forecast errors are calculated. Then, few regression models were developed. The independent variables were population size and growth rate; the dependent variable was forecast error.

The following questions are addressed in this study:

- What is the effect of population size on forecast accuracy and growth rate on forecast accuracy?
- Which is the best regression model that explains the relationship, between forecast accuracy and population size & between forecast accuracy and growth rate?

II. RELATED STUDY

One of the most common statistical techniques in population analysis is regression model.The Regression analysis examines and explains the influence of independent variables on the dependent variableusing a sample population. Often it showed the relationship between independent variables on dependent variable.

Numerous studies used regression analysis in their methodology. In 2014, Weston et al. proposed a model with a regression tree for estimating population density. [15] In 2015, Lama Nayal developed a robust regression model andforecasted the demand of new single family houses. [8]Sunghae Jun and his team members proposed a divided regression model for bigdata analysis to reduce the computation burden. [12]

In 2015, Samir Mazidbhai highlights the problem of

Revised Manuscript Received on June 10, 2019.

A Poongodai, CMR College of Engineering & Technology, Hyderabad, Telangana, India. (Email: apoongodai@cmrcet.org)

R Suhasini, CMR College of Engineering & Technology, Hyderabad, Telangana, India.

R Muthukumar, ICFAI University, Hyderabad, Telangana, India.

population growth in India using the theory of Demographic Transition that explains the effect of economic development on the size of the population and its growth rate.[10]Manjuland their team members studied the population dynamics in India - study of how the population changes over time. [6]

In 2017, Geetanjali D analysed the relationship of income, output, and employment with population growth in the context of India. In this paper, the relationship among the variables is established using Multivariate regression analysis and ARIMA model is used in estimating the population growth. [3]

However, only few studies examined population forecast accuracy using regression analysis. Those studies used aggregate data of individual places for analysis. Taymanetal.analyzed differences in average forecast error based on the aggregate data of population size and growth rate. [11] Lenze used data for individual places for 67 counties in Florida, US, but did not develop and investigate alternative regression models, but analyzeda set of 5-yearforecasts. [1]

In this study,an alternative form of the regression models designed for investigatingto what extent forecast accuracy is affected by growth rate and population size.Also, the direction and size of error covering several time periods and forecast horizons is examined.

III. CHARACTERISTICS OF DATA SETS& RESULTS

This study uses population data that covers almost 35 states in India from the period of 1901 to 2011. 10^{th} percentile is the value (or score) below which 10% of the observations are found. 90% of data values lie below the 90^{th}percentile.Table 1 summarizes the characteristics of population size for all the states in India. The mean population size in 2011 is five times the mean population size in 1901 and the median from 1901 to 2011 is increased around 8 times. The 90th percentile population size grew by 5 times.

Table 1 Characteristics of population size

		Y2011	Y2001	Y1991	Y1981	Y1971	Y1961	Y1951	Y1941	Y1931	Y1921	Y1911	Y1901
Mean		34.5959	29.3925	24.1835	19.5237	15.6617	12.5496	10.2326	9.1046	7.9708	7.1806	7.2027	6.8113
Median		16.7879	13.8505	9.4206	4.2808	3.4604	2.6586	3.2539	.9179	.6362	.4885	.4868	.4755
Percentiles	10	.3658	.3019	.2238	.0241	.0191	.0145	.0230	.0000	.0000	.0000	.0000	.0000
	90	96.4054	81.3051	67.1360	57.8615	46.7521	37.4116	30.5175	27.0154	24.0570	21.5037	21.4583	19.3082

Table 2 Characteristics of population growth rate

		G2001	G1991	G1981	G1971	G1961	G1951	G1941	G1931	G1921	G1911	G1901
Mean		19.2992	26.3855	24.6862	29.7146	32.0589	32.8481	10.9216	13.7745	12.7479	2.7935	8.5972
Median		18.8111	23.3562	25.7306	26.7417	27.9609	23.5955	11.3125	12.7492	11.9412	.0000	6.6972
Percentiles	10	6.6368	14.9585	.0000	10.4986	11.8666	-9.8258	-6.0615	.0000	.0000	-5.5543	-4.6091
	90	28.0014	50.5074	44.7775	51.6622	60.0957	75.8271	23.5188	22.8594	25.9428	13.7601	26.0161

Table 2 summarizes the characteristics of growth rate and population size for all the states in India. Almost in every year, the mean growth rate is higher than the median growth rate.The growth rate of population is irregular and moderate till 1921.

The growth rate was just 0.2 percent from 1901 to 1921 and the population size increased by 16 million for the whole period of 2 decades. But suddenly there wasa huge increase in the population size from 1921 onward. This was mainly due to the decline in the death rate. According to the census commission in the year 1951, the rate of growth in the population was alarming. The population increasedby 110 million from 1921 to 1951 and from the year 1951onwards for the next 40 decades, there was a rapid growth in population size and another 485 million were added to the population.

IV. FORECAST DATA

Two data sets were used for the study of population forecast accuracy: 10-year forecast (launch years from 1911 to 2011) and 20-year forecast from 1921 to 2001. For each and every state, population size and growth rate, forecast for 10- and 20-year horizons are collected. The following two extrapolation techniques were used for obtaining forecasts for each launch year: linear and exponential. In Excel, trend function is used to implement the linear technique and growth function is used as the exponential technique. The forecasts refer solely to population size, and the average of the forecast values calculated from the extrapolation methods.

Forecast error refers to the percentile difference between forecast and the actual population size at the given year. Then, Two errors – Absolute Percent Error (APE) and Algebraic Percent Error (ALPE) are measured. APE is the measure of precision and it ignores the direction of error and ALPE is used to measure bias, accounting for the direction of error.In this work, SPSS is used for population analysis.

V. SIMPLE REGRESSION MODEL

The proposed regression model is used to examine the effect of growth rate and population size on APE and ALPE. A simple regression model is constructed to find the statistical relationship between errors (APE and ALPE) and population size

and among growth rate and errors(APE and

ALPE). In SPSS statistical software, the linear regression model is chosen, dependent (APE and ALPE) and independent variables (population size and growth rate) are selected accordingly and the method selected is "enter" method.

The following hypotheses aretested in this regression analysis:

1. H1: Population size has no effect on absolute percent error
2. H2: Growth rate has no effect on absolute percent error
3. H3: Population size has no effect on the algebraic percent error
4. H4: Growth rate has no effect on the algebraic percent error

The regression coefficient and adjusted R^2 values for the model with onevariable is given in Table 3 and Table 4. Table 3shows APE and Table 4 shows ALPE.

In all the instances of Table 3, the independent variable had the expected sign and was statistically significant except growth rate for 10-year horizon. An increase in the population size reduces the error and increase in the absolute growth rate, raises error. However, as shown,miniscule adjusted R^2 values, the variables did not describe the forecast errorsvariation. The 20-year forecast horizon showed a little effect on either adjusted R^2 values or the regression coefficient.

Table 3 Absolute Percent Error: Regression Coefficient and Adjusted R^2 value

	Horizon length			
	10-year Horizon	Sig Val	20-year Horizon	Sig Val
Population size	-3.363E-7	0.000	-2.498E-7	.033
Adjusted R^2	0.033		0.023	
Growth Rate	-.004	0.966	.093	0.000
Adjusted R^2	-0.003		0.116	

*significant at 0.05

In Table 4, algebraic percent error is shown in detailed manner. Population size showed positive effect on algebraic errors for 20-year forecast horizon, contradicting the hypothesis. However, the regression coefficient is very smaller than coefficient of absolute errors and the adjusted R^2 values were also very small. The results witnessed that population size had little effect on bias. The growth rate showed a negativeimpact on algebraic errors, contradicting the hypothesis. An increase of the forecast horizon length has a slighteffect on regression coefficient and adjusted R^2 values.

Table 4 Algebraic Percent Error: Regression Coefficient and Adjusted R^2 value

	Horizon length			
	10-year horizon	Sig. Val	20-year Horizon	Sig. Val
Population size	1.618E-7	.136	3.059E-7	0.016
Adjusted R^2	0.004		0.031	
Growth Rate	-0.200	.063	-.105	0.000
Adjusted R^2	.007		0.126	

*significant at 0.05

VI. CONCLUSION

In this paper, simple univariate regression model is constructed to relate growth rate and population size with forecast accuracy. It estimates the statistical relationship and effect of growth rate &population size on absolute and algebraic errors. As a future work, the multivariate model can be constructed to improve the performance of simple regression model. Various parameters such as launch year, state division can be included to find their effect on the regression result ofgrowth rate and population size.

REFERENCES

1. D.G Lenze, 2000, "Forecast accuracy and efficiency: an evaluation of ex ante substate long-term forecasts", International Regional Science Review, 23, 201–226.
2. Divisha S, "Population projection: Meaning, Types and Importance", http://www.sociologydiscussion.com/demography/population-projections/population-projections-meaning-types-and-importance/3058", Accessed on 1/10/2018
3. GeetanjaliDiwani, 2017, "Nexus between Population, Income, Output and Employment: Econometric Evidence from India", International Journal of Multidisciplinary and Current Research, ISSN: 2321-3124.
4. J Tayman, 1996, "The accuracy of small area population forecasts based on a spatial interaction land-use modeling system", Journal of the American Planning Association, 62, 85–98.
5. Jeff Tayman, Stanley K. Smith and Stefan Rayer, 2011, "Evaluating Population Forecast Accuracy: A Regression Approach Using County Data", Population Res Policy Rev, Springer, 30:235–262.
6. ManjulMayankPandey, RupamTiwari and AnupamaChoubey, 2015, "Population Dynamics in India",International Journal of Scientific & Engineering Research, Vol 6, Issue-1, ISSN 2229-5518.
7. N Keyfitz, 1981, "The limits of population forecasting", Population and Development Review, 7, 579–593.
8. Nayal, Lama, 2015, "Regression Analysis To Forecast the Demand of New Single Family Houses in USA", Culminating Projects in Mechanical and Manufacturing Engineering, Paper 15.
9. Rayer S, 2008, "Population forecast errors: A primer for planners", Journal of Planning Education and Research, 27, 417–430.
10. Samir MazidbhaiVohra, 2015, "Population Growth – India's Problem", PARIPEX – Indian Journal of Research.
11. Schafer E, Tayman J, Carter L, 1998, "The role of population size in the determination and prediction of population forecast errors: An evaluation using confidence interval for subcounty areas", Population Research and Policy Review, 17(1), 1–20.
12. Seung-Joo, Lee Sunghae Jun and Jea-Bok Ryu, 2015, "A Divided Regression Analysis for Big Data", International Journal of Software Engineering and Its Applications
13. Smith S K, 1987, "Tests of forecast accuracy and bias for county population projections", Journal of the American Statistical Association, 82, 991–1003.
14. Stoto M, 1983, "The accuracy of population projections", Journal of the American Statistical Association, 78, 13–20.
15. Weston Anderson, Seth Guikema, Ben Zaitchik, William Pan, 2014, "Methods for Estimating Population Density in Data-Limited Areas: Evaluating Regression and Tree-Based Models in Peru", PLoS ONE 9(7): e100037