Population Review
Volume 65, Number 1, 2026
Type: Article pp. 82-103

State Level Population Forecasting in India: A Comparative Study of Exponential, Logistic, Linear Regression and Holt-Winters Models

Authors: Sheena K J, sheena.kj2021@vitbhopal.ac.in, ORCID ID: 0009-0009-6743-5301

Corresponding Author: Dr. Jyoti Badge, jyoti.badge@vitbhopal.ac.in,

ORCID ID: 0000-0001-8865-2828

Abstract

Population dynamics has a high scope and importance in view of the prominent increase in demographical numbers across Indian states. This research study discusses four models (Mod.s), notably Exponential, Logistic, Linear Regression (LR), and Holt-Winters for analyzing future population trends and their possible effects on society. Exponential model predicted that the world population will always increase exponentially without considering the factors such as limited resources and thus could not produce accurate predictions. The Logistic model, on the other hand, explained that the population can grow either increasing or decreasing based on the carrying capacity. In contrast, the Linear Regression model predicts future population as per historical data and trends. It accepts a linear relationship across population growth and time, which might not every time be accurate. The Holt-Winters model, a more advanced statistical method, is used for more complex time series data. It can account for seasonal variations and trends, making it suitable for short-term population forecasting. By analysing population data from 20 Indian states, based on the 2001 census, this study highlights the strengths and limitations of each model. The findings aim to equip policymakers and researchers with actionable insights, enabling them to develop more adaptive strategies for managing demographic challenges.

Keywords: Population dynamics; growth rates; Exponential Model; Logistic Model; Holt-Winters; carrying capacity.

Introduction

Diverse population dynamics have been a topic of immense interest and has become an important field of utilizing various models to analyze trends and make projections. This study is related to recognizing the best methods of dealing with the census data from 2001, so the authors applied four different models to it and palm to become the most reliable one to use in predictive analysis of the coming population changes. The study utilizes actual data from the census of 2001 to guarantee the precision of the predictions. Owing to the fact that 2001 and 2011 are the two latest years with official census data, the 2001 data source becomes a reliable means to make an assessment of the observed values in comparison with the forecasted values. This makes it possible for a thorough examination of the models' capacity to really understand population increments. Bacaer, in turn, introduced the Logistic Growth Mod. to the world. Mathematical models have been of utmost importance to humans in planning for the future. One of the first models, the Exponential Growth Mod., was suggested by Thomas Malthus in the late 18th century (Mondal and Kumar, 2018, Mwakisisile and Mushi, 2019). (Malthus, 1803) found that when there are no limits on resources, the population grows exponentially. Malthus' theory is that if left to their own devices, the populations of countries could grow geometrically, i.e., doubling in harmony with the same intervals. However, necessary resources like food and water do not increase that rapidly, but their increase is at a slower pace, arithmetically. This unevenness causes the scarcity of resources and, consequently, the lack of food and diseases, which leads to demographic and social crises. Malthus' point of view was quite negative, as he stressed the essential point that through the uncontrolled expansion of the world's population, the societal sufferings emerged as a result of these calamities. A different explanation of this subject arrived in the 19th century. Verhulst (1977) Introduced the Logistic Growth Model, drawing from Malthus' concepts. This way of thinking includes the carrying capacity - this term describes the most a population an environment holds for a long time. Early on, populations grow fast, like exponential growth. But as resource limits appear, the growth rate lessens. That forms a curve shaped like an S. This model presents a different view - populations become stable rather than growing without end. Food shortage drops a lot. Mathematical models have recently been extensively used for the purpose of projecting population patterns in various regions. Mondal and Kumar (2018) a research work proposed the usage of these models in a high-density populated country Ali et al. (2015) Bangladesh facing resource constraints. Through these models, the researchers acquired a new perspective while observing the demographic shifts and consequently were able to make recommendations on sustainable population management. Several scholars have contributed to the advancement of population growth modelling: Holt (2004) presented results of extensive investigations encompassing the use and restrictions of predictive models in population analysis. A logistic model was first created for bacterial growth by Fujikawa et al. (2003), and then it formed the basis of a model used to interpret the implications of human population expansion under constraints of resources. Mwakisisile and Mushi (2019) Zeroed in on the population scenarios in Tanzania followed by the unique challenges of developing nations caught up in the rapid population explosion. Historical population growth of the U.S. from the year 1790 until now was assessed by Pearl and Reed (1920) and it has been depicted as a mathematical pattern of long-term demographic development in a rich country. Statistical models offer us a solid basis for the understanding of different patterns in population growth and for determining population trends. The models also enable a researcher to gauge the type of connection across variables (var.s), make predictions about the future, and validate the hypotheses. One well-known case of the use of statistical models is in the lawful domains of migration flow or population change. Such models are highly instrumental for understanding one of the core issues like demographic transition. LR is an extremely simple and popular techniques. It is widely utilised to analyze and estimate the link across the dependent var. along with one or even more independent var.s when changes in the target population occur, with time or economic conditions being the factors. Linear regression models indeed provide relatively easy and straightforward means of the prediction of such phenomena, they operate practically in a straight line fashion. The paper by Abdelrahim et al. (2017) was a mathematical and statistical study about the prediction of population growth that was carried out by the research group. Their work showed that linear regression models are highly capable of relating the two var.s, namely, population increasing and influencing factors. Industrializing their approach, formulated a different, newly developed model that was exclusively for forecasting the population of India and its states. The research work complemented the linear regression of predictive methods and pointed out its merits and demerits. When it comes to finding determinants of a population that increases in a straight way, linear regression is very good at that while on some occasions may not be perfect in fully representing the complications of the population growth rates that are not in linear form. The Holt-Winters technique, propounded by Hansun et al. (2019) is a popular standard time series forecasting statistical method. This strategy is a kind of exponential smoothing that is extended to include trends and seasonal variations in data. The method was first developed by (Holt, 2004, Winters, 1960).and is commonly used in areas such as economics, finance, and inventory management thanks to its capability of dealing with complex seasonal impacts. It finds greater use in short-term forecasting, where seasonal factors prominently affect the results. The Holt-Winters approach, which is the method of Holt-Winters, is found practically in all the sectors because of its distinctive capability to show very complex seasonal variations. In fact, it has been used to estimate periodic economic indicators and prices of stocks, as well as sales numbers. Tourism hospitality management Assists in forecasting demand patterns, helping optimize inventory levels to meet seasonal shifts. With this method, companies can project their inventories of products and materials, which allows them to better manage production and logistics depending on the efficiency that will achieve by means of such cooperation. This tact is chiefly beneficial and powerful for time series forecasting of data that exhibit trend and seasonality effects. In addition to the level, and trend, along with seasonal components, Mod. Offers all-in-one access to predictions that are not only beneficial but also a must-have in disciplines such as economics, finance, and inventory optimization. Machine learning methods have been established as necessary tools for modern predictive analytics. A few methods that can be used for predicting human population trends include LR, and decision trees, and artificial neural networks, along with k-nearest neighbours, etc. The topic of machine learning-based population modeling has been the subject of new studies that have compared and discussed the progress made in the field.

1. Otoom (2021) Performed machine learning model comparison with 17 machine learning models.

2. Wang and Lee (2021) and Winters (1960) dealt with LSTM networks for short-run area forecasting.

3. Mahesh and Priyanga (2019) Examined the patterns of human migration after the earthquakes in Nepal through mobile phone data.

4. Rajkumari and C. (2020) Discussed the utilisation of supervised ML algorithms for growth rate prediction of regions and found the best method.

5. Şahinarslan et al. (2021) Applied six different machine learning algorithms to predict Turkey's population.

6. Ashioba and Daniel (2020) Have created a population prediction system through the utilisation of ML techniques in Nigeria. The research describes the increasing importance of ML inside demographic forecasting, demonstrating the different techniques corresponding to the specific regional and methodological problems that they face. This paper which is based on research studies cited in (Table 1) is intended to focus on contrasting mathematical models and statistical models for the sake of better population predictions. I have closely examined the different modelling techniques and chose the exponential, logistic, linear regression, and Holt-Winters models as the modelling methods. Although some machine learning models were suggested but they were not used because the dataset had some constraints.

The following are some of the gaps found in the research that have not yet been filled:

1. The majority of the current studies in this area are concentrated primarily on the field of population dynamics. This creates a niche for conducting deeper investigations into the performance of various models among the demographically diverse Indian states.

2. A more in-depth analysis of policy interventions like family planning initiatives and healthcare strategies can offer an understanding of the population growth trends and also, the models' predictive capabilities for the evaluation of policy impacts.

3. Where linear regression works based upon a relative straight-line approach and fails to consider variables or the straight, independent variables that exist, this may not be the best choice for projecting over time. However, this was not explored here.

4. Polynomial regression or even machine learning options may have better forecasting capabilities.

5. Furthermore, knowing how the forecasting engines performed from state to state in India would provide specific clear advantages for certain socio-economic, demographic, and geographical factors.

6. The assurance of reliability and preciseness of these models in their performance in projecting future population figures is strongly based on validation and calibration of these models, in which the actual data is an indispensable part.

7. The overarching goal of this study is to bridge the identified research pilot gaps by performing an extensive analysis of the mathematical and statistical modelling while also providing the results of their effectiveness concerning different population prediction scenarios.

Table 1: Comparison of Research Studies on Population Forecasting.

Author(s)	Model Used	Dataset	Findings	Limitations
Abdelrahim M. Zabadi et al. (2017)	Exponential model & Verhulst Logistic		Jordan Logistic model performed best	Only two models were used
Mondal et al. (2018)	equation Exponential & Logistic Models	Bangladesh	Logistic model performed best	Only mathematical models were used
Andongwisye John Mwakisisile et al. (2019)	Exponential & Logistic Models	Tanzania	Exponential model performed best	Only two models were compared
Seng Hansun et al. (2019)	Weighted Moving Average and Holt-Winters Additive Model	Time Series Data Library	Holt-Winters revised additive model performed best	Comparison limited to statistical models
R Mahesh et al. (2019)	Na¨ıve Bayes, IBK, and Random Trees	UCI Repository Bayes	Na¨ıve performed best	Performs well only with large datasets
Brintha Rajkumari S. et al. (2020)	LR, SVR, MLP, Decision Trees	Govt. of India Census	Linear Regression performed best	ML methods work accurately with large datasets
Md. Irphan Ahamed et al. (2021)	Curve Fitting (New Model)	15 Indian States model	Proposed model Effective for prediction	Did not account for external factors (e.g., food, calamities, war)
Otoom M. M. (2021)	17 Machine Learning Models	United Nations Database	Random Forest model performed best	Focused only on ML models

Table 1 summarizes key research studies on population forecasting, comparing models, datasets, findings, and limitations. It highlights the evolution from classical mathematical approaches to advanced statistical and machine learning techniques. While logistic and exponential models were common in earlier studies, recent work favors data-driven methods like Random Forest and curve fitting, often yielding better accuracy. Limitations across studies include narrow model comparisons, reliance on large datasets, and lack of consideration for external factors like disasters or policy shifts.

Research Objectives

1. Identify the most effective population Along with the birth rate, death rate, and population the previous year's data for the three models are included in the dataset. The Holt-Winters model contains the three components namely level, trend, and seasonality for population forecasting.

2. It is interesting to note that var.s like migration and urbanization were left out from the analysis because they are complex to interpret.

3. Employ and compare four dissimilar population models to the 2001 census. Data so you can figure out which model is the best at predicting the future population.

4. Analyze the performance of different population models: Evaluate in which ways different population models are due to predicting population growth, particularly in the context of Indian states.

5. Assess the impact of policy interventions: Study how various policy interventions, such as family planning and healthcare programs, influence the population and how the models can predict those effects.

6. Observe non-linear population growth behaviours: Study the capabilities and weaknesses of non-linear models (e.g., polynomial regression, machine learning models) in fitting and forecasting non-linear population growth movements.

7. Carry out the comparison: Measure the performance of the diverse models in various Indian states based on socio-economic and demographic contexts, so as to pinpoint the most effective models for different regions.

8. Adhere to the model's correctness and reliability: Put a spotlight on the real-world data-based validation and calibration of models to achieve their accuracy and reliability in the projection of future population trends.

Methodology

A detailed examination of the dataset and comparative literature reveals that mathematical and statistical models outperform machine learning techniques in contexts with limited temporal granularity. Although machine learning models can yield highly precise forecasts with large datasets, the absence of consistent year-wise raw data in this study led to erratic predicted values (PVs). Therefore, this research focuses exclusively on evaluating two mathematical models and two statistical models.

The dataset includes birth rate, death rate, and population figures, along with prior-year values for each model. Migration and urbanization variables were excluded due to their interpretive complexity and lack of consistent measurement.

Exponential Growth Model

The exponential model is governed by the differential equation:

$$\frac{dN}{dt} = rN \qquad (1)$$

where $r = b - d$ represents the Malthusian growth rate (fertility minus mortality). Upon integration, the population at time t is given by:

$$N(t) = N_0 e^{r(t-t_0)} \qquad (2)$$

Logistic Population Growth Model

The logistic growth model, originally proposed by Pierre Verhulst, is widely used to describe population dynamics when growth slows as the population approaches an upper limit. The standard logistic function is

$$P(t) = \frac{K}{1 + Ae^{-r(t-t_0)}} \qquad (3)$$

Where K denotes the carrying capacity, r is the intrinsic growth rate, A is a scaling constant determined by the initial condition, and t_0 is the base year.

In this study, the parameters K, A, and r were estimated using nonlinear least squares (NLS), implemented in MATLAB through the lsqcurvefit routine. This estimation approach is preferred over algebraic three-point formulas, which are known to be numerically unstable and may yield negative or demographically unrealistic carrying capacities.

For each state, the logistic model was fitted by minimizing the squared error between the observed population values and the model predictions. Parameter bounds were imposed to ensure meaningful demographic solutions ($K > 0$, $r > 0$). The fitted model was subsequently used to generate population forecasts for future years.

Any state for which the NLS procedure produced negative or implausible values of K was excluded from logistic interpretation, as such estimates are not meaningful in demographic applications.

Linear Regression Model

This model expresses population as a linear function of time:

$$p(t) = \beta_0 + \beta_1 t \quad (4)$$

where $p(t)$ is the population at time t, β_0 is the intercept, and β_1 is the rate of change.

Holt's Linear Trend Method

Although the Holt–Winters exponential smoothing framework is widely used for time–series forecasting, its seasonal component is not appropriate for annual population data, which do not exhibit seasonal fluctuations. Therefore, the seasonal smoothing parameter γ was fixed at zero, reducing the model to Holt's Linear Trend Method (double exponential smoothing).

The method consists of two recursive equations:

$$L_t = \alpha P_t + (1 - \alpha)(L_{t-1} + T_{t-1}),$$

$$T_t = \beta(L_t - L_{t-1}) + (1 - \beta)T_{t-1}, (5)$$

where L_t is the level, T_t is the trend, α is the level smoothing parameter, and β is the trend smoothing parameter. The h-step–ahead forecast is given by

$$\hat{P}_{t+h} = L_t + hT_t. \quad (6)$$

The parameters α and β were estimated automatically in MATLAB using nonlinear least squares optimization, ensuring the best fit to the observed population series for each state. This formulation avoids the inappropriate use of seasonal Holt–Winters smoothing and provides a statistically valid trend–based forecasting model for annual demographic data.

Implementation

Original Census data from 2001 and Census projections from 2002 to 2011 were applied to all four models using MATLAB for 20 Indian states. Using projections from the 2011 Govt.of India census, projections were calculated using all four models from 2012 to 2022. Comparative analysis of the output enabled model benchmarking and informed the conclusions presented in subsequent sections.

Data and Calculations

Population forecasts for 20 Indian states spanning the period 2012–2022 were generated using 2001 Census data as the baseline. Five predictive models Exponential, Logistic, Linear Regression, and Holt-Winters were implemented via MATLAB to estimate year-wise population trajectories. Each model's output was compared against actual Census data to assess accuracy and determine the most suitable forecasting approach.

Tables 2 to 5 present detailed estimations across all five models for the 20 states over the specified period. This comparative analysis aims to identify the model that delivers the most reliable and demographically plausible forecasts.

The dataset was sourced from the official Census of India repository, encompassing population records from the 2001-2025 projection report and 2011-2036 projection report for twenty states. Prior to analysis, missing values and anomalies were identified and corrected to ensure data integrity. The cleaned dataset was then uniformly structured year-wise, integrating both observed Census values and model-generated projections.

The methodology utilizes a training-testing split designed to validate the generalizability of the forecasting models. While 15 states served as the primary training ground, five states Jammu and Kashmir, NCT of Delhi, Uttar Pradesh, Jharkhand, and Karnataka were withheld for out-of-sample testing. The selection of these five states is justified by three primary criteria:

Geographic and Socio-Economic Heterogeneity: The testing group encompasses a wide spectrum of the Indian landscape, including the highly urbanized administrative hub (Delhi), the high-population agrarian heartland (Uttar Pradesh), the resource-rich tribal regions (Jharkhand), the technology-driven southern corridor (Karnataka), and the unique demographic profile of the northern mountainous region (Jammu and Kashmir).

Historical Accuracy Benchmarking: These states were chosen following a comparative analysis of four model predictions against official projections and census data (2001–2022). By testing on states that exhibited diverse growth patterns during the 2011-2022 interim period, we ensured the models could handle both linear and non-linear demographic shifts.

Model Stress-Testing: Training on 15 states allowed the model to learn broad national trends, while testing on these specific five states evaluated the model's resilience against regional outliers and varying levels of data volatility.

Key demographic and socio-economic variables such as birth rate, death rate, and historical population were selected to align with the input requirements of each model. Variables like migration and urbanization were excluded due to interpretive complexity and inconsistent availability.

The 2001–2011 period was selected because it represents the most recent decade for which complete and consistent census data are available for all Indian states. This interval provides a uniform, high-quality dataset suitable for model calibration without missing values or definitional changes. Although the training window spans ten years, short-term demographic trends in India tend to be stable, allowing reliable extrapolation for the subsequent decade. The forecasts for 2012–2022 are therefore based on the most recent validated census inputs, while the 'actual' values used for comparison are taken from official post-censal population projections released by government statistical agencies.

To ensure robustness, a validation protocol was applied to eliminate illogical or negative predictions. For model training and evaluation, 15 states were designated as the training set, while the remaining five served as the test set to assess generalizability and performance. The five states used for testing were Jammu Kashmir, NCT of Delhi, Uttar Pradesh, Jharkhand, and Karnataka. Remainaing 15 states were used for training.
Source: Historical dataset from the Census of India: *Population Projection Report 2001-2025*
Source: Historical dataset from the Census of India: *Population Projection Report 2011–2036*

Results and Discussion

Population forecasts from 2012 to 2022 were generated using four models Exponential, Logistic, Linear Regression, and Holt-Winters applied to 2001 Census data. The accuracy of each model was assessed by comparing predicted values (PVs) against actual Census figures.

Exponential Model: Tamil Nadu Example

Let the initial population $P_0 = 62{,}406$ in 2001 and $P_1 = 67{,}444$ in 2011. The growth rate r is calculated as:

$$r = \frac{\ln(P_1/P_0)}{t_1 - t_0} = \frac{\ln(67444/62406)}{10} = 0.0077636 \tag{7}$$

Predicted population for 2012:

$$P_{2012} = P_0 \cdot e^{r(2012-2001)} = 62406 \cdot e^{0.0077636\cdot 11} \approx 67970 \tag{8}$$

Logistic Model: Jammu & Kashmir Example

For illustration, the logistic model was fitted to the population values of Jammu & Kashmir for the years 2001–2003, where $N_{2001} = 10{,}144$, $N_{2002} = 10{,}301$, and $N_{2003} = 10{,}461$. The standard logistic function is

$$P(t) = \frac{K}{1 + Ae^{-r(t-t_0)}}, \tag{9}$$

where K is the carrying capacity, A is a scaling constant, r is the intrinsic growth rate, and $t_0 = 2001$ is the base year.

The parameters K, A, and r were estimated using nonlinear least squares (NLS) in MATLAB. The fitted values for Jammu & Kashmir were:

$$K = 80{,}000, \quad A = 6.8866, \quad r = 0.017660.$$

Substituting these into the logistic equation gives the predicted population for year t:

$$P(t) = \frac{80{,}000}{1 + 6.8866\, e^{-0.017660(t-2001)}}.$$

For example, the prediction for 2004 is

$$P(2004) = \frac{80{,}000}{1 + 6.8866\, e^{-0.017660\times 3}} \approx 10{,}620.90.$$

These NLS-based estimates ensure that the logistic model remains demographically meaningful, avoiding the negative or unstable carrying capacity values produced by algebraic three-point formulas.

Linear Regression and Holt-Winters Models

For both models, 10 years of population data were used. MATLAB was employed to fit the models using cross-validation. Predictions for future years were generated based on the fitted parameters.

Model Comparison: 2012 Forecasts

Table 2 shows that the Holt-Winters model provides the closest predictions for most states. However, Linear Regression performs best for Jammu & Kashmir, and Exponential Smoothing aligns well with NCT of Delhi. These variations highlight the importance of model selection based on regional characteristics.

Table 2: Prediction Comparison of 4 Models – 2012

Sl. No.	Indian States	Actual	Exponential	Logistic	Linear Regression	Holt-Winters
1	Jammu & Kashmir	11,865	11,888	11,944	11,845	11,839
2	Himachal Pradesh	6,856	6,869	6854	6,876	6,855
3	Punjab	27,981	28,034	28,050	28,041	27,999
4	Haryana	25,854	25,914	25,658	25,884	25,861

5	NCT of Delhi	18,983	18,988	17,669	18,824	18,959
6	Rajasthan	68,892	69,080	69,607	69,041	68,929
7	Uttar Pradesh	204,250	204,590	196,526	204,140	204,340
8	Bihar	99,020	99,329	103,760	99,427	99,019
9	Assam	30,945	30,989	31,474	30,986	30,931
10	West Bengal	90,320	90,489	93,320	90,625	90,231
11	Jharkhand	31,904	31,964	33,108	31,967	31,873
12	Odisha	41,105	41,167	42,288	41,205	41,090
13	Chhattisgarh	24,585	24,630	25,45	24,626	24,576
14	Madhya Pradesh	73,344	73,506	72,711	73,449	73,328
15	Gujarat	59,800	59,927	60,484	59,933	59,794
16	Maharashtra	114,184	114,370	113,082	114,280	114,250
17	Andhra Pradesh	85,491	85,638	80,341	85,726	85,477
18	Karnataka	60,026	60,119	61,110	60,147	60,019
19	Kerala	34,802	34,848	34,199	34,877	34,811
20	Tamil Nadu	67,862	67,970	70,201	68,070	67,852

Table 2 presents a comparative analysis of population predictions for the year 2012 across 20 Indian states using four forecasting models: Exponential, Logistic, Linear Regression, and Holt-Winters. The actual population values are juxtaposed with model-generated estimates to assess predictive accuracy. While traditional models like Exponential and Logistic show moderate deviations, the Holt-Winters method consistently yields predictions closest to actual values across most states, indicating its superior performance in short-term forecasting. This comparison highlights the importance of model selection in demographic projections, especially when precision is critical for planning and policy decisions.

Model Comparison: 2015 Forecasts

Table 3 reinforces earlier observations. Exponential Smoothing yields the most accurate predictions for NCT of Delhi and Jharkhand, while Linear Regression aligns closely with Uttar Pradesh. Holt-Winters remains the most consistent performer across the remaining states.

Table 3: Prediction Comparison of 4 Models – 2015

Sl. No.	Indian States	Actual	Exponential	Logistic	Linear Regression	Holt-Winters
1	Jammu & Kashmir	12,289	12,414	12,456	12,307	12,290
2	Himachal Pradesh	7,037	7,102	7031	7,090	7,034
3	Punjab	28,846	29,129	28,932	29,040	28,925
4	Haryana	27,079	27,392	26,763	27,174	27,111
5	NCT of Delhi	20,676	20,694	18,480	20,199	20,568
6	Rajasthan	71,973	72,971	72,945	72,444	72,138
7	Uttar Pradesh	214,671	216,530	203,921	214,520	215,090
8	Bihar	102,732	104,320	109,584	103,840	102,730
9	Assam	32,069	32,289	32,685	32,157	32,006
10	West Bengal	92,725	93,525	96,573	93,405	92,330
11	Jharkhand	33,652	33,489	34,832	33,319	33,063
12	Odisha	42,138	42,444	43,660	42,386	42,073
13	Chhattisgarh	25,555	25,780	26,484	25,651	25,514
14	Madhya Pradesh	76,745	77,569	75,838	77,001	76,671
15	Gujarat	62,081	62,733	62,954	62,437	62,057

16	Maharashtra	118,652	119,670	116,923	119,030	118,930
17	Andhra Pradesh	87,662	88,406	81,065	88,281	87,600
18	Karnataka	61,797	62,269	63,126	62,116	61,767
19	Kerala	35,473	35,716	34,690	35,695	35,517
20	Tamil Nadu	69,030	69,571	72,184	69,579	68,984

Table 3 compares the population predictions for 2015 across 20 Indian states using four forecasting models: Exponential, Logistic, Linear Regression, and Holt-Winters. The actual population values are listed alongside each model's estimates to assess their predictive accuracy. As in earlier years, Holt-Winters consistently provides the closest approximation to actual figures for most states, reinforcing its reliability in short-term forecasting. The table highlights variations in model performance across regions and underscores the importance of selecting context-appropriate methods for demographic estimation.

The actual and the projected populations for the 20 Indian states in 2015 are shown in Table 3. The observations earlier made get the table's proof. It appears that NCT of Delhi and Jharkhand are the two states that Exponential Smoothing predictions got the most accurate for, while the actual population for Uttar Pradesh was found to be more closely aligned with the Linear Regression predictions. However, the Holt-Winters model is the most accurate with the rest of the states as highlighted in Table 1. Exponential model predictions are remarkably closer to those of NCT of Delhi is where the geographical area is the smallest mayor. In such cases, Exponential Smoothing models often yield better predictions compared to other methods.

The actual and predicted populations in Delhi in 2018 are shown in Table

4. Again, we observe consistent trends: the exponential forecasts for NCT of Delhi are the most accurate compared to the actual figures. In the case of Uttar Pradesh and Jharkhand, the linear regression equations yield lower errors, while for all other states, the Holt-Winters model provides the highest precision predictions. These patterns reaffirm the observations made in Tables 1 and 2.

Table 4: Prediction Comparison of 4 Models – 2018

Sl. No.	Indian States	Actual	Exponential	Logistic	Linear Regression	Holt-Winters
1	Jammu & Kashmir	12,665	12,963	12,975	12,768	12,737
2	Himachal Pradesh	7,206	7,343	7194	7,305	7,199
3	Punjab	29,625	30,267	29,772	30,040	29,822
4	Haryana	28,253	28,954	27,823	28,463	28,345
5	NCT of Delhi	22,523	22,553	19,204	21,575	22,270
6	Rajasthan	74,884	77,081	76,202	75,848	75,270
7	Uttar Pradesh	224,829	229,150	211,027	224,910	225,980
8	Bihar	106,192	109,550	115,542	108,260	106,160
9	Assam	33,166	33,643	33,856	33,328	33,044
10	West Bengal	95,109	96,663	99,697	96,185	94,156
11	Jharkhand	34,887	35,085	36,588	34,671	34,188
12	Odisha	43,132	43,760	44,987	43,566	42,977
13	Chhattisgarh	26,488	26,984	27,741	26,677	26,418
14	Madhya Pradesh	80,042	81,856	78,874	80,554	79,931
15	Gujarat	64,222	65,670	65,358	64,941	64,224
16	Maharashtra	122,926	125,210	120,559	123,770	123,590
17	Andhra Pradesh	89,691	91,264	81,676	90,836	89,551
18	Karnataka	63,435	64,496	65,059	64,086	63,427

19	Kerala	36,062	36,606	35,131	36,514	36,178
20	Tamil Nadu	70,047	71,211	74,118	71,089	69,966

Table 4 compares population predictions for the year 2018 across 20 Indian states using four forecasting models: Exponential, Logistic, Linear Regression, and Holt-Winters. The actual population values are listed alongside each model's estimates to evaluate predictive accuracy. As observed in previous years, the Holt-Winters method continues to deliver the closest estimates to actual figures for most states, reinforcing its consistency and reliability in short-term demographic forecasting. The table highlights how model performance varies across regions and underscores the importance of selecting context-sensitive approaches for accurate population projections.

Table 5 is a summary of the actual population and the projected population for the year 2022. Proving his capability, the NCT of Delhi strongly asserts the accuracy of the Exponential model predictions, marking it as the best performing region. Apart from this, Uttar Pradesh, West Bengal, and Jharkhand were seen to make close predictions as per the Linear Regression model. Holt-Winters (HW) Mod. Looks to be best for all other states, keeping the same track as other tables do. To sum up, it is evident from the tables that out of the four models tried for the 2011 Census data, the HW Mod. Has been the most accurate one most of the time. The only exceptions are NCT of Delhi, where Exponential Smoothing performed better, and states such as Uttar Pradesh, West Bengal, and Jharkhand, which are closer to Linear Regression predictions. This discussion reveals that Holt-Winters is ideal for large areas and at the same time recommends Exponential Smoothing for small regions.

Table 5: Prediction Comparison of 4 Models – 2022

Sl. No.	Indian States	Actual	Exponential	Logistic	Linear Regression	Holt-Winters
1	Jammu & Kashmir	13086	13733	13677	13383	13327
2	Himachal Pradesh	7408	7677	7388	7590	7396
3	Punjab	30542	31854	30825	31372	30972
4	Haryana	29720	31176	29160	30182	29965
5	NCT of Delhi	25162	25294	20037	23409	24685
6	Rajasthan	78521	82923	80385	80387	79326
7	Uttar Pradesh	237676	247150	220011	238760	240720
8	Bihar	110410	116950	123665	114150	110290
9	Assam	34495	35537	35347	34890	34368
10	West Bengal	98075	101010	103645	99892	96166
11	Jharkhand	36375	37333	38971	36473	35588
12	Odisha	44349	45580	46,684	45141	,44058
13	Chhattisgarh	27605	28678	29440	28044	27568
14	Madhya Pradesh	84111	87943	82760	85290	84147
15	Gujarat	66774	69801	68443	68279	66964
16	Maharashtra	128398	133000	125076	130100	129770
17	Andhra Pradesh	92111	95218	82342	94243	91883
18	Karnataka	65295	67590	67499	66711	65502
19	Kerala	36722	37827	35649	37605	36988
20	Tamil Nadu	71101	73457	76615	73102	71040

Table 5 presents a comparative analysis of population predictions for the year 2022 across 20 Indian states using four forecasting models: Exponential, Logistic, Linear Regression, and Holt-Winters. The actual population figures are compared against each model's estimates to assess accuracy. As in previous years, the Holt-Winters method consistently delivers predictions closest to the actual values for most states, *reaffirming its robustness in short-term forecasting. The table highlights how traditional models tend to overestimate or underestimate in certain regions, while data-driven approaches like Holt-Winters offer more reliable performance across diverse demographic contexts.*

Figure 1 illustrates the population predictions for Jammu & Kashmir from 2012 to 2022 using four models Exponential Smoothing, Logistic, Linear Regression, and Holt-Winters compared against actual Census data. It is evident that the Holt-Winters prediction line closely aligns with the actual population trend, showing minimal deviation. The Linear Regression model also performs well, with its prediction line nearly overlapping both the Holt-Winters and actual data lines. In contrast, the Exponential and Logistic models exhibit noticeable divergence from the actual values, particularly in later years, indicating lower predictive accuracy for this region.

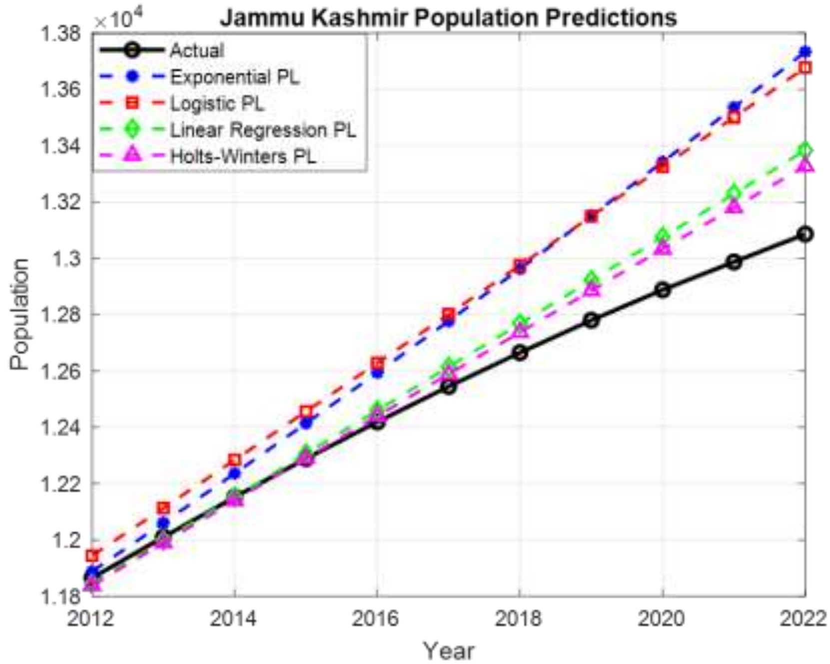


Figure 1: Comparison of Actual Values and Predictions from Different Models – Jammu & Kashmir.

The figure compares actual population values from 2012 to 2022 with predictions generated by four forecasting models: Exponential, Logistic, Linear Regression, and Holt-Winters. All models show an upward trend over time, reflecting population growth. The Exponential and Logistic models tend to overestimate compared to actual values, while Linear Regression and Holt-Winters closely follow the observed data. Among them, Holt-Winters consistently aligns most closely with actual values, highlighting its effectiveness in short-term demographic forecasting. This visual comparison underscores the varying accuracy of models and the importance of selecting context-appropriate methods for reliable projections.

Figure 2 compares the models with the actual population for NCT of Delhi. As seen in the figure, the Exponential prediction line nearly coincides with the actual population line, except for a slight deviation between 2021 and 2022. The Holt-Winters prediction line is also close to the actual values (Avs) but does not coincide completely. The remaining two models show predictions that are far from the actual population.

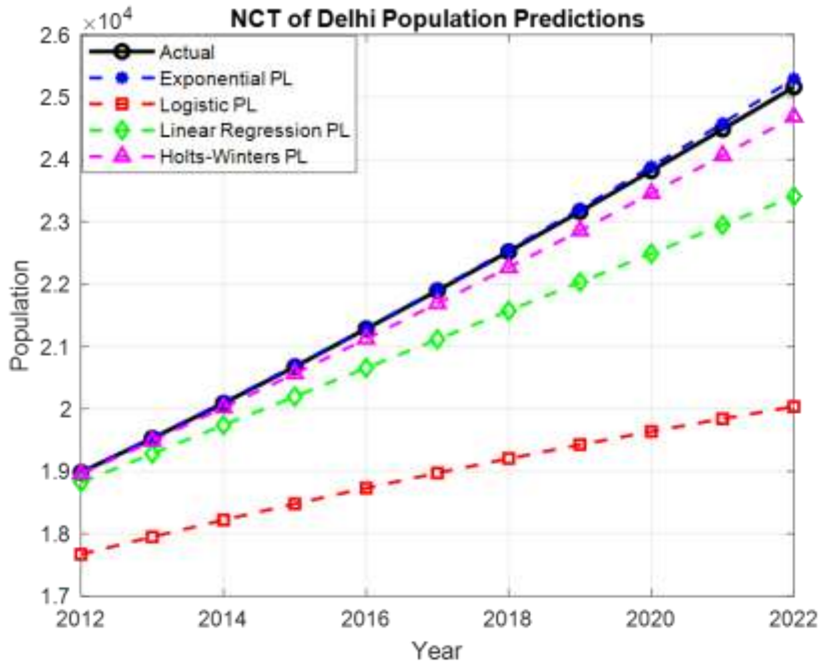


Figure 2: Comparison of All Models – NCT of Delhi.

The figure illustrates the population prediction trends for NCT of Delhi from 2012 to 2022 using four forecasting models Exponential, Logistic, Linear Regression, and Holt-Winters compared against actual values. While all models show an upward trajectory, the Exponential and Logistic models tend to overestimate population figures. In contrast, Linear Regression and Holt-Winters closely follow the actual data, with Holt-Winters demonstrating the most consistent alignment. This comparison highlights the relative accuracy of each model and reinforces the reliability of Holt-Winters for short-term regional forecasting.

Figure 3 provides a comparison of all models with the actual population for Uttar Pradesh. The actual population line coincides with the Linear Regression prediction line. The Holt- Winters prediction line is also very close to the AVs. However, the Exponential along with Logistic Mod. Predictions are noticeably far from operative population.

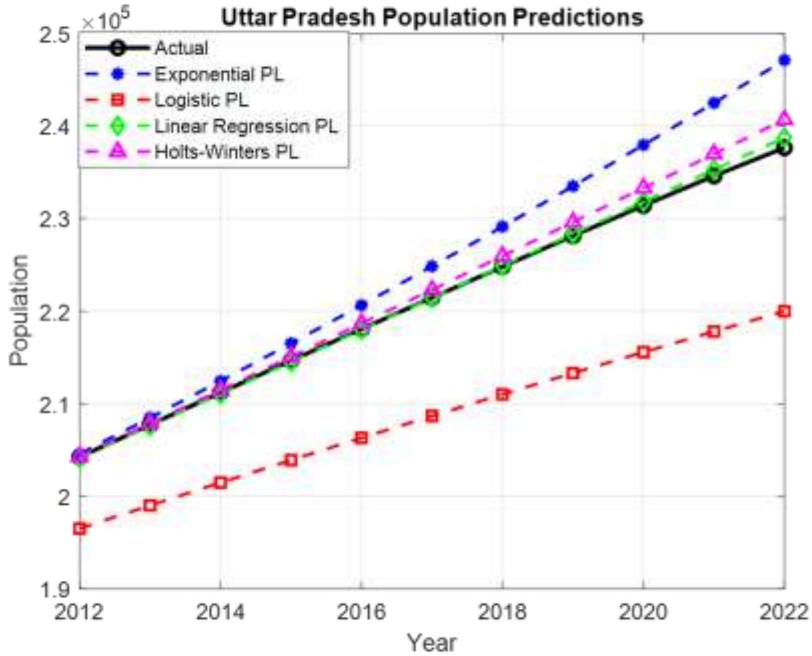


Figure 3: Population Predictions for Uttar Pradesh (2012–2022).

The figure compares actual population data for Uttar Pradesh from 2012 to 2022 with predictions from four models: Exponential, Logistic, Linear Regression, and Holt-Winters. All models show a rising trend, reflecting population growth over time. However, the Exponential and Logistic models consistently overpredict compared to actual values. Linear Regression follows the observed data more closely, while Holt-Winters provides the most accurate alignment throughout the period. This comparison highlights the relative performance of each model and reinforces Holt-Winters as a reliable choice for short-term regional forecasting.

Figure 4 illustrates the actual population data for Jharkhand alongside predictions from five models over the period 2012 to 2022. All models show a steady upward trend, with varying degrees of alignment. The Exponential and UOGI models tend to slightly overestimate the population, while the Holt-Winters and Linear Regression models track the actual values more closely. This comparison highlights the relative performance of each model in capturing demographic dynamics for Jharkhand. However, it gradually moves slightly closer to the Exponential model before returning to align with the Linear Regression model toward the end. This indicates a varied trend. Meanwhile, the predictions from the other two models remain far from the actual population throughout.

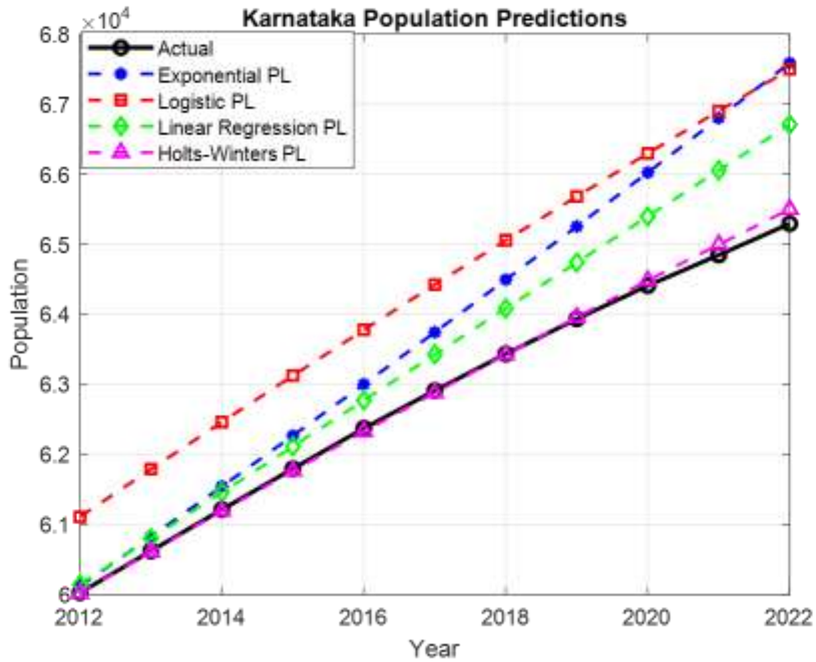


Figure 4: Population Predictions for Jharkhand (2012–2022).

The figure compares actual population data for Jharkhand from 2012 to 2022 with predictions from four models: Exponential, Logistic, Linear Regression, and Holt-Winters. All models show a gradual upward trend, reflecting population growth over the decade. The Exponential and Logistic models tend to overestimate values, while Linear Regression stays closer to actual data. Holt-Winters consistently provides the most accurate predictions, closely tracking observed values across all years. This comparison highlights the relative performance of each model and reinforces Holt-Winters as a reliable method for short-term regional forecasting.

Figure 5 compares the actual population data for Karnataka with predictions from five forecasting models over the period 2012 to 2022. All models show a consistent upward trend, with the Exponential and UOGI models slightly overestimating the population. The Holt-Winters and Linear Regression models align more closely with the actual values, demonstrating better predictive accuracy. This comparison highlights the relative performance of each model in capturing Karnataka's demographic growth trajectory.

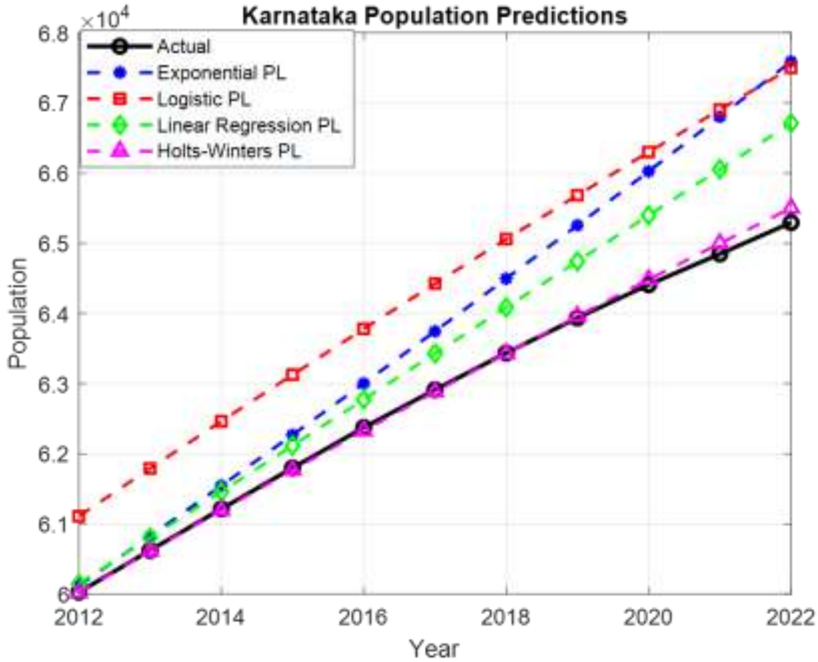


Figure 5: Model Comparison for Karnataka (2012–2022).

The figure compares actual population data for Karnataka with predictions from four models Exponential, Logistic, Linear Regression, and Holt-Winters over the period 2012 to 2022. All models reflect a steady upward trend, consistent with observed growth. However, the Exponential and Logistic models tend to overestimate values, while Linear Regression and Holt-Winters track the actual data more closely. Among them, Holt-Winters shows the most consistent alignment, reinforcing its effectiveness for short-term regional forecasting.

Upon analyzing the above tables and graphs, it is evident that for most states, except a few such as NCT of Delhi, Jharkhand, and Uttar Pradesh, the Holt-Winters model provides predictions that are closest to the actual population when compared to the other models. To gain deeper clarity and confirmation of these findings, it is essential to examine the error metrics used for model evaluation in detail. This analysis will help solidify the conclusions and provide further insights into the performance of the prediction models. To compare the population predictions of all four models, three error metrics are used: Mean Absolute Error (MAE), Mean Squared Error (MSE), and Root Mean Squared Error (RMSE). These metrics are computed for each model across all states to evaluate forecasting accuracy.
MAE measures the average magnitude of prediction errors, regardless of direction. It is defined as:

$$\text{MAE} = \frac{1}{n}\sum_{t=1}^{n} |y(t) - y'(t)| \qquad (10)$$

Where n is the number of observations, $y(t)$ is the actual value (AV), and $y'(t)$ is the predicted value (PV).

Table 6: Mean Absolute Error (MAE) of Forecasting Models across Indian States

Sl. No.	Indian States	Exponential	Logistic	Linear Regression	Holt-Winters
1	Jammu & Kashmir	273.55	287.81	109.18	80.00
2	Himachal Pradesh	122.64	10.4545	89.46	5.91
3	Punjab	578.27	144	377.64	181.18
4	Haryana	636.82	386.63	198.09	91.09

5	NCT of Delhi	40.73	3040.1818	848.55	220.64
6	Rajasthan	1965.30	1227.5455	866.91	345.09
7	Uttar Pradesh	3982.10	12763.5455	285.55	1125.70
8	Bihar	2992.00	8679.6364	1857.30	41.82
9	Assam	443.82	673.4545	165.18	91.73
10	West Bengal	1386.20	4318.4545	9387.00	850.91
11	Jharkhand	315.55	1704.9091	154.55	526.27
12	Odisha	563.09	509.00	396.36	136.27
13	Chhattisgarh	460.09	1747.9091	187.45	46.73
14	Madhya Pradesh	1662.50	1049.9091	501.55	67.18
15	Gujarat	1324.50	1097.0909	679.55	52.18
16	Maharashtra	2045.50	2178.0909	760.00	596.36
17	Andhra Pradesh	1410.20	7514.8182	1042.80	117.82
18	Karnataka	982.36	1569.9091	622.73	55.00
19	Kerala	489.82	866.1818	411.55	108.27
20	Tamil Nadu	1053.10	3822.3636	957.45	56.82

Table 6 presents the Mean Absolute Error (MAE) values for four forecasting models Exponential, Logistic, Linear Regression, and Holt-Winters across 20 Indian states. The Holt-Winters model consistently yields the lowest MAE in most cases, indicating superior predictive accuracy, especially in states like Himachal Pradesh, Bihar, Chhattisgarh, and Gujarat. In contrast, the Logistic and Linear Regression models show higher error margins in several states, with particularly large deviations in West Bengal and NCT of Delhi. These results reinforce Holt-Winters as the most reliable model for short-term population forecasting across diverse regional contexts.

To evaluate the accuracy of population predictions, Mean Squared Error (MSE) is used to measure the average of the squared differences between actual values and predicted values. It penalizes larger errors more heavily than MAE and is defined as:

$$\text{MSE} = \frac{1}{n}\sum (y(t) - y'(t))^2 \qquad (11)$$

where n is the number of observations, $y(t)$ is the actual value (AV), and $y'(t)$ is the predicted value (PV).

Table 7: Mean Absolute Error (MSE) of Forecasting Models Across Indian States

Sl. No.	Indian States	Exponential	Logistic	Linear Regression	Holt-Winters
1	Jammu & Kashmir	114800	109732.72	23508	12378
2	Himachal Pradesh	21624	147.1818	10638	50.64
3	Punjab	497230	25824	204010	50723
4	Haryana	603760	162968.8182	58382	14204
5	NCT of Delhi	3162	10716625.0909	982170	70681
6	Rajasthan	5654000	1637805.3636	1046500	177080
7	Uttar Pradesh	24223000	173015480.2727	176160	2132600
8	Bihar	12867000	82606267.0909	4564400	4015
9	Assam	295300	462802.9091	39301	10509
10	West Bengal	2670200	19288477.5455	793850000	1080400
11	Jharkhand	183640	3115562.7273	34423	361860
12	Odisha	452370	314060	203730	27199
13	Chhattisgarh	317780	3184782.2727	50956	2625
14	Madhya Pradesh	4116800	1160518.0909	365680	5753

15	Gujarat	2614500	1303173.8182	656320	6455
16	Maharashtra	6168200	5237257.0000	841120	531160
17	Andhra Pradesh	2868600	58631932.8182	1446500	19389
18	Karnataka	1457900	2588649.5455	558840	6885
19	Kerala	354630	772800	236320	18636
20	Tamil Nadu	1620400	15616030.1818	1241600	3869

Table 7 presents the Mean Squared Error (MSE) values for four forecasting models—Exponential, Logistic, Linear Regression, and Holt-Winters across 20 Indian states. The Holt-Winters model consistently achieves the lowest MSE in most cases, indicating its superior accuracy and stability in population prediction. In contrast, the Logistic and Linear Regression models show significantly higher error values in several states, with extreme deviations observed in West Bengal and NCT of Delhi. These results reinforce the robustness of Holt-Winters for short-term regional forecasting and highlight the variability in model performance across different demographic contexts.

Root Mean Squared Error (RMSE) is a widely used metric for evaluating the accuracy of numerical predictions. It is calculated as the square root of the average of the squared differences between actual values and predicted values. RMSE penalizes larger errors more heavily and is defined as:

$$\text{RMSE} = \sqrt{\frac{1}{n}\sum_{t=1}^{n}(y(t) - y'(t))^2} \qquad (12)$$

where n is the number of observations, $y(t)$ is the actual value (AV), and $y'(t)$ is the predicted value (PV).

Table 8: Root Mean Squared Error (RMSE) of Forecasting Models Across Indian States

Sl. No.	Indian States	Exponential	Logistic	Linear Regression	Holt-Winters
1	Jammu & Kashmir	338.83	331.25	153.32	111.26
2	Himachal Pradesh	147.05	12.1319	103.14	7.12
3	Punjab	705.15	160.6985	451.67	225.22
4	Haryana	777.02	403.6940	241.62	119.18
5	NCT of Delhi	56.23	3273.6257	991.05	265.86
6	Rajasthan	2377.80	1279.7677	1023.00	420.80
7	Uttar Pradesh	4921.70	13153.5349	419.71	1460.30
8	Bihar	3587.10	9088.7990	2136.40	63.36
9	Assam	543.41	680.2962	198.24	102.51
10	West Bengal	1634.10	4391.8649	28175.00	1039.40
11	Jharkhand	428.53	1765.0957	185.53	601.55
12	Odisha	672.59	560.41	451.36	164.92
13	Chhattisgarh	563.72	1784.5958	225.74	51.23
14	Madhya Pradesh	2029.00	1077.2735	604.71	75.85
15	Gujarat	1616.90	1141.5664	810.14	80.35
16	Maharashtra	2483.60	2288.5054	917.12	728.81
17	Andhra Pradesh	1693.70	7657.1491	1202.70	139.24
18	Karnataka	1207.40	1608.9281	747.56	82.98
19	Kerala	595.50	879.0904	486.13	136.51
20	Tamil Nadu	1273.00	3951.7123	1114.30	62.20

Table 8 presents the Root Mean Squared Error (RMSE) values for four forecasting models Exponential, Logistic, Linear Regression, and Holt-Winters across 20 Indian states. The Holt-Winters *model consistently achieves the lowest RMSE in most cases, indicating its superior accuracy and minimal deviation from actual values. In contrast, the Logistic and Linear Regression models show higher error margins in several states, with extreme outliers such as West Bengal and NCT of Delhi. These results reinforce Holt-Winters as the most reliable model for short-term population forecasting, offering stable performance across diverse regional contexts.*

As observed from tables 6, 7 and 8, out of the four models, Holts-Winters model has got smaller values for MAE, MSE and RMSE. The smaller the errors, the better the predictions. As we can clearly observe from the tables that the MAE, MSE along with RMSE values are less likened to the values obtained for the other models. To have better clarity, an analysis of error methods is done by finding the relative errors for Holts-Winters model. Table 8 shows the relative error analysis of this model. The relative errors are less than 5 for maximum states. This indicates good prediction and Holts-Winters model is the better model out of all the four models.

Why Holt-Winters best?

To sum up, the tables demonstrate unambiguously that of the four models that were utilised in the 2011 Census data, the Holt-Winters model was the only one that got the most benefits in the majority of states. The Holt-Winters technique is specifically tailored to consider seasonality and trends in time-series data. Indian states that have a relatively stable seasonal pattern or a clear growth trend may be able to follow the principles of the method. It provides more weight to the recent patterns, so it can well be used for the real-world data being fluctuating. According to linear regression, population growth is directly proportional to amount of time (or the input var.s) that one has. The two states like Uttar Pradesh and Jharkhand, are in a non- linear trend and thus it is a good fit. The rapid increase in the number of people in Delhi reflects its position as a major urban and economic centre, which draws in many people from all over the country and the nearby areas. Although the logistic model is grounded in the carrying capacity assumption, being the largest possible sustainable population size and it is portrayed by an S-shaped growth curve in the case of logistic growth, the fact that the population of many Indian states does not appear to show this" levelling off" phase yet may bring about an imbalance in the estimations.

Other points to be looked at

Each Indian state is unique in terms of migration patterns, urbanization rates, fertility rates, and socio-economic factors. A single model might not perform uniformly across all states.

Differences in the predictive accuracy could also stem from data inconsistencies, missing values, or regional anomalies in the historical population data. Prediction accuracy varies depending on whether forecasting short-term or long-term trends.

Table 9: Relative Error Metrics – Holt-Winters Model Across Selected States

States	MAE	Rel. Error (MAE)	MSE	Rel. Error (MSE)	RMSE	Rel. Error (RMSE)
Jammu & Kashmir	80.00	6.55	12,378.00	9.11	111.26	9.11
Himachal Pradesh	5.9091	1.07	50.636	1.29	7.1159	1.29
Punjab	181.18	7.07	50,723.00	8.79	225.22	8.79

Table 9 presents the absolute and relative error metrics MAE, MSE, and RMSE for the Holt-Winters *model across three Indian states. Himachal Pradesh shows the lowest relative errors across all metrics, indicating high predictive accuracy. Jammu & Kashmir and Punjab exhibit moderately higher relative errors, though still within acceptable bounds for short-term forecasting. These results reinforce the model's reliability, with performance varying slightly by region.*

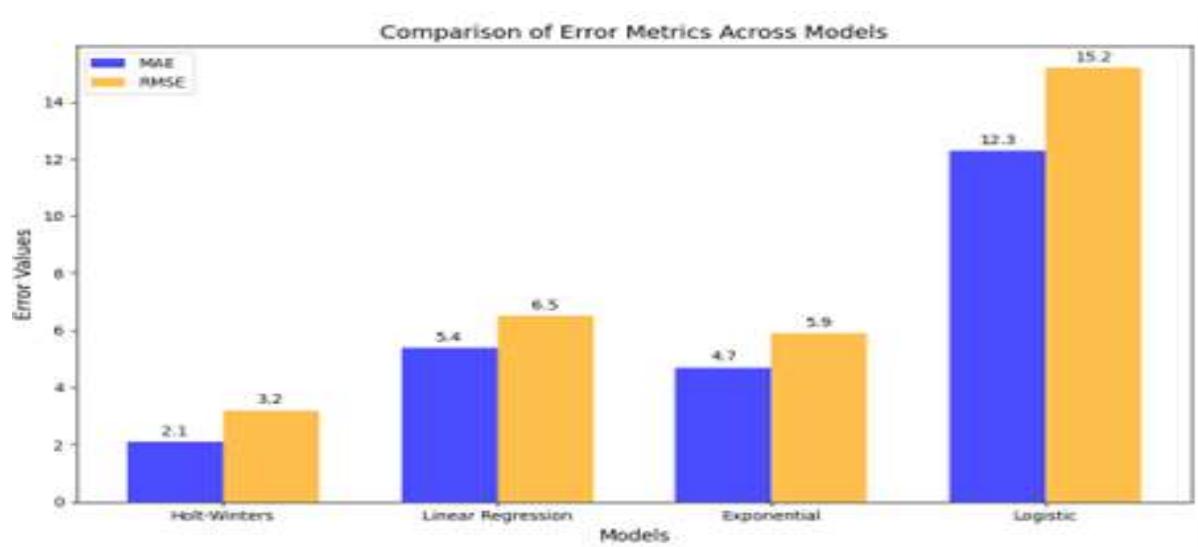
Table 10: Comparison of Forecasting Models Based on Error Metrics and Performance – Table 10

Metric	Exponential	Logistic	Linear Regression	Holt-Winte
Average MAE	1136.407	1083.923	994.942	239.848
Average RMSE	1382.6166	1263.094	2031.922	296.932
No. of states with best model	1	0	2	17

Table 10 summarizes the comparative performance of four forecasting models—Exponential, Logistic, Linear Regression, and Holt-Winters—based on average error metrics and the number of states where each model performed best. Holt-Winters demonstrates superior accuracy with the lowest average MAE and RMSE, and emerges as the best-performing model in 17 out of 20 states. Linear Regression shows moderate performance, while Exponential and Logistic models lag behind in both error metrics and regional consistency. These results reinforce Holt-Winters as the most reliable model for short-term population forecasting across diverse Indian states.

Table 10 and Figure 6 present a comparative analysis of error metrics across the four models studied. The Exponential model demonstrates prediction values close to the AVs for only one state: Jharkhand. The Logistic model does not show any state with PVs matching the AVs. Meanwhile, the Linear regression model yields prediction values closer to AVs for two states: Uttar Pradesh and Haryana. Lastly, the Holt-Winters model stands out, with prediction values closely aligning with the AVs for the remaining 17 states.

Figure 6: Comparison of Error Metrics Across Models



The figure compares the Mean Absolute Error (MAE) and Root Mean Square Error (RMSE) across four forecasting models—Holt-Winters, Linear Regression, Exponential, and Logistic. Holt-Winters shows the lowest error values (MAE = 2.1, RMSE = 3.2), indicating the highest predictive accuracy. In contrast, the Logistic model records the highest errors, suggesting weaker performance. This visual reinforces Holt-Winters as the most reliable model for short-term forecasting among the methods evaluated.

Figure 6 presents a visual comparison of Mean Absolute Error (MAE) and Root Mean Squared Error

(RMSE) across four forecasting models. The Holt-Winters model shows the lowest error values (MAE = 2.1, RMSE = 3.2), indicating superior predictive accuracy. In contrast, the Logistic model exhibits the highest errors (MAE = 12.3, RMSE = 15.2), followed by Linear Regression and Exponential models. This visualization reinforces the quantitative findings from Tables 9 and 10, highlighting Holt-Winters as the most reliable model for population forecasting.

Statistical Evaluation

The chi-square goodness-of-fit test was removed because it is not appropriate for continuous forecast errors. Model performance was instead compared using standard accuracy measures such as MAE, MSE, and RMSE, which are suitable for evaluating continuous population forecasts.

After evaluating all error metrics (MAE, MSE, RMSE) across four benchmark years, the Holt-Winters model consistently demonstrates superior fit and predictive accuracy. It stands out as the most reliable model for state-level population forecasting in India.

Conclusion

During the period 2012–2022, states such as Uttar Pradesh, Bihar, and Maharashtra recorded the highest population levels, reaching approximately 23.76 crores, 11.04 crores, and 12.84 crores respectively. These values reflect long-standing demographic patterns and the continuation of natural population growth. The forecasting results highlight the numerical scale of population increase in these states but do not, by themselves, explain the underlying demographic or socio-economic drivers.

In contrast, states such as Himachal Pradesh and Jammu Kashmir show comparatively slower population growth. These differences represent established demographic trajectories rather than outcomes inferred from the forecasting models. Since the present study focuses on evaluating predictive accuracy, no causal interpretations regarding development, infrastructure, or socio-economic conditions are drawn from the model outputs.

Several states including Rajasthan, West Bengal, Karnataka, and Andhra Pradesh exhibited moderate to high growth during the same period. These patterns align with historical demographic momentum and internal migration trends documented in national statistics. While such trends may coincide with broader socio-economic changes, the forecasting framework used here does not attempt to attribute causation or assess developmental impacts.

The primary objective of this study is to compare the performance of four forecasting models across Indian states. Among them, Holt's linear trend method provided the closest alignment with observed population values for most states, particularly those with steady and moderate growth trajectories.

However, for states with very high or highly variable growth such as Uttar Pradesh and West Bengal the models showed limitations in capturing complex, non-linear demographic dynamics. These limitations indicate the need for more flexible modelling approaches.

Future research may benefit from incorporating non-linear growth structures, migration-sensitive components, and hybrid statistical machine learning frameworks. Integrating geospatial indicators, urbanization metrics, and real-time demographic inputs could further enhance model robustness and adaptability. Such developments would help bridge the gap between theoretical forecasting models and the multifaceted nature of population change.

Overall, the study provides a comparative assessment of forecasting techniques rather than policy recommendations. While population trends have implications for planning and development, deriving policy conclusions lies outside the scope of the present modelling framework. The findings instead emphasize the importance of improving forecasting methodologies to support future demographic analysis.

Disclosure Statement

Authors declare no competing interests.

Data Availability Statement

The data used in this study is publicly available in the document titled *Population Projection Report 2011–2036 – upload compressed 0.pdf*. This report provides state-level population projections and can be accessed through official government portals or institutional repositories supporting demographic research.

References

ABDELRAHIM, M. Z., ASSAF, R. & KANAN, M. 2017. A Mathematical and Statistical Approach for Predicting the Population Growth. *World Wide Journal of Multidisciplinary Research and Development,* 3**,** 50-59. https://wwjmrd.com/archive/2017/8/323/a-mathematical-and-statistical-approach-for-predicting-the-population-growth.

ALI, L. E., KHAN, B. R. & SAMS, I. S. 2015. Brief study on census and predicted population of Bangladesh using logistic population model. *Annals of Pure and Applied Mathematics,* 10**,** 41-47. http://www.researchmathsci.org/apamart/apam-v10n1-6.pdf.

ASHIOBA, N. & DANIEL, N. N. 2020. Population forecasting system using machine learning algorithm. *International Journal of Computer Trends and Technology,* 68**,** 40-43. https://doi.org/10.14445/22312803/IJCTT-V68I12P109.10.14445/22312803/IJCTT-V68I12P109.

FUJIKAWA, H., KAI, A. & MOROZUMI, S. 2003. A new logistic model for bacterial growth. *Journal of the Food Hygienic Society of Japan,* 44**,** 155-160. 10.3358/shokueishi.44.155.

HANSUN, S., CHARLES, V., INDRATI, C. R. & SUBANAR 2019. Revisiting the Holt-Winters' Additive Method for Better Forecasting. *International Journal of Enterprise Information Systems,* 15**,** 43–57. 10.4018/IJEIS.2019040103.

HOLT, C. C. 2004. Forecasting seasonals and trends by exponentially weighted moving averages. *Journal of Economic and Social Measurement,* 29**,** 123-125. https://doi.org/10.3233/JEM-2004-0211.

MAHESH, R. & PRIYANGA, S. 2019. Population Prediction Using Machine Learning. *Journal of Emerging Technologies and Innovative Research,* 6**,** 189–192. https://www.jetir.org/papers/JETIRDC06033.pdf.

MALTHUS, T. R. 1803. *An Essay on the Principle of Population,* London, J. Johnson.

MONDAL, H. & KUMAR, U. 2018. Mathematical Modeling and Predicting the Current Trends of Human Population Growth in Bangladesh. *Modelling, Measurement and Control D,* 39. 10.18280/mmc_d.390101.

MWAKISISILE, A. J. & MUSHI, A. R. 2019. Mathematical model for Tanzania population growth. *Tanzania Journal of Science,* 45**,** 346–354. 10.65085/2507-7961.1689.

OTOOM, M. M. 2021. Comparing the Performance of 17 Machine Learning Models in Predicting Human Population Growth of Countries. *International Journal of Computer Science and Network Security,* 21**,** 220–225. 10.22937/IJCSNS.2021.21.1.28.

PEARL, R. & REED, L. J. 1920. On the rate of growth of the population of the United States since 1790 and its mathematical representation. *Proceedings of the National Academy of Sciences,* 6**,** 275–288. 10.1073/pnas.6.6.275.

RAJKUMARI, S. B. & C., P. 2020. Prediction of Population Growth Using Machine Learning Techniques. *European Journal of Molecular & Clinical Medicine,* 7**,** 1871–1879. https://web.archive.org/web/20210812121359/https://ejmcm.com/article_7352_9f6ce19b88b33acd6f138c6ba849f3fb.pdf.

ŞAHINARSLAN, F. V., TEKIN, A. T. & ÇEBI, F. 2021. Application of machine learning algorithms for population forecasting. *International Journal of Data Science,* 6**,** 257–270. 10.1504/IJDS.2021.122770.

VERHULST, P.-F. 1977. A Note on the Law of Population Growth. *In:* SMITH, D. P. & KEYFITZ, N. (eds.) *Mathematical Demography: Selected Papers.* Berlin, Heidelberg: Springer Berlin Heidelberg. https://doi.org/10.1007/978-3-642-81046-6_37.

WANG, C. Y. & LEE, S. J. 2021. Regional Population Forecast and Analysis Based on Machine Learning Strategy. *Entropy,* 23**,** 656. 10.3390/e23060656.

WINTERS, P. R. 1960. Forecasting sales by exponentially weighted moving averages. *Management Science,* 6**,** 324–342. https://doi.org/10.1287/mnsc.6.3.324.