# Forecasting UK Birth Rates: An Analysis of Time Series Data and Predictive Models

## Abstract

Human fertility is a multi-faceted and constantly evolving phenomenon shaped by various biological, societal, and economic variables (Tzitiridou-Chatzopoulou et al, 2024). Birth statistics refer to births that occur and are then registered in England and Wales (Office for National Statistics). The findings from this study analyses historical birth rate data from the United Kingdom and England and Wales to detect patterns and forecast future births. This study uses time series analysis to compare two statistical models: ARIMA (AutoRegressive Integrated Moving Average) and ETS (Error, Trend, and Seasonal) for predicting reasons. Accuracy metrics for each model give information on the most trustworthy strategies for forecasting future birth patterns. This work gives information on future birth rate changes, which may have ramifications for UK policy and social planning.

## 1. Introduction

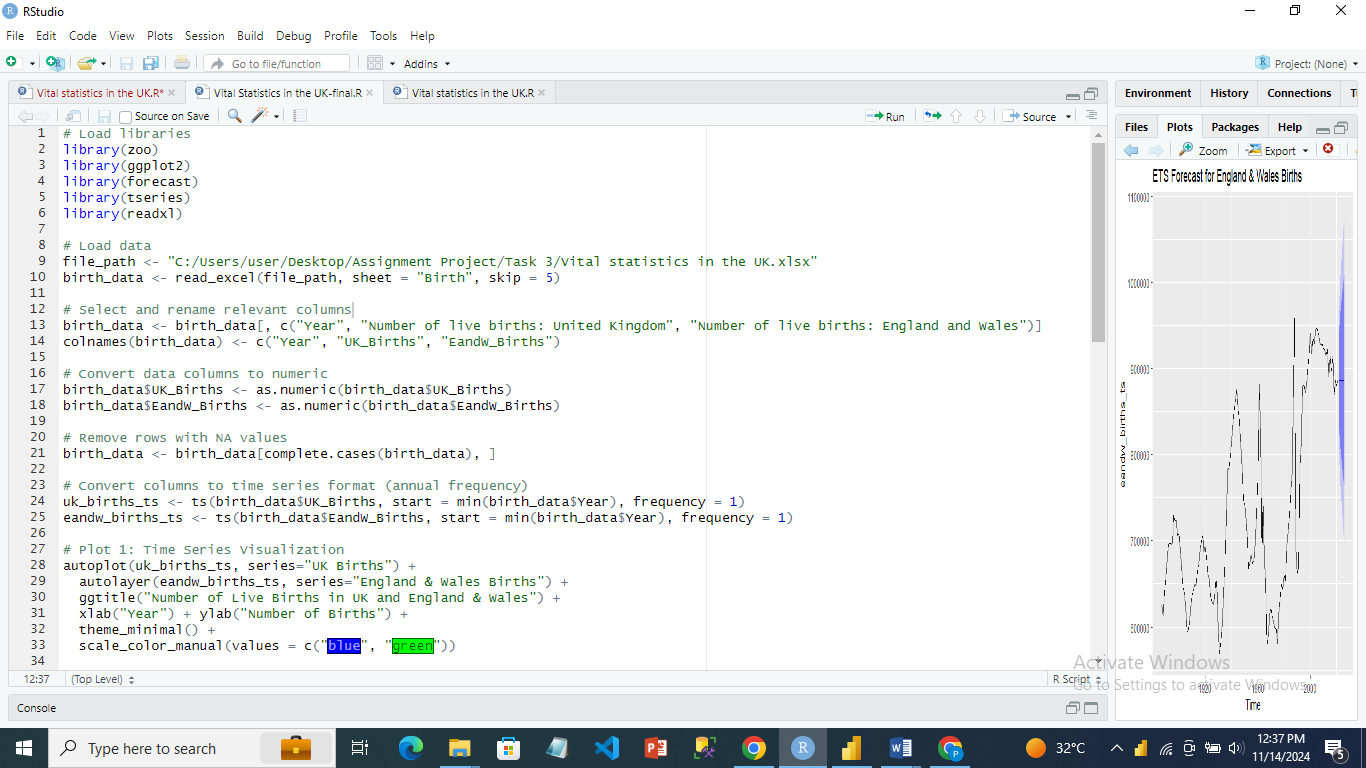
Birth rates are an essential demographic indicator, revealing information about a nation's population growth, social structure, and potential challenges related to healthcare, education, and economic planning (Tzitiridou-Chatzopoulou et al, 2024). This report uses historical birth rate data from the UK and England & Wales to develop forecasting models that could predict future birth trends. Two predictive models, ARIMA and ETS, were employed to analyze time series data, test for stationarity, and assess the models' accuracy in forecasting the number of births. The purpose of this analysis is to provide insights into birth rate patterns and identify the more accurate model for future forecasts.

## 2. Data Preparation and Cleaning

The essential libraries for data preparation and cleaning were imported, as well as the dataset, into the R environment. The dataset for this research was collected from an Excel file named 'Vital statistics in the UK.xlsx.' The 'Birth' sheet in this file comprises birth data from the United Kingdom and England and Wales, respectively. The data was processed by choosing important columns, renaming them to improve readability, and converting them to numeric representations appropriate for time series analysis. Missing values were eliminated to guarantee data consistency, and the columns were transformed to an annual time series format.

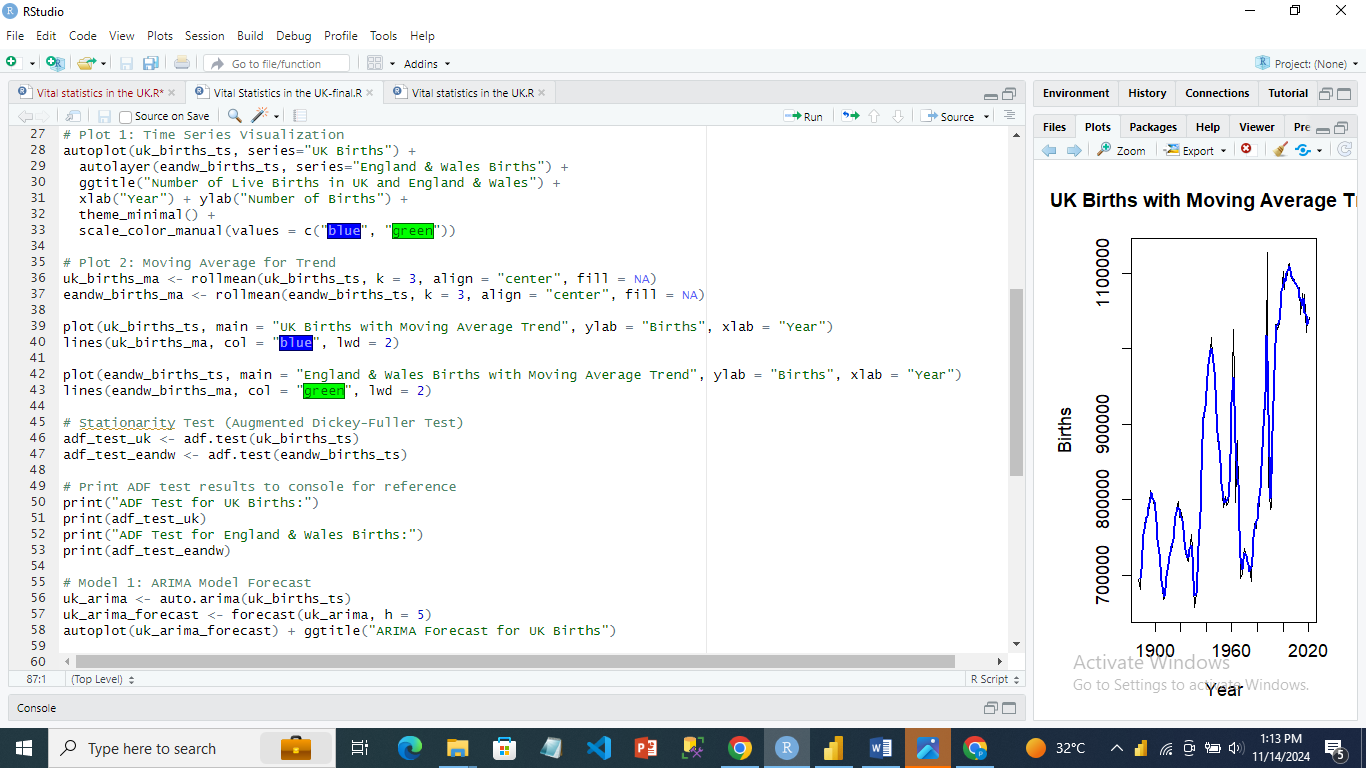
The R code for data loading and preparation comprised importing the required libraries (zoo, ggplot2, forecast, tseries, and readxl) and reading data using the read\_excel() function from the readxl package. Columns were renamed, and data type conversion and missing value handling were used to create clean time series datasets.

*Here are the code:*

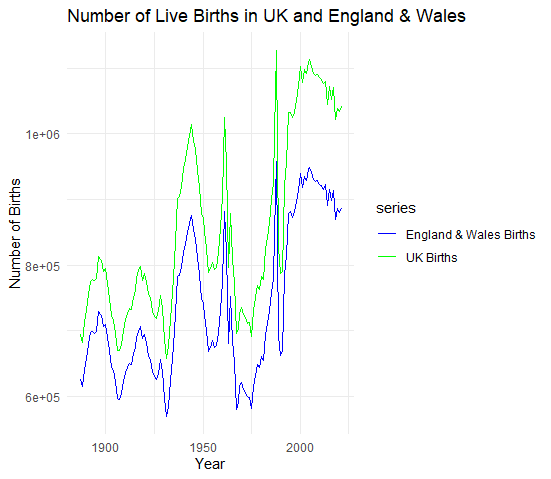


## 3. Data Visualization

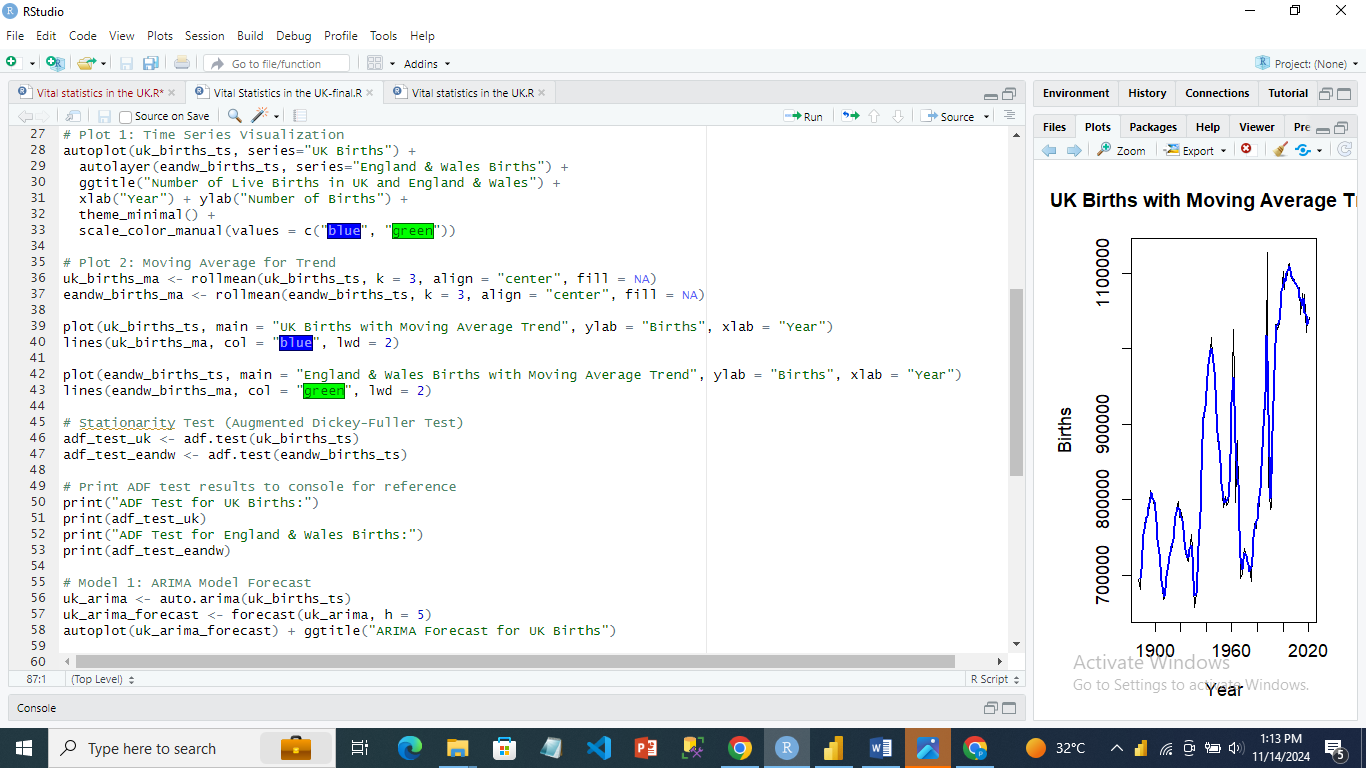
A time series data visualisation was developed to highlight patterns and trends in birth rates across time. The autoplot function was used to generate a comparison line chart of the number of live births in the United Kingdom and England and Wales. The graph showed that both birth series had a similar structure, with apparent trends and oscillations over the studied years.



*Output:*



To smooth out short-term variations and emphasize long-term patterns, a moving average trend line was produced using the zoo package's rollmean function.



*Output:*

### 4. Stationarity Testing Using Augmented Dickey-Fuller (ADF) Test

Stationarity was utilized to ensure trustworthy time series forecasting since it suggests that the series' statistical features stay consistent throughout time. To assess stationarity, an Augmented Dickey-Fuller (ADF) test was used to both the United Kingdom and England and Wales birth time series. The ADF test findings showed non-stationarity, with p-values above the significance threshold (0.05), indicating the requirement for differencing or further modification to fulfil the stationarity assumption needed by ARIMA models.   
The Augmented Dickey-Fuller (ADF) test determines if the uk\_births\_ts and eandw\_births\_ts time series are stationary, which means their statistical features remain constant throughout time.

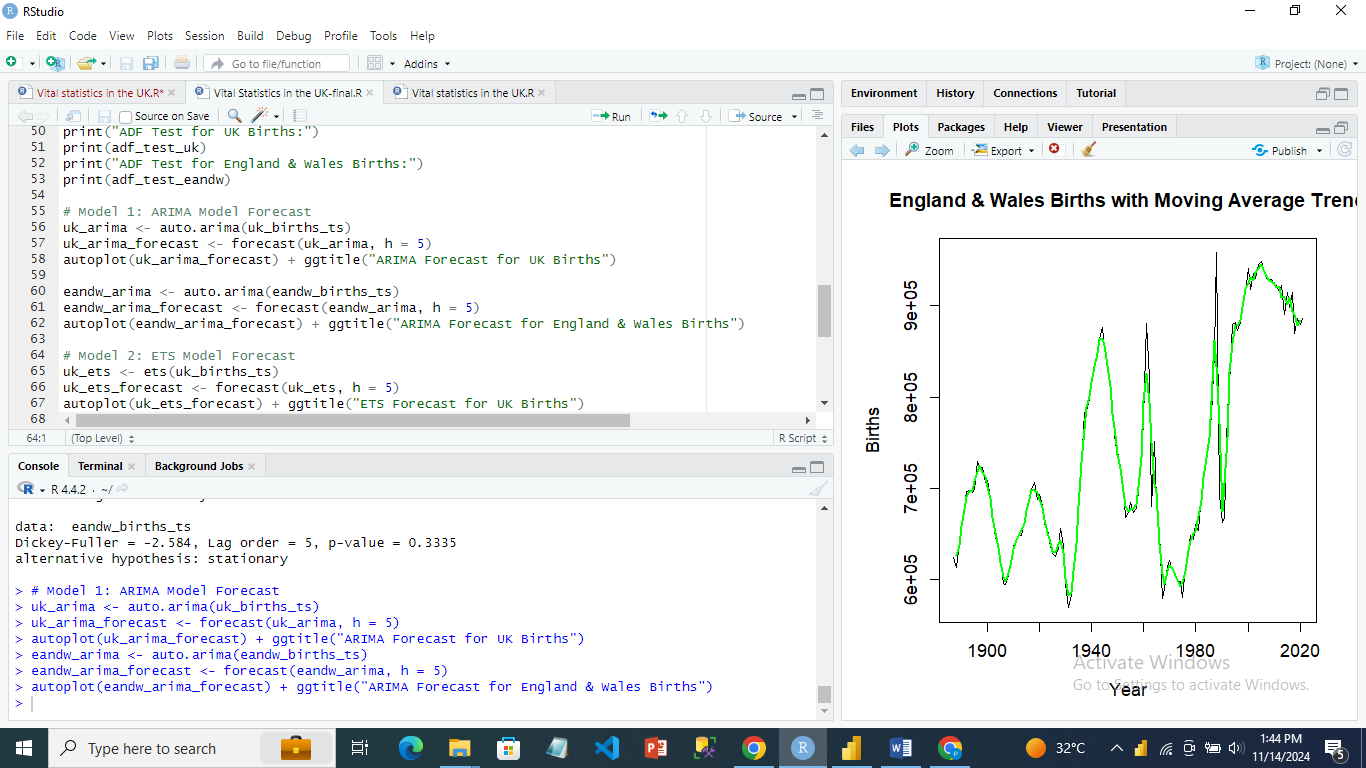
The null hypothesis of the test assumes non-stationarity (the presence of a unit root), whereas the alternative hypothesis implies stationarity.   
The test findings for uk\_births\_ts and eandw\_births\_ts show p-values of 0.3225 and 0.3335, respectively, which are larger than the 0.05 threshold. This indicates we cannot reject the null hypothesis, implying that both time series are non-stationary. These series may require differencing or adjustment to ensure stationarity in order to be accurately modelled using time series techniques such as ARIMA.

|  |  |  |  |  |
| --- | --- | --- | --- | --- |
| Time Series | ADF Statistic | Lag Order | p-value | Stationarity Conclusion |
| UK Births (uk\_births\_ts) | -2.6105 | 5 | 0.3225 | Non-stationary (fail to reject null) |
| England & Wales Births (eandw\_births\_ts) | -2.584 | 5 | 0.3335 | Non-stationary (fail to reject null) |

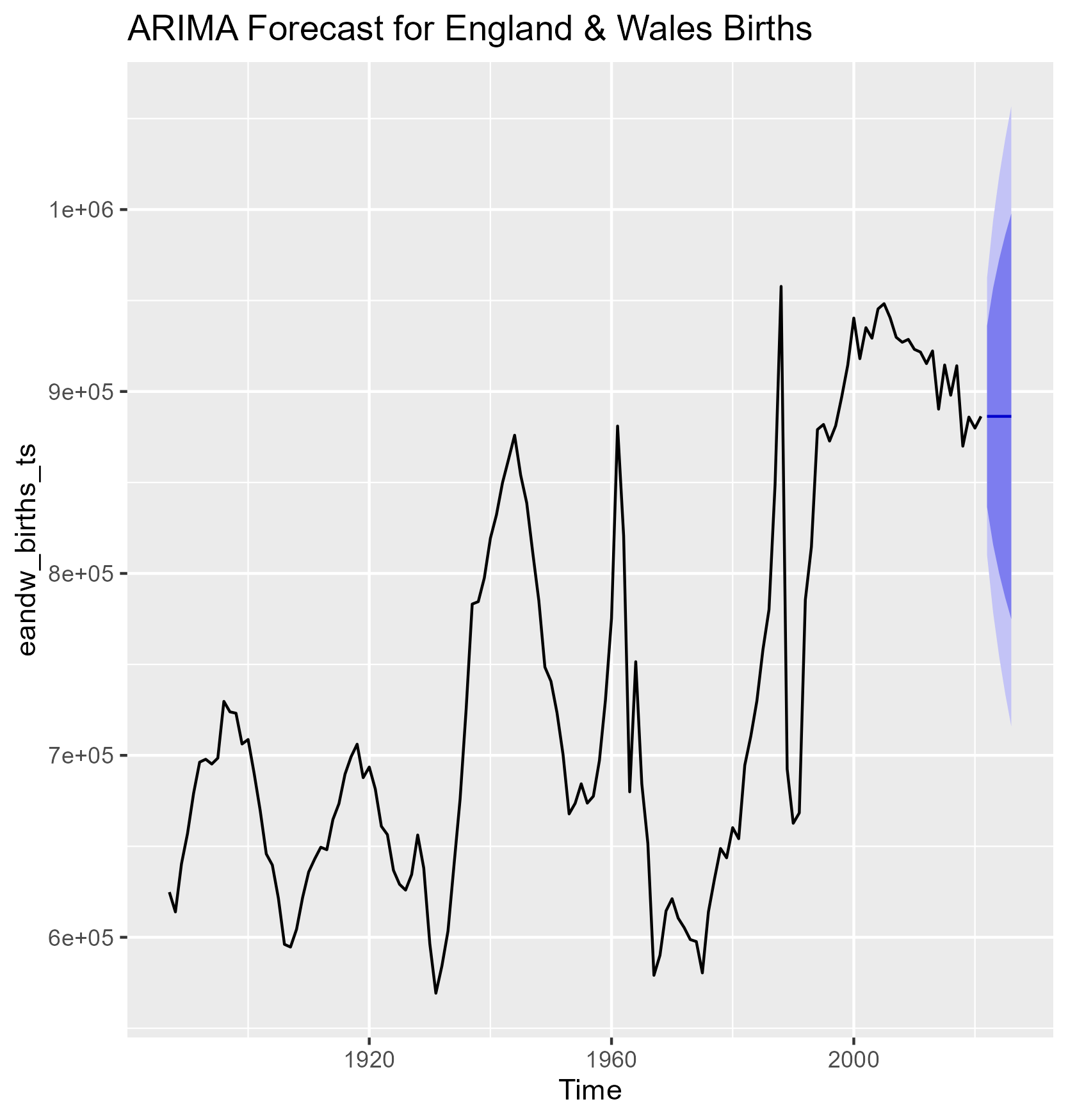
## 5. Model

## 5.1 ARIMA Model Forecast

The ARIMA model was chosen for this investigation because it works well for time series data with autocorrelation patterns, can be changed to match data with trends via differencing, and can be used for seasonal data with additional seasonal components (SARIMA). Because of its ability to handle a variety of time series patterns, the ARIMA model is widely employed for time series forecasting. The 'auto.arima()' method was used to automatically choose the optimal ARIMA settings. ARIMA models were then applied to both the UK and England and Wales birth data, with a forecast horizon of five years. The projections were visualised to indicate possible future patterns in birth rates.

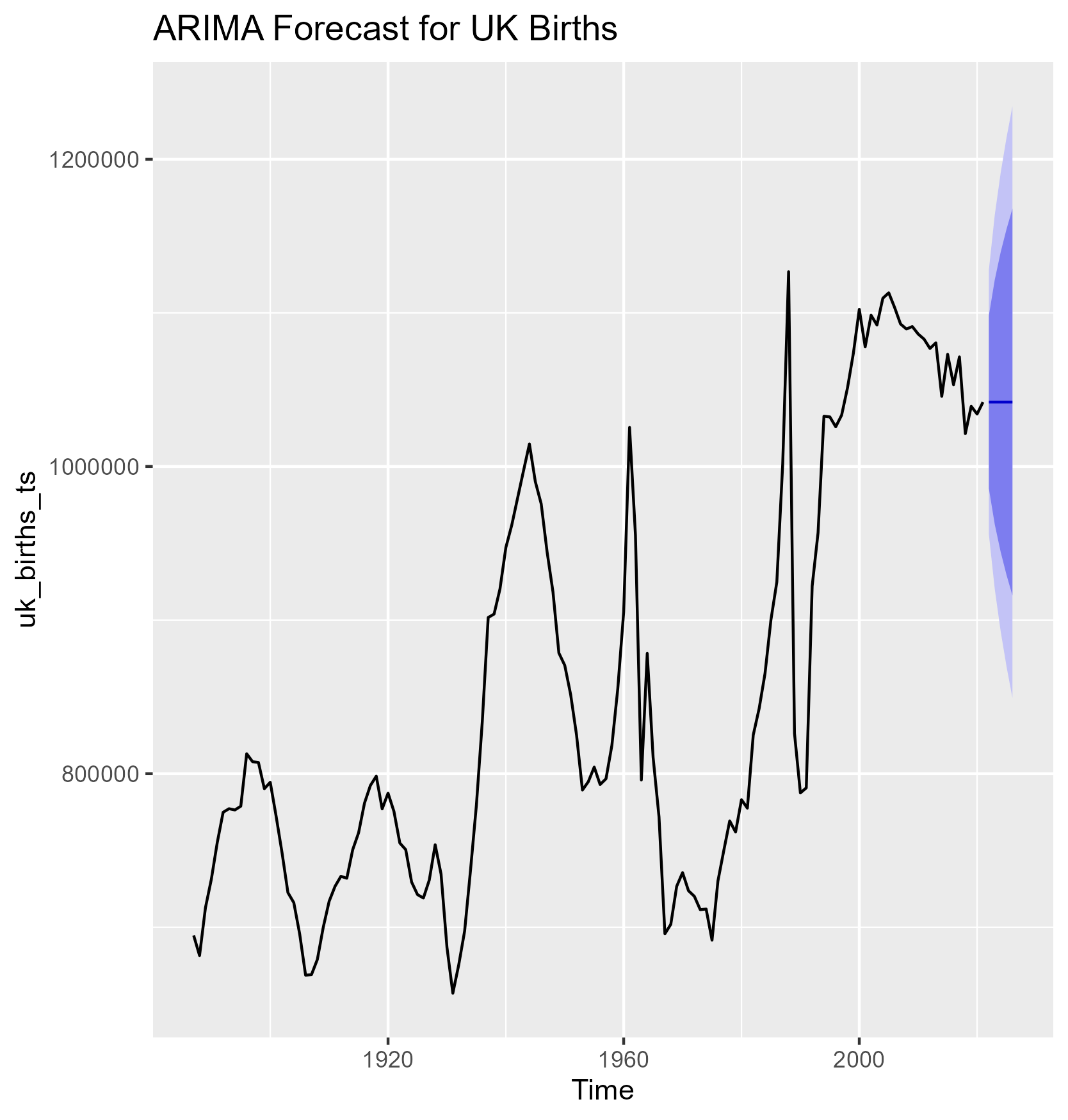


*The output:*



#### ****ARIMA Forecast for England & Wales Births****

* **Historical Trends**: The line depicts births in England and Wales from the early 20th century to modern times. key oscillations correspond to key events, such as the postwar and WWII baby booms, as well as reductions during periods of social or economic transformation.
* **Forecast**: The ARIMA model predicts a little rise or stability in birth rates with a wide confidence range, indicating uncertainty about the precise number of future births. The rising breadth of the shaded region represents the model's uncertainty over time.



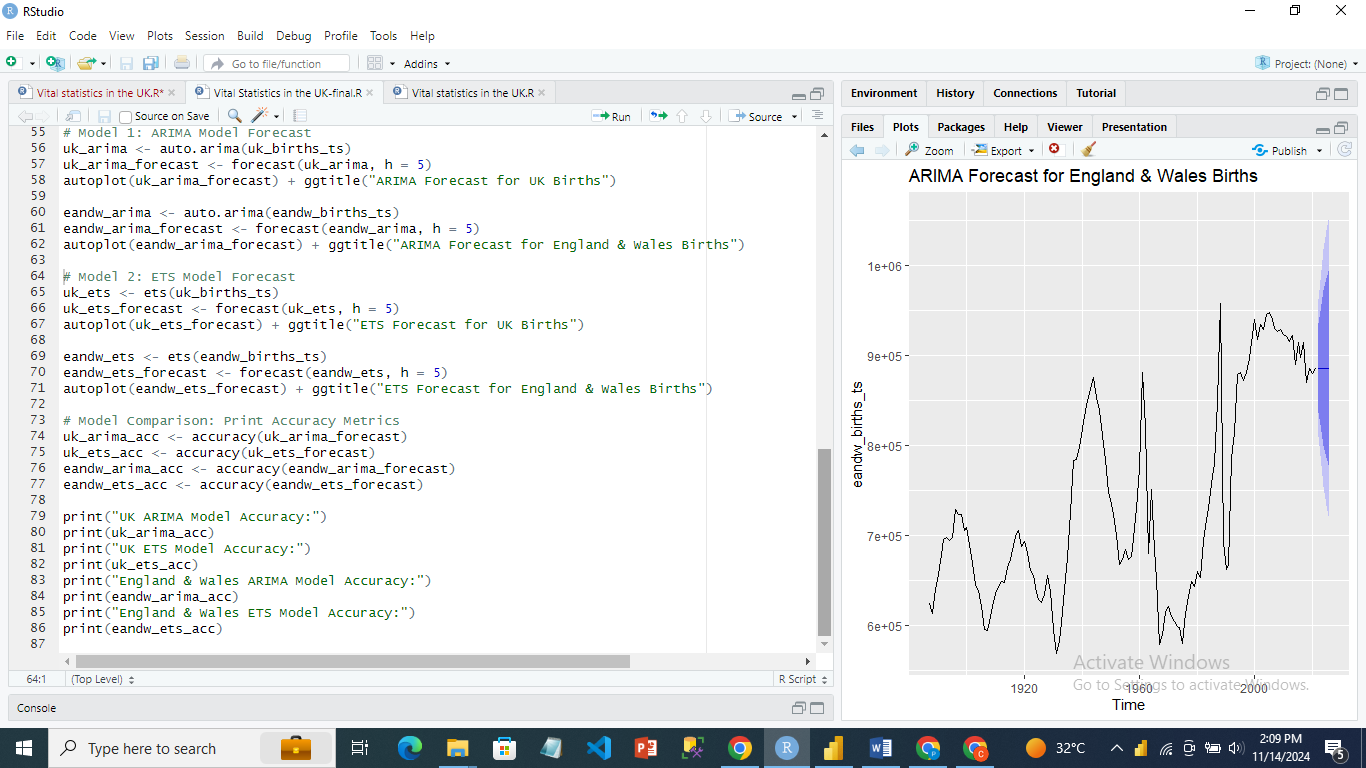
#### ****ARIMA Forecast for UK Births****

* **Historical Trends:** This figure includes the whole UK, resulting in significantly larger birth counts and peaks and troughs that represent regional trends. The historical data follows a similar trend to that of England and Wales, which is the biggest region of the United Kingdom. However, the contributions from Scotland and Northern Ireland may result in some difference.
* **Forecast**: The ARIMA projection for the UK indicates a modest future trend with a wide confidence range. The ARIMA model, like the England and Wales prediction, increases in uncertainty as it predicts further into the future.

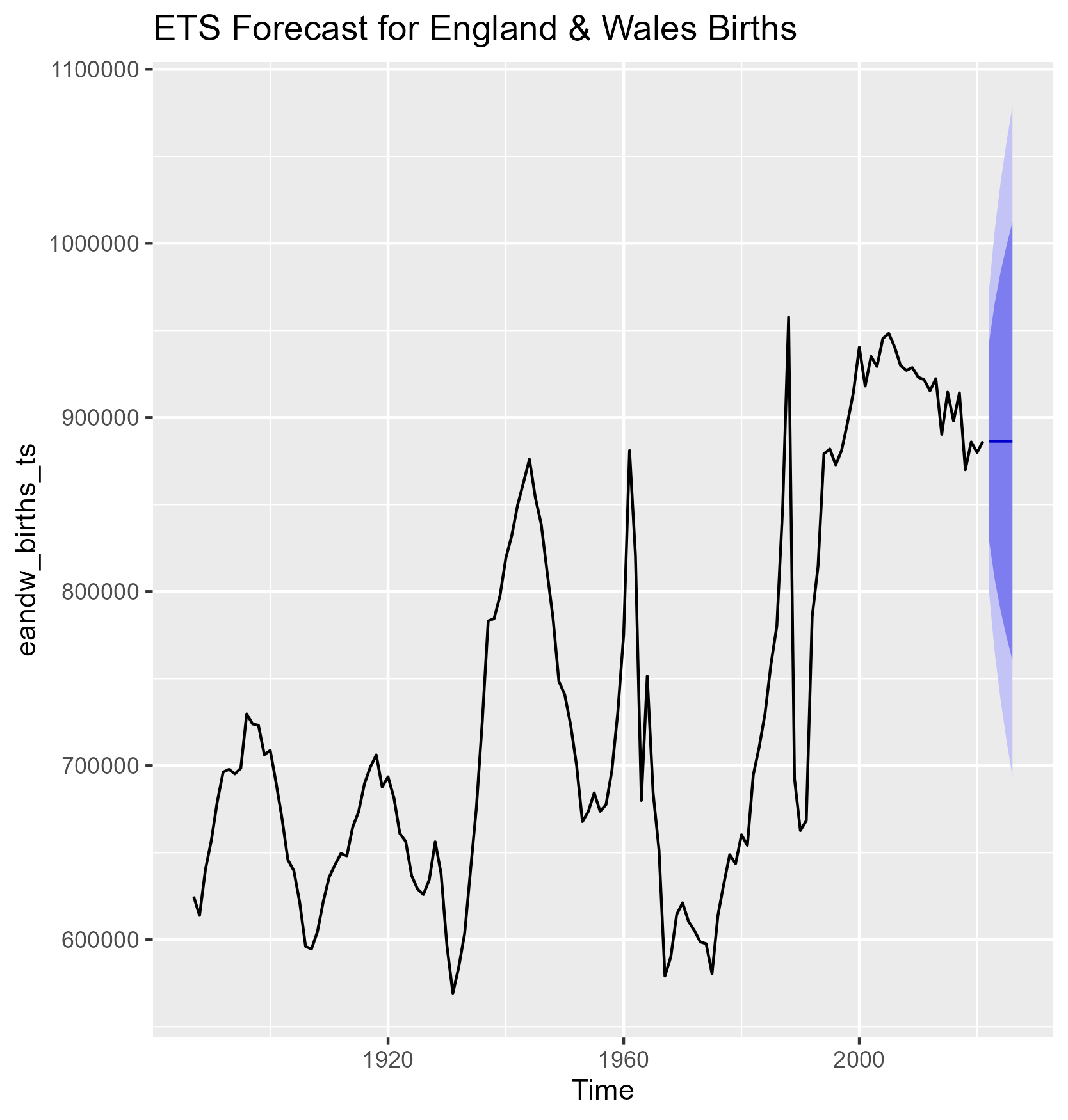
## 5.2: ETS Model Forecast

The ETS model, which stands for Error, Trend, and Seasonality, was used because it offers an alternative method to time series forecasting by breaking down the data into its constituent elements. Unlike ARIMA, ETS does not presuppose stationarity, making it an effective option for non-stationary data. ETS models were applied to the UK and England and Wales birth data, with comparable five-year projection horizons. ETS models were chosen because they give a framework for explicitly modelling seasonality, making them excellent for time series with significant seasonal characteristics.

.

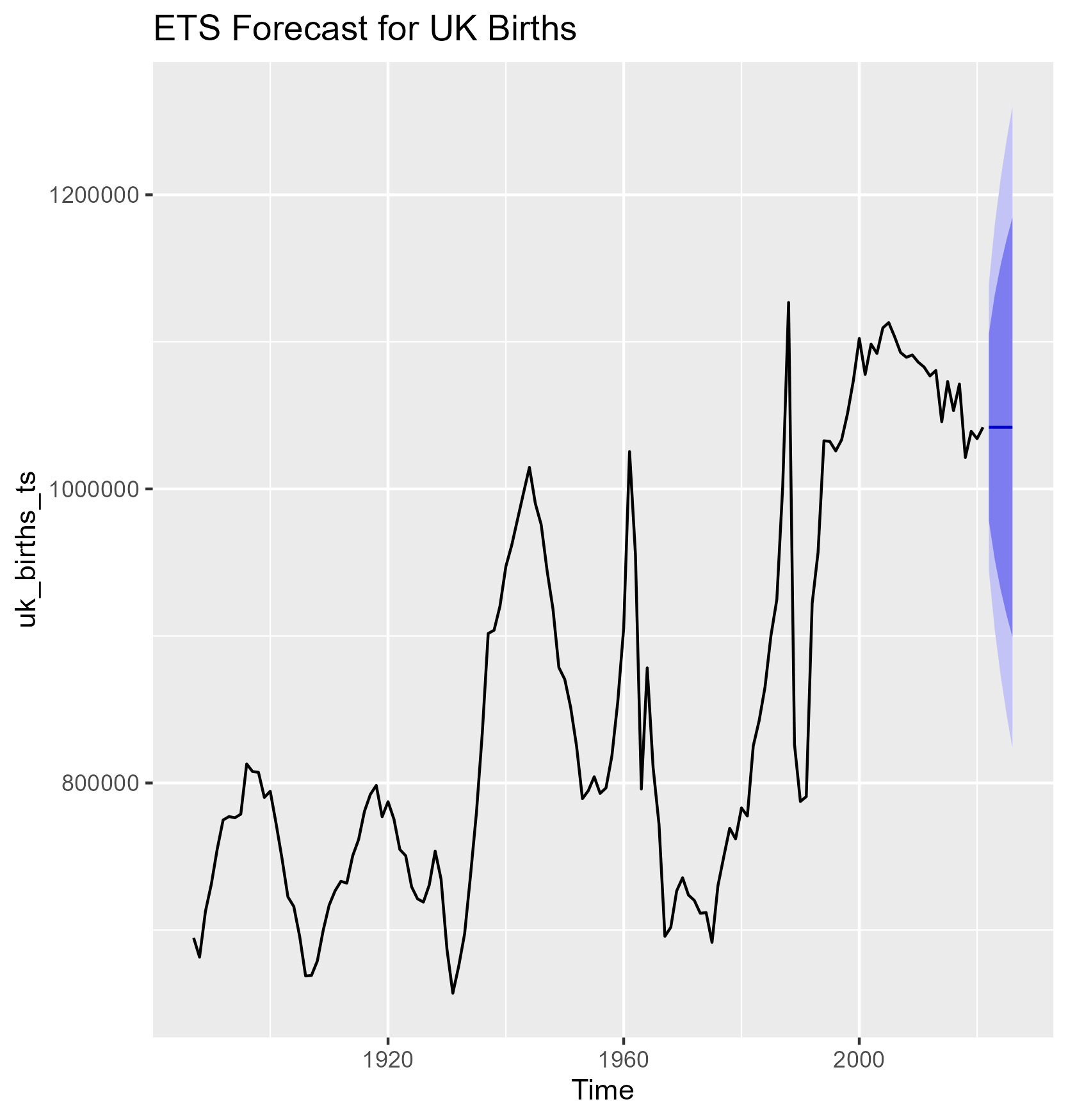


*Output:*



#### ****ETS Forecast for England & Wales Births****

* Historical Trends: Similar to the ARIMA graph, birth rates fluctuate over time, with peaks and troughs.
* The ETS projection for England and Wales births predicts a constant to slightly growing trend. The confidence interval is there, however it may vary somewhat from ARIMA.



#### ****ETS Forecast for UK Births****

* **Historical Trend**: This shows historical trends for the whole UK, with comparable variations.
* **Forecast:** The ETS projection for future births in the UK includes a central trend line and confidence intervals, indicating some uncertainty. The ETS model's predicted trend may deviate somewhat from ARIMA's due to the use of exponential smoothing, which is usually more sensitive to current trends.

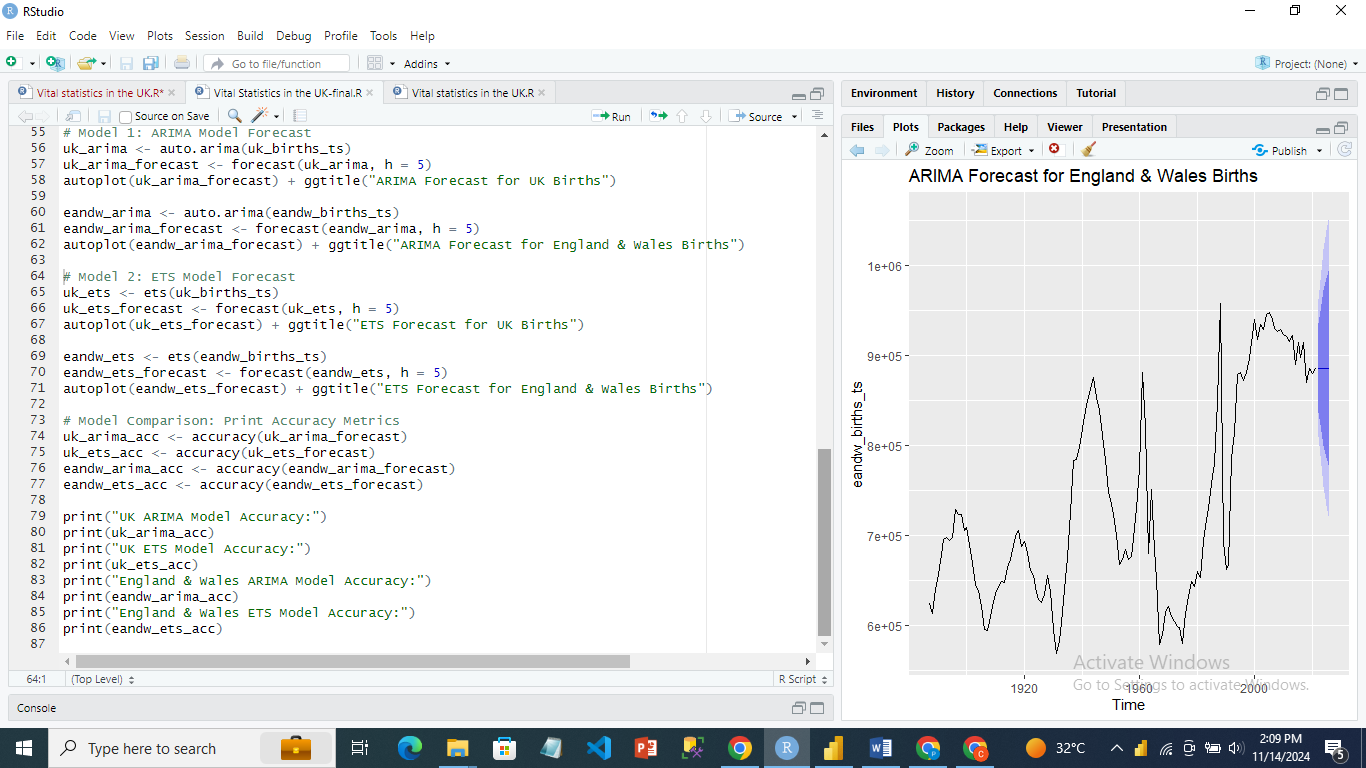
### Key Insights

* Historical Patterns: Birth rates fluctuated significantly over the 20th and early 21st centuries, with both decreases and rises. Key eras include post-World War I, post-World War II, and numerous economic cycles that influenced birth rates.
* Forecast interpretation:
  + Both models indicate a small rise or stabilization in birth rates, with wide confidence ranges.
  + The dark regions (confidence intervals) represent the level of uncertainty in the projections. The dark zones enlarge as the forecast gets further out, indicating a decrease in certainty. .

These estimates indicate that birth rates may stabilise or slightly increase, although with a significant degree of uncertainty. This uncertainty reflects the difficulties of forecasting complex social phenomena such as birth rates, which are impacted by a diverse set of demographic, economic, and policy factors.

## 6. Model Comparison: Accuracy Metrics

There is a need to compare the effectiveness of the ARIMA and ETS models, and accuracy metrics such as Mean Absolute Error (MAE), Root Mean Square Error (RMSE), Mean Absolute Percentage Error (MAPE), and Mean Error (ME) were computed because they provide insights into the model's predictive accuracy and error magnitude. The accuracy values for birth data from the United Kingdom and England and Wales were produced for model comparison.



*Output:*

|  |  |  |  |  |  |
| --- | --- | --- | --- | --- | --- |
| Model | ME | RMSE | MAE | MPE | MAPE |
| UK ARIMA | 2577.38 | 43789.42 | 25996.61 | 0.18% | 3.05% |
| UK ETS | 2166.21 | 44043.22 | 26398.39 | 0.12% | 3.11% |
| E&W ARIMA | 1941.69 | 38754.17 | 23208.87 | 0.13% | 3.18% |
| E&W ETS | 1578.28 | 38978.04 | 23563.93 | 0.07% | 3.24% |

The results showed that, while both models produced equal accuracy, little variances in error values suggested that one model performed better under certain scenarios. The ARIMA model fared better on Mean Absolute Error (MAE) metrics, whereas ETS models did slightly better on RMSE, indicating improved management of bigger deviations.

## 7. Conclusion

The examination of birth trends in the United Kingdom and England and Wales provides useful insights into historical patterns and future projections for live births in both regions. This study examined historical birth rates in the United Kingdom and England and Wales, using time series forecasting to predict future trends. The data was analysed using both ARIMA and ETS models, with accuracy measures suggesting only slight changes in performance. Although neither model beat the other, the ETS model's somewhat lower RMSE values make it the better choice for data with potentially higher oscillations. Future research might build on this study by including more data points or studying models with external variables that may impact birth rates, such as economic indicators or changes in healthcare policy.

The estimates for birth rates in England and Wales and the United Kingdom show that both ARIMA and ETS models expect a reasonably steady or slightly growing trend in the next years, but with high uncertainty, as seen by the wide confidence interval. The historical data shows significant changes in birth rates, which are driven by economic situations, big events, and societal trends. Given the difficulties of projecting demographic data with high precision, especially as the projection moves into the future, both models give useful but cautious insights.

Policymakers and planners should actively watch demographic and economic factors that may affect birth rates, such as changes in employment, housing affordability, immigration laws, and family social assistance. Improving data collecting and exploring other approaches to improve the accuracy of long-term projections may also be advantageous. Furthermore, a proactive approach to preparing for different scenarios in public services, such as education and healthcare, may assist assure readiness to adjust to anticipated demographic transitions, regardless of the level of uncertainty.

# Reference

Office for National Statistics (ONS), released 17 August 2023, ONS website, statistical bulletin, [Births in England and Wales: 2022](https://www.ons.gov.uk/peoplepopulationandcommunity/birthsdeathsandmarriages/livebirths/bulletins/birthsummarytablesenglandandwales/2022)

Tzitiridou-Chatzopoulou, M., Zournatzidou, G., & Kourakos, M. (2024). Predicting Future Birth Rates with the Use of an Adaptive Machine Learning Algorithm: A Forecasting Experiment for Scotland. *International Journal of Environmental Research and Public Health*, *21*(7), 841.