The Impact of COVID-19 on U.S. Birth Volumes: A Comprehensive Analysis from 2007 to 2023

Abstract

This paper presents a comprehensive analysis of U.S. birth volumes from 2007 to 2023, focusing on the potential impact of the COVID-19 pandemic. Our initial analysis, conducted in mid-2023, used birth data from 2007 to 2022 and forecasting models to predict birth volumes for 2023 and beyond. The models were evaluated based on their performance, and insights were drawn regarding long-term birth trends and seasonality. By mid-2024, we updated our analysis using actual birth volumes from 2023 and re-evaluated the forecasting models' effectiveness. This paper details the steps taken, the significant deviations in birth trends during the pandemic, and insights derived from comparing pre-pandemic and post-pandemic trends. The analysis confirms a long-term decline in birth volumes since 2007, with significant anomalies during the COVID-19 period. Future work should explore additional socio-economic factors influencing birth volumes and update forecasting models to reflect post-pandemic realities.

Introduction

The United States, like many developed nations, has experienced a declining birth rate for over a decade. This trend has been driven by various socio-economic factors, including changing family planning norms, economic uncertainty, and evolving societal values. The COVID-19 pandemic, beginning in early 2020, introduced new dynamics that may have accelerated or temporarily altered these long-term trends.

Our initial analysis, conducted in the summer of 2023, utilized data from 2007 to 2022 to forecast birth volumes for 2023. This forecast was based on established models like ARIMA, SARIMA, Holt-Winters, and Prophet. The analysis revealed potential shifts in birth patterns due to the pandemic and hinted at a possible short-term rebound in birth volumes following the significant drop in 2020.

Now, with the availability of actual birth data for 2023, we revisit the earlier analysis to assess how the pandemic impacted birth volumes, whether the forecasts held up, and what insights can be drawn from comparing pre- and post-pandemic data. The findings of this study have broad implications for public health, economic policy, and social services planning.

Methodology

Data Source and Preprocessing

For this analysis, we collected monthly U.S. birth volume data from the CDC's natality database for the period **2007 to 2023**. The data for each year includes the number of births for each month, forming a comprehensive time series for the entire period.

- **2007-2022** data was used for the initial forecast conducted in mid-2023.
- **2023** actual data was integrated into the analysis in mid-2024, allowing us to re-evaluate the forecasting models and gain deeper insights into the impact of COVID-19.

Forecasting Models

We employed the following forecasting models for our initial analysis:

- **ARIMA (Auto-Regressive Integrated Moving Average)**
- **SARIMA (Seasonal Auto-Regressive Integrated Moving Average)**
- **Holt-Winters Exponential Smoothing**
- **Prophet**

Each model was trained on birth data from 2007 to 2022, with forecasts generated for the 24-month period from January 2023 to December 2024. These forecasts were then compared with actual birth volumes from 2023 when available in 2024.

Statistical Comparisons

To evaluate the models' effectiveness, we calculated two performance metrics for each model:

- **Mean Absolute Error (MAE)**: The average of the absolute errors between actual and forecasted birth volumes.
- **Root Mean Squared Error (RMSE)**: The square root of the average squared errors, penalizing larger errors more heavily.

For further insights into the birth trends, we performed a **paired t-test** and **Wilcoxon signed-rank test** to assess whether the differences in birth volumes between 2022 and 2023 were statistically significant.

Long-Term Trend Analysis

We decomposed the birth volume time series into its **trend**, **seasonal**, and **residual** components using classical time series decomposition. This allowed us to capture the long-term declining trend, seasonal fluctuations, and any unexplained anomalies, particularly focusing on the potential effects of COVID-19.

COVID-19 Impact Analysis

We created a **counterfactual forecast** based on pre-pandemic data (2007-2019) to estimate what birth volumes would have been without the pandemic. This allowed us to compare the actual birth volumes post-2020 with the forecasted values and quantify any deviations caused by the pandemic.

Results

1. Initial Forecast Performance (Based on 2007-2022 Data)

The models' initial performance was evaluated using 2016-2022 data, with **Prophet** emerging as the most accurate model at that time:

- **Prophet**: Provided the lowest MAE and RMSE, accurately capturing both trends and seasonality. It was recommended as the best-performing model for future forecasting.
- **ARIMA & SARIMA**: These models provided reasonable predictions, but they underperformed compared to Prophet, especially when dealing with seasonal variations.
- **Holt-Winters**: Performed the worst in the initial analysis, with the highest MAE and RMSE values, indicating that it struggled to capture the long-term and seasonal components effectively.

2. Performance with Actual 2023 Data

Once the 2023 actual data became available, we compared the models' forecasts against real birth volumes. The results were surprising, as they differed significantly from the initial recommendations:

- **Holt-Winters**: Emerged as the best-performing model when evaluated against the actual 2023 data. It had the lowest MAE (7,068.13) and RMSE (8,098.14), indicating a much more accurate forecast than the other models.
- **ARIMA & SARIMA**: Both models performed equally, with an MAE of 10,144.34 and RMSE of 11,228.18. These models provided a reasonable approximation but failed to outperform Holt-Winters.
- **Prophet**: The best-performing model in the initial analysis performed the worst when applied to 2023 data. Its MAE of 11,289.15 and RMSE of 12,794.57 indicated large forecasting errors.

3. Statistical Tests: 2022 vs. 2023 Comparison

We performed a paired t-test and Wilcoxon signed-rank test to assess whether the difference between birth volumes in 2022 and 2023 was statistically significant:

- **Paired t-test**: The t-test resulted in a p-value of **0.009078**, indicating a statistically significant difference between 2022 and 2023 birth volumes. The average difference in birth volumes was **-5,979.5**, meaning there were approximately 5,979 fewer births per month in 2023 compared to 2022.
- **Wilcoxon signed-rank test**: The Wilcoxon test yielded a p-value of **0.009277**, further confirming the significant decline in birth volumes between 2022 and 2023.

4. Long-Term Trends and COVID-19 Impact

Our analysis of the birth volumes from 2007 to 2023 revealed a steady long-term decline in U.S. birth rates. This trend was briefly interrupted during the COVID-19 pandemic:

- **COVID-19 Impact**: Birth volumes dropped sharply in 2020, followed by a brief rebound in 2021-2022. However, this rebound did not fully offset the long-term decline, and by 2023, birth volumes had again declined sharply.

The **counterfactual forecast** based on pre-pandemic data (2007-2019) showed that actual birth volumes from 2020 onward deviated significantly from the expected values. In particular, the drop in birth volumes during 2020 was much larger than the forecasted trend, suggesting that COVID-19 had a substantial negative impact on birth rates.

Discussion and Insights

Model Performance

The major takeaway from our updated analysis is that **Holt-Winters**, which was initially dismissed due to poor performance, emerged as the most accurate model for 2023. This shift highlights the need for continuous model evaluation as real-world conditions change, especially during periods of significant disruption like the COVID-19 pandemic.

Impact of COVID-19 on Birth Rates

COVID-19 had a measurable impact on U.S. birth volumes. The sharp decline in 2020, followed by a modest recovery in 2021-2022, suggests that families may have delayed childbirth due to uncertainty, healthcare disruptions, and economic instability. However, the subsequent decline in 2023 suggests that the long-term downward trend in birth volumes has resumed, potentially exacerbated by the pandemic.

Seasonality and Long-Term Decline

The decomposition of birth volume data confirmed the presence of strong seasonal patterns, with birth volumes peaking during the summer months. However, the long-term trend of declining birth volumes, which began in 2007, appears to have continued unabated, with COVID-19 serving as an additional stressor.

Conclusion and Future Work

The analysis of U.S. birth volumes from 2007 to 2023 provides important insights into both the long-term decline in birth rates and the short-term disruptions caused by the COVID-19 pandemic. While initial forecasts identified **Prophet** as the most effective model, the evaluation of 2023 data revealed that **Holt-Winters** was more accurate in the post-pandemic environment.

Future Work:

1. **Incorporating Socio-Economic Variables**: Future analyses should incorporate additional factors such as unemployment rates, healthcare access, and economic indicators to better understand the underlying drivers of birth rate changes.
2. **Post-Pandemic Forecasting Models**: The forecasting models should be updated and fine-tuned to account for the post-pandemic realities, including potential shifts in societal behaviors.
3. **Long-Term Projections**: With the pandemic's effects still unfolding, it is crucial to continue monitoring birth trends and refine long-term projections for demographic and policy planning.

By adapting forecasting techniques and considering broader societal factors, this research can provide valuable insights for policymakers, healthcare providers, and social service agencies.

References

- CDC WONDER Natality Data (2007-2023).
- Hyndman, R. J., & Athanasopoulos, G. (2018). Forecasting: Principles and Practice.
- Taylor, S., & Letham, B. (2018). Prophet: Forecasting at Scale.
