MSc program in Business Administration and Data Science
Department of Digitalization

FORECASTING BIRTHS IN DENMARK: SEASONAL ARIMA AND DYNAMIC METHODS APPROACH

Predictive Analytics (CDSCO1005U)

Examiners:
Herdis Steingrimsdottir
Thomas Einfeldt

Student:
Eduard Aguado (176199)

Number of pages: 9
Word count: 2376
Submission date: 15-08-2025

Abstract

This study forecasts monthly births in Denmark for 2025 using official data from January 2010 to December 2023, incorporating structural breaks linked to the 2014 “Do it for Denmark” fertility campaign and the COVID-19 pandemic. The problem addressed is whether including such exogenous events enhances the accuracy of demographic forecasts and finding the best model to forecast births. The research applies Seasonal ARIMA (SARIMA) and Seasonal ARIMAX (SARIMA with exogenous regressors) models, supported by Box-Cox transformation, STL decomposition, stationarity testing (ADF, KPSS), and structural break detection. The dataset used comprises 180 monthly observations from Statistics Denmark from 2010 to 2025, with two added dummy variables capturing the post-campaign and post-COVID periods. Results show that the SARIMAX(1,1,1)(0,1,1)[12] model, including both dummies, achieves the lowest RMSE, MAE, MAPE, and MASE while maintaining parsimony. The findings demonstrate that accounting for structural breaks improves forecast performance in this situation, supporting the use of dynamic SARIMA models with relevant exogenous regressors for demographic and policy planning.

Keywords: Time Series Forecasting; SARIMAX; Dynamic Regression Models; Structural Breaks; Fertility; Denmark.

List of Abbreviations

Abbreviation	**Definition**
ADF	Augmented Dickey-Fuller Test
KPSS	Kwiatkowski–Phillips–Schmidt–Shin Test
QLR	Quandt Likelihood Ratio Test
SAR	Seasonal Autoregressive
ARIMA	AutoRegressive Integrated Moving Average
SARIMA	Seasonal AutoRegressive Integrated Moving Average

Contents

1 Introduction

Birth trends play a critical role in shaping long-term demographic structures, influencing public policy, labor markets, and welfare planning. In Denmark, recent demographic shifts have raised concerns about declining fertility rates, particularly in the aftermath of COVID-19 and state-led fertility campaigns. Understanding and forecasting monthly births is essential for anticipating future societal and economic pressures.

This paper aims to forecast monthly births in Denmark for the year 2025 using time series methods. The study uses SARIMA and SARIMAX models to evaluate the impact of structural breaks, specifically those caused by COVID-19 and a national fertility campaign. The forecast relies on official birth data from 2010 to 2024 and introduces two exogenous dummy variables to capture these breakpoints.

The investigation follows an empirical strategy based on seasonal ARIMA modeling with and without exogenous regressors. Among the four models assessed, SARIMAX(1,1,1)(0,1,1) performs best in terms of accuracy and parsimony. The findings suggest that including structural breaks improves forecast performance and can inform policy design aimed at reversing fertility decline.

2 Related Work

Previous research has examined birth trends, their seasonal and cyclical patterns, and the impact of structural breaks. Fertility is characterised by seasonal and cyclical variations (Balan and Jaba 2016), whereby corresponding months across years exhibit strong seasonal correlations. These patterns interact with trends, which may shift due to structural breaks, for example, some studies suggest that the COVID-19 pandemic and the lockdown caused reduced the fertility temporarily by changing the trend. Specifically in Europe, there was a 14.1% decline in live births in January 2021, 9 to 10 months after the epidemic peaks and first lockdowns (Pomar et al. 2022).

3 Dataset Description

The dataset was extracted from the Statistics Bank of Denmark (StatBank.dk) and originally contained live births by child's gender, mother's age, and month from 1973 to 2025. For the purpose of this analysis, the variables for gender and mother's age were removed as they were not relevant. The data were then aggregated to obtain the total monthly births in Denmark from 2010 to 2025 as it serves as a big enough dataset. Additionally, the date variable was then converted to format to YYYY-MM-DD, which is the default format recognized by RStudio. All of this preprocessing process was performed using Microsoft Excel.

The final dataset is shown in Figure 1. It contains 180 observations of monthly births from January

2010 to December 2024. The number of births does not appear stable across months and varies significantly depending on the season, as clearly illustrated in the seasonal plot shown in Figure 1. Despite this seasonality, the variance appears stable, suggesting that the amplitude of seasonal fluctuations remains consistent over time.

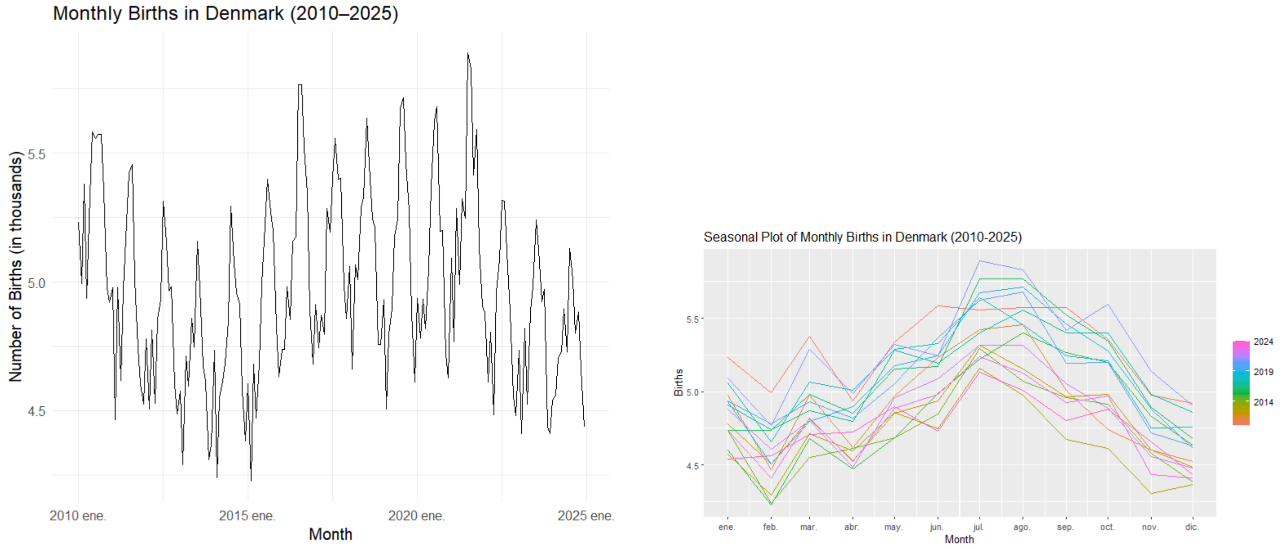

Figure 1: Timeseries plot and seasonal plot of raw data

The series shows several trend and level shifts, as seen in the time series plot in Figure 1. Initially, births exhibit a negative trend that reverses around 2015, shortly after the start of the ”Do it for Denmark” campaign and the follow-up ”Do it for Mum” campaign. These government-led media campaigns aimed to increase the birth rate between 2014 and 2016, and Denmark’s birth rate did rise following them (Sims 2016). In 2021, another shift occurs, with a slight delay after the onset of the COVID-19 pandemic, coinciding with a a decline in births. These hypothetical structural breaks will be statistically tested in Section 4.3.

4 Methodology

4.1 Data Overview and Series Transformation

Understanding the data in terms of structure, trends, and seasonality is key to performing a forecast. Mathematical transformations like log transformations increase the interpretability of the values and stabilize the variance of time series. Box-Cox transformations help find a λ value that makes the seasonal variation consistent across the series and therefore simplifies the forecasting model (Hyndman and Athanasopoulos 2018). Although the raw data seems to have stable variance, the series was transformed using the Box-Cox method with an optimal λ of -0.9. The transformed data has a similar structure to the original (see Appendix), with large values slightly compressing and small values expanding a bit. The Box-Cox transformed series was selected for the forecast, as in most cases, transformed data is preferred over raw data (Hyndman and Athanasopoulos 2018).

Breaking down the series into components helps reveal the underlying patterns. STL is a filtering procedure for decomposing a time series into trend, seasonal, and remainder components (Cleveland et al. 1990). This method can handle any type of seasonality, allowing the seasonal component to evolve over time, while the trend and residuals adapt accordingly. The Figure shows the STL decomposition graphs where it can be first observed that the trend component changes direction multiple times throughout the period, suggesting potential structural breaks to be discussed in Section 4.3. Additionally, a clear and consistent seasonal pattern is also present, reflecting the cyclical nature of monthly births. Finally, the remainder shows low variance and no visible pattern, indicating that most of the variation is explained by the trend and seasonal components, and that the residuals are likely homoskedastic given the no variance across time and randomly distributed.

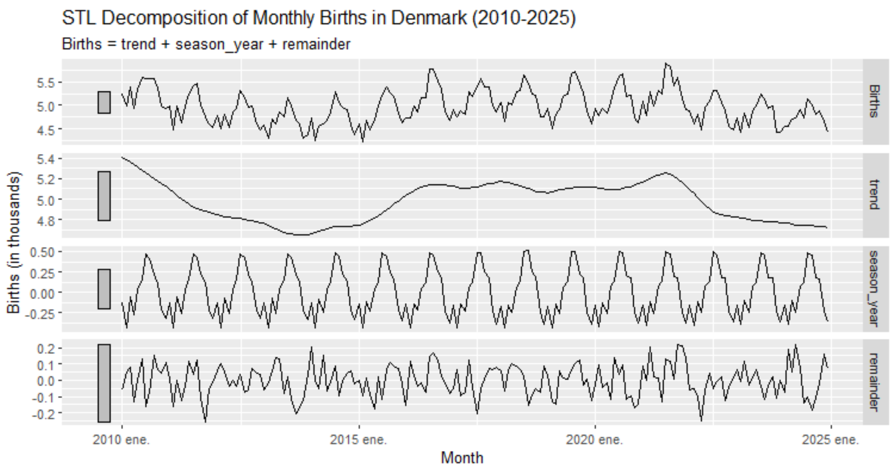

Figure 2: STL Decomposition of raw data

Autocorrelation analysis helps identify the presence and type of seasonality in the series. As shown in Figure 3, the ACF exhibits significant positive autocorrelation at lags 1 and 12, consistent with yearly seasonality, and a negative autocorrelation at lag 6, possibly indicating a phase reversal. Minor surrounding fluctuations are interpreted as noise. A similar pattern is observed in the PACF, with significant spikes at lags 1 and 12 and reduced noise. Additionally, lags 9 and 10 remain significant in the PACF, suggesting they may also carry explanatory power.

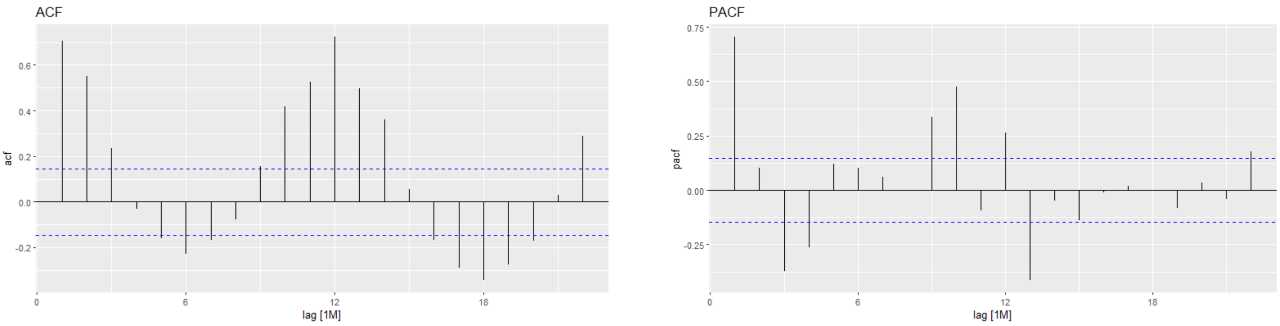

Figure 3: ACF and PACF correlograms of raw data

4.2 Stationarity

A stationary time series is one whose properties do not depend on the time at which the series is observed (Hyndman and Athanasopoulos 2018). This is, the mean, variance and distribution of residuals and autocorrelation remain stable over time. This is why, when working with time series data, it is important to ensure that the data are stationary before proceeding with any type of analysis (C.W.J. Granger 1974). To check this, the Augmented Dickey-Fuller (ADF) and the Kwiatkowski–Phillips–Schmidt–Shin (KPSS) tests are used to assess the stationarity of the time series. The ADF test has a null hypothesis that the series is non-stationary (contains a unit root), while the KPSS test assumes the series is stationary as its null hypothesis. The best practice is to combine both tests to reduce the risk of rejecting the null hypothesis (Kwiatkowski et al. 1992).

At first glance, the Box-Cox transformed data appears stationary, but this was formally tested to confirm it. All tests can be found in the Appendix. The series was evaluated as a seasonal time series with a deterministic trend, using a 5% significance level. Based on the *tau3* component of the ADF test with trend, the null hypothesis of a unit root is rejected since the test statistic is smaller than the critical value ($-4.39 < -3.99$). This result is supported by the KPSS test of type *tau*, where the test statistic is below the critical value ($0.19 < 0.216$), indicating that the null hypothesis of stationarity with trend is not rejected. The series was also tested for a seasonal structure with a constant mean (drift) using the the KPSS *mu* test, which rejects the H0 supporting the non-stationarity around a mean. ($0.19 > 0.14$).

Together with the ADF and KPSS test results and visual inspection of the series, it appears to be locally trend-stationary, meaning it is trend-stationary but likely subject to structural changes (further explored in Section 4.3) that alter the direction of the trend. Additionally, the R commands did not suggest any need for regular or seasonal differencing to achieve stationarity, which aligns with the test results.

4.3 Structural Breaks

To test for structural breaks, the Quandt Likelihood Ratio (QLR) test was used. It is a modified version of the Chow test used when the break date is unknown (Hanck et al. 2024). The H0 of this test is that there are no structural breaks in the model parameters over time. The QLR test is rejected at a high confidence with a F-statistic of 22.01 and a p-value of 0.001, indicating at least one structural break changing the direction of the trend.

Figure 4 shows two points exceeding the QLR test's critical value. Using the *breakpoints()* function, these correspond to May 2015 and November 2021. These trend changes reflect the impact of the "Do it for Denmark" fertility campaign starting in May 2014 (see Section 3) and the COVID-19 pandemic beginning in March 2020. The 9-12 month lag between the events and the breaks aligns with pregnancy duration, reflecting the delayed effect on birth rates. This aligns with the visual analysis in Section 3, where the series begins with a negative trend until 2015, followed by birth growth until around 2021, and then returns to the initial negative trend after 2021.

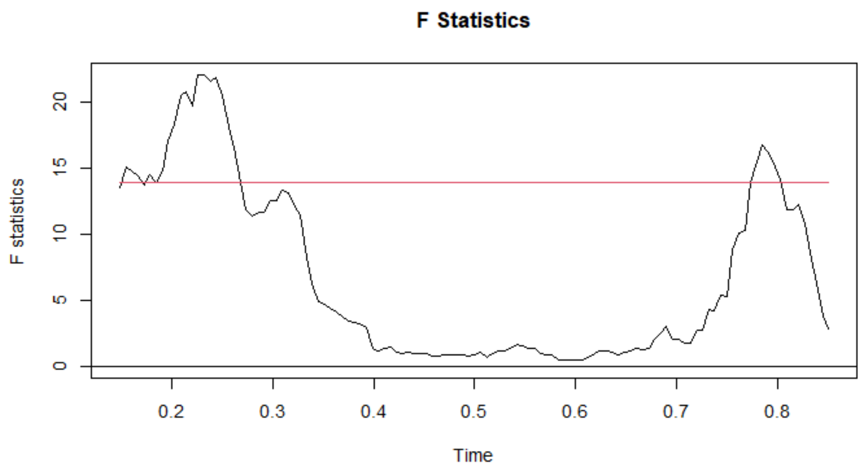

Figure 4: QLR Test

ARIMA models use information from past observations of a series but do not incorporate other external information potentially relevant (Hyndman and Athanasopoulos 2018). Structural breaks can be accounted for in forecasts by adding dummy variables to ARIMA models, resulting in ARIMAX models. Two dummy variables were created: the *post-campaign* dummy, which equals 1 for the period from May 2015 to November 2021, and the *post-covid* dummy, which equals 1 for the period after November 2021, and 0 otherwise for both. Including these dummies helps isolate the trend variation caused by structural breaks, allowing the ARIMA terms to better capture the underlying process and thereby improving forecast accuracy by reducing the noise in the coefficients.

5 Model Selection

5.1 Modeling Framework

As outlined in Section 1, this project uses Seasonal ARIMA models (SARIMA) and a dynamic variation of SARIMA (SARIMAX). These models include three parameters: p (the order of the autoregressive part), d (the degree of differencing to achieve stationarity), and q (the order of the moving average part). Both non-seasonal and seasonal components are specified using this structure. Finally, the SARIMAX models also include exogenous regressors, in this project the structural breaks described in Section 4.3.

$$\Phi_P(B^s)\,\phi_p(B)\,(1-B)^d(1-B^s)^D y_t = \Theta_Q(B^s)\,\theta_q(B)\,\varepsilon_t \tag{1}$$

$$\Phi_P(B^s)\,\phi_p(B)\,(1-B)^d(1-B^s)^D y_t = \beta^\top x_t + \Theta_Q(B^s)\,\theta_q(B)\,\varepsilon_t \tag{2}$$

The equation 1 represents a general SARIMA model. The backshift operator (B) operating on y_t shifts the series back one period, enabling a compact representation of autoregressive and moving average terms (Hyndman and Athanasopoulos 2018). The left-hand side captures both seasonal and non-seasonal dynamics through differencing and lag polynomials, while the right-hand side models the error structure. The SARIMAX model shown in Equation 2 extends this framework by including t exogenous regressors x_t, allowing the model to account for external influences.

5.2 Model Comparison

Following these two approaches, four models were selected using distinct methodologies (see Table 2). The `auto.arima()` function in R was first used to estimate a SARIMA model without structural breaks, yielding SARIMA(1,0,1)(2,1,0). When exogenous regressors were included, it returned SARIMAX(1,1,1)(2,1,0). However, as `auto.arima()` has limited capacity to incorporate domain knowledge or explore alternative seasonal specifications, additional models were manually selected. This process was guided by the insights obtained from the STL decomposition, ACF, PACF, and the improved performance observed when structural break dummies were included. Specifically, SARIMAX(1,1,1)(0,1,1) and SARIMAX(1,1,1)(1,1,1) were also estimated for forecasting.

Specification	AIC	AICc	BIC	Selection Method
A) SARIMA(1,0,1)(2,1,0)	197	196	181	auto.arima()
B) SARIMAX(1,1,1)(2,1,0)	198	197	175	auto.arima()
C) SARIMAX(1,1,1)(0,1,1)	**220**	**219**	**201**	**Manual Search**
D) SARIMAX(1,1,1)(1,1,1)	218	217	196	Manual Search

Table 1: Model selection metrics

First, comparing the two models selected by the `auto.arima()` command (A and B), both yield similar AIC and AICc values, but Model B, which includes structural breaks, slightly outperforms Model A. This indicates that the exogenous variables add useful information for forecasting. Although Model B has a higher BIC due to the inclusion of additional regressors, these variables were retained in the manual model selection process. Based on the selection metrics, SARIMAX(1,1,1)(0,1,1) clearly emerges as the best model. SARIMAX(1,1,1)(1,1,1) was also included to assess the impact of a different Seasonal Autoregressive (SAR) component on the forecast, despite its slightly higher information criteria.

SARIMAX(1,1,1)(0,1,1)[12] stands as the best model. It includes one non-seasonal autoregressive term, one non-seasonal difference, and one non-seasonal moving average term, along with one seasonal difference and one seasonal moving average term with a yearly cycle, as well as structural break variables. It maintains parsimony while incorporating the most relevant information. To validate that the model captured the predictable structure, residual diagnosis was performed. The residuals appear to fluctuate randomly around zero, with no clear pattern, trend, or large-scale deviations. This is also reflected in the histogram, which shows a slight departure from normality due to mildly heavy tails, but no significant skewness or outliers.

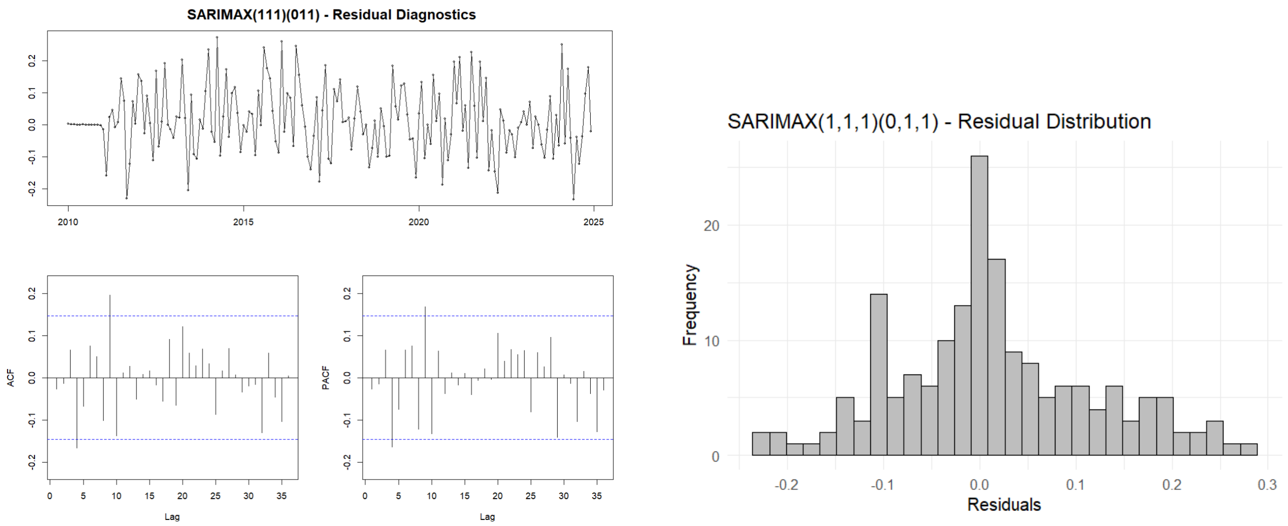

Figure 5: Residual Diagnosis for the best model

Additionally, the Ljung-Box test was applied to assess whether the autocorrelations in the model residuals are jointly zero, as recommended by Hyndman and Athanasopoulos (2021). The null hypothesis, which states that the residuals are independently distributed (i.e., no autocorrelation), is not rejected at the 5% significance level for this model, with a p-value of 0.103. This result is consistent with the residual ACF and PACF shown in Figure 5. Overall, the residual diagnostics suggest that the model is adequate for forecasting.

6 Results

Table 2 presents the selection metrics for the four models evaluated. Among them, the SARIMAX(1,1,1)(0,1,1) model achieves the lowest RMSE, MAE, MAPE, and MASE, closely followed by SARIMAX(1,1,1)(1,1,1), indicating superior forecasting accuracy. Despite the nearly identical errors, Model C is more parsimonious and therefore preferred. Additionally, the non-dynamic model (A) exhibits the highest error values, suggesting that the inclusion of structural break dummies improves forecast performance. Overall, SARIMAX(1,1,1)(0,1,1) stands as the best model due to its balance between forecasting accuracy and model simplicity, making it the preferred specification for generating the 2025 prediction.

Specification	**RMSE**	**MAE**	**MAPE**	**MASE**	**Selection Method**
A) SARIMA(1,0,1)(2,1,0)	0.123	0.095	2.58	0.547	auto.arima()
B) SARIMAX(1,1,1)(2,1,0)	0.121	0.092	2.50	0.531	auto.arima()
C) SARIMAX(1,1,1)(0,1,1)	**0.105**	**0.079**	**2.15**	**0.458**	**Manual Search**
D) SARIMAX(1,1,1)(1,1,1)	0.105	0.079	2.16	0.459	Manual Search

Table 2: Model selection metrics

Figure 6 shows the 12-month forecast for 2025 using the SARIMAX(1,1,1)(0,1,1) model. The blue shaded areas may seem too wide, but they indicate the 80% and 95% confidence intervals. The intervals widen gradually across the year, reflecting increasing forecast uncertainty over time. The forecast maintains the historical seasonal pattern, and it shows a stable and reliable model performance over the short horizon.

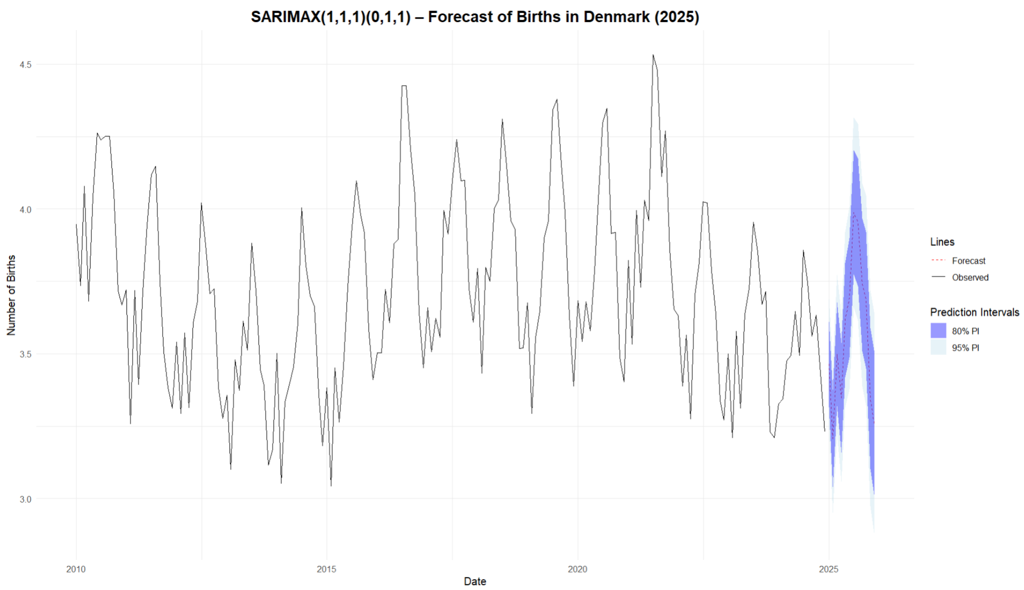

Figure 6: SARIMAX(1,1,1)(0,1,1) forecast of births in Denmark 2025

7 Conclusions

The aim of this study was to forecast monthly births in Denmark for 2025 using time series models, incorporating structural changes from a fertility campaign and the COVID-19 pandemic. The analysis showed that including these structural break dummies improved forecasting performance. SARIMAX(1,1,1)(0,1,1) achieved the best results across both forecast accuracy metrics (RMSE, MAE, MAPE, MASE) and model selection criteria (AIC and BIC), while maintaining model simplicity.

These findings suggest that accounting for exogenous structural shifts enhances short-term birth forecasts. A reasonable course of action would be to prioritize SARIMAX models with relevant regressors when structural changes are known. Future work could integrate additional demographic or economic indicators to extend forecasting capacity beyond 2025.

References

Balan, Christiana and Elisabeta Jaba (Mar. 2016). “Birth Seasonality Patterns in Central and Eastern Europe during 1996-2012”. In: *Romanian Statistical Review, ISSN 1018-046X* 1, pp. 9–20.

C.W.J. Granger, P. Newbold (1974). “Spurious regressions in econometrics”. In: *Journal of Econometrics*. URL: https://www.sciencedirect.com/science/article/abs/pii/0304407674900347.

Cleveland, Robert B. et al. (1990). “STL: A Seasonal-Trend Decomposition Procedure Based on Loess”. In: *Journal of Official Statistics* 6.1, pp. 3–73. URL: https://www.wessa.net/download/stl.pdf.

Hanck, Christoph et al. (2024). *Introduction to Econometrics with R*. Accessed: 2025-08-09. x. URL: https://www.econometrics-with-r.org/ITER.pdf.

Hyndman, R.J. and George Athanasopoulos (2018). *Forecasting: Principles and Practice*. 2nd. Accessed: 2025-08-09. OTexts. URL: https://otexts.com/fpp2/.

Kwiatkowski, Denis et al. (1992). “Testing the null hypothesis of stationarity against the alternative of a unit root: How sure are we that economic time series have a unit root?” In: *Journal of Econometrics*. URL: https://www.sciencedirect.com/science/article/abs/pii/030440769290104Y.

Pomar, Léo et al. (2022). “Impact of the first wave of the COVID-19 pandemic on birth rates in Europe: a time series analysis in 24 countries”. In: *The Lancet Regional Health - Europe* 17, p. 100363. DOI: 10.1016/j.lanepe.2022.100363. URL: https://pmc.ncbi.nlm.nih.gov/articles/PMC9619770/.

Sims, Alexandra (2016). *Denmark’s bizarre series of sex campaigns lead to baby boom*. Accessed: 2025-08-09. URL: https://www.independent.co.uk/news/world/europe/denmark-s-bizarre-series-of-sex-campaigns-lead-to-baby-boom-a7062466.html.

Appendix

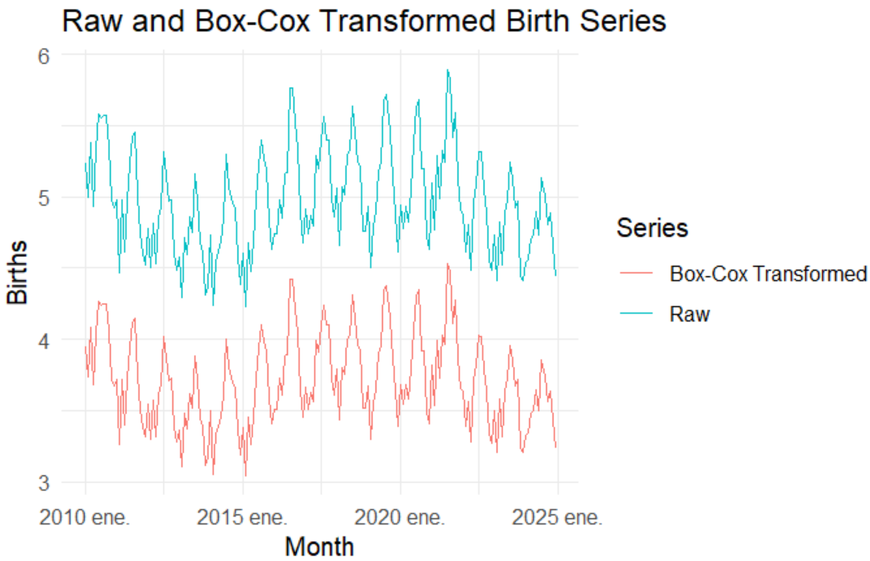

Figure 7: Box-Cox transformed and raw data timeseries plot

Table 3: Stationarity Test Results

Test	Type	Test Statistic	5% Critical Value
ADF	Trend (tau3)	-4.394	-3.431
ADF	Drift (tau2)	-4.121	-2.881
KPSS	Level (mu)	0.192	0.463
KPSS	Trend (tau)	0.146	0.194

Table 4: Ljung-Box Test Results on Model Residuals

Model	Ljung-Box Statistic	p-value
SARIMA(1,0,1)(2,1,0)	45.12	0.062
SARIMAX(1,1,1)(2,1,0)	46.57	0.027
SARIMAX(1,1,1)(0,1,1)	41.25	0.103
SARIMAX(1,1,1)(1,1,1)	41.23	0.083