Analysis of Seasonal Birth Rate Trends Using Time Series Models

1st Sarah Thuman
Stevens Institute of Technology
Hoboken, USA
sthuman@stevens.edu

2nd Syed Aziz
Stevens Institute of Technology
Hoboken, USA
saziz2@stevens.edu

3rd Victor Alvarez
Stevens Institute of Technology
Hoboken, USA
valvarez1@stevens.edu

***Abstract*—This comprehensive report presents a dual approach in time series analysis, exploring both the intricate seasonal patterns in United States birth rates and the volatile movements of Bitcoin prices. Utilizing ARIMA, SARIMA, and advanced GARCH modeling techniques, this study delves into the distinct characteristics of these datasets. The birthrate analysis employs ARIMA and SARIMA models to unravel seasonal trends and forecast future rates, enhanced by a detailed exploration of the data through various visualization tools. Meanwhile, the Bitcoin price analysis leverages ARIMA modeling coupled with GARCH techniques to capture its erratic market behavior, aiming to predict future price movements with a nuanced understanding of volatility. The synthesis of these diverse methodologies demonstrates the robust versatility of time series analysis in decoding complex temporal phenomena, offering insightful forecasts and deepening our understanding of both biological and financial domains.**

I. INTRODUCTION

This comprehensive report is an exploration into the dynamic world of time series analysis, focusing on two contrasting datasets: the historical birthrate data in the United States and the volatile Bitcoin market.

For the birthrate data, stretching from 1969 onwards, we delve deep into identifying patterns and fluctuations using ARIMA and SARIMA models. Our objective is not just to discern the overall trends, but to thoroughly understand the seasonal aspects that punctuate these data. The intricate seasonal variations present an opportunity to refine our forecasting techniques and to reveal the underpinnings of these societal rhythms.

Turning our attention to the Bitcoin market, we enter the realm of financial time series analysis. Here, the focus is on predicting future price movements in a highly volatile and unpredictable environment. By employing a combination of ARIMA and GARCH models, we aim to capture and explain the erratic behavior of cryptocurrency prices. The study of Bitcoin's price movements presents a different set of challenges and nuances compared to the birthrate analysis, mainly due to its non-seasonal and speculative nature.

Through this juxtaposition of two distinct datasets, our report underscores the adaptability and robustness of time series analysis. The ability to extract meaningful insights from both the predictably cyclical birth rates and the mercurial world of cryptocurrency trading showcases the versatility of statistical modeling in time series analysis. Our findings not only contribute to the academic understanding of these two domains but also offer practical applications and forecasting tools relevant to policy-makers, financial analysts, and social planners.

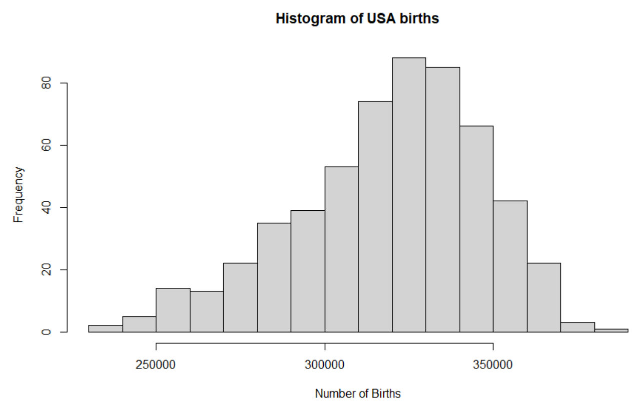

Fig. 1. Histogram of USA births data.

II. BIRTHRATE DATASET ANALYSIS

The birthrate dataset analysis entails a systematic exploration of data trends, seasonality, and potential irregularities that can affect model accuracy and forecasting reliability.

A. Data Preprocessing and Visualization

The initial step in our analysis involved thorough preprocessing of the data. To focus our analysis on a consistent demographic context, we specifically narrowed the dataset to encompass only the data from the United States, as opposed to using worldwide data. This preprocessing included checking for and handling missing values, filtering out irrelevant rows, and transforming the dataset into a time series object for analysis.

1) Histogram of Births: The histogram of birth counts revealed a roughly normal distribution but with a slight left skew potential outliers on the higher end, suggesting that certain periods or years may have experienced unusually high birth rates.

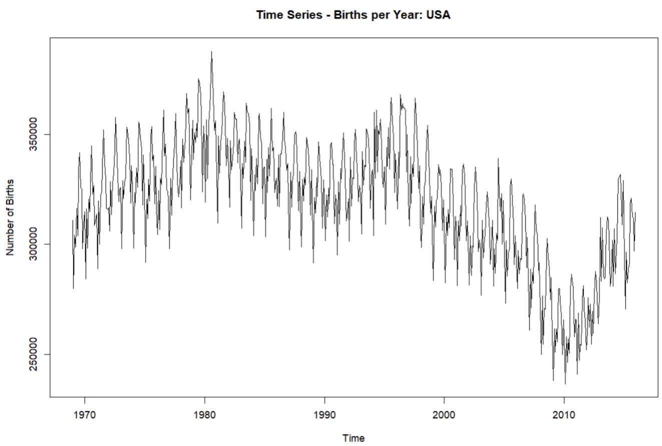

Fig. 2. Time series plot of USA births data.

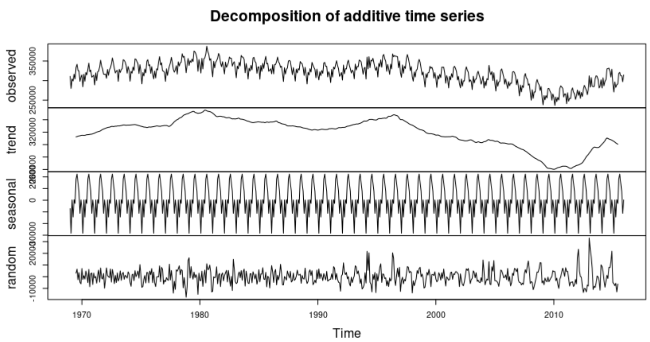

Fig. 3. Decomposition of USA Births data.

2) Time Series Decomposition: The decomposition of the birthrate time series into its constituent components illustrated a clear seasonal pattern, as well as a long-term trend that was subjected to further analysis to discern its implications for public policy and resource allocation.

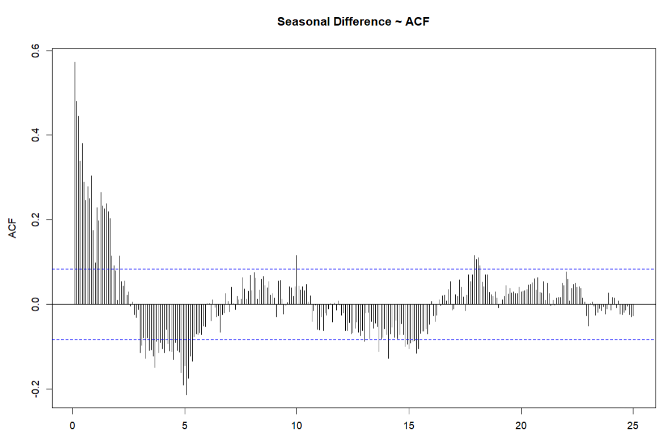

Fig. 4. ACF plot of differenced data.

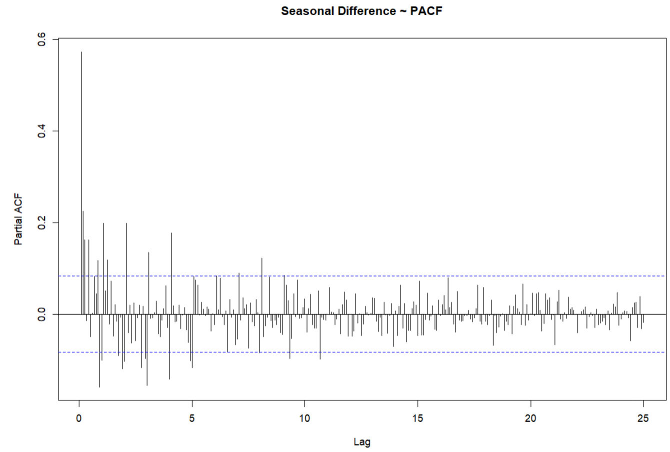

Fig. 5. PACF plot of differenced data.

Model Number	Model	AIC
1	SARIMA(0,1,0)x(2,0,3)	11423.53
2	SARIMA(0,1,0)x(3,0,3)	11422.00
3	SARIMA(0,1,0)x(4,0,4)	11423.66
4	SARIMA(0,1,3)x(4,0,4)	11315.07
5	SARIMA(3,1,0)x(2,0,3)	11198.92
6	SARIMA(3,0,0)x(3,1,3)	11198.38
7	SARIMA(0,1,0)x(5,0,5)	11268.99
8	SARIMA(0,0,3)x(5,1,5)	11321.22
9	SARIMA(0,0,5)x(5,1,5)	11280.38
10	SARIMA(3,0,5)x(5,1,5)	**11150.65**

Fig. 6. Births dataset model comparison.

3) ACF and PACF Analysis: Initially, the Autocorrelation Function (ACF) and Partial Autocorrelation Function (PACF) plots of the original USA births data were examined to visually assess stationarity. Upon observing the gradual decline in the ACF and a significant cut-off in the PACF, non-stationarity was suspected. This was subsequently confirmed with the Augmented Dickey-Fuller test. After making the data stationary by taking one difference, the ACF and PACF plots were revisited to determine the appropriate autoregressive terms for the ARIMA model, which indicated the number of autoregressive terms to include.

III. Model Fitting and Evaluation

Our modeling efforts focused on capturing the seasonal aspects inherent in the data. We leveraged SARIMA models, which allowed us to incorporate both non-seasonal and seasonal components into our analysis. The parameters were selected based on AIC criteria and diagnostic checks to ensure the model adequately captured the underlying processes. After extensive analysis, the final model was determined to be

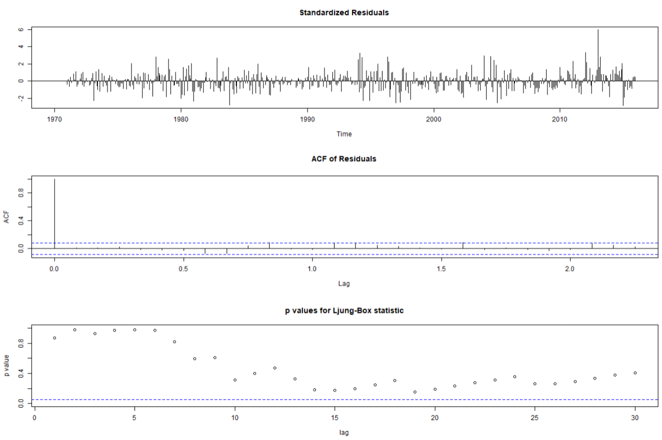

Fig. 7. Results of residuals checks on best model selection for birth data.

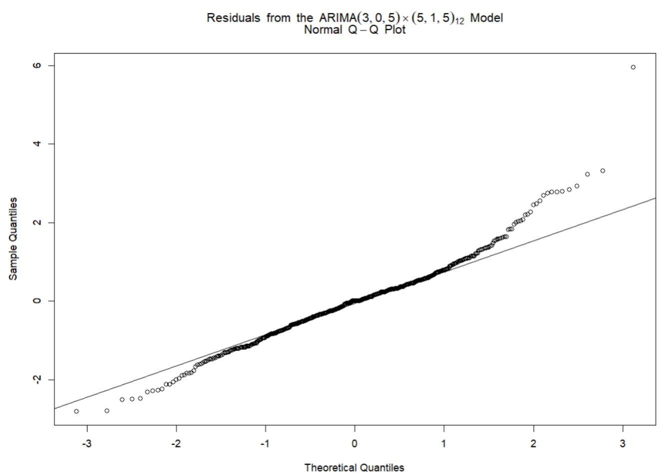

Fig. 8. Q-Q plot of residuals for births data.

$SARIMA(3, 0, 5) \times (5, 1, 5)_{12}$, which incorporates both seasonal and non-seasonal components identified in the dataset.

A. Residual Analysis

An analysis of the residuals from the fitted SARIMA model confirmed that the residuals behaved like white noise, indicating a good fit. There were no apparent patterns or autocorrelation in the residuals, as evidenced by the ACF and Ljung-Box tests. Additionally, the 'tsdiag' function was utilized to provide a deeper assessment of the model's adequacy, further ensuring that the model fits the data well.

B. Forecast Evaluation

The validation of our forecasted model was substantiated by analyzing the standardized residuals to ensure that they behaved as white noise, which indicates a well-fitting model. The plot of standardized residuals demonstrates a random pattern centered around zero without any evident trends or seasonality (see Figure 7). The accompanying Autocorrelation Function (ACF) of residuals and p-values for Ljung-Box statistic further corroborate the absence of autocorrelation,

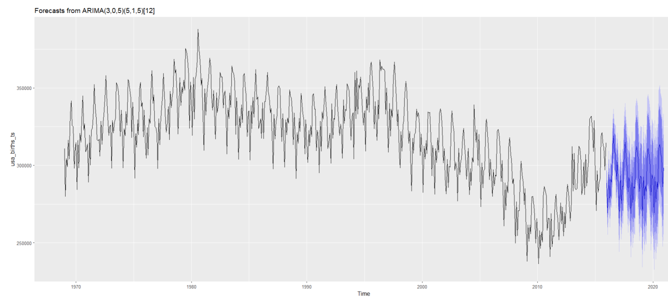

Fig. 9. Forecast of USA births.

confirming that the residuals do not exhibit systematic patterns that the model failed to capture.

Following the residual analysis, forecasts were generated using the best model parameters. The evaluation of these forecasts relied on both visual inspection and quantitative metrics such as Mean Absolute Error (MAE), Root Mean Squared Error (RMSE), and Mean Absolute Percentage Error (MAPE). These measures are critical for assessing the model's accuracy in predicting future values and for benchmarking against other forecasting models.

Although specific forecast plots are not shown here, the consistency of the residual behaviors gives us confidence in the reliability of the forecasts produced by the model. Future research may focus on presenting detailed forecast accuracy metrics alongside the validation of residuals to present a comprehensive view of the model's performance.

IV. Limitations and Further Research

While the models selected provide significant insights into the datasets, there are inherent limitations that must be acknowledged, and which open avenues for further research.

A. Limitations

The primary limitation in our analysis stems from the models' assumptions, such as linearity and stationarity, which may not fully capture the complexities of real-world data. Additionally, the exclusion of potential external variables that could influence the trends in birthrates might limit the scope of our conclusions.

B. Further Research

Future research could explore the inclusion of external regressors in the models, such as demographic factors for birthrates. Furthermore, exploring non-linear models or machine learning approaches could provide deeper insights, especially in capturing complex patterns that traditional time series models might overlook.

Fig. 10. Bitcoin Closing Prices Over Time plot. The plot shows a general upward trend with significant volatility and notable peaks and troughs.

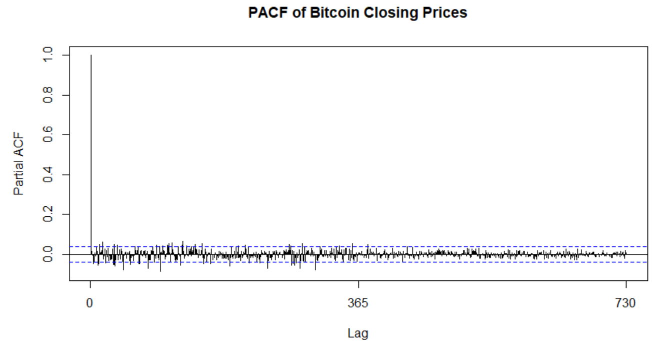

Fig. 12. PACF of Bitcoin Closing Prices plot. The plot shows no significant spikes, implying minimal autoregressive behavior is present in the data.

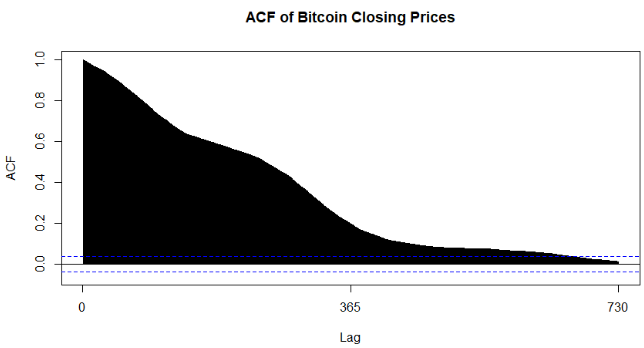

Fig. 11. ACF of Bitcoin Closing Prices plot. The slow decay indicates a non-stationary series with significant autocorrelation at short lags.

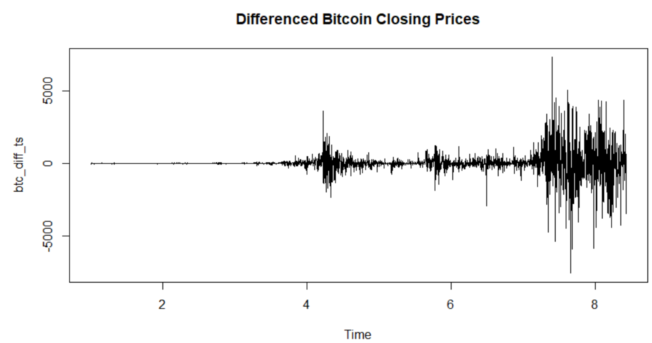

Fig. 13. Differenced Bitcoin Closing Prices plot. The absence of a trend suggests that the differencing process has stabilized the mean of the series.

V. Bitcoin Dataset Analysis

A. Data Preprocessing and Visualization

The preprocessing of the Bitcoin dataset was meticulous, ensuring the integrity and quality of the data for analysis. The date column was standardized to a Date format to facilitate chronological analysis. A detailed examination of Bitcoin's closing prices over time highlighted the characteristic volatility of the cryptocurrency market. The initial plot (Figure 10) depicts a general upward trend starting around 2017 with sharp fluctuations, indicative of both bullish and bearish markets.

1) Initial Trends and Volatility: An exploration of the Bitcoin closing prices' ACF (Figure 11) reveals a high level of autocorrelation, especially at shorter lags. The slow decay of the ACF suggests a non-stationary time series, implying that past prices are indicative of future prices.

The PACF plot (Figure 12) assists in determining the order of the autoregressive terms for our ARIMA model. Most of the spikes within the confidence bounds suggest that autoregressive terms are not significant after the first few lags.

B. Stationarity and Differencing

To address non-stationarity, differencing was applied to the Bitcoin prices, as confirmed by the Augmented Dickey-Fuller test. The differenced prices plot Figure 13 exhibits no visible trend, which suggests stationarity has been achieved. This step is crucial to ensure the suitability of the ARIMA model.

C. Bitcoin Data Model Selection

The Bitcoin price analysis required models that could handle the high volatility and non-seasonal nature of financial data. An ARIMA(0,1,1) model was initially fit to address the series' non-stationarity. To adequately model the pronounced volatility of Bitcoin prices, a GARCH(1,1) component was integrated with the ARIMA model. This combination was chosen for its effectiveness in capturing the 'volatility clustering' often observed in financial markets, where high volatility periods are followed by high volatility and low by low.

D. Model Selection and Fitting

Based on the ACF and PACF plots of the differenced data, an ARIMA model was selected. Post-model fitting diagnostics on the ARIMA residuals and the ACF of residuals (Figure 14) indicate that the residuals behave like white noise, confirming the model's adequacy.

The forecasts generated from the ARIMA model (Figure 15) were evaluated for accuracy against historical data, showing the model's proficiency in predicting future values within a specified confidence interval.

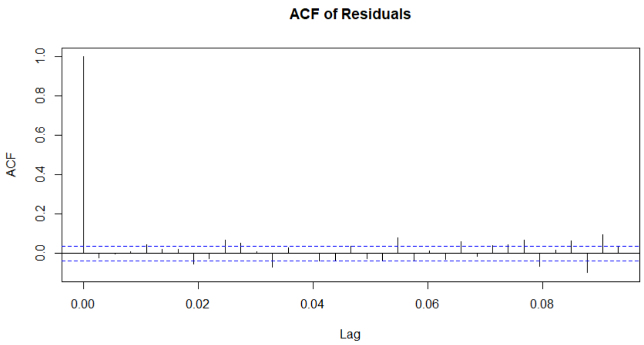

Fig. 14. ACF of Residuals plot. The plot suggests that the residuals are uncorrelated, as they fall within the confidence bounds.

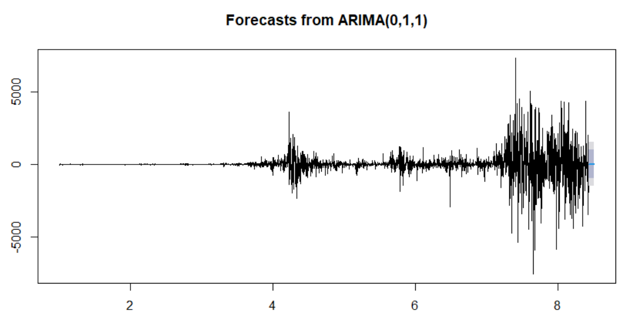

Fig. 15. Forecasts from ARIMA Model plot. The future values predicted by the ARIMA model are depicted, showing the model's range of expected future values.

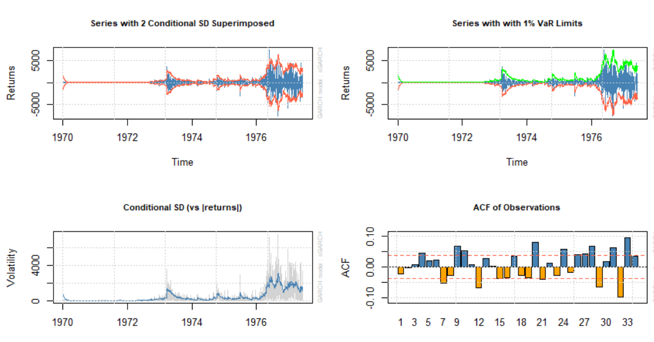

Fig. 16. Conditional Volatility from GARCH Model plot. This plot shows the volatility clustering phenomenon, with periods of high volatility followed by periods of relative calm.

E. Volatility Modeling with GARCH

To account for the pronounced volatility in Bitcoin's price, a GARCH model was fitted. The model's conditional volatility plot (Figure 16) highlights periods of high and low volatility, which are crucial for understanding market risk.

VI. Conclusion

This report presented a comprehensive analysis of two very different datasets using advanced time series models to unearth underlying patterns and predict future trends. In the analysis of U.S. birthrates, we utilized ARIMA and SARIMA models to identify and forecast seasonal and non-seasonal patterns, providing insights that are crucial for social planning and public policy. Contrastingly, the Bitcoin price analysis employed both ARIMA and GARCH models to tackle the erratic behavior typical of financial markets, specifically highlighting the challenges and strategies in predicting volatile market movements.

Our findings illustrate the versatility and robustness of time series analysis in handling data with distinct characteristics: the predictable, cyclic nature of birthrates, and the unpredictable volatility of Bitcoin prices. The juxtaposition of these analyses not only demonstrates the adaptability of statistical models to various domains but also enhances our understanding of the factors driving changes in both biological and financial contexts.

The insights gained from this dual analysis approach reinforce the importance of tailored modeling strategies to address specific characteristics of the dataset under study. As we continue to refine these models and adapt them to new datasets, the potential for time series analysis to contribute to diverse fields remains promising and expansive.