**Final Project**

**ARIMA, SARIMA, ARCH, GARCH, and State-space Model**

**Name:** Krishna Prasad Palvai

**Student ID:** 922672795

**ARIMA**

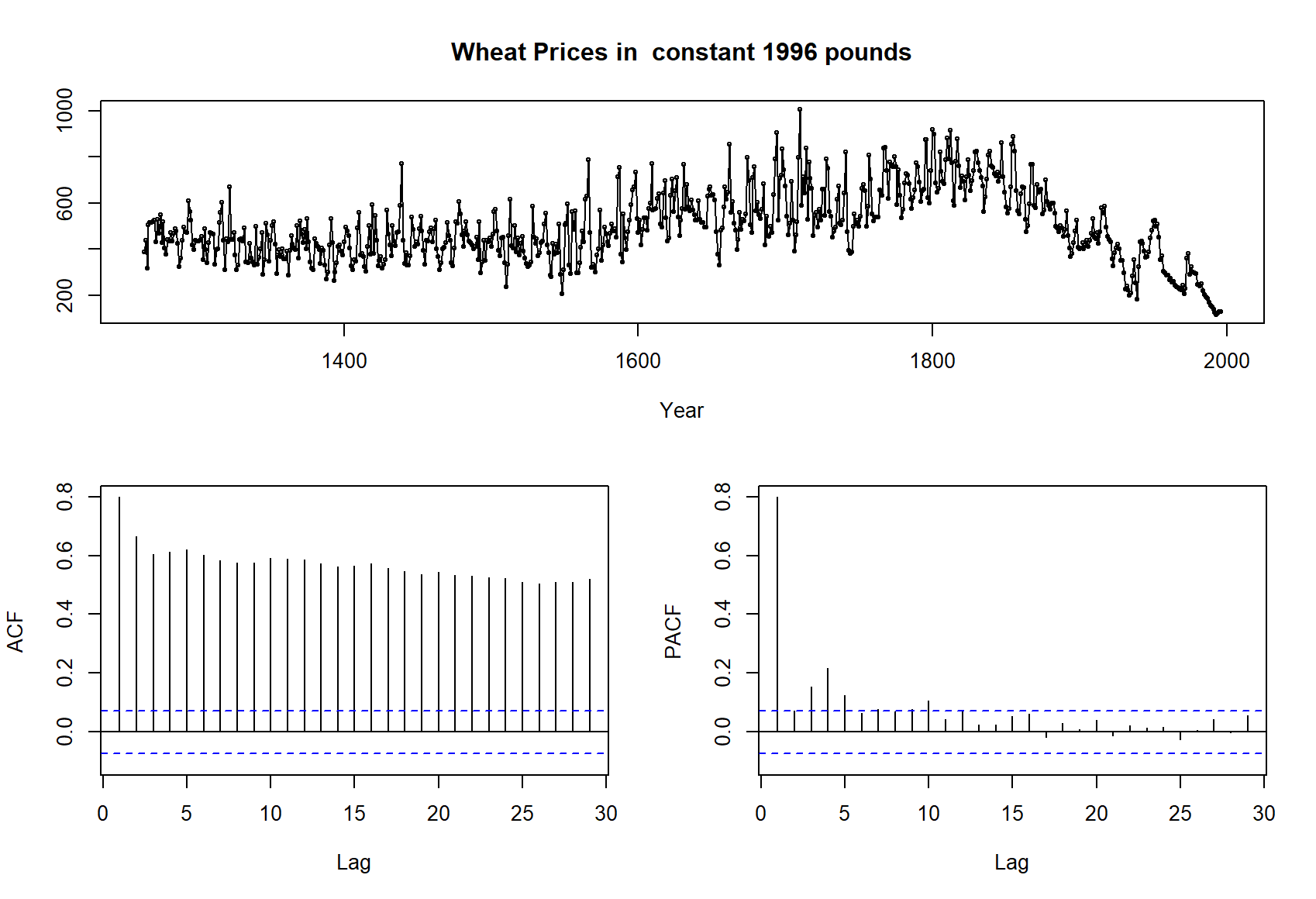
ARIMA (Autoregressive Integrated Moving Average) models are fundamental time series models used for forecasting without incorporating external factors. They rely on past observations to predict future values, being a subset of linear regression models for time series analysis. "AutoRegressive" signifies regression against the series' past values, while "Integrated" indicates applying differencing to achieve stationarity. These models do not inherently consider external variables.

However, ARIMA models have limitations. They may struggle with nonlinear relationships and might not effectively capture complex patterns or sudden changes in data. Assumptions for ARIMA include stationarity post differencing, constant variance, and independence of residuals.

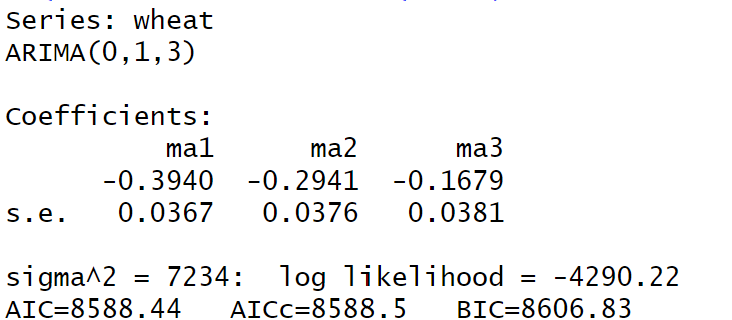
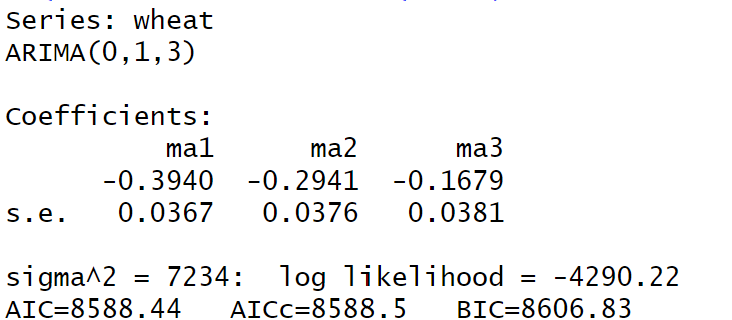
Despite these limitations, ARIMA models have various applications in sales and revenue forecasting. Their simplicity and effectiveness in capturing linear relationships in historical data make them suitable for short-term predictions in scenarios with stable, linear patterns.

Mathematically, ARIMA is represented by ARIMA(p, d, q), where p represents autoregression order, d denotes differencing, and q signifies the moving average order. The model can be expressed as y’(t) = c + ϕ1\* y′(t−1) +⋯ + ϕp\*y′(t−p) + θ1\*ε(t−1) +⋯ + θq\*ε(t−q) + εt

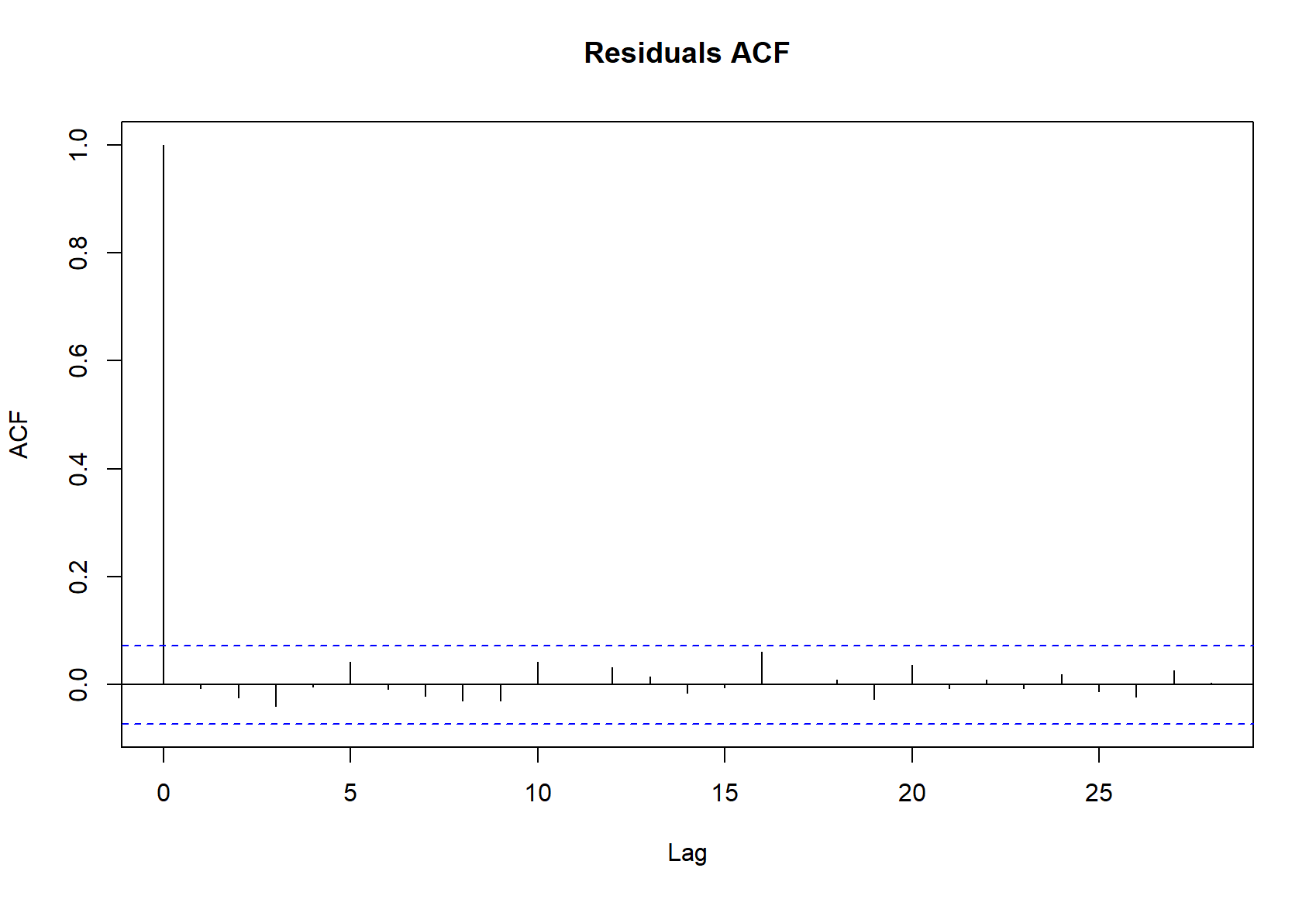
In essence, ARIMA models provide a basic yet effective approach for short-term forecasting in scenarios where historical patterns exhibit linearity and require no external influence considerations.

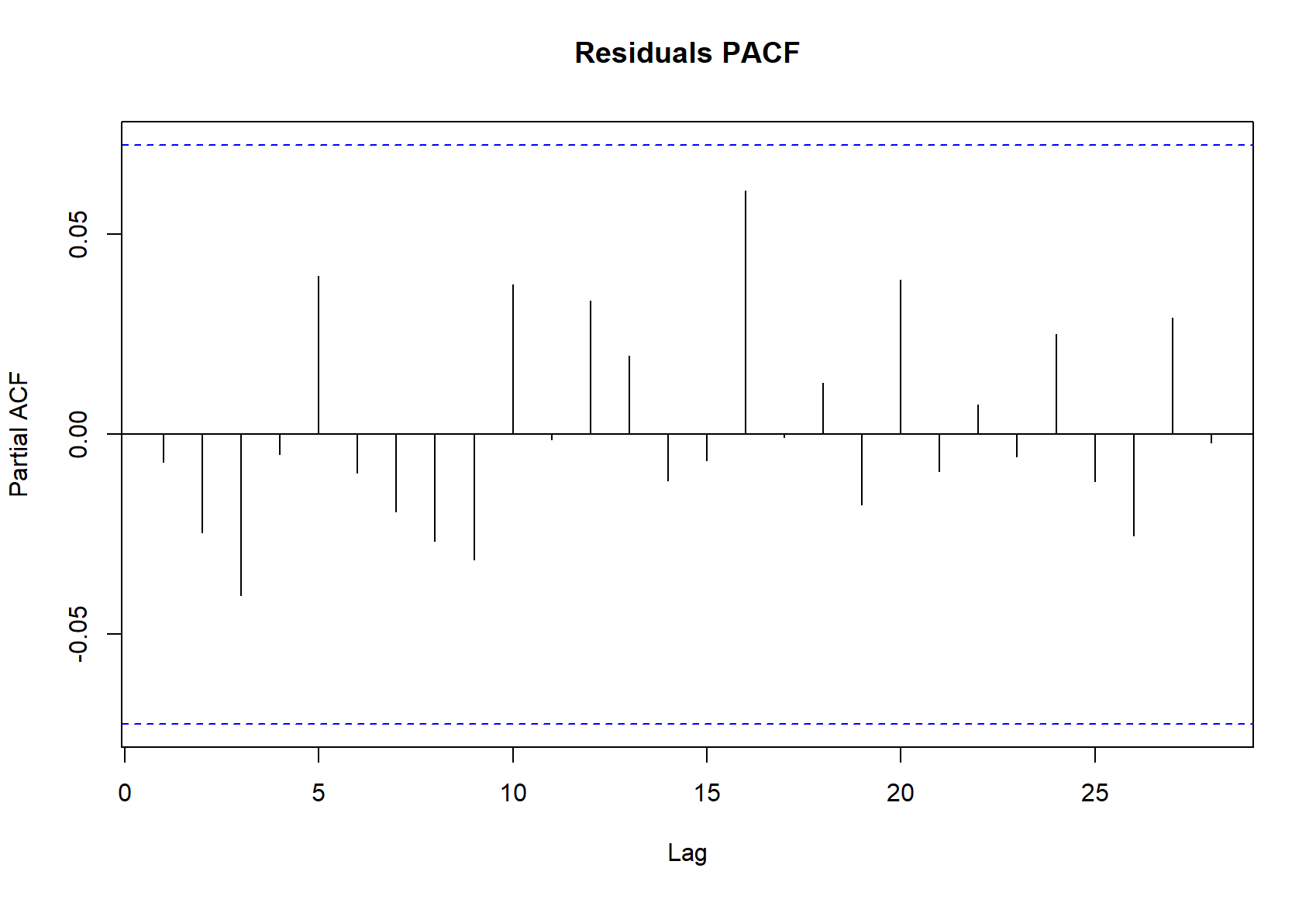


Timeseries containing wheat prices is used for ARIMA modelling. ACF plot of original time series indicate there is high autocorrelation.

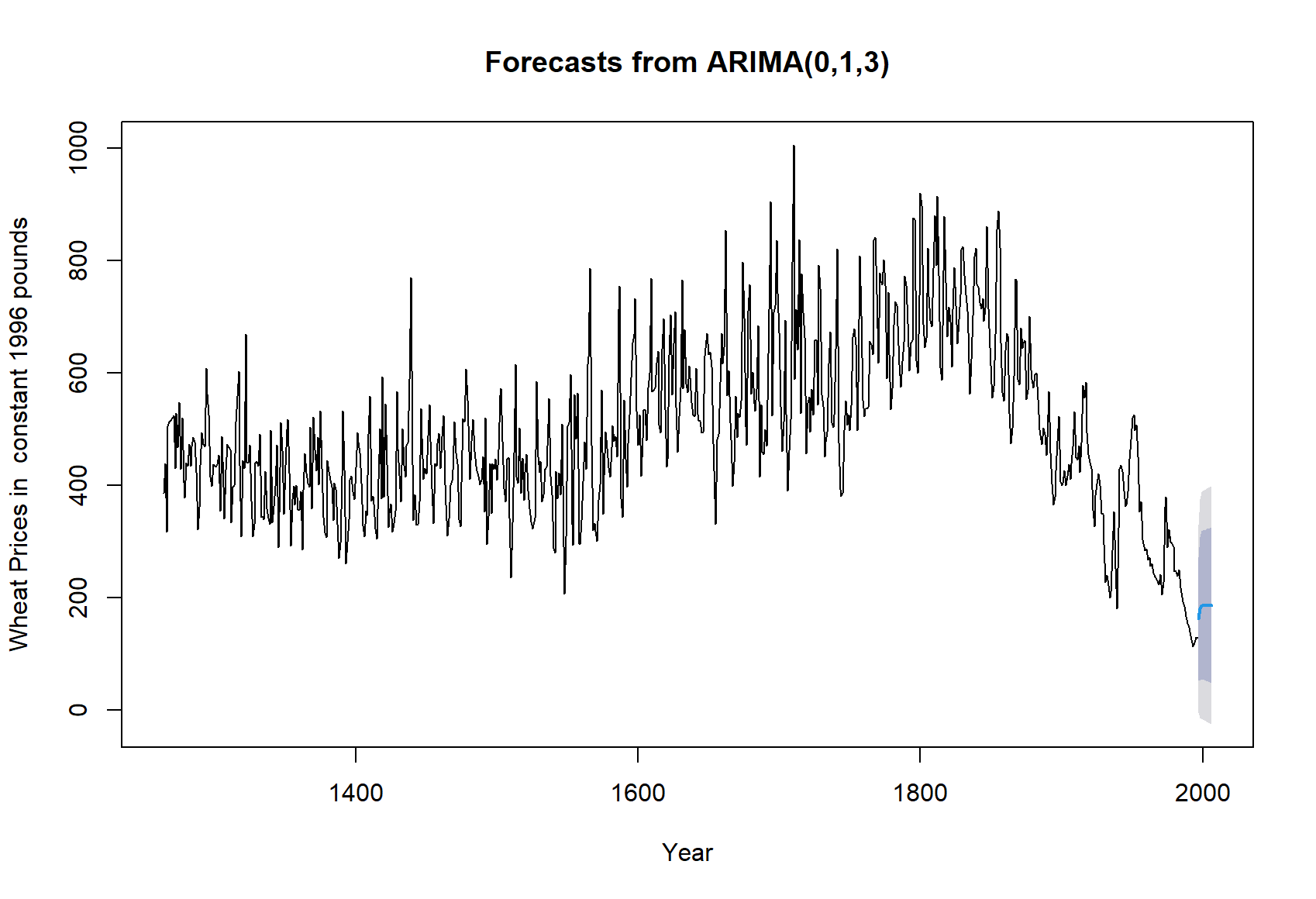
 

ARIMA(0,1,3) is the best model to forecast wheat prices based on Information criteria value criteria.





The ACF and PACF plots for residuals exhibit no significant out-of-bound lags, indicating a clean and satisfactory pattern.



Forecast of wheat prices with ARIMA(0,1,3) model also indicating 80% and 95% confidence interval for the point estimate.

**SARIMA**

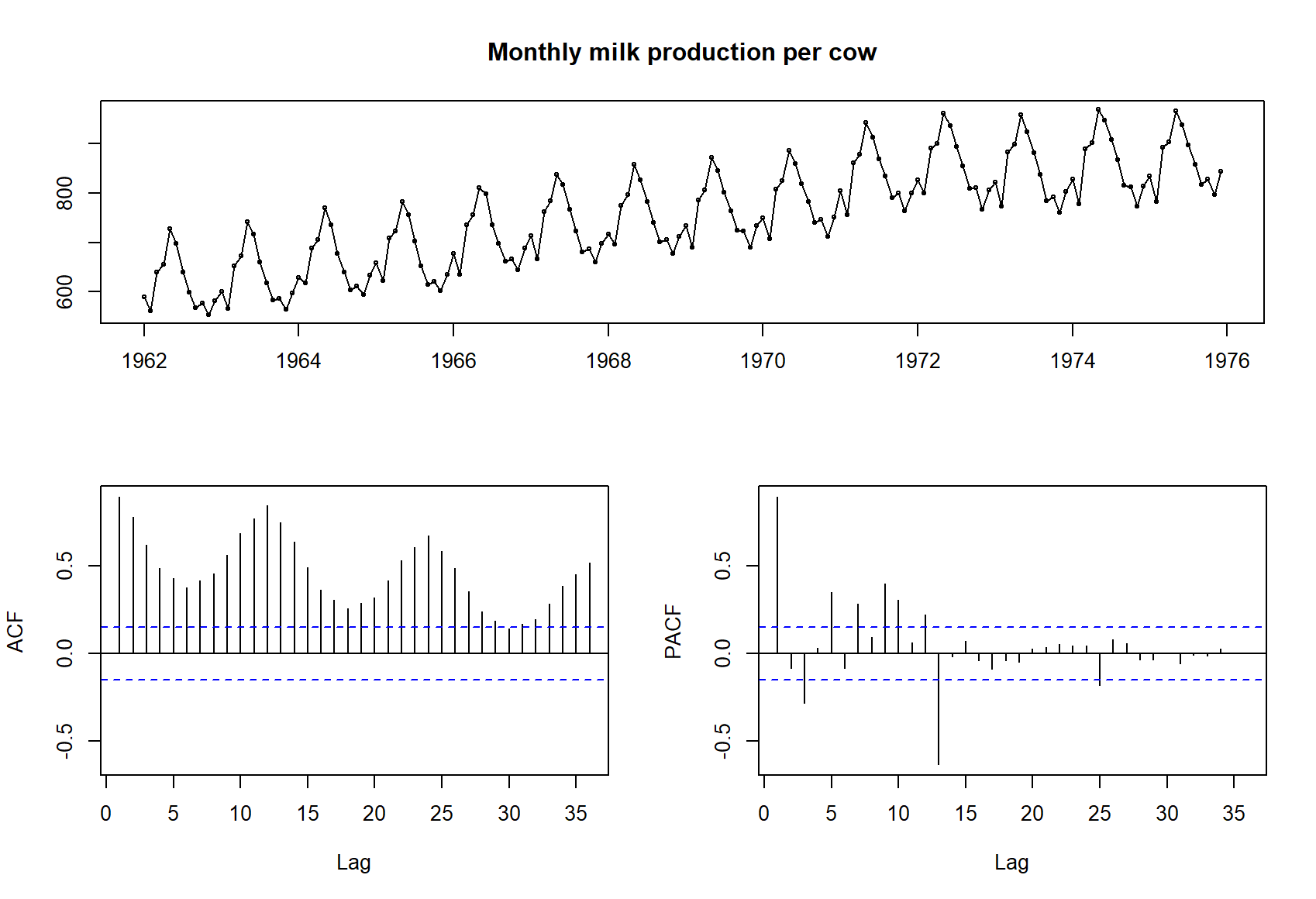
SARIMA (Seasonal AutoRegressive Integrated Moving Average) models are potent statistical tools tailored for forecasting time series data characterized by periodic patterns. Widely applied in various domains like retail sales, stock prices, and climate data analysis, SARIMA models are instrumental in predicting patterns influenced by seasonality.

The model is defined by six parameters: p, d, q, P, D, and Q representing autoregressive, differencing, and moving average orders, both for the non-seasonal and seasonal components. These parameters guide the identification of the model structure and assist in capturing the inherent seasonal patterns within the data.

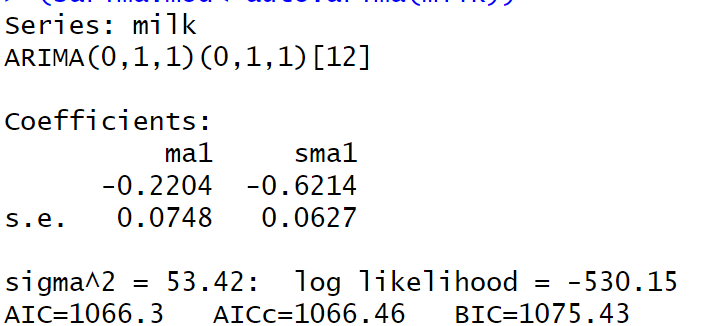
However, SARIMA models come with limitations. They might struggle with capturing complex nonlinear relationships or irregular patterns within the data. The assumptions for SARIMA encompass stationarity post differencing, constant variance, and independence of residuals.

Despite limitations, SARIMA models find significant application in forecasting fields influenced by seasonality or cyclic trends. Their ability to account for periodic patterns makes them particularly useful in predicting time series data with repetitive seasonal fluctuations.

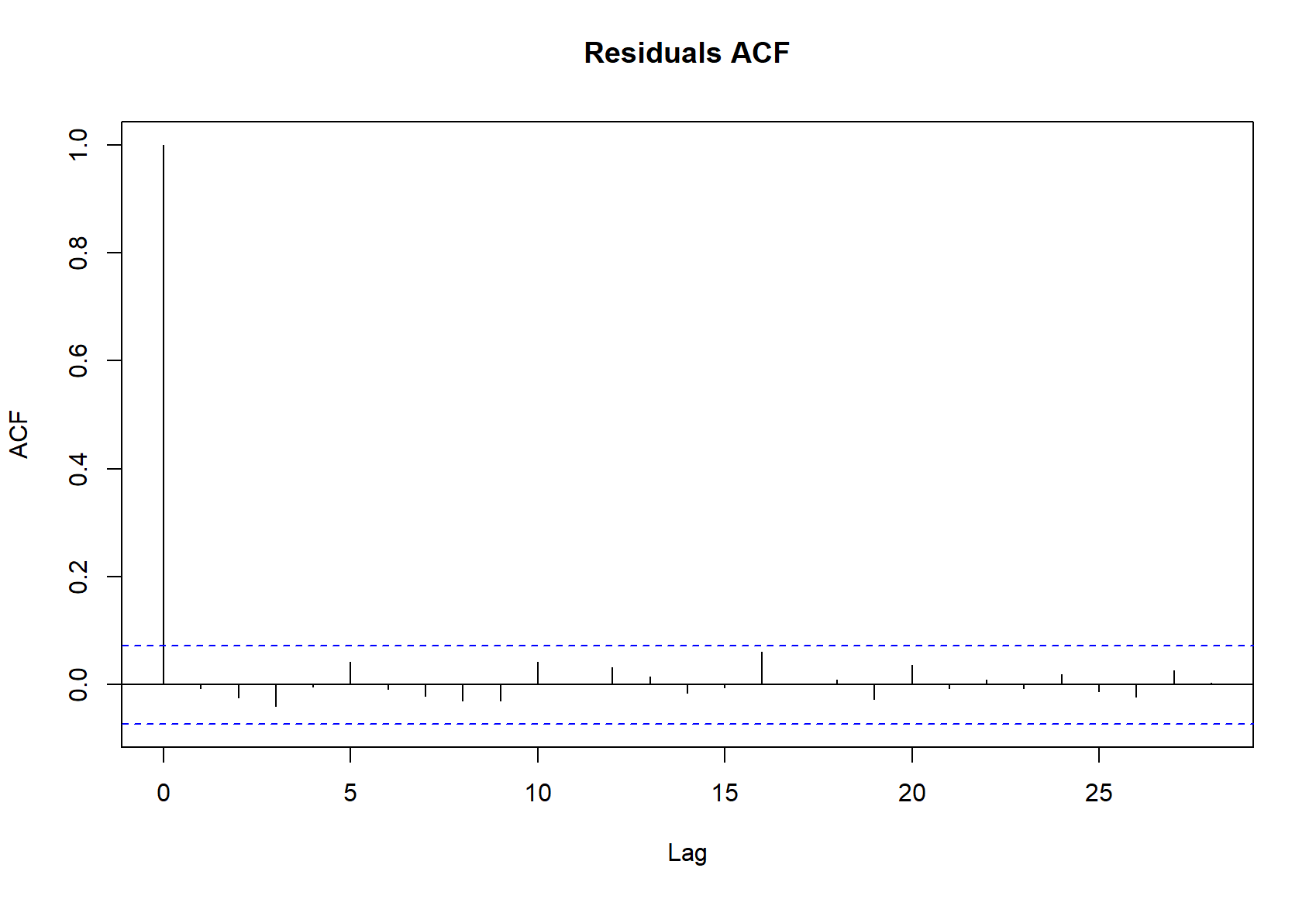
Mathematically, the SARIMA model is represented as SARIMA(p, d, q) (P, D, Q)m, where p, d and q represent non-seasonal ARIMA components, and P, D and Q denote seasonal ARIMA components with period m. The model's expression incorporates both non-seasonal and seasonal components to capture seasonal patterns and variations in the time series data.

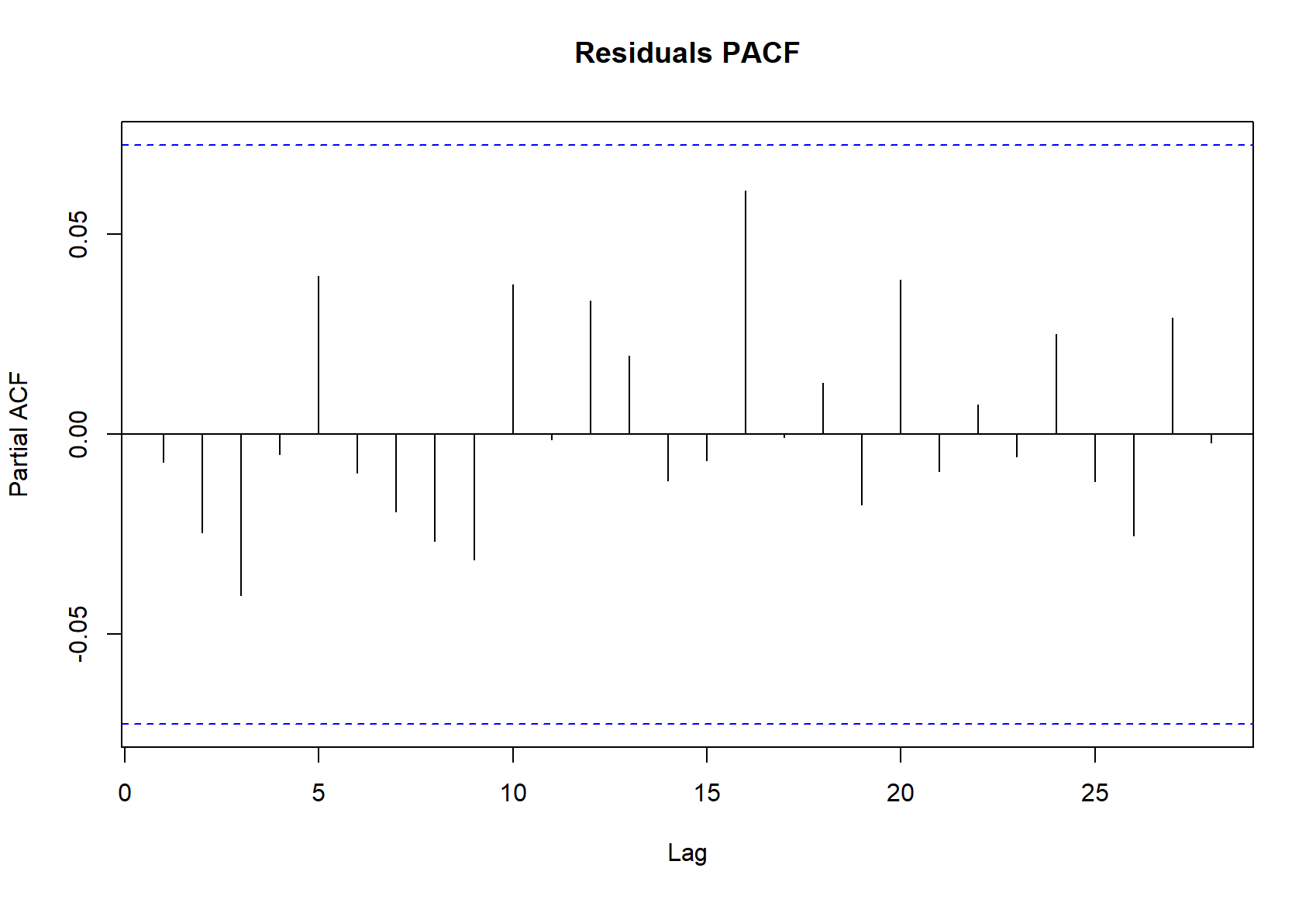


Timeseries containing monthly milk production is used for SARIMA modelling. ACF plot of original time series indicate there is high autocorrelation.

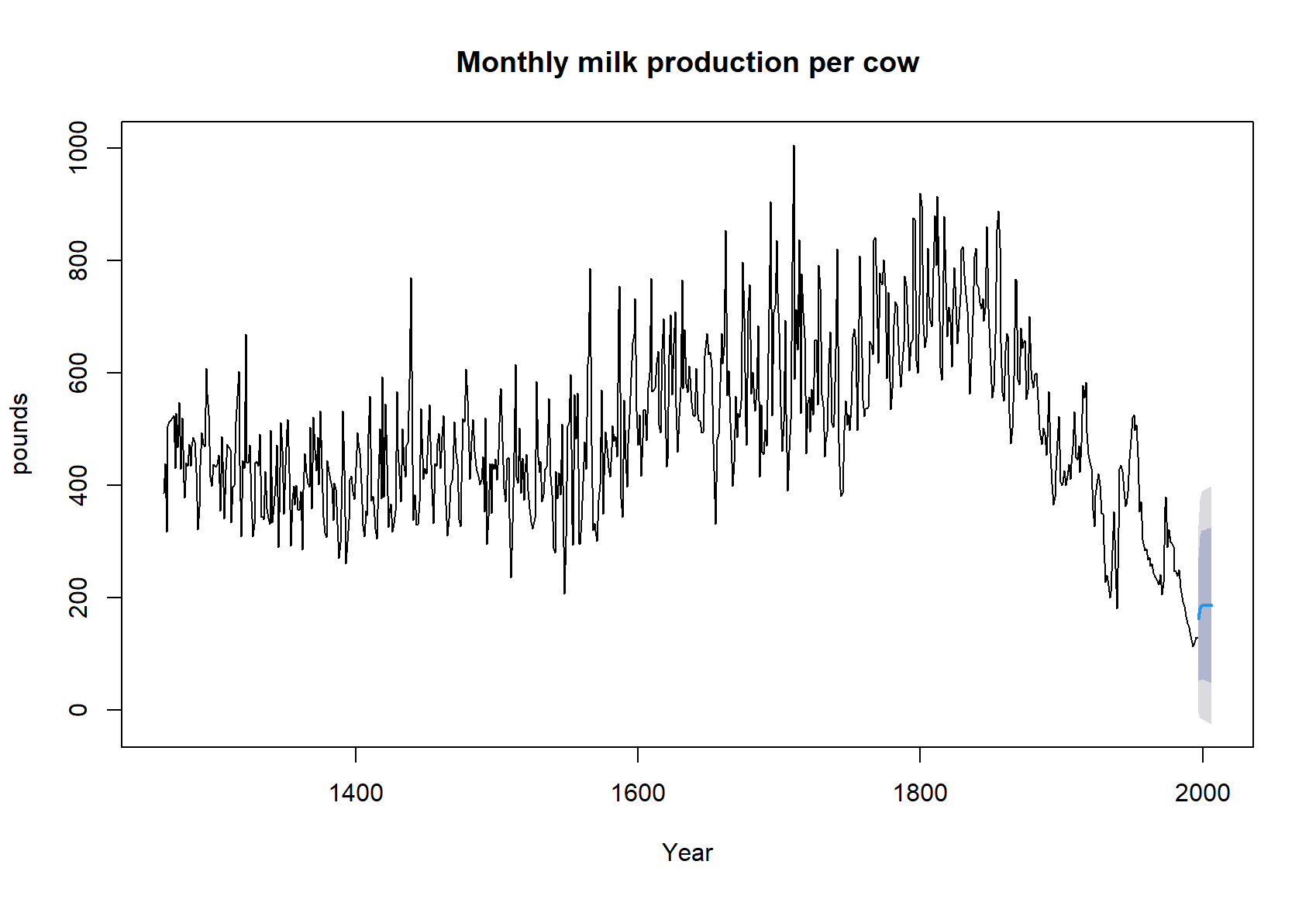


ARIMA(0,1,1)(0,1,1)[12] is the best SARIMA model to forecast monthly milk production based on Information criteria value criteria.





The ACF and PACF plots for residuals exhibit no significant out-of-bound lags, indicating a clean and satisfactory pattern.



Forecast of monthly milk production with ARIMA(0,1,1)(0,1,1)[12] model also indicating 80% and 95% confidence interval for the point estimate.

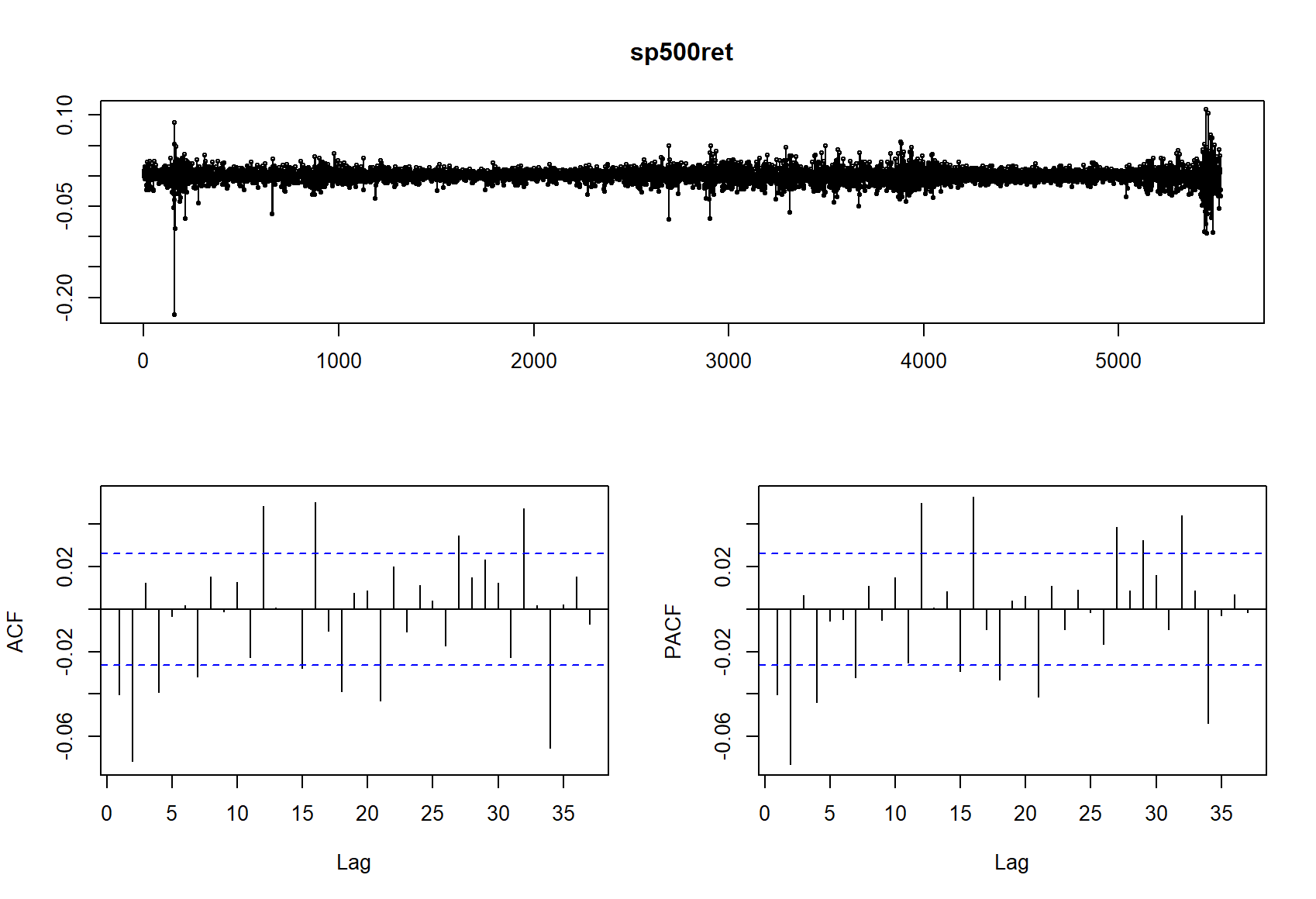
**ARCH**

The Autoregressive Conditional Heteroskedasticity (ARCH) model, conceptualized by Robert Engle in 1982, addresses changing variance or volatility patterns observed in financial and economic time series data. This model posits that the variance of a series relates to its past squared residuals, signifying that high residual values correspond to increased volatility periods.

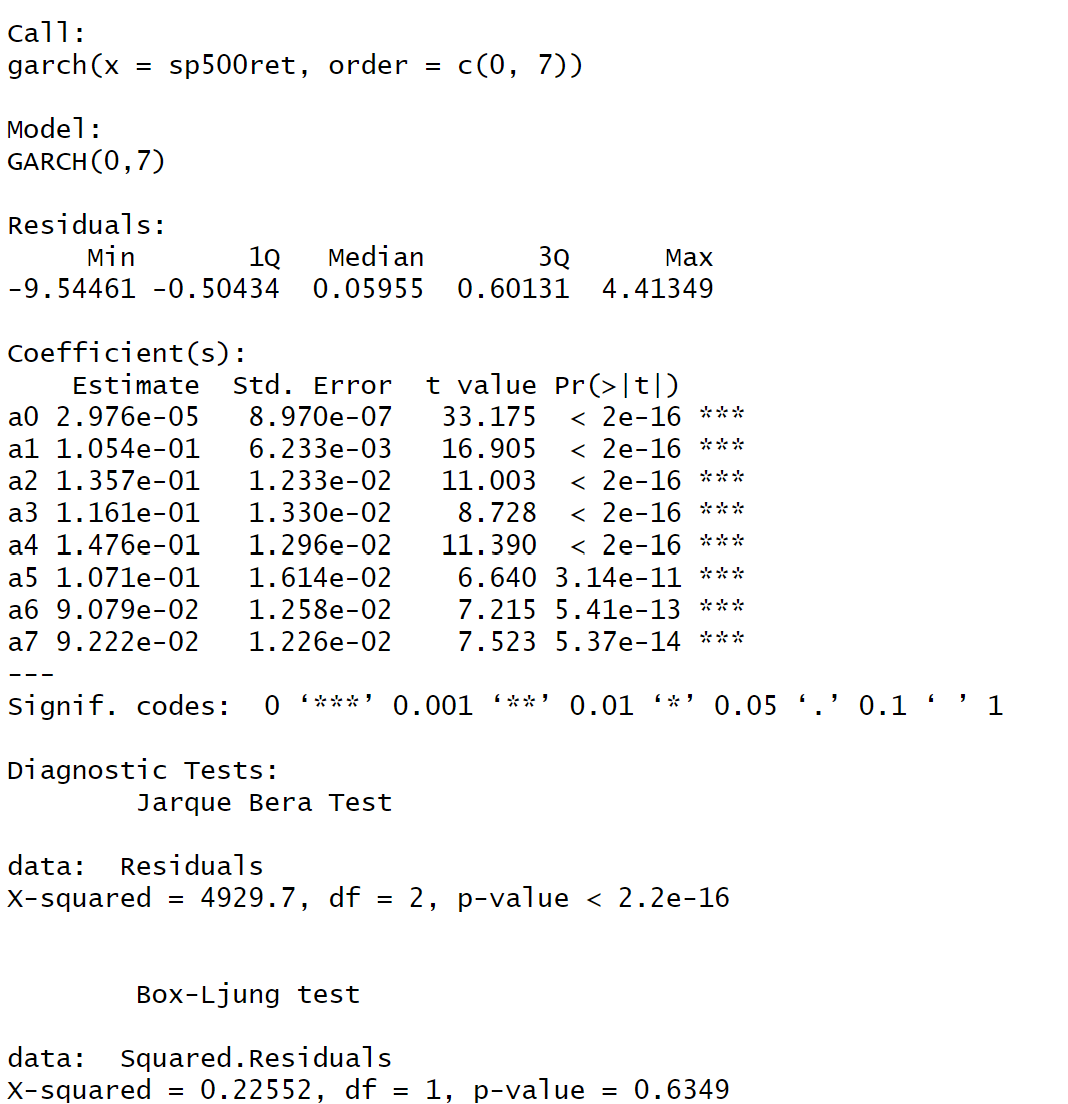
Primarily utilized in finance, particularly in risk management, the ARCH model forecasts asset return volatility, estimates Value at Risk (VaR), and assesses risks associated with diverse financial instruments and portfolios. Its applications extend beyond finance; it finds utility in economics to gauge economic uncertainty and in meteorology for modelling weather volatility. Additionally, ARCH is beneficial in engineering for analysing variability in processes and systems.

However, the ARCH model has limitations. It assumes no structural changes in volatility patterns, and its effectiveness might diminish if underlying assumptions regarding residual distributions are violated. Yet, its flexibility allows for diverse applications and provides insights into data with varying volatilities.

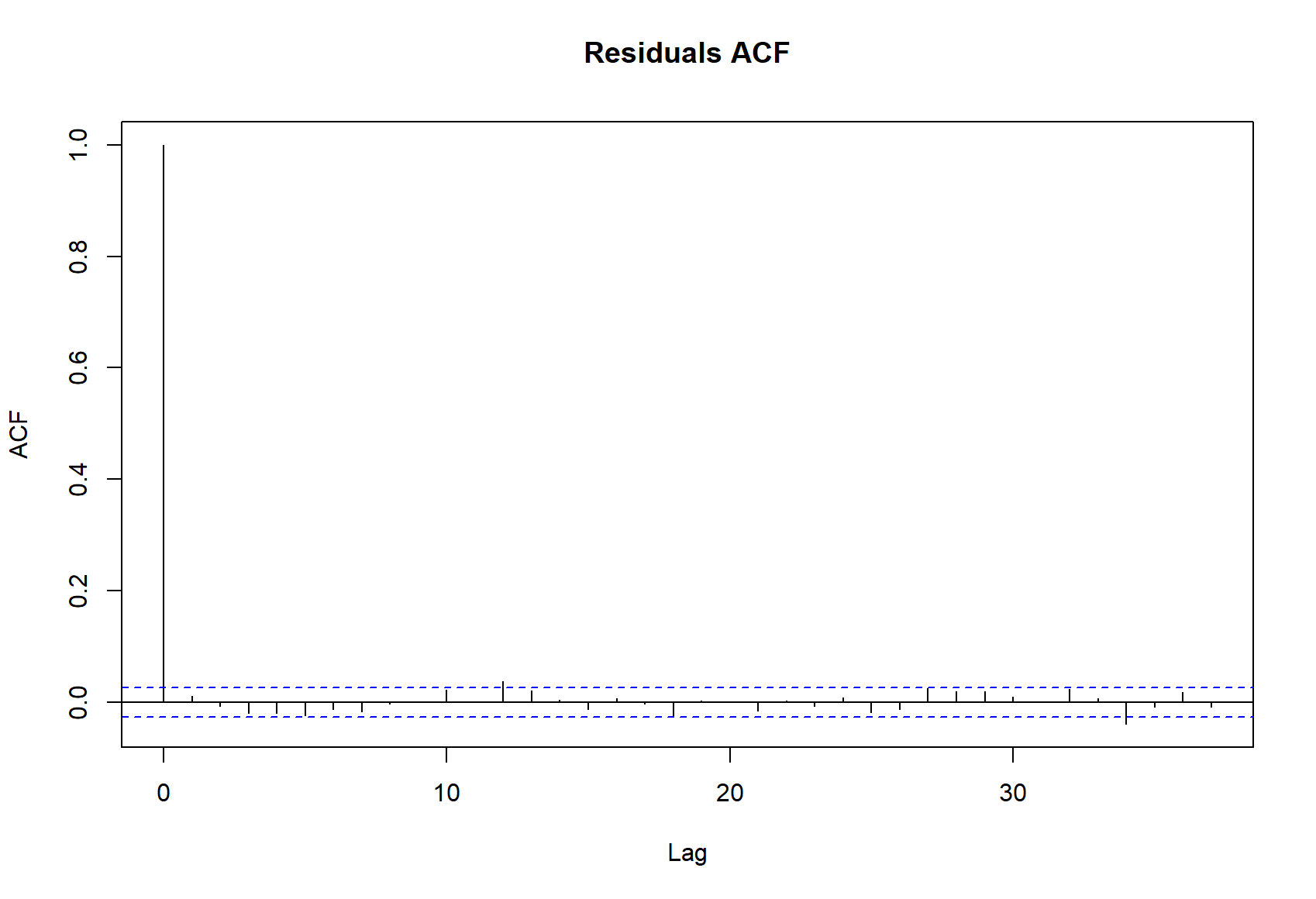
Mathematically, ARCH is represented as ARCH(q), where q denotes the number of lagged squared residuals used to model current volatility. The model's formulation captures time-varying volatility, aiding in comprehending and managing volatility dynamics in different fields.

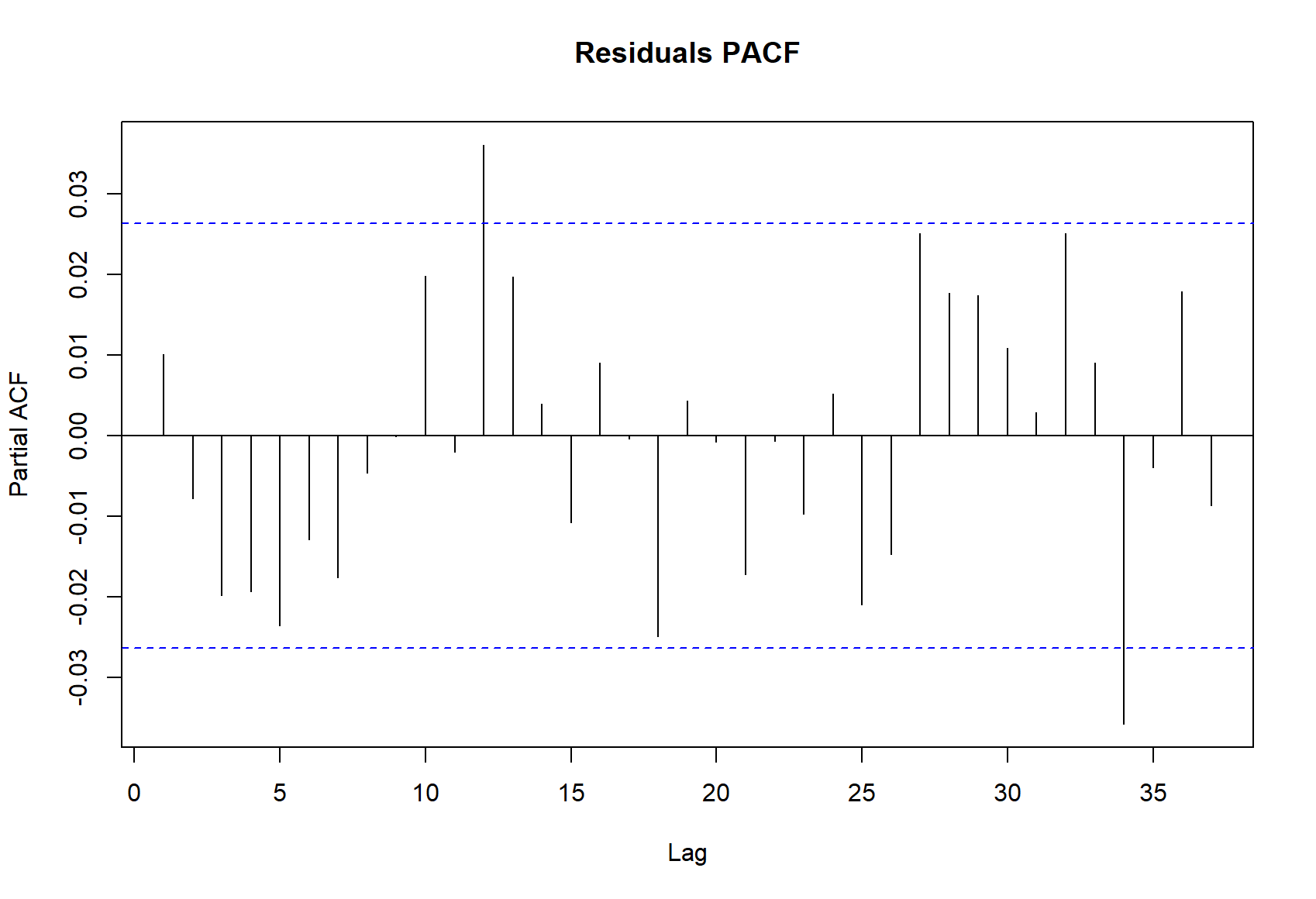


Timeseries containing SP500 returns is used for ARCH and GARRCH modelling.



ARCH(7) model is fit to estimate the variance in SP500 using the first seven lag of squared residuals. All seven lags of squared residuals are statistically significant.





The ACF plot of residuals exhibit no significant out-of-bound lags but PACF has two out-of-bound lags.

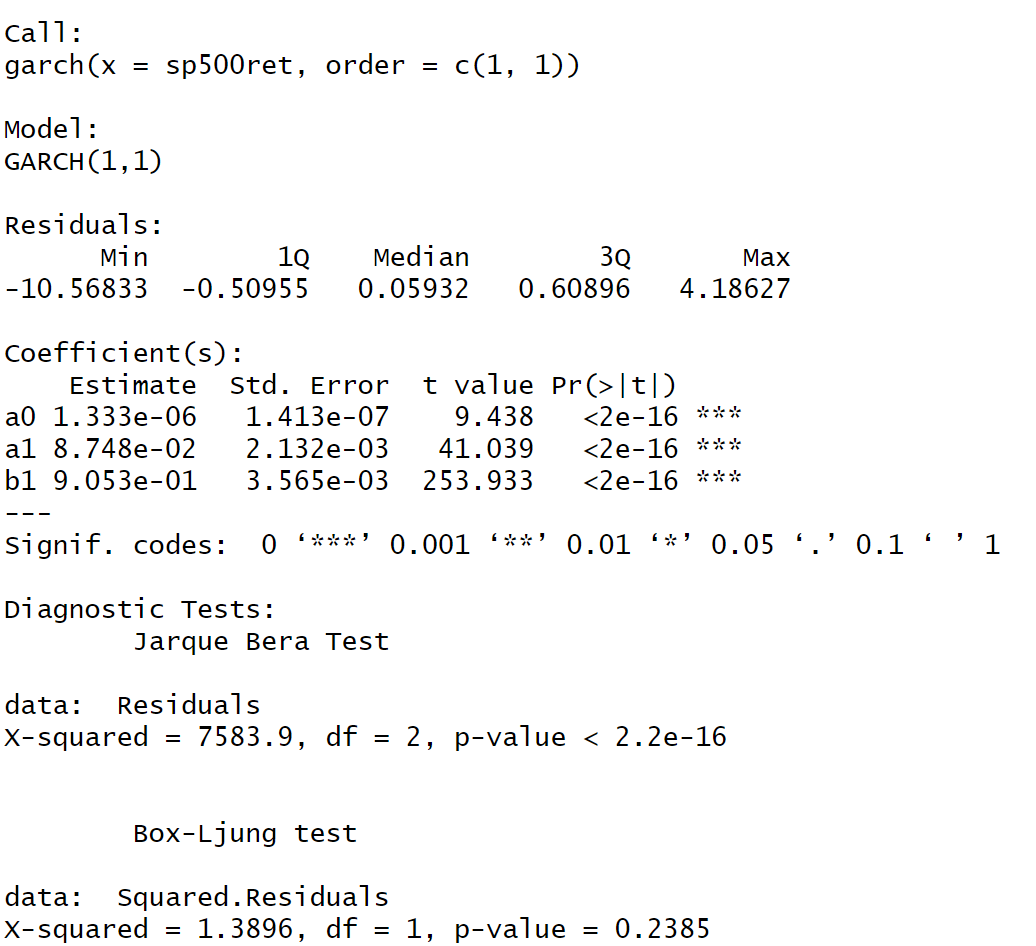
**GARCH**

The Generalized Autoregressive Conditional Heteroskedasticity (GARCH) model, introduced by Tim Bollerslev in 1986, aims to capture volatility clustering and time-varying variances in diverse datasets, both financial and non-financial. GARCH extends the Autoregressive Conditional Heteroskedasticity (ARCH) model by integrating past squared residuals and lagged conditional variances to model time series variance.

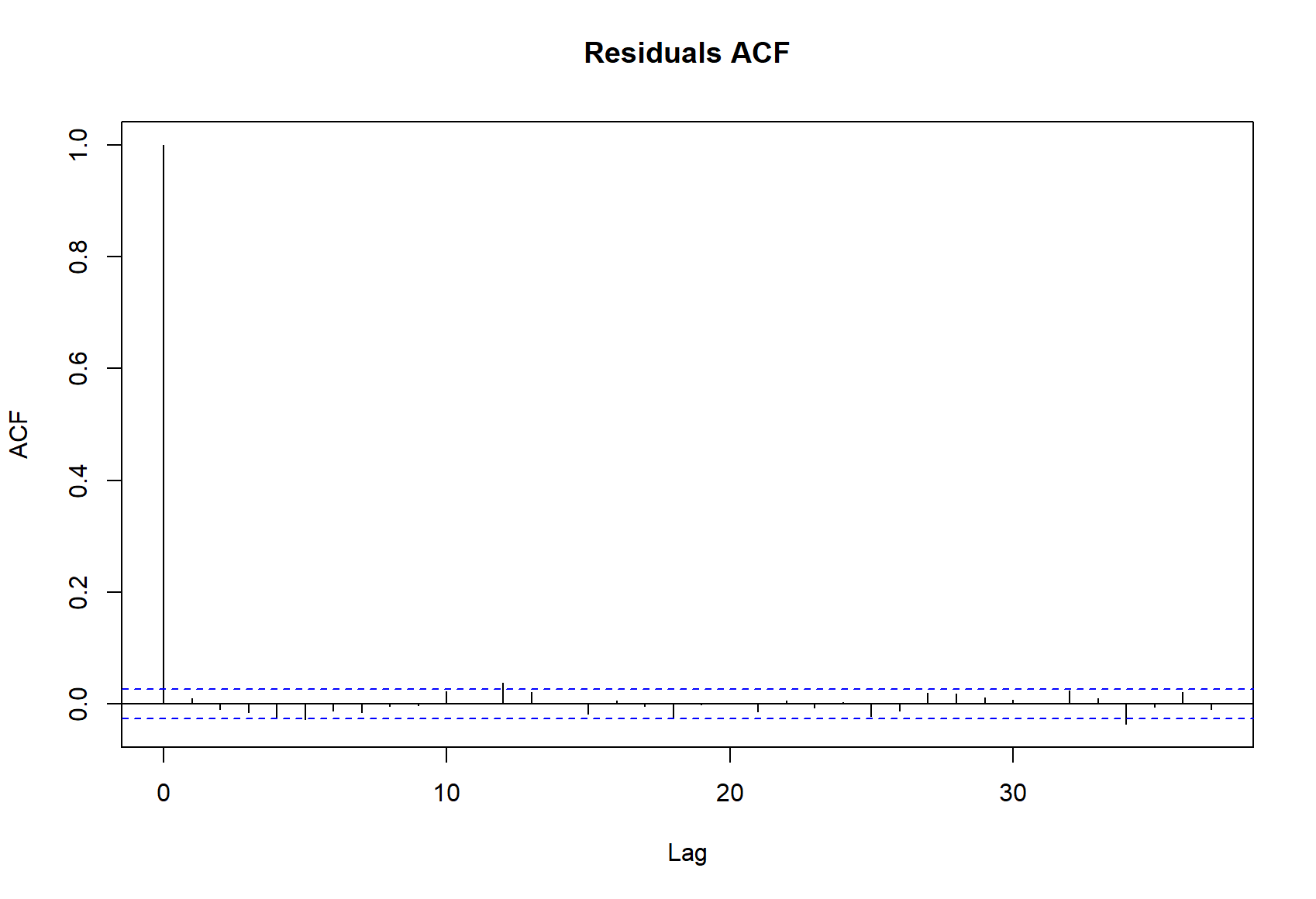
GARCH's primary application lies in finance, serving multiple purposes. It is pivotal in forecasting volatility, estimating Value at Risk (VaR), and optimizing portfolio allocations. Its role aids traders and investors in comprehending asset-related risks for making informed decisions. GARCH is also crucial in options pricing, where volatility serves as a critical input.

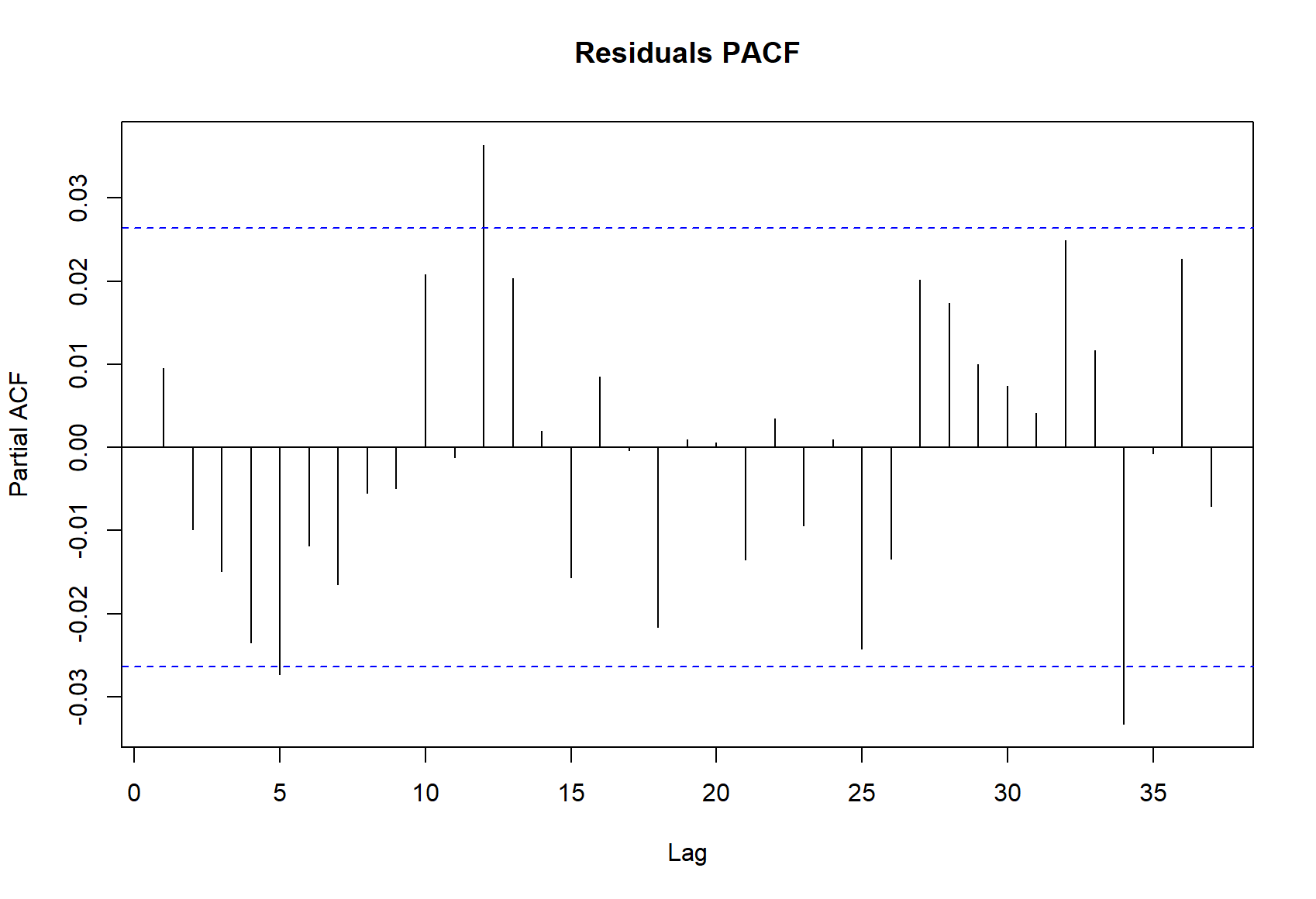
Beyond finance, GARCH finds utility in meteorology for modelling weather volatility, engineering for quality control, and economics to model inflation and economic uncertainty. Its adaptability in various fields is attributed to its capability to address changing volatility over time.

However, GARCH has limitations as it assumes volatility patterns to persist indefinitely and might struggle if assumptions about residual distributions are violated. Mathematically, GARCH is denoted as GARCH(p, q), where p represents autoregressive terms and q stands for moving average terms in the model. GARCH's adaptability makes it an invaluable tool for comprehending and managing risks across dynamic and uncertain environments in diverse fields.



GARCH(1,1) model is fit to estimate the variance in SP500 using the first lag of squared residual and first lag of variance. Both coefficients of lag of squared residual and lag of variance are statistically significant.





The ACF plot of residuals exhibit no significant out-of-bound lags but PACF has three out-of-bound lags.

**State-space Model**

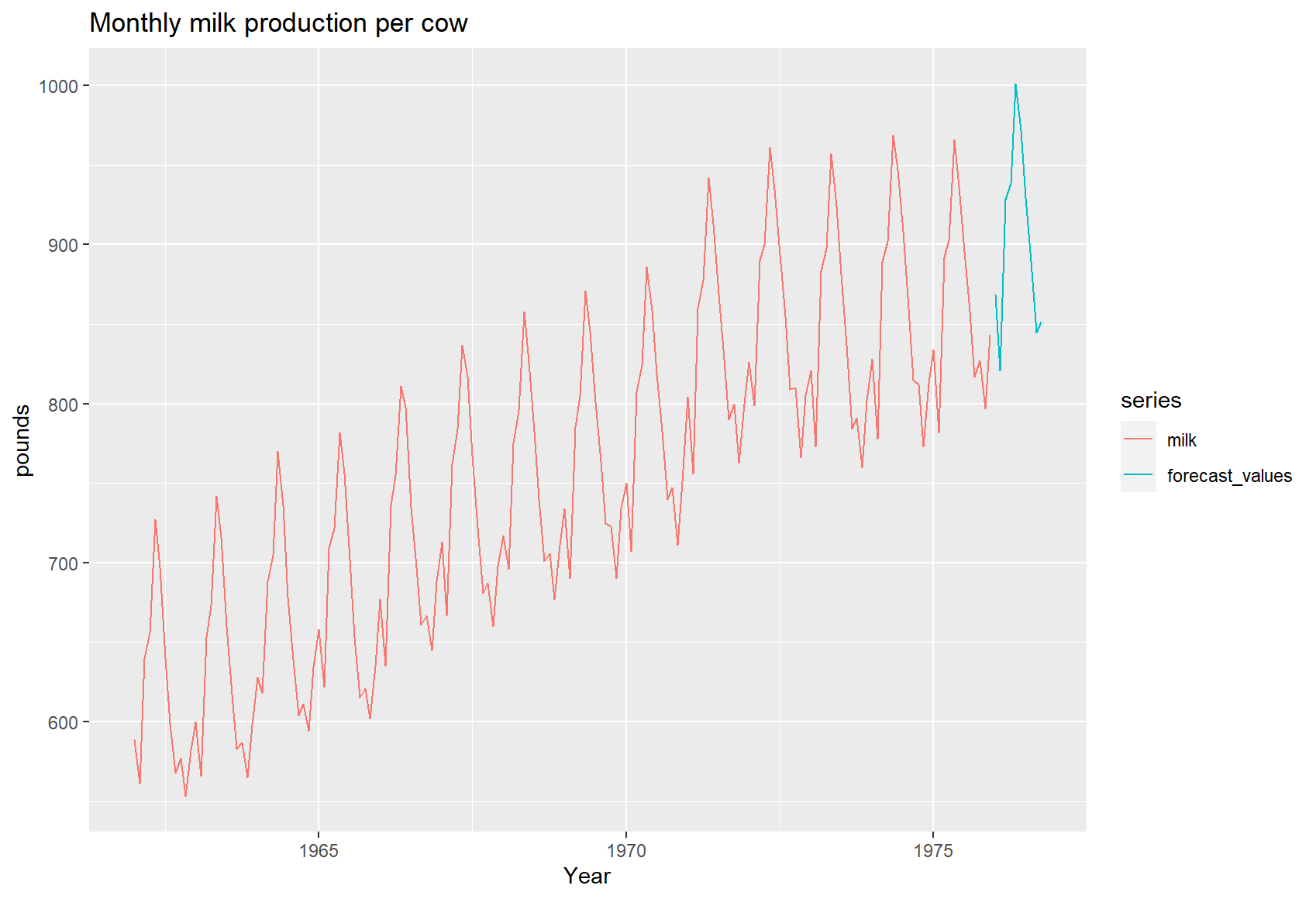
State space models (SSMs) are statistical tools used to describe and forecast time series data by separating observed measurements from unobserved (hidden) states. They encapsulate a system's dynamics, comprising a state equation and an observation equation. These models find applications in various fields like economics, finance, engineering, and biology.

The basic concept involves a state equation representing the underlying process, while the observation equation links the unobservable states to the observed data. SSMs can handle various complexities, such as trends, seasonality, and dynamic patterns, providing flexibility in modelling diverse data.

SSMs have a wide range of applications, including economic forecasting, signal processing, weather prediction, and robotics. They are beneficial in situations where hidden processes influence observable outcomes, allowing for more accurate predictions and understanding of underlying mechanisms.

Limitations of SSMs include the challenge of accurately specifying the model and handling high-dimensional data, which can impact computational complexity and model performance. Assumptions revolve around linearity, normality, and stationary conditions in the state and observation equations.

Mathematically, an SSM is expressed as xt+1 = Fxt + vt (state equation) and yt = Hxt + wt (observation equation), where xt denotes the state vector, yt the observation vector, F the state transition matrix, H the observation matrix, and vt and wt as state and observation noise, respectively. Despite limitations, SSMs provide a flexible framework for modelling and forecasting time series data influenced by hidden dynamics.



State space model having local, trend and seasonal components is fit to forecast monthly milk production. Forecast values are shown in green.