Time Series Analysis and Forecasting on JJ & Amazon dataset - ARIMA, LSTM and GRU Models

Student Name – Boga Sreevarshini
Student ID – 22095406
Github Link – https://github.com/varshiniboga/Time_Series_Analysis_Forecasting.git

1. Introduction

This report analyses two time series datasets using advanced forecasting techniques:

- Johnson & Johnson (JJ): sales records (1960–1980)
- Amazon (AMZN): stock prices (2018–2023)

We implement LSTM, GRU, and ARIMA models to address key challenges:

- JJ: severe seasonal effect and exponential development
- AMZN: Involvement of non-stationarity and difficult volatility

***Goal*:** Compare model performance and generate 24-month forecasts.

2. Methodology Overview

2.1 Why LSTM/GRU?

- **Strengths:** Recurrent neural networks (RNNs) such as LSTM and GRU are capable of long-term dependencies and sequence information within the time series.
- **Limitations**: Intensive computing takes a lot of data.
- **Model Setup**:
  - **JJ:** LSTM & GRU with 2 hidden layers (128 → 64 units), dropout (0.3–0.4), 8-quarter sliding window.
  - **AMZN**: 60-day look-back window, data scaled with MinMax normalization, single 100-unit layer for both GRU and LSTM.

2.2 Why ARIMA?

- **Strengths**: Suitable to short-to-medium term linear trends, less data, interpretable parameters.
- **Limitations**: It is not ideal in capturing complex nonlinear or long range dependencies.
- **Modeling Process**:
  - **JJ:** Chosen model ARIMA(4,1,3) based on lowest AIC (-144.2). Log transform + differencing used to achieve stationarity.
  - **AMZN:** ARIMA(0,1,0) selected via auto_arima; daily log-differencing handled non-stationarity.

2.3 Evaluation Metrics

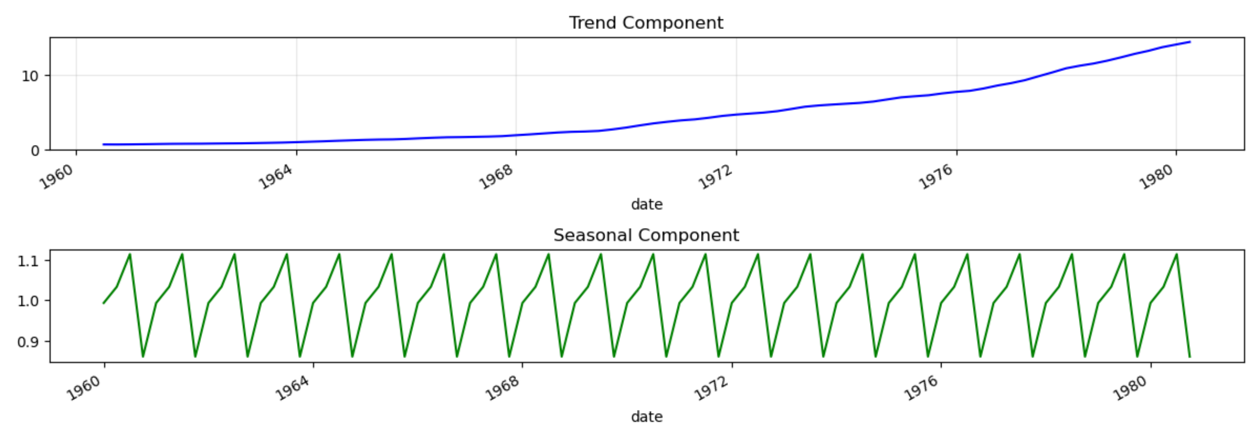
- **MAE** : Average absolute deviation.
- **RMSE** : Punishes large errors.
- **MAPE** : Relative accuracy.

3. Johnson & Johnson Analysis

3.1 Data Characteristics

- **Trend**: Exponential growth (Fig 1a).
- **Seasonality**: Strong quarterly cycles (Fig 1b).
- **Non-Stationarity**: ADF p-value=1.0 → Fixed by log-transform + seasonal differencing (ADF p=0.0004).

Figure 1:
(a) Original sales series shows exponential growth.
(b) Seasonal decomposition reveals stable quarterly seasonality.



3.2 Model Performance

- **LSTM** and **GRU** outperform ARIMA in terms of **MAPE**, indicating better generalization on test data.
- **ARIMA** excels in in-sample metrics but generalizes poorly in forecast horizon due to overfitting.

Model	**MAE**	**RMSE**	**MAPE**
LSTM	0.566	0.794	4.12%
GRU	0.673	0.861	4.90%
ARIMA	0.28	0.40	8.21%

3.3 Forecasts

Fig 2: 24-Month Forecast

- LSTM predicts higher sales ($20.35M by 1982) vs GRU ($19.86M).

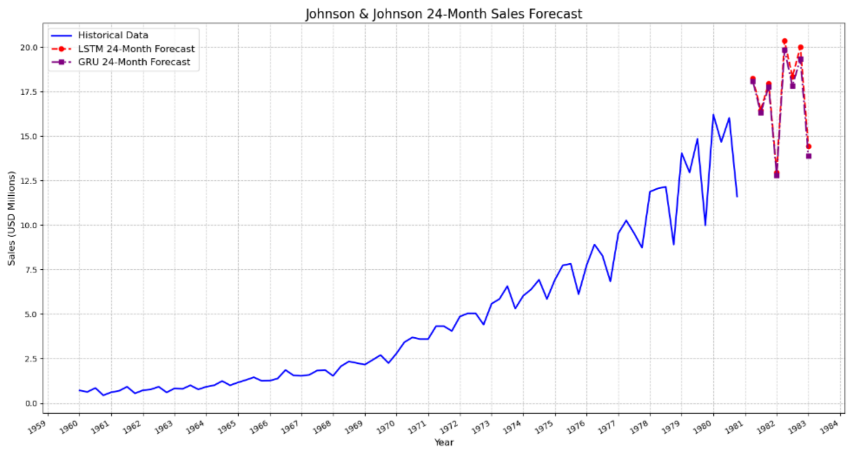
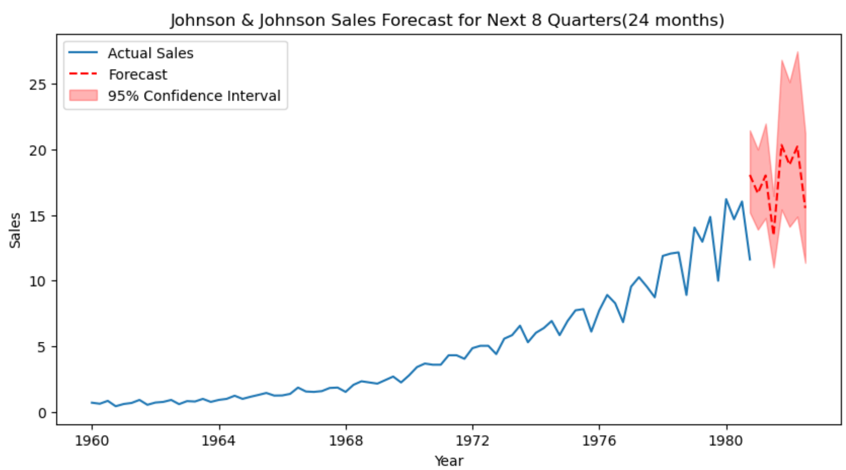


Fig 3: ARIMA Forecast with Confidence Intervals

- ARIMA shows wider confidence intervals (Fig 3).



3.3 Forecast Interpretation (Next 8 Quarters)

Figure 2: LSTM Forecast - Predicts sales reaching $20.35M by 1982.
Figure 3: ARIMA Forecast - Forecasts sales around $19.80M, but shows wide confidence intervals.

- LSTM shows smoother growth aligned with historical trend.
- GRU predicts slightly lower values, peaking at $19.86M.
- ARIMA shows higher uncertainty, potentially underestimating growth in nonlinear environments.

4. Amazon Stock Analysis

4.1 Data Characteristics

- **Volatility:** High-frequency fluctuations, no seasonal structure.
- **Stationarity**:
  - ADF test (original): $p = 0.453$ → Non-stationary.
  - After log-differencing: $p < 0.0001$ → Stationarity achieved.

Key Challenge: Capturing complex noise patterns without overfitting.

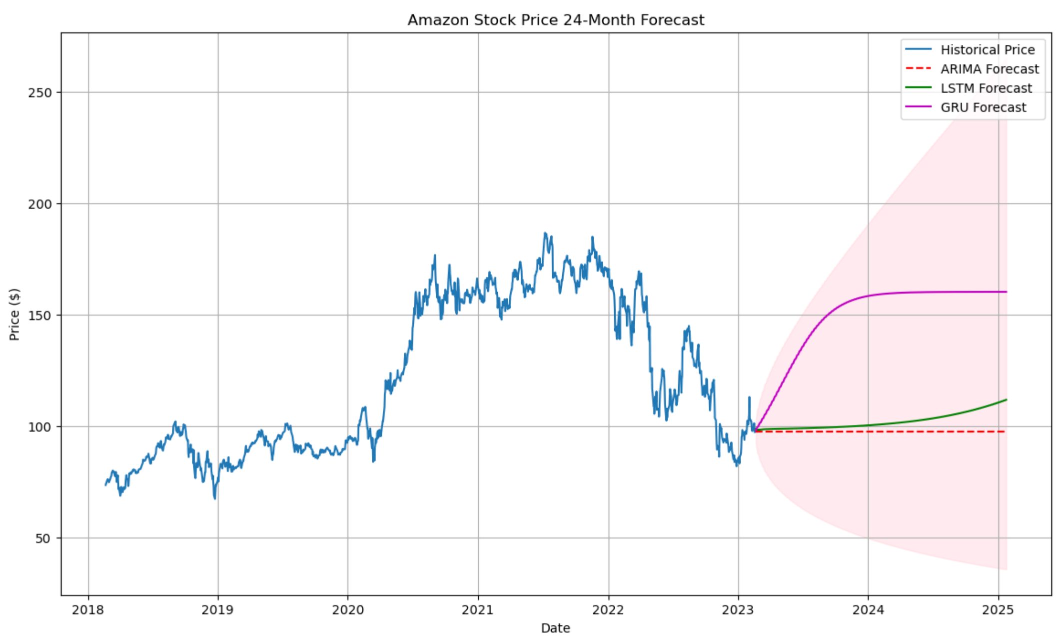
4.2 Model Performance

- **ARIMA** surprisingly performs best suggesting that simpler models are more robust under volatile, non-seasonal noise.
- **LSTM** struggles with erratic fluctuations and tends to over-predict.
- **GRU** balances complexity and performance, closely matching ARIMA.

Model	RMSE	MAE	MAPE
ARIMA	3.68	2.76	2.37%
GRU	3.80	2.89	2.51%
LSTM	4.56	3.59	3.13%

4.3 Forecasts

Fig 4: 24-Month Stock Forecast



• GRU forecast projects Amazon stock stabilizing around **$165** by 2025, indicating strong upward momentum.
• LSTM predicts slower growth, leveling off near **$130**, with less volatility.
• ARIMA remains flat near **$98**, with wide confidence intervals reflecting high uncertainty.

5. Key Insights

1. **Seasonal Data (JJ)**:
   - LSTM/GRU excel at capturing multi-year cycles.
   - *Inference*: Neural nets outperform ARIMA for nonlinear growth.
2. **Volatile Data (AMZN)**:
   - ARIMA/GRU handle randomness better than LSTM.
   - *Inference*: Simpler models generalize better for noisy financial data.
3. **Confidence Intervals**:
   - ARIMA provides interpretable uncertainty bounds (Fig 3).
   - LSTM/GRU lack native uncertainty estimates.
4. **Compute Trade-offs**:
   - ARIMA trains in seconds; LSTM/GRU require GPU acceleration.

6. Recommendations

Use Case	Best Model	Justification
Seasonal Series (e.g., Sales)	LSTM	Best at capturing periodic growth and lagged dependencies
Volatile Series (e.g., Stocks)	GRU or ARIMA	Robust against noise, faster training
Explainability Needed	ARIMA	Provides AIC-based tuning and confidence intervals

7. Conclusion

- **LSTM/GRU** dominate for seasonal, trend-heavy series (JJ: MAPE = 4.5%).
- **ARIMA** is ideal for parsimonious forecasts on volatile data (AMZN: MAPE = 2.4%).
- Hybrid approaches (e.g., ARIMA-LSTM) may bridge statistical and deep learning strengths.

This discussion shows that the selection of a model is conditioned by the nature of data. LSTM + GRU has worked best on seasonal JJ data to capture the complex trends with less MAPE whereas ARIMA performed better in the data of volatile Amazon stock, it proved to be more robust and interpretable. In brief, the model of deep learning is better adapted to trend-heavy data, with structured time series, but ARIMA is applicable to non-seasonal data, with noise. Future studies may consider hybrid models that merge the virtues of each to get a more parallel situation.

References

1. Junyoung Chung, Caglar Gulcehre(2014). *Empirical Evaluation of Gated Recurrent Neural Networks on Sequence Modeling* https://arxiv.org/abs/1412.3555
2. Analytics Vidhya(2020) *LSTMs Explained: A Complete, Technically Accurate, Conceptual Guide with Keras* https://medium.com/analytics-vidhya/lstms-explained-a-complete-technically-accurate-conceptual-guide-with-keras-2a650327e8f2
3. Amazon Stock Data: Yahoo Finance (2023).
4. Sima Siami-Namini, Akbar Siami Namin (2018*).Forecasting Economics and Financial Time Series: ARIMA vs. LSTM* https://arxiv.org/abs/1803.06386