Technical Document on Forecasting Financial Metrics

1. Introduction & Approach

Objective

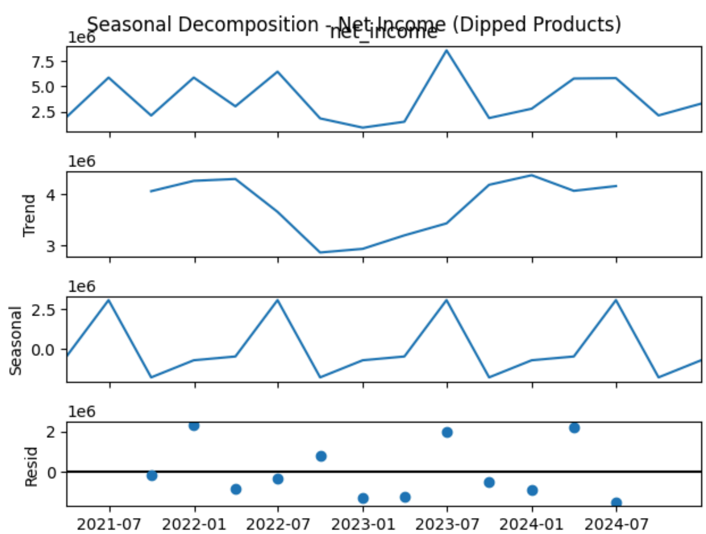
- **Goal:** Forecast quarterly financial metrics (Revenue and Net Income) for two companies – DIPPED PRODUCTS PLC and Richard Pieris Exports PLC.

- **Data:** We have 16 quarters of data per company. The first 14 quarters are used for training, and the last 2 quarters are reserved for testing.

Approach

- **Methodology:**
  We used Seasonal ARIMA (SARIMA) models to forecast each metric because quarterly data tend to exhibit seasonality (with a period of 4).
  Our Exploratory Data Analysis (EDA) included:

  - **Visual Inspection:** We plotted time series line graphs to observe trends, seasonal patterns, and volatility.



- **Stationarity Checks:** The Augmented Dickey-Fuller (ADF) test was performed to determine which series were stationary. For non-stationary series, we considered differencing or transformations.

- **Model Selection:** We used `auto_arima` to select the best model orders `(p, d, q)` and seasonal parameters `(P, D, Q, 4)`. In cases where the training series (14 quarters) was too short and produced singular matrices, we forced seasonal differencing (setting D=0) to stabilize the estimation.

- **Assumptions:**

  - Historical trends and seasonal patterns will persist into the near future.

  - The data have been preprocessed to resolve issues such as missing values and bracketed negative values.

  - External factors are assumed to remain relatively constant over the forecast horizon.

---

2. Data Limitations & Handling of Inconsistencies

Limitations

- **Short Series:**
  With only 16 quarters (14 for training and 2 for testing), the available data limit the model's complexity and the reliability of error metrics.

- **Volatility:**
  The financial metrics (especially revenue) are highly volatile and exhibit large swings. This volatility makes it challenging for complex models to reliably capture underlying patterns.

Data Handling

- **Missing/Inconsistent Data:**

  - Missing numeric values were filled using appropriate imputation (often zero or derived from adjacent periods).

- Values in brackets (e.g., "(100)") were correctly converted to negative numbers.

- **Stationarity Adjustments:**
  - ADF tests revealed that while DIPD's net income was stationary (p-value ~ 0.0), REXP's net income was not (p-value ~ 0.187).
  - For non-stationary series, differencing or forced parameter adjustments were applied.
- **Transformations Considered:**
  Log transformations were discussed (and can be applied) to stabilize variance, especially for revenue data with a broad range.

---

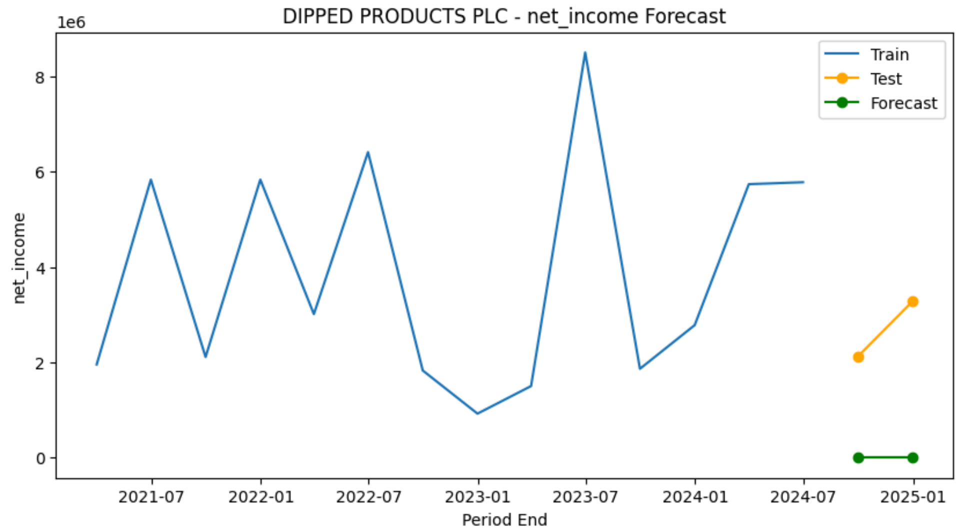
3. Methods & Final Results

Models Considered and Chosen

A. DIPPED PRODUCTS PLC – Net Income

1. **EDA Findings:**
   - The net income series, though volatile, showed a pattern that could be approximated by a constant mean once stationarity was confirmed.
2. **Chosen Model:**
   - **ARIMA(0,0,0)(0,0,0)[4] intercept**
     Note: Auto-ARIMA initially encountered singular matrix errors and, when forced with D=0, selected a naive intercept model.
3. **Results:**
   - **MAE:** ≈ 2,697,988
   - **RMSE:** ≈ 2,759,362.95
   - The model forecasts an average net income for future periods, but the forecast underestimates the test values (which are around 5–6 million).

4. **Graph:**



   ○

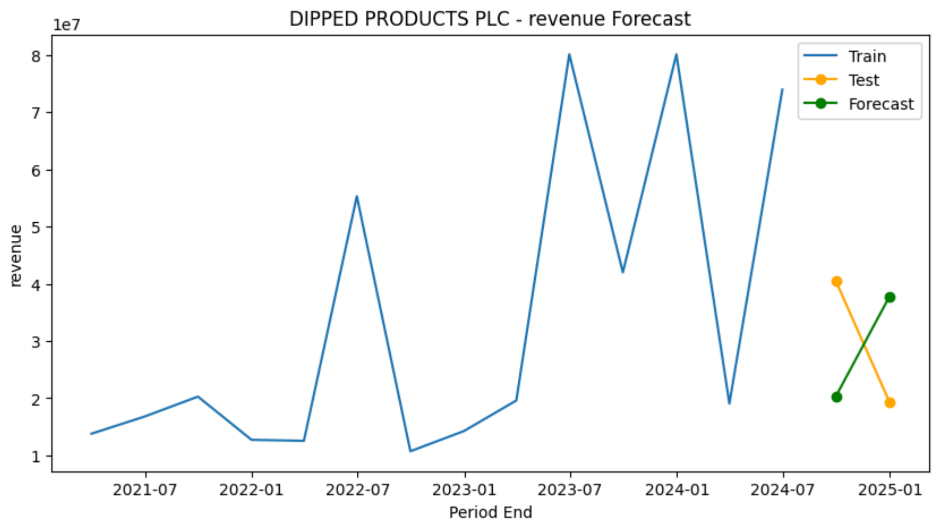
5. **Improvement Ideas:**
   - Increase data frequency (e.g., monthly) or historical span.
   - Use log transformations to stabilize variance.
   - Explore alternative models (e.g., Exponential Smoothing, ETS) or incorporate exogenous variables if available.

B. DIPPED PRODUCTS PLC – Revenue

1. **EDA Findings:**
   - Revenue is highly volatile with a wide range (1M to 8M+), showing significant seasonal fluctuations.
2. **Chosen Model:**
   - **ARIMA(0,0,0)(0,0,1)[4] intercept** – a minimal seasonal MA(1) model.
3. **Results:**
   - **MAE:** ≈ 19,277,212
   - **RMSE:** ≈ 19,296,170

- The forecast is essentially constant with a slight seasonal component, failing to capture large fluctuations.

4. **Graph:**

- 

5. **Improvement Ideas:**

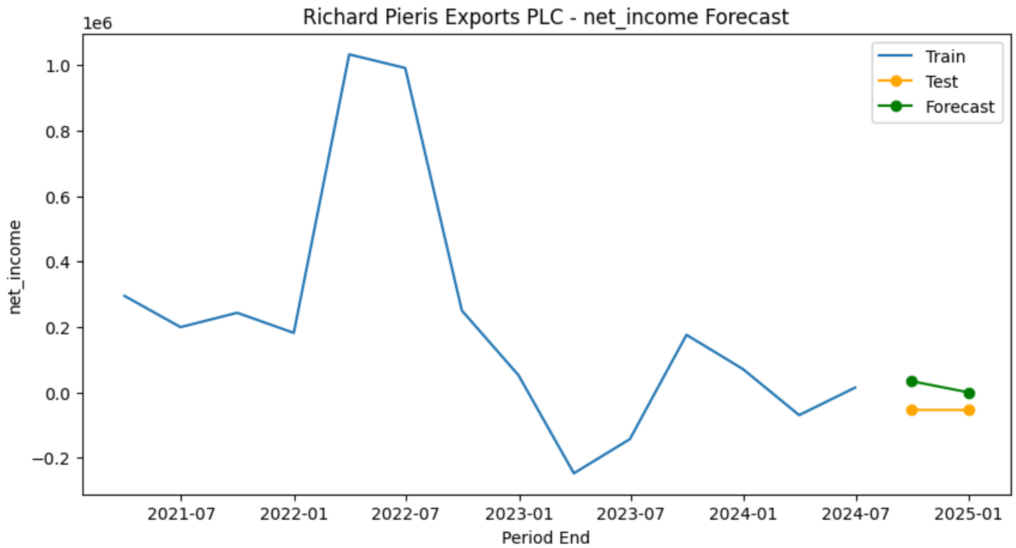
- Use a log transformation to capture relative changes.
- Obtain more historical data to capture volatility.
- Consider exogenous regressors that might drive revenue fluctuations.

C. RICHARD PIERIS EXPORTS PLC – Net Income

1. **EDA Findings:**
   - The net income series is on a smaller scale (hundreds of thousands) with moderate fluctuations and some seasonality.
2. **Chosen Model:**
   - **ARIMA(0,0,1)(0,0,0)[4] intercept** – a simple MA(1) model.
3. **Results:**
   - **MAE:** ≈ 70,916

- **RMSE:** ≈ 72,908
- The model underestimates the actual test net income (around 200k), and warnings about near-singular covariance matrices suggest instability.

4. **Graph:**



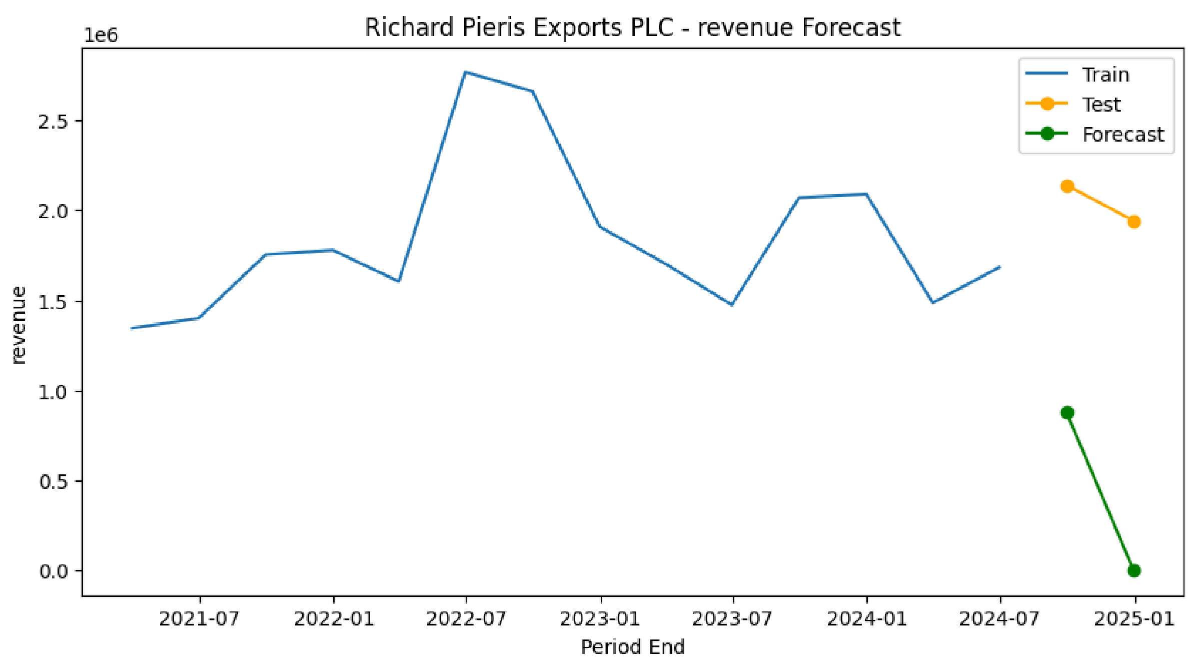
5. **Improvement Ideas:**
   - Consider first differencing or log transformations.
   - Increase the data set size to improve parameter estimation.
   - Explore simpler models such as naïve or seasonal naïve approaches.

D. RICHARD PIERIS EXPORTS PLC – Revenue

1. **EDA Findings:**
   - The revenue series, although less volatile than DIPD's, still exhibits significant swings and seasonal behavior.
2. **Chosen Model:**
   - **ARIMA(0,0,1)(0,0,0)[4] intercept** – capturing a one-lag MA effect.
3. **Results:**

- **MAE:** ≈ 1,601,745
- **RMSE:** ≈ 1,637,463
- The forecast underestimates the test revenue (around 2.5 million), reflecting limitations due to the short series.

4. **Graph:**



5. **Improvement Ideas:**
   - Use seasonal differencing if a quarter-to-quarter pattern is detected.
   - Apply log transformation.
   - Consider including exogenous factors or additional data points.

Evaluation Summary

- **High Error Metrics:** The relatively high MAE and RMSE in certain cases indicate that the models are not capturing the volatility or trends effectively.

- **Simple Models:** The models default to very simple structures (often just a constant or a basic MA term) due to the limited training data.
- **Data Constraints:** With only 14 training quarters, it is challenging to capture complex seasonal or trend behavior, and the test set is too small to robustly validate the forecast.

---

4. Final Takeaways & Recommendations

Are These Models "Good" or "Bad"?

- **Current Models Are Weak:**
  Large forecast errors and warnings about near-singular matrices demonstrate that with only 14 data points the models have limited reliability.
- **Key Limitations:**
  - **Short Historical Series:** Only 16 quarters, with 14 for training, restrict the ability to model complex patterns.
  - **High Volatility:** Especially in revenue, the variability makes it hard for the model to produce accurate forecasts.
- **Model Adequacy:**
  - The constant or near-constant forecasts indicate that the models are primarily reflecting an average rather than true dynamic behavior.

Recommendations for Improvement

1. **Acquire More Data:**
   - Extend the historical data beyond 16 quarters or switch to a higher-frequency dataset (e.g., monthly) to provide more observations.
2. **Simpler Forecasting Methods:**
   - Consider naïve or seasonal naïve forecasts or exponential smoothing (ETS), which might perform better with limited data.

3. **Data Transformations:**
    - Apply log transformations or differencing to stabilize variance and improve stationarity.
4. **Include Exogenous Variables:**
    - Incorporate external factors (economic indicators, commodity prices, etc.) using SARIMAX if they are known drivers of the metrics.
5. **Robust Validation:**
    - Use rolling or expanding window validation to better understand model performance given the small sample size.

This report outlines our approach, underlying assumptions, data limitations, methods, and final results. The attached charts support our conclusions and highlight areas for future improvement. The primary conclusion is that the current models are limited by the small dataset and high volatility, and further data collection or simpler methods should be considered for more robust forecasting.