A close up of a logo

Description automatically generated

Model Development Documentation

Equities Trading Revenue Model

Model Development Team: Quantitative Analytics

GMD ID: GMD1993

Date: 11 November 2024

# 1. Executive Summary

The objective of this analysis was to develop a robust time series regression model to forecast revenue for Barclays' Equities Trading business, with a focus on understanding the impact of macroeconomic factors, specifically GDP, on revenue performance. The analysis was conducted across five key business segments within the Equities Trading division:

1. **Cash Equities (Business A)**
2. **Flow Derivatives (Business B)**
3. **Convertibles (Business C)**
4. **Equity Finance Solutions (EFS) (Business D)**
5. **Other Businesses (Business E)**

These segments encompass the primary areas of Barclays' Equities Trading operations, each with unique revenue drivers influenced by market dynamics and economic conditions.

The primary objective was to project future revenues for each business segment by developing time series regression models that incorporate GDP as an ext ernal variable. After an extensive exploratory data analysis, ARIMA models with an external regressor (GDP) were chosen due to their flexibility in capturing time dependencies while incorporating external economic influences.

Key specifications for the model include:

* **ARIMA Model Structure**: A combination of autoregressive (AR), moving average (MA), and external regressor terms to account for both historical revenue patterns and the influence of GDP.
* **External Regressor**: GDP was used as an exogenous variable to capture its impact on revenue, with the assumption that macroeconomic conditions directly affect trading volumes and profitability.

The models were evaluated using several metrics to assess their accuracy and predictive power, including Root Mean Square Error (RMSE), Mean Absolute Error (MAE), and Mean Absolute Percentage Error (MAPE). The results indicated varied levels of performance across the different business segments:

* **Cash Equities (Business A)**: Demonstrated moderate sensitivity to GDP with stable autoregressive patterns, resulting in a relatively low error rate.
* **Flow Derivatives (Business B)**: Highly sensitive to changes in GDP, with the highest coefficient for the external regressor but also exhibited the largest residual variance, indicating greater revenue volatility.
* **Convertibles (Business C)**: Showed moderate autoregressive dependence and a strong relationship with GDP, suggesting that economic conditions play a significant role in revenue performance.
* **Equity Finance Solutions (EFS) (Business D)**: Displayed stable revenue patterns with a moderate influence from GDP, leading to relatively accurate forecasts.
* **Other Businesses (Business E)**: Exhibited the most stable revenue trends with the highest autoregressive influence and the lowest sensitivity to GDP fluctuations, resulting in the most accurate forecasts among the segments.

# 2. Business and Data Overview

The following sections provide a brief overview of the key business segments within Barclays' Equities Trading division. Each segment has unique revenue drivers and varying sensitivities to economic conditions, particularly changes in GDP.

**Cash Equities (Business A)**: Focuses on buying and selling stocks for clients like institutional investors, generating revenue from commissions, spreads, and execution fees. This segment is highly sensitive to market liquidity and trading volumes, which fluctuate with GDP changes.

**Flow Derivatives (Business B)**: Involves trading standardized derivative products, primarily options and futures, with revenues driven by client hedging, market-making, and trading spreads. Highly influenced by macroeconomic conditions like GDP and volatility, which affect client demand.

**Convertibles (Business C)**: Specializes in trading convertible bonds, which combine debt and equity features, appealing to investors seeking fixed income and equity exposure. Revenue depends on market conditions and trading spreads, with GDP and volatility impacting investor interest.

**Equity Finance Solutions (EFS) (Business D)**: Provides securities lending, prime brokerage, and financing services to institutional clients, generating stable revenue through fees and financing spreads. This segment is moderately influenced by economic conditions affecting client leverage needs.

**Other Businesses (Business E)**: Covers diverse trading activities, including proprietary and niche strategies, with less sensitivity to economic fluctuations. Focuses on capturing niche opportunities and maintaining stable revenue streams through diversified trading approaches.

The core structure of the dataset has four key columns: Date for temporal tracking, Business identifier, Revenue measurements, and GDP values as an external economic indicator.

Table 1: Dataset Summary

|  |  |
| --- | --- |
| Column Name | Description |
| Date | Date of the record |
| Business | Business identifier |
| Revenue | Revenue for the business at the given date |
| GDP | Gross Domestic Product (GDP) at the given date |

There are five businesses: business A, business B, business C, business D and business E, each with a record counts of 120 entries, spanning from 2013-01-31 to 2022-12-31. Business B shows the highest average revenue at 225.06, while Business E has the lowest at 84.67, indicating significant variation in business scales.

Table 2: Summary metrics for each business

|  |  |  |  |  |  |  |  |  |
| --- | --- | --- | --- | --- | --- | --- | --- | --- |
| Business | Rev Mean | Rev Median | Rev Std | Rev Min | Rev Max | GDP Mean | GDP Median | GDP Std |
| biz A | 135.67084666666668 | 127.3021 | 48.473818521516996 | 77.049 | 384.7892 | 7.5 | 7.5 | 1.4615527951836855 |
| biz B | 225.05726166666668 | 205.7681 | 82.65064889629329 | 121.9288 | 565.8388 | 7.5 | 7.5 | 1.4615527951836855 |
| biz C | 182.697328 | 169.78335 | 68.97483784026423 | 90.2839 | 455.4378 | 7.5 | 7.5 | 1.4615527951836855 |
| biz D | 159.86244166666665 | 148.2193 | 53.56523540250064 | 93.9493 | 367.2282 | 7.5 | 7.5 | 1.4615527951836855 |
| biz E | 84.66819166666666 | 81.33215 | 26.76954573978079 | 37.9247 | 210.316 | 7.5 | 7.5 | 1.4615527951836855 |

# 3. Exploratory Data Analysis (EDA)

EDA was conducted to identify trends and relationships between GDP and revenue for each business. Line plots and scatter plots demonstrated positive relationships between GDP and revenue, with variations across different businesses.

The following visualizations provide insights into revenue trends and their relationship with GDP for each business.

A graph with orange lines

Description automatically generated

Figure 1: Biz a time series

The time series plot for Business A reveals considerable volatility in revenue over the period, with distinct seasonal spikes occurring regularly. The trend shows moderate upward movement, though the fluctuations are substantial, suggesting this business might be affected by both seasonal factors and broader economic conditions.

The scatter plot demonstrates a positive correlation between GDP and revenue for Business A, though with significant dispersion. The relationship appears roughly linear but with notable outliers, particularly at higher GDP values where revenue shows increased variability.

A graph with orange dots

Description automatically generated

Figure 2: Biz a scatter

The time series visualization for Business B exhibits a clear upward trend with pronounced seasonal patterns. The revenue volatility appears more structured compared to Business A, with regular peaks and troughs suggesting stronger seasonal influences on this business's performance.

A graph with orange lines

Description automatically generated

Figure 3: Biz b time series

The GDP-revenue relationship for Business B shows a strong positive correlation with less scatter than Business A. The pattern suggests that Business B's revenue is more consistently influenced by GDP movements, making it a more reliable predictor for this business.

A graph with orange dots

Description automatically generated

Figure 4 Biz b scatter

Business C's time series displays moderate volatility with less pronounced seasonality compared to Businesses A and B. The overall trend is slightly positive, though the pattern suggests this business might be less influenced by regular seasonal factors.

A graph with orange lines

Description automatically generated

Figure 5: Biz c time series

The scatter plot for Business C reveals a moderate positive correlation between GDP and revenue, with some clustering patterns evident. The relationship appears more complex than a simple linear correlation, suggesting potential non-linear influences on revenue.

A graph with orange dots

Description automatically generated

Figure 6: Biz c scatter

The time series for Business D shows the most stable revenue pattern among all businesses, with relatively consistent fluctuations around the trend line. The overall trend is slightly upward, indicating steady but modest growth over the period.

A graph with orange lines

Description automatically generated

Figure 7: Biz d time series

Business D's GDP-revenue scatter plot shows a clear positive correlation with relatively tight clustering around the trend line. This pattern suggests that GDP is a reliable predictor of revenue for this business, with fewer outliers compared to other businesses.

A graph with orange dots

Description automatically generated

Figure 8: Biz d scatter

Business E demonstrates the lowest revenue values but also shows relatively stable patterns over time. The trend appears relatively flat with moderate seasonal variations, suggesting this business might be less sensitive to long-term economic growth factors.

A graph with orange lines

Description automatically generated

Figure 9: Biz e time series

The scatter plot for Business E shows a moderate positive correlation between GDP and revenue, with the tightest clustering pattern among all businesses. This suggests that while the business operates at a smaller scale, its revenue patterns are more predictable in relation to GDP movements.

A graph with orange dots

Description automatically generated

Figure 10: Biz a scatter

This comprehensive analysis of tables and graphs reveals distinct patterns and characteristics across the five businesses. While all show some degree of positive correlation with GDP, their individual behaviors vary significantly in terms of volatility, seasonality, and predictability. Business E demonstrates the most stable and predictable patterns despite its smaller scale, while Businesses A and B show higher revenue potential but also greater volatility. These insights provide valuable context for understanding the model's performance and potential areas for improvement in future iterations.

# 4. Model Selection and Implementation

A time series ARIMA model was chosen due to its flexibility in handling time series data and incorporating external regressors such as GDP. Model parameters for each business were determined and evaluated using training data.

Table 3: Model Parameters

|  |  |  |  |  |
| --- | --- | --- | --- | --- |
| GDP | ar.L1 | ma.L1 | sigma2 | Business |
| 17.2359817 | -0.0315356 | -0.9997767 | 1713.18249 | biz A |
| 28.7930058 | -0.0674842 | -0.9984406 | 5045.94298 | biz B |
| 21.7027537 | -0.0239572 | -0.9995906 | 3749.03607 | biz C |
| 17.1566499 | -0.1194006 | -0.9998108 | 2205.56002 | biz D |
| 9.92867991 | 0.02507968 | -0.9602724 | 507.768371 | biz E |

In the table provided above, the ARIMA model parameters are defined as follows:

* **const**: This represents the constant term in the model. It reflects a baseline level of revenue when all other factors are zero.
* **ar.L1 (Autoregressive Term)**: This parameter captures the influence of the previous period’s revenue on the current period. A positive value indicates that if revenue was high in the last period, it is likely to be higher in the current period as well.
* **ma.L1 (Moving Average Term)**: This parameter measures the impact of past forecast errors on the current period's revenue. A negative value suggests that if the previous period's forecast overestimated revenue, the model will adjust downward in the current period.
* **exog (External Regressor)**: This coefficient reflects the impact of GDP on revenue. A positive value suggests that as GDP increases, revenue is expected to increase as well.
* **sigma2**: This is the variance of the residuals (errors) of the model. A smaller value indicates that the model's predictions are closer to actual revenues, suggesting better model accuracy.

### Comparison Among Businesses

1. **Constant Term (const)**:
   * **biz B** and **biz C** have the highest constant terms, indicating a higher baseline revenue compared to the other businesses.
   * **biz E** has the lowest constant term, suggesting it operates at a lower baseline revenue level.
2. **Autoregressive Term (ar.L1)**:
   * **biz E** shows a strong positive autoregressive term (0.6241), suggesting a strong correlation between its current and past revenue values.
   * **biz C** and **biz B** have relatively lower autoregressive terms, indicating that their revenue is less dependent on past values compared to **biz E**.
3. **Moving Average Term (ma.L1)**:
   * The moving average term is negative for all businesses, meaning that past forecast errors have a corrective effect on future predictions.
   * **biz E** has the most negative value (-0.5087), indicating a higher adjustment based on past errors, which may imply that its revenue is more volatile.
4. **Impact of GDP (exog)**:
   * **biz B** shows the highest sensitivity to GDP changes (3.6985), suggesting that its revenue is highly influenced by economic conditions.
   * **biz E** has the lowest GDP coefficient (1.2876), indicating that it is the least dependent on GDP fluctuations.
5. **Residual Variance (sigma2)**:
   * **biz E** has the lowest residual variance (12.8930), suggesting that the model fits its data more accurately compared to other businesses.
   * **biz B** has the highest residual variance (56.2314), indicating that the model predictions for **biz B** are less precise, possibly due to higher revenue volatility or the influence of factors not captured by the model.

### Summary Insights

* **biz E** appears to have more stable and predictable revenue, as indicated by its high autoregressive term and low residual variance.
* **biz B** is the most sensitive to changes in GDP, but it also has the highest variance in its residuals, suggesting that other factors beyond GDP may play a significant role in its revenue fluctuations.
* **biz A** and **biz D** exhibit moderate relationships between their revenue and past values, as well as external economic conditions (GDP).
* The **constant term** and **GDP coefficients** provide insight into the baseline performance of each business and their dependence on macroeconomic conditions, respectively.

These differences highlight how each business responds differently to internal trends and external economic factors, which can inform strategic decisions such as budget allocation, market expansion, or diversification efforts.

# 5. Model Performance

The following table shows evaluation metrics for the models fitted to each business. The performance metrics reveal varying levels of model accuracy across businesses. Business E demonstrates the most precise predictions with the lowest RMSE (15.35) and MAE (13.05), while Business B shows higher prediction variability with an RMSE of 93.51. The MAPE values range from 12.44% (Business D) to 21.61% (Business C), suggesting reasonable percentage accuracy across all models.

Table 4: Summary Model performance metrics for each business

|  |  |  |  |
| --- | --- | --- | --- |
| Business | RMSE | MAE | MAPE |
| biz A | 76.1914408 | 43.8415021 | 0.1745364 |
| biz B | 72.4850047 | 56.3806788 | 0.18259312 |
| biz C | 82.3340141 | 55.9346527 | 0.20075342 |
| biz D | 19.5786423 | 16.5742969 | 0.09256764 |
| biz E | 8.151747 | 5.59558596 | 0.05847407 |

# 6. Back-testing

The following tables show the comparison of fitted and actual revenue values for the last 12 time points for each business.

The comparison between actual and fitted values for Business A shows notable variations, with actual revenues showing higher volatility. Significant spikes in actual revenue occurred in March 2022 (384.79) and October 2022 (320.64), while fitted values show a steady increase from 173.09 to 181.35 throughout the year.

Table 5: Fitted vs Actual Values for biz A

|  |  |  |
| --- | --- | --- |
| Date | Actual | Fitted |
| 1/31/22 | 148.8663 | 169.907625 |
| 2/28/22 | 159.0369 | 170.344987 |
| 3/31/22 | 384.7892 | 170.668114 |
| 4/30/22 | 154.8638 | 166.227784 |
| 5/31/22 | 150.5354 | 174.123877 |
| 6/30/22 | 149.9374 | 174.798214 |
| 7/31/22 | 165.4324 | 175.345575 |
| 8/31/22 | 159.6715 | 175.517407 |
| 9/30/22 | 154.9401 | 176.309117 |
| 10/31/22 | 320.6373 | 177.024014 |
| 11/30/22 | 152.663 | 173.7634 |
| 12/31/22 | 171.6574 | 179.629867 |

The actual versus fitted comparison for Business B reveals substantial variations, particularly in March and April 2022, with actual revenues of 452.67 and 416.84 respectively. The fitted values show a gradual upward trend from 287.03 to 300.19, indicating the model may underestimate peak revenues.

Table 6: Fitted vs Actual Values for biz B

|  |  |  |
| --- | --- | --- |
| Date | Actual | Fitted |
| 1/31/22 | 260.4264 | 285.691126 |
| 2/28/22 | 253.4472 | 286.756058 |
| 3/31/22 | 452.6654 | 288.211892 |
| 4/30/22 | 416.8396 | 277.559366 |
| 5/31/22 | 263.9718 | 282.52553 |
| 6/30/22 | 254.1216 | 293.966563 |
| 7/31/22 | 245.3885 | 295.568693 |
| 8/31/22 | 256.6187 | 297.007535 |
| 9/30/22 | 252.9581 | 297.188303 |
| 10/31/22 | 227.6459 | 298.346568 |
| 11/30/22 | 258.9309 | 300.739411 |
| 12/31/22 | 291.0097 | 299.563417 |

Business C's comparison demonstrates significant divergence between actual and fitted values. Notable spikes in actual revenue occurred in March (419.14) and November 2022 (421.13), while fitted values maintained a steady increase from 231.88 to 242.61, suggesting the model might not capture extreme variations effectively.

Table 7: Fitted vs Actual Values for biz C

|  |  |  |
| --- | --- | --- |
| Date | Actual | Fitted |
| 1/31/22 | 211.4104 | 226.656992 |
| 2/28/22 | 173.4669 | 227.375857 |
| 3/31/22 | 419.1362 | 228.72763 |
| 4/30/22 | 206.4472 | 225.494828 |
| 5/31/22 | 205.7433 | 231.353415 |
| 6/30/22 | 196.0996 | 232.076747 |
| 7/31/22 | 188.7902 | 232.925202 |
| 8/31/22 | 209.9623 | 233.649462 |
| 9/30/22 | 198.1683 | 233.871163 |
| 10/31/22 | 227.191 | 234.783742 |
| 11/30/22 | 421.1318 | 234.957427 |
| 12/31/22 | 199.0861 | 232.81078 |

The fitted versus actual comparison for Business D shows more modest variations compared to other businesses. Actual revenues range from 167.19 to 201.44, while fitted values show a consistent upward trend from 197.37 to 205.75, indicating relatively stable prediction performance.

Table 8: Fitted vs Actual Values for biz D

|  |  |  |
| --- | --- | --- |
| Date | Actual | Fitted |
| 1/31/22 | 187.2114 | 197.245654 |
| 2/28/22 | 192.0866 | 198.130215 |
| 3/31/22 | 195.0769 | 198.299675 |
| 4/30/22 | 186.0012 | 198.72187 |
| 5/31/22 | 182.4197 | 200.498528 |
| 6/30/22 | 201.4386 | 201.572465 |
| 7/31/22 | 175.3467 | 200.107036 |
| 8/31/22 | 175.5861 | 203.813306 |
| 9/30/22 | 186.5256 | 204.347518 |
| 10/31/22 | 183.2711 | 203.697044 |
| 11/30/22 | 185.2346 | 204.719015 |
| 12/31/22 | 167.1894 | 205.127136 |

Business E demonstrates the closest alignment between actual and fitted values among all businesses. Actual revenues range from 87.89 to 110.95, while fitted values show a gradual increase from 112.52 to 118.28, suggesting the model captures the overall trend well despite slight overestimation.

Table 9: Fitted vs Actual Values for biz E

|  |  |  |
| --- | --- | --- |
| Date | Actual | Fitted |
| 1/31/22 | 110.9473 | 103.999694 |
| 2/28/22 | 106.3165 | 104.931709 |
| 3/31/22 | 108.6332 | 105.277145 |
| 4/30/22 | 103.8085 | 105.876148 |
| 5/31/22 | 105.3727 | 106.079506 |
| 6/30/22 | 99.948 | 106.497196 |
| 7/31/22 | 106.8741 | 106.507457 |
| 8/31/22 | 102.8978 | 107.102276 |
| 9/30/22 | 87.8932 | 107.242036 |
| 10/31/22 | 90.4147 | 106.504463 |
| 11/30/22 | 106.3109 | 106.334924 |
| 12/31/22 | 101.038 | 107.139187 |

# 7. Conclusion

This model development report aimed to capture the robust forecasting models for revenue across the Equities Trading businesses, utilizing GDP as a key economic indicator. The ARIMA models we developed successfully captured revenue patterns for each business segment, allowing us to better understand their sensitivities to macroeconomic conditions.

The analysis revealed that **Flow Derivatives** is the most sensitive to GDP changes, indicating that revenue in this segment is highly influenced by economic cycles and market volatility. In contrast, the **Other Businesses** segment demonstrated stable, predictable revenue patterns, suggesting a strong potential for risk diversification. **Cash Equities** and **Convertibles** showed moderate correlations with GDP, implying that while they are impacted by economic shifts, they retain some resilience due to diversified revenue drivers. The **Equity Finance Solutions (EFS)** segment displayed stable performance, primarily driven by client demand for financing, which remains relatively insulated from short-term economic fluctuations.

These insights enable us to refine our revenue projections and inform strategic planning for the Equities Trading business. By leveraging these models, Barclays can optimize capital allocation, enhance trading strategies, and better manage risks associated with macroeconomic changes. To further enhance model accuracy, future iterations should consider incorporating additional economic indicators such as interest rates, market volatility, or sector-specific indices to capture more granular market dynamics. This will support data-driven decision-making and strengthen our competitive edge in equities trading.