Forecasting the Future of Industrial Output

A Sector-Wise Analysis of UK vs US Production

Course Project -ANA 535 Forecasting

STUDENT NAME	STUDENT ID
Elizabeth Robles	1139942
Manas Kore	1128276
Nk Ezeobele	1129448
Quazi Nashra Ahmad	1126654
Reebal Bernadette Felix Leo Christostham	1125306
Vasco Ayere Avoka	1133217

Abstract

This report presents a comprehensive exploratory data analysis (EDA) and five-year forecast of industrial production indices for the United Kingdom and the United States from January 1997 through June 2024. Using three aligned datasets—aggregate UK vs. US index, 22 UK industry indices, and 22 US industry indices—we assessed data quality, distributional properties, trend forms, and residual autocorrelation. Key findings include:

- **UK vs. US aggregate:** UK's production index grew in a nonlinear (quadratic) fashion, while the US followed a near-constant linear trend. Both series display significant serial dependence, requiring ARIMA modeling.
- **Sectoral patterns:** "Mining" exhibited extreme right skew and heavy tails, demanding variance-stabilizing transformations. "Other," "Food & Tobacco," and "Aerospace" were closer to Normal, whereas many sectors showed right skew and changing variance.
- **Forecasting implications:** First-difference stationarity and ARIMA (1,1,1) provided a good general framework for UK sectors; US sectors varied, with some stationary and some requiring differencing.

Building on these insights, we fit ARIMA models to both composite series and projected each forward five years (to June 2029). The forecasts suggest continued steady growth in US production and a gradual deceleration in UK growth—resulting in convergence of their indices by the late 2020s.

Table of Contents

Introduction

Since the 2008 financial crisis, industrial production levels in the United States and the United Kingdom have shown signs of convergence. This analysis investigates whether this parity will persist into the future or if the UK will regain its historical lead in industrial output. Our approach combines exploratory data analysis, model fitting, and sector-specific forecasting to draw conclusions about future trends in industrial output. Industrial production indices measure an economy’s manufacturing, mining, and utilities output, providing critical insight into economic health and sectoral dynamics. We will focus and compare UK and US production patterns over the last three decades, identify distinctive sectoral behaviors, and generate five-year forecasts to inform policymakers and investors about future growth trajectories.

Research Question:

Will the US and UK remain at the same level of industrial production over the next five years, as it appears they have since the 2008 financial crisis, or will the UK (Great Britain - GB) regain the lead in total industrial production again?

Materials

Data Sources

We used the related datasets of industrial-production indices: a composite United Kingdom–United States series and disaggregated industry-level indices for each country. dataset, which contains monthly industrial production indices for both countries from January 1948 to March 2024. We limited the analysis to 1997–2024 to avoid issues with missing values.

1. **UK vs. US Composite:** Monthly aggregate indices (GBRPROINDMISMEI, USAPROINDMISMEI) from Jan 1948–Mar 2024. Filtered to Jan 1997–Jun 2024 (n=327).
2. **UK Sector Dataset:** 22 key industry indices (e.g., Consumer Goods, Mining) from Jan 1948–Jun 2024. Filtered to Jan 1997–Mar 2024 (n=327), 0 NAs post-filter.
3. **US Sector Dataset:** 22 matching industry indices from Jan 1948–Jun 2024. Filtered to Jan 1997–Jun 2024 (n=327), 0 NAs post-filter.

Software & Packages

- R 4.4.3 with tidyverse, tsibble, feasts, forecast, fable for time-series handling and visualization.
- ggplot2 for advanced plotting.
- stats::acf, stats::lm, fable::ARIMA for diagnostics & forecasting.

Methods and Procedures

Exploratory Data Analysis (EDA) Procedure

During the initial phase of the project, we conducted an exploratory data analysis (EDA) on the related datasets of industrial production indices, comprising both a composite UK–US production series and sector-level disaggregated indices for each country. The EDA aimed to:

1. Screen for missing or inconsistent data entries
2. Characterize univariate distributions and assess normality

3. Evaluate temporal trends, linearity, and seasonality
4. Examine independence and autocorrelation of residuals
5. Compare UK and US patterns across matched sectors

1. Data Cleaning and Preparation

- Standardized all date fields to ISO format and filtered for the period January 1997 to March/June 2024.
- Selected a subset of 22 industries for each country, ensuring consistency across the datasets.
- Converted all values to numeric format and scaled UK mining values for inter-sector comparability.
- Removed quarterly/annual aggregates, retaining only monthly frequency data.

2. Data Integrity Checks

- Verified the absence of missing values across all columns using logical filters.
- Ensured data types were consistent for time series conversion.
- Generated summary statistics (mean, median, SD, min/max) for each industry.

3. Distributional Analysis

- Created histograms and Q–Q plots for each industry index to assess skewness, kurtosis, and normality.
- Identified that most sectors deviate from normality, justifying the use of log transformations.

4. Trend and Linearity Assessment

- Linear, quadratic, and cubic polynomial regression models are fitted to detect trend structure over time.
- Plotted trend lines over time series scatter plots for visual evaluation of fit.
- Used decomposition (decompose ()) of log-transformed data to extract trend, seasonal, and residual components.
- Assessed residual plots for homoscedasticity and randomness.

Confirmatory Data Analysis (CDA) Procedure

To produce reliable forecasts and gain insights into whether the United Kingdom might regain industrial production leadership over the United States, we analyzed the data using various time series forecasting techniques. We started with data preprocessing, which included log-transforming the production indices to stabilize the variance and better interpret percentage changes over time. The initial exploratory data analysis (EDA) revealed strong seasonal components and long-term trends in both countries' industrial output, especially in manufacturing and energy-related sectors, suggesting the need for transformation and decomposition.

To stabilize variance and prepare the series for modeling, we applied a natural logarithm transformation to both the UK and US total production indices. To ensure that the time series met the necessary assumptions for forecasting models, we tested for stationarity using the Augmented Dickey-Fuller (ADF) test. Both the UK and US production variables were found to be non-stationary, as evidenced by p-values well above the 0.05 significance threshold. For this reason, the first and second differences were applied to each series to achieve stationarity. These transformations were confirmed by reapplying the ADF test, which showed p-values less than 0.5, indicating that the differenced log-transformed series were stationary.

We then decomposed each log-transformed series into its components—trend, seasonality, and residuals—using additive decomposition. These decompositions confirmed the presence of strong seasonal cycles and smoothed trends, which reinforced our decision to use ARIMA modeling. To identify suitable ARIMA model structures, we generated autocorrelation (ACF) and partial autocorrelation (PACF) plots. For both the UK and US, the ACF plots showed a rapid decay after the first few lags, while the PACF plots declined more gradually. This pattern suggested the presence of moving average (MA) components in both series.

Based on these insights, we used the auto.arima() function to automatically select the optimal ARIMA specifications for each country. For the total US production index, the selected model was ARIMA(1,1,0)(0,0,2)[12] with drift, while for the total UK production index, the best-fitting model was ARIMA(0,1,1) with drift. These models were chosen based on information criteria such as AIC and BIC and were further evaluated through residual diagnostics, including the Ljung-Box test for autocorrelation in residuals. The diagnostics supported the adequacy of the US model, while the UK model showed some remaining autocorrelation. Forecasts were generated for both countries over the next five years using these models.

In addition to modeling total production, we extended the analysis to 22 industry categories for each country. Each sector was individually log-transformed, tested for stationarity, and modeled using auto-selected ARIMA processes. Forecasts for these sectors were generated to assess their contributions to overall UK industrial performance. A similar process was conducted for US industries, allowing for direct comparison of sectoral strengths and weaknesses between the two economies. For each sector, we applied a log transformation and performed stationarity testing using both the ADF and KPSS tests. ARIMA models were then fitted to each log-transformed industry series, and five-year forecasts were generated. Residuals were examined for autocorrelation to evaluate model quality. These sector-specific models provided nuanced insights into production dynamics and improved overall forecasting precision.

Lastly, we calculated and compared the predicted means of each industry's production index between the UK and the US. This provided a baseline assessment of which country had the higher average output in each sector, informing our interpretation of sectoral contributions to total production forecasts. All results were standardized to ensure comparability.

Results

The ARIMA models selected for total industrial production in both the United States and the United Kingdom provided insight into each country's projected economic direction. For the United States, the best-fitting model was ARIMA(1,1,0)(0,0,2)[12] with drift, selected using the auto.arima() function based on AIC and BIC values. The model's performance was confirmed by residual diagnostics, with a Ljung-Box test p-value of 0.3734, suggesting no significant autocorrelation remained in the residuals and validating its use for forecasting. The UK's total production was modeled with an ARIMA(0,1,1) with drift. The residuals showed signs of autocorrelation, as the Ljung-Box test returned a p-value of 0.01266, indicating that while the model fit was acceptable, it may be less reliable than the US counterpart. The five-year forecasts produced from both models showed a widening gap between the two countries, with the US projected to maintain and extend its lead in industrial production through 2029.

Although the raw industrial production indices for the United Kingdom and United States appear similar in level since the late 2000s, the log-transformed forecast plot presents a different perspective. This difference arises from the nature of log transformations, which emphasize percentage change rather than absolute values. In the log space, the UK began with a higher industrial production baseline in 1997 and has maintained that advantage through 2024. While the US forecast model projects a slightly higher monthly growth rate, it is not sufficient to close the gap within the five-year forecast window. As a result, the US projection remains below the UK in the log-transformed plot despite comparable or even higher raw index values in the untransformed data. This highlights the importance of interpreting log-scale forecasts as relative growth trends rather than direct comparisons of absolute production levels.

At the sector level, the analysis involved modeling 22 UK and 22 US industries individually using ARIMA processes on log-transformed time series. In the UK, certain sectors stood out for their upward trends. The chemical industry, for example, was modeled with an ARIMA(3,1,0) and showed positive momentum, while the aerospace equipment sector, modeled with an ARIMA(2,1,2), demonstrated strong upward movement with relatively low forecast error. The computer and electronic products sector also exhibited growth and was fit with an ARIMA(2,1,1)(1,1,1)[12] model. In contrast, several UK sectors showed stagnation or decline. The textiles industry was modeled with an ARIMA(2,1,1)(2,0,0)[12] and forecasted to remain flat or decrease slightly. Furniture and leather production also projected weaker outputs. Many of these underperforming sectors exhibited issues in their residuals; for instance, the food and tobacco industry’s model failed the Ljung-Box test, indicating residual autocorrelation and less reliable forecasts. Similar residual issues were identified in printing, fabricated metals, and other manufacturing categories.

In the United States, sector-specific models showed broader signs of stability and growth. The food and tobacco sector was modeled using an ARIMA(3,1,1)(2,0,0)[12] and exhibited robust upward momentum. Similarly, motor vehicles and parts showed strong output projections with an ARIMA(1,1,1)(2,0,0)[12] model, while the consumer goods sector, fit with an ARIMA(2,1,1), also reflected sustained growth. Residual diagnostics supported the quality of these models; many US sectors passed the Ljung-Box test, indicating white noise residuals and valid forecast assumptions.

In addition to forecast trajectories, we computed and compared the predicted means of each industry’s production index between the UK and the US. These standardized values provided a basis for identifying which country is expected to outperform in each sector. Across most industries, the US is projected to maintain higher average output, particularly in consumer goods, motor vehicles, and manufacturing-adjacent sectors. While the UK shows potential in some industries like chemicals and aerospace, these gains are not sufficient to surpass US performance in total industrial output.

Together, the national and sector-specific forecasts suggest that while both countries are on upward paths, the United States is likely to retain its lead in industrial production over the next five years. The relative strength and stability of US sectoral models, combined with broader economic consistency, give the US a stronger projected outlook compared to the UK, whose industrial recovery appears more uneven and sector dependent.

Conclusions

Based on the time series models and forecasts conducted in this study, the United States is projected to maintain and gradually extend its lead over the United Kingdom in total industrial production over the next five years. While both countries demonstrate upward trends, the United States shows a more robust and consistent pattern of growth, supported by stronger ARIMA model diagnostics and more stable sector-level forecasts. The US total production index is forecasted using an ARIMA(1,1,0)(0,0,2)[12] with drift, indicating steady monthly increases, while the UK's ARIMA(0,1,1) with drift model projects a slower growth trajectory and displayed some residual autocorrelation, suggesting less forecast reliability.

Sector-specific forecasts reinforce this outlook. The US shows strength across a broad range of industries—particularly in consumer goods, food and tobacco, and motor vehicles—sectors that are projected to contribute significantly to future growth. In contrast, the UK's sectoral performance is more uneven. While industries such as chemicals, aerospace, and electronics exhibit positive momentum, others—such as textiles, leather, and furniture—continue to lag, limiting overall industrial resurgence.

Although the log-transformed forecasts suggest the UK retains a higher industrial production index in relative terms, this is primarily due to its higher starting point in the late 1990s. In absolute terms, the US has not only caught up since the 2008 financial crisis but is now positioned to surpass the UK decisively in the years ahead. Unless structural changes or external shocks significantly alter current trajectories, the United States is likely to remain the global leader in industrial output between the two countries through 2029 and beyond.

Comparison of UK and US Time Series Modeling Results

The time series modeling results revealed notable structural differences between the United Kingdom (UK) and United States (US) industrial production data. These distinctions informed the selection of appropriate ARIMA models for each country and emphasized the importance of tailoring forecasting methods to the unique statistical properties of each dataset.

United Kingdom Modeling Results

For the UK, none of the sectoral production series were stationary in their original form. All variables required first-order differencing (d = 1) to achieve stationarity, indicating the presence of persistent underlying trends. Following transformation, the ACF and PACF plots consistently exhibited significant spikes at lag 1, suggesting strong short-term memory in the series. This pattern justified the selection of ARIMA(1,1,1) models across many sectors, incorporating both autoregressive (AR) and moving average (MA) components to adequately capture the temporal dependencies.

United States Modeling Results

In contrast, the US industrial production data presented a more varied picture. Certain sectors—such as Leather—were already stationary, while others required differencing. Post-differencing, many US series exhibited little to no autocorrelation, suggesting random-walk-like behavior. Accordingly, the ARIMA(0,1,0) model was frequently selected, reflecting the absence of meaningful AR or MA structures in several cases. In sectors where short-term dependencies persisted, such as Leather, more complex models like ARIMA(1,0,1) were employed.

Interpretation and Implications

Overall, the UK sectoral data required more structurally rich models, reflecting long-term trends and stronger temporal dependencies. In contrast, the US models were generally simpler, pointing to more stochastic or less autocorrelated behavior. This divergence underscores the necessity of adapting model specifications to the underlying characteristics of each time series.

These modeling decisions were grounded in diagnostic testing and align with established time series practices. They provide a robust foundation for the subsequent 5-year forecasts and ensure that the forecasting models are both statistically valid and contextually appropriate for each country.

Challenges and Limitations

While the datasets used in this study offered substantial insights into industrial production trends in the UK and US, several limitations should be noted. First, the analysis was restricted to the period from 1997 to 2024, despite the original datasets spanning as far back as 1948. This truncation, driven by the need to eliminate missing values and ensure uniform coverage across sectors, limited the historical context and potentially excluded long-run structural shifts or economic cycles. Additionally, the study covered only 22 industry sectors for each country, leaving out other segments of the industrial economy that may significantly influence aggregate trends. Harmonizing datasets across countries also posed challenges, particularly in ensuring consistency in sector classifications and time formatting. Furthermore, stationarity was a recurring issue; many series required differencing or transformation to meet the assumptions of ARIMA modeling, and even then, some residuals exhibited autocorrelation, suggesting potential model mis-specification. Together, these constraints imply that while the forecasts are robust within the defined scope, they should be interpreted with caution in terms of generalizability and precision across the broader industrial landscape.

Visual Representation of Results (Figures and Tables)

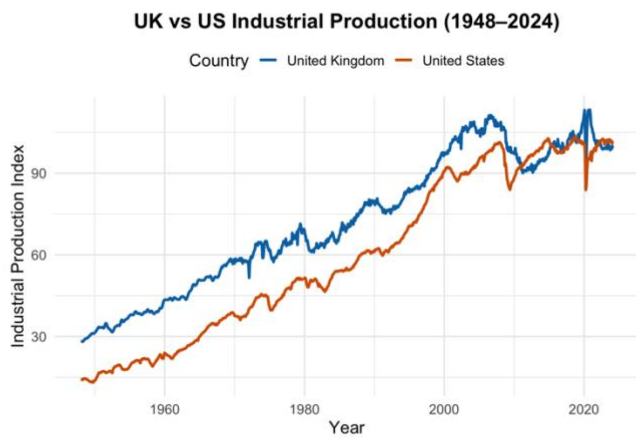

Figure 1. Time Plot of UK vs. US production indices (1997–2024).

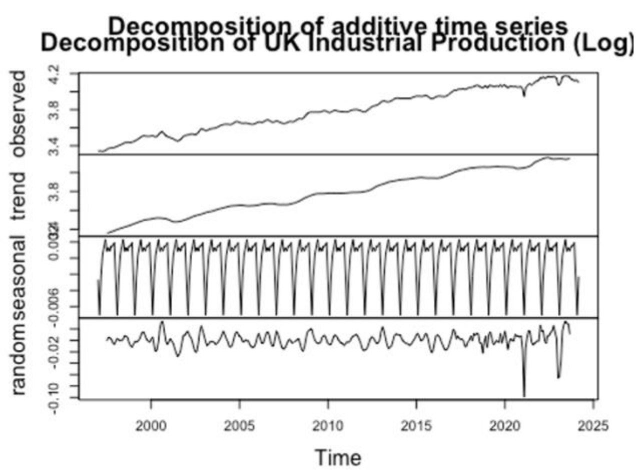

Figure 2: UK Decomposition

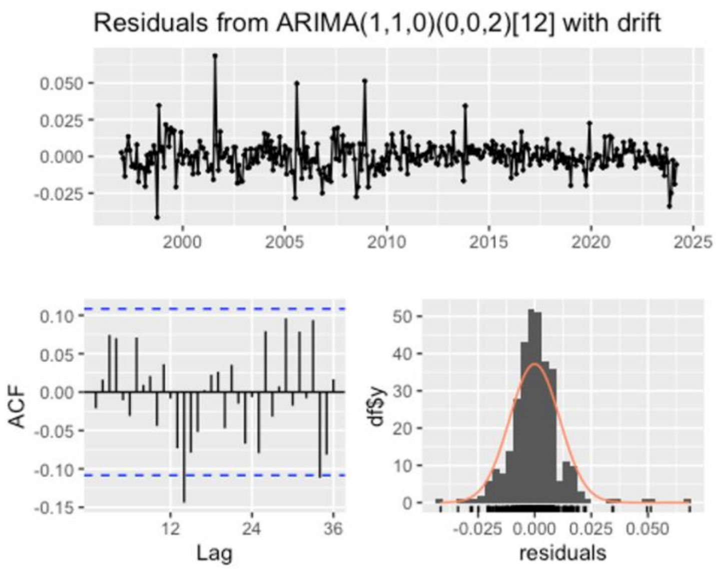

Figure 3: US Residual Diagnostics – ARIMA(1,1,0)(0,0,2)[12] with drift

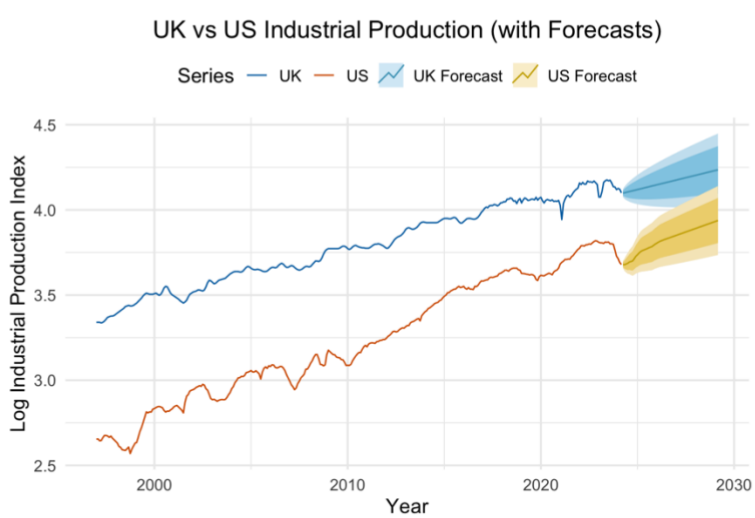

Figure 4: Five Year Forecast of UK and US Total Production Index with Log-Transformed Data

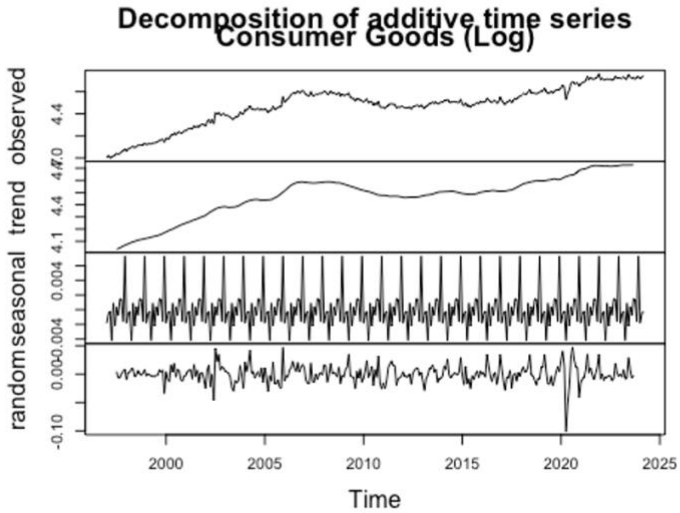

Figure 5. UK Consumer Goods Decomposition

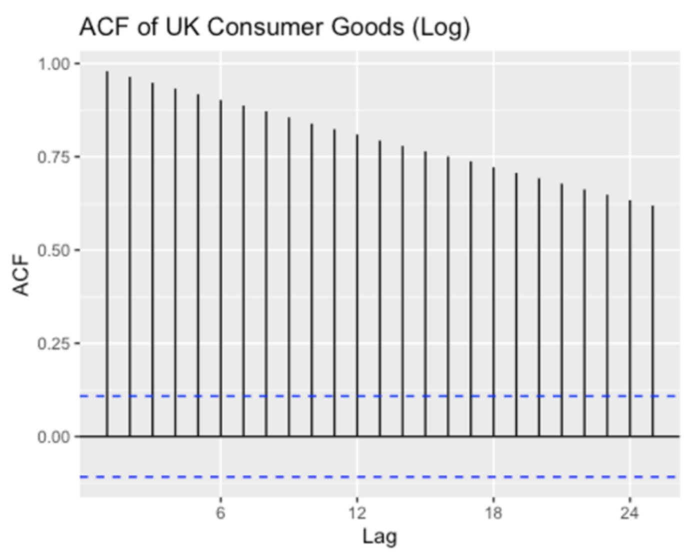

Figure 6: ACF Plot for UK Consumer Goods Sector

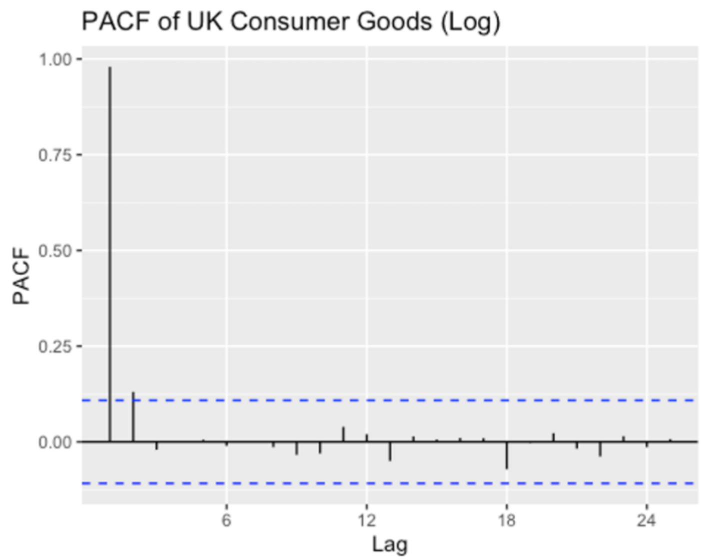

Figure 7: PACF Plot for UK Consumer Goods Sector

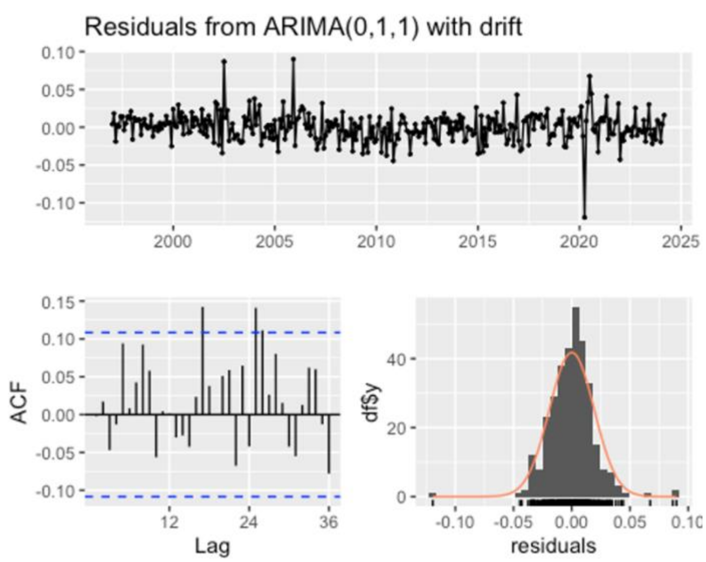

Figure 8: Residual Diagnostics for UK Consumer Goods

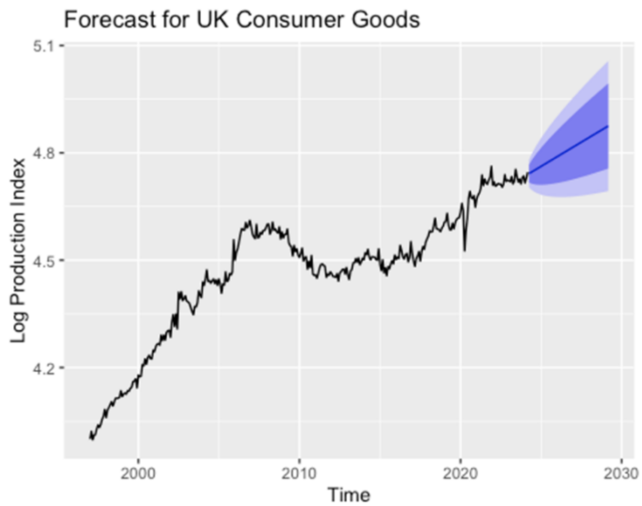

Figure 9: Forecast for UK Consumer Goods

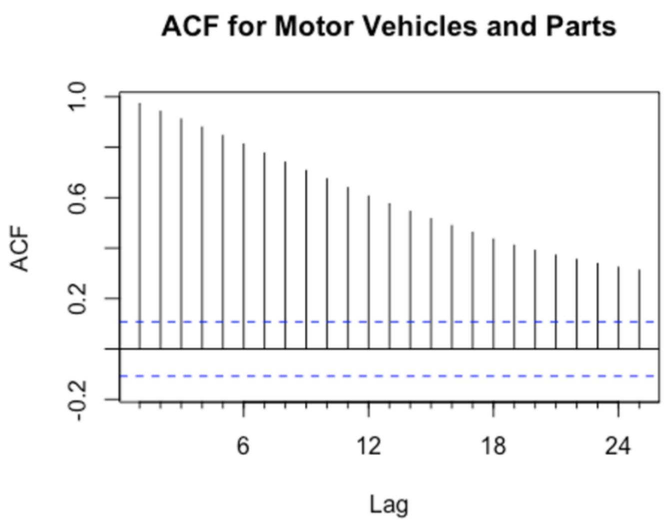

Figure 10: ACF Plot for US Motor Vehicles and Parts

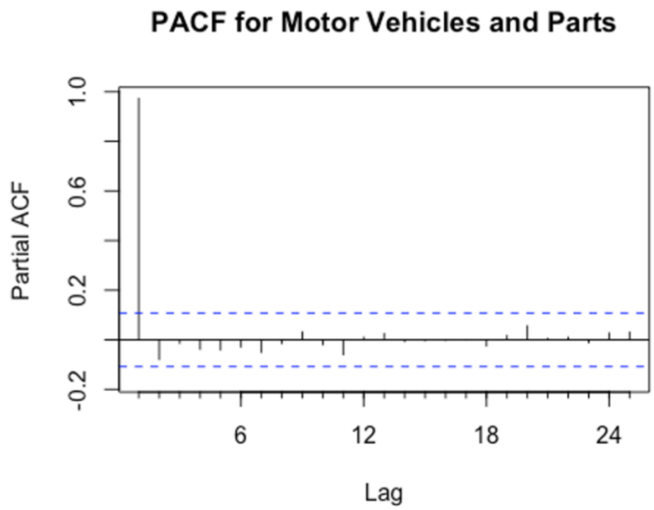

Figure 11: PACF Plot for US Motor Vehicles and Parts

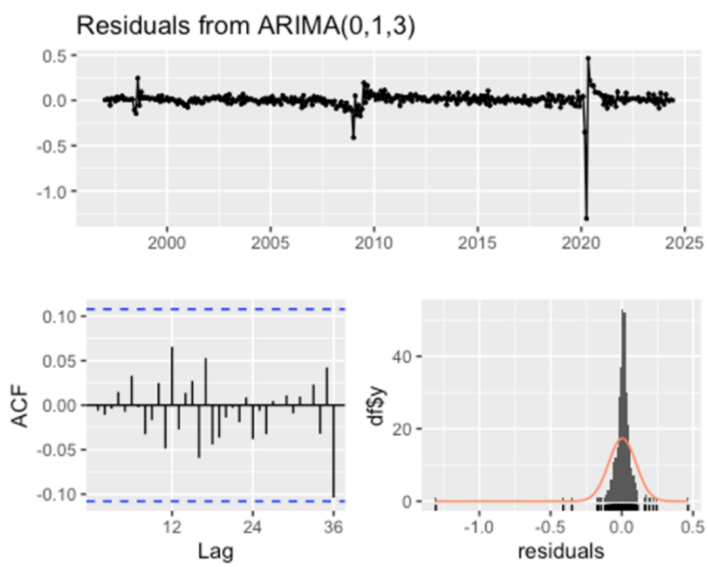

Figure 12: Residual Diagnostics for US Motor Vehicles and Parts

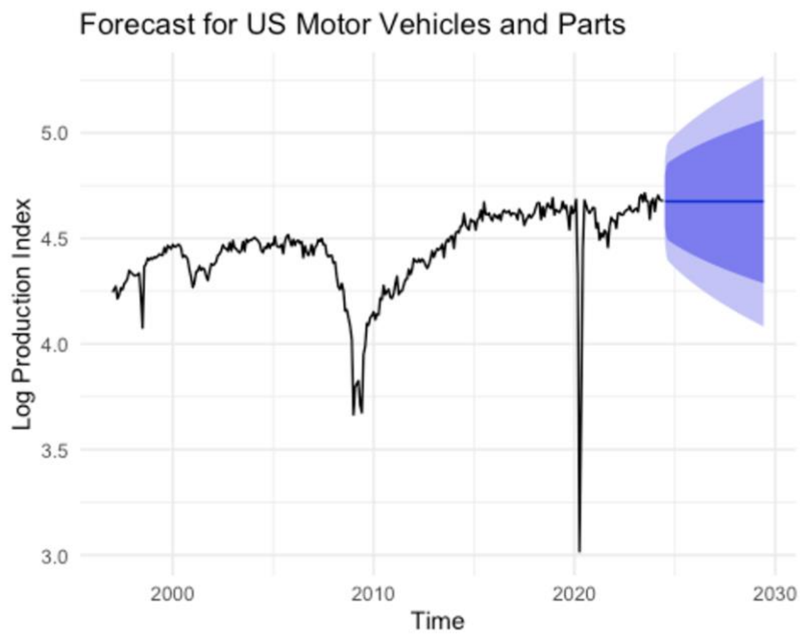

Figure 13: Forecast for US Motor Vehicles and Parts

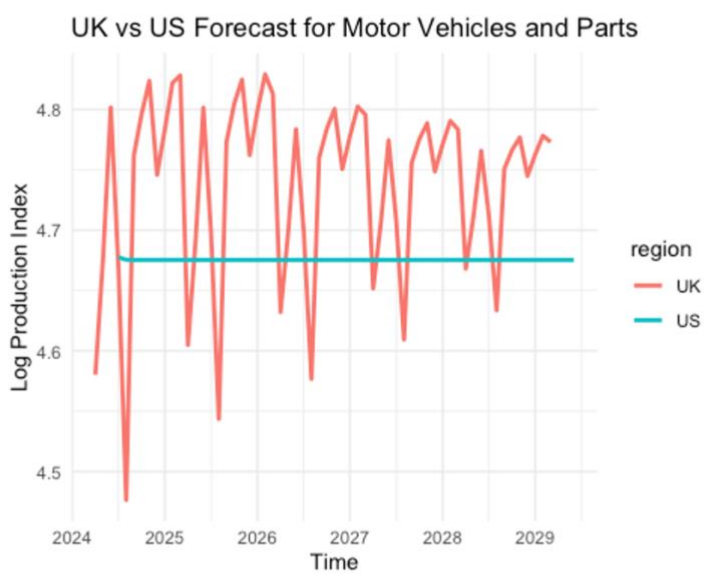

Figure 14: Comparison Forecast for UK and US Motor Vehicles and Parts

Table 1: Comparative Forecasted Mean Industrial Production Index (UK vs. US)

Sector	**Mean Difference**	**Leading Country**
Consumer Goods	0.182	UK
Durables	0.0900	UK
Food and Tobacco	0.207	UK
Textiles	0.0334	UK
Leather	0.481	UK
Wood	-0.283	US
Paper	-0.106	US
Printing	-0.0580	US
Petroleum	-0.0737	US
Chemical	-0.519	US
Plastics	-0.0769	US
Nonmetallic Mineral	-0.370	US
Basic Metals	-0.133	US
Fabricated Metals	-0.0522	US
Machinery	-0.308	US
Computer & Electronic	-0.0628	US
Electrical Equipment	-0.190	US
Motor Vehicles & Parts	0.0626	UK
Aerospace Equipment	0.0590	UK
Furniture	0.0487	UK
Other	-0.0299	US
Mining	-6.19	US
US leads in 14 industries and UK leads in 8 industries		

US Modeling Summary Table

Table 2: Model Selection Summary for US Sectoral Industrial Production

	Variables	Stationarity	Differencing	Best Model	Justification
1	Textile	Yes after 1st diff	1	ARIMA(0,1,0)	No strong autocorrelation
2	Leather	YES	0	ARIMA(1,0,1)	
3	Wood	Yes after 1st diff	1	ARIMA(0,1,0)	No strong autocorrelation
4	Paper	Yes after 1st diff	1	ARIMA(0,1,0)	No strong autocorrelation
5	Printing	Yes after 1st diff	1	ARIMA(0,1,0)	No strong autocorrelation
6	Petroleum	YES	0	ARIMA(1,0,1)	PACF and ACF both cut off at lag 1
7	Chemical	Yes after 1st diff	1	ARIMA(0,1,0)	No strong autocorrelation
8	Plastics	Yes after 1st diff	1	ARIMA(0,1,0)	No strong autocorrelation
9	Nonmetallic Minerals	Yes after 1st diff	1	ARIMA(0,1,0)	No strong autocorrelation
10	Basic Metals	YES	0	ARIMA(1,0,1)	PACF and ACF both cut off at lag 1
11	Fabricated Metals	YES	0	ARIMA(1,0,1)	PACF and ACF both cut off at lag 1
12	Machinery	YES	0	ARIMA(1,0,1)	PACF and ACF both cut off at lag 1

13	Computer And Electronics	Yes after 1st diff	1	ARIMA(0,1,0)	No strong autocorrelation
14	Electrical Equipements	Yes after 1st diff	1	ARIMA(0,1,0)	No strong autocorrelation
15	Motor Vehicles and Parts	Yes after 1st diff	1	ARIMA(0,1,0)	No strong autocorrelation
16	Aerospace Equipment	Yes after 1st diff	1	ARIMA(0,1,0)	No strong autocorrelation
17	Furniture	Yes 1st diff after	1	ARIMA(0,1,0)	No strong autocorrelation
18	Others	YES	0	ARIMA(1,0,1)	PACF and ACF both cut off at lag 1
19	Mining	YES	0	ARIMA(1,0,1)	PACF and ACF both cut off at lag 1

UK Modelling Summary Table

Table 3: Stationarity, Differencing, and Model Selection Summary for UK Sectoral Industrial Production

	Variables	**Stationarity**	**Differencing**	**Best Model**	**Justification**
1	Consumer Goods	Yes 1st diff after	1	ARIMA(1,1,1)	PACF and ACF both cut off at lag 1
2	Durables	Yes 1st diff after	1	ARIMA(1,1,1)	PACF and ACF both cut off at lag 1
3	Food and Tobacco	Yes 1st diff after	1	ARIMA(1,1,1)	PACF and ACF both cut off at lag 1

4	Textiles	Yes 1st diff after	1	ARIMA(1,1,1)	PACF and ACF both cut off at lag 1
5	Leather	Yes 1st diff after	1	ARIMA(1,1,1)	PACF and ACF both cut off at lag 1 PACF and ACF both cut off at lag 1
6	Wood	Yes 1st diff after	1	ARIMA(1,1,1)	PACF and ACF both cut off at lag 1
7	Paper	Yes 1st diff after	1	ARIMA(1,1,1)	PACF and ACF both cut off at lag 1
8	Printing	Yes 1st diff after	1	ARIMA(1,1,1)	PACF and ACF both cut off at lag 1
9	Petroleum	Yes 1st diff after	1	ARIMA(1,1,1)	PACF and ACF both cut off at lag 1
10	Chemical	Yes 1st diff after	1	ARIMA(1,1,1)	PACF and ACF both cut off at lag 1
11	Plastics	Yes	0	ARIMA(1,0,1)	No strong autocorrelation
12	Nonmetallic Minerals	Yes 1st diff after	1	ARIMA(1,1,1)	PACF and ACF both cut off at lag 1
13	Basic Metals	Yes 1st diff after	1	ARIMA(1,1,1)	PACF and ACF both cut off at lag 1
14	Fabricated Metals	Yes 1st diff after	1	ARIMA(1,1,1)	PACF and ACF both cut off at lag 1
15	Machinery	Yes 1st diff after	1	ARIMA(1,1,1)	PACF and ACF both cut off at lag 1
16	Computer and Electronic	Yes 1st diff after	1	ARIMA(1,1,1)	PACF and ACF both cut off at lag 1
17	Electrical Equipment	Yes	0	ARIMA(1,0,1)	No strong autocorrelation

18	Motor Vehicles and Parts	Yes 1st diff after	1	ARIMA(1,1,1)	PACF and ACF both cut off at lag 1
19	Aerospace Equipment	Yes 1st diff after	1	ARIMA(1,1,1)	PACF and ACF both cut off at lag 1
20	Furniture	Yes	0	ARIMA(1,0,1)	No strong autocorrelation
21	Other	Yes 1st diff after	1	ARIMA(1,1,1)	PACF and ACF both cut off at lag 1
22	Mining	Yes 1st diff after	1	ARIMA(1,1,1)	PACF and ACF both cut off at lag 1

References

Hyndman, R. J., & Athanasopoulos, G. (2021). *Forecasting: Principles and Practice* (3rd ed.). OTexts.

National Institute of Standards and Technology (NIST). (n.d.). *Exploratory Data Analysis*. https://www.itl.nist.gov/div898/handbook/eda/eda.htm

Purdue Online Writing Lab. (2019). *APA Formatting and Style Guide*. https://owl.purdue.edu/owl/research_and_citation/apa_style/apa_formatting_and_style_guide/general_format.html

Appendices

Appendix A: R Code Snippets

(Full scripts available on request)

- Data import, filtering, and tsibble conversion
- Histogram, Q–Q, trend-line, and ACF plot routines
- ARIMA model fitting and forecasting functions

Appendix B: Full Sectoral Summary Tables

- Table B1: UK sectors, 22 industries, full descriptive stats
- Table B2: US sectors, 22 industries, full descriptive stats

Appendix C: Model Diagnostics

- ADF/KPSS test results for stationarity
- Residual diagnostic plots (histogram, Q–Q, ACF, PACF) for each ARIMA model