Condensed report

Max Austin - Mada2

2025-04-21

Introduction

This paper analyzes the United Kingdom's international trade activity from 1995 to 2023 using data from the CEPII BACI database. The aim is to test whether future UK trade values—total trade, product-level flows, and country-level trade patterns—can be accurately forecasted using current data-based techniques.

To answer this, the report focuses on three significant questions: Will UK volumes of trade rise or fall in coming years? What goods and trading partners will set future trends? And what is difficult when attempting to forecast long-term trade?

The original data comprises over 258 million observations of trade between 238 nations and 5,000 products. It is reduced to UK trade only as an importer or exporter. The work is conducted in four stages: initial exploration and visualization, correlation analysis, product and country clustering, and finally, deep learning models for prediction.

Additional exploratory plots and additional figures are provided in the supplementary PowerPoint appendix due to space constraints.

Exploration and Visualisation

Before analysis, the raw CEPII BACI dataset of more than 258 million records of international trade was pre-filtered to keep only rows where the UK imported or exported. This reduced the dataset to 13.4 million observations.

Product categories were originally 6-digit Harmonized System (HS) codes and country codes were numeric and both replaced with descriptive labels. Products were too broad in their detail (e.g., 21 varieties of meat, multiple varieties of coffee and eggs), so these were merged into more general product categories where feasible.

After cleansing and consolidation post, the dataset came down to 4.96 million observations around equal to 1.9233751% of the initial dataset. They consist of 3.39 million exports and 1.59 million imports.

In spite of additional export deals of quantity, this does not include greater trading value. There are greater total values of import over the period for UK compared to exports, with its net trading balance being -5.1058488^{9}, which is equal to continuous deficit.

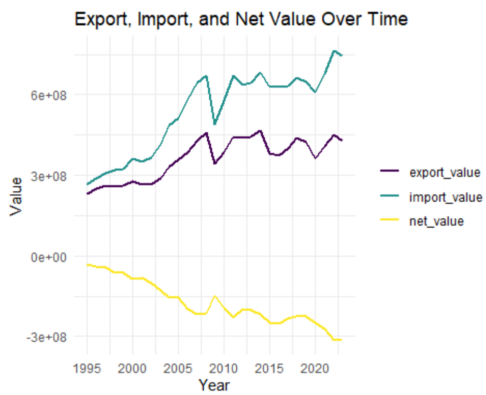
Visualisation

Trade value over time

Figure X illustrates the UK's net trade, import, and export totals between 1995 and 2022. Exports (purple) and imports (teal) exhibit steady growth up to 2008, when they varied depending on world economic turning points. Net trade (yellow), always in negative figures, shows a persistent and growing trade deficit—especially after 2015.

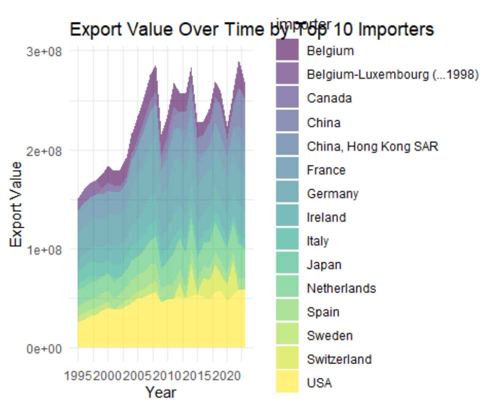
Severe economic shocks such as the 2008 financial crisis and pandemic 2020 triggered sharp declines in trade, then partial recoveries. Post-Brexit (2016) onwards, export growth plateaus while imports remain unchanged, which indicates emerging new trade frictions.

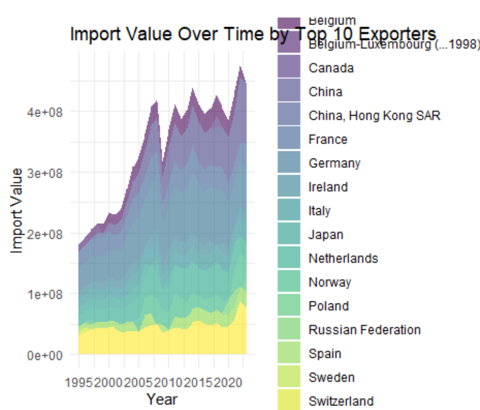
This historical trade deficit bears witness to UK dependence on imports and highlights the challenge of export competitiveness.



Stacked Area Plots

The stacked area plots indicate the UK's export and import values with its top ten trading partners between 1995 and 2023. Export values rose steadily until the late 2010s, led consistently by the US, Germany, France, and the Netherlands. Imports also rose, with Germany, China, and the US being leading suppliers. Notably, China's role expanded exponentially in the 2000s. Both graphs identify the UK's historic reliance on its key European and world partners, and the persistent import and export value deficit is a signal of a long-term structural trade deficit.

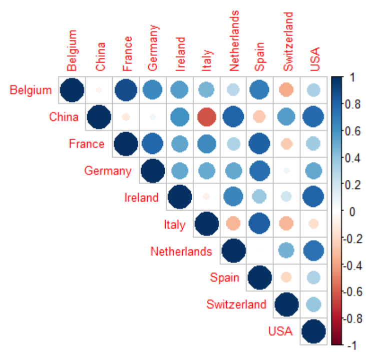


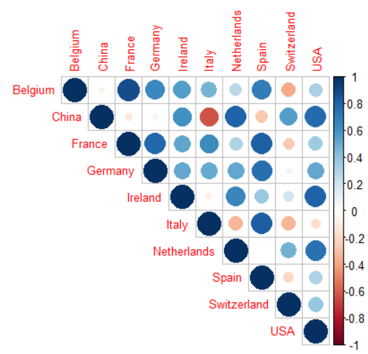


Correlations

Correlation between trade partners

These two correlation matrices highlight the UK's prime trading partners segregated according to whether they are being exported to (left) or imported from (right). Exporting, the UK is extremely positively correlated with most of Europe's partners—Germany, France, Ireland, and the Netherlands—sitting in synchronised export trends into these countries. The USA, being an important partner, is less correlated, likely reflecting differences in economic cycles. On the import side, the relationships are closer still, especially between EU countries, and could reflect coordinated supply chains and integrated buying from the single market. On aggregate, the matrices reveal strong regional trade linkages, useful for identifying co-movements in trade volume that impact future projections.

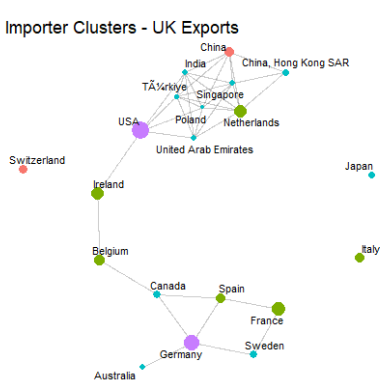


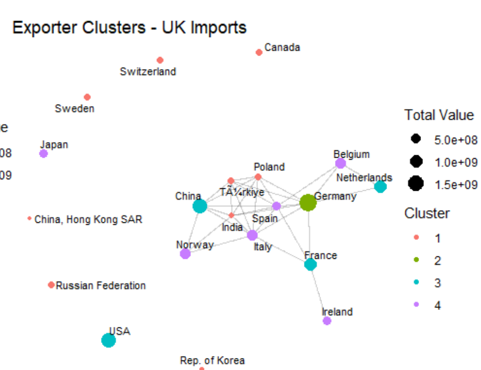


Clustering

The cluster plots reveal natural clusters of the UK's trading partners based on similarity in trade. European countries like Germany, France, and Spain are tightly networked groups within UK export clusters (importers), whereas partners like the USA and Ireland are isolated, which indicates different trade patterns.

The exporter clusters (UK imports) also show analogous tendencies, a core European group isolated from the likes of the USA, Canada, and China. The inherent structures demonstrate trade cohesion at a regional level and vindicate the use of deep learning models that can pick out such complex, not-so-clear patterns in trading behavior.





Deep Learning

Methods

Two separate deep learning models were developed to forecast the United Kingdom's trade values—one for exports (UK as exporter) and one for imports (UK as importer). Both models were run using the Keras API in R, with TensorFlow backing, and particularly used Long Short-Term Memory (LSTM) layers to capture the temporal dynamics inherent in the trade data most effectively. Prior to modeling, trades with trade values less than 2500 were removed to simplify and avoid overfitting. Trade values were then log-transformed using log1p to address skewness and stabilize variance. Three lag features (lag_1, lag_2, lag_3) were also derived based on previous trade values summarized by trading partner (importer/exporter), product, and year, which are specifically tailored to identify sequential trade patterns.

Categorical features, i.e., trading partners (exporter/importer countries) and product types, were factor-encoded as integer representations. The corresponding factor levels were stored separately to enable proper decoding after prediction. These integer-encoded features were converted to dense embedding vectors through learnable embedding layers so that the model is capable of learning useful representations of countries and products. Numeric variables like the year of trade and lagged trade attributes were standardized (zero mean and unit variance) for improved model performance and convergence.

For each model, the categorical embeddings (for products and trading partners) were concatenated, projected via dense layers, and repeated over time to be correctly aligned with the numeric time-series features. This tensor was then input into an LSTM layer that is able to learn complex sequential patterns. The output of the LSTM was passed through additional dense layers to generate a continuous prediction of trade value.

The importer model and exporter model were both individually trained for 20 epochs at a batch size of 216 and mixed-precision training (mixed_float16) for optimization of computational resources. Mean Squared Error (MSE) was the principal loss function with Mean Absolute Error (MAE) additionally tracked as an assessment metric. A 20% validation split and epoch-wise data shuffling were employed during training, with training curves monitored closely to maintain convergent stability. Future predictions for the 2020–2030 time period were then created by producing prediction grids of all potential combinations of partners and products. Predictions were then re-decoded back into their native units using retained scaling attributes and factor-level vectors to enable correct interpretation.

In contrast with conventional feedforward networks, such an LSTM-grounded architecture introduces a strong sequence-aware methodology in forecasting the UK's future trade values accurately.

Exporter moded training

The exporter model trains fast, with both the training loss and MAE falling off sharply in the initial 5 epochs and flattening out soon after. The validation measures follow nearly the same trajectory, with no divergence between the validation and training curves. This indicates excellent generalization and no risk of overfitting. Both loss and MAE flatten early and remain flat from epoch 20, which indicates that the model reaches optimum performance early and maintains it for a long time.

Importer Model Training

The importer pattern has a linear, smooth reduction in training loss and MAE over the 20 epochs. Validation measures are larger to start and fluctuate more widely, especially after epoch 10, but level off in general. The consistent difference between training and validation curves suggests slight underfitting or limited ability to identify complex patterns in some import relationships. However, the model is stable in performance and does not overfit, suggesting good—but conservatively so—generalization.

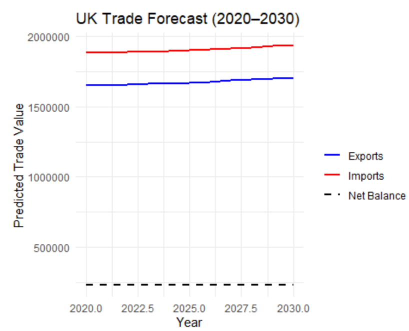
Summary of training

Both exporter and importer models have steep declines in training loss and mean absolute error (MAE) during early training. The exporter model converges smoothly, with validation metrics tracking training closely, showing good generalization and no sign of overfitting. The importer model also converges well, though with a slightly wider and more oscillating validation gap, which likely is a sign of the richness of import relationships. While there are some fluctuations in the late period, both models are stable and suitable for precise trade forecasting.

Results

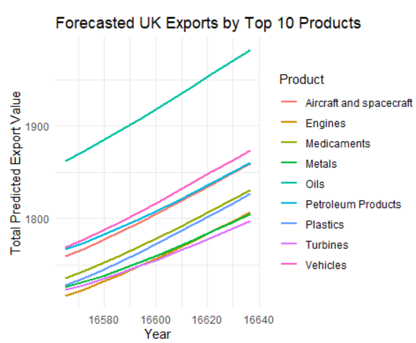
Forecasted Total Trade Overview

Figure X graphically forecasts UK imports, exports, and net trade balance from 2020 to 2030 using an autoregressive LSTM model. Both imports (red) and exports (blue) are stable in the long run with imports consistently higher than exports. This forms a persistent trade deficit (black dashed line). The autoregressive framework—employing lagged values of trade—signifies that past trends in trade are strong predictors of future direction, sustaining the established structural trade deficit in UK trade.



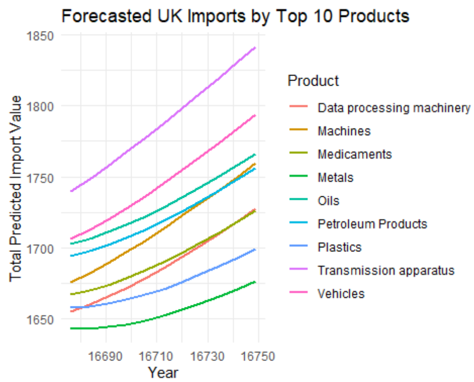
Export top 10 producs

Figure X presents the forecasted export values for the UK's top 10 product categories for the period 2020–2030. All the major goods—metals, petroleum products, engines, and vehicles—are found to increase steadily over time. The increasing trends are smooth as a result of the autoregressive LSTM model, which extrapolates based on recent growth trends and smooths over short-run volatility.



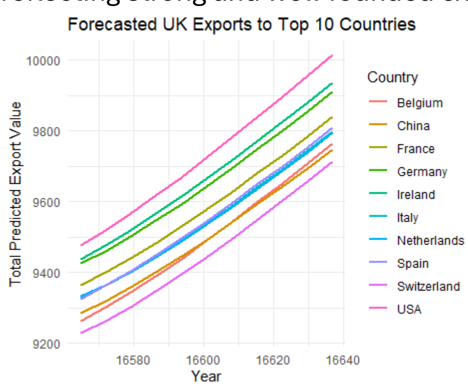
Import top 10 producs

Figure X shows estimated UK import values for the leading 10 product categories from 2020 through 2030. Most of the large categories—like vehicles, data processing equipment, and petroleum goods—are likely to rise steadily. While some like metals and plastics rise at a lesser rate, aggregate demand for imports appears to be on the rise in all sectors. This is an indication of the autoregressive LSTM model's tendency to sustain current patterns of growth while suppressing short-run variability.



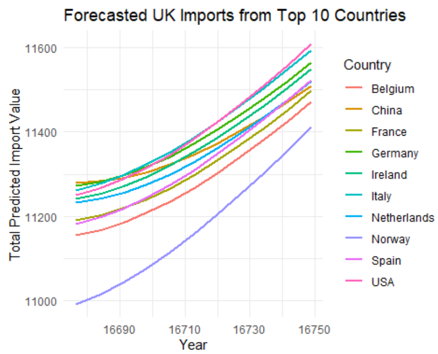
Export top 10 Countries

Figure X shows projected UK export values to its top 10 major international trading partners from 2020 to 2030. The USA is poised to remain the largest export market, with Germany and the Netherlands following in second and third places. All countries show steady upward trends, reflecting strong and well-founded export growth in the UK's largest trade relationships.



Import top 10 countries

Figure X shows forecast UK import values from its top 10 trading nations from 2020 to 2030. China and Germany remain top suppliers, but all countries—particularly Norway and the USA—have steady growth. The general pattern upwards suggests a growing dependence on key global sources and no evidence of falling import business.



Deep learning conclusion

The deep learning models constructed to forecast UK trade flows—separately for exports and imports—produced robust, interpretable results. Leveraging embedded representations for products and countries, together with lag-based numerical features, the two models generalized well over training and validation.

The autoregressive structure, forecasting from close proximate history, was responsible for smooth increasing trends in most goods and partners. Although this makes the model powerful, it diminishes responsiveness to surprise shifts in policy or economics.

Overall, the models supply reasonable high-level forecasts and confirm long-term existing trade imbalances, but their conservative results point to a use best suited for long-range planning rather than minute short-run prediction.

Final conclusion

This article sought to determine whether or not UK trade—total, by product category, and by destination/origin country—can be properly forecast using existing data-based techniques. Starting with 1995-2023 historical facts, visual observation revealed stable patterns such as the UK's chronic trade deficit, its excessive dependence on a few leading trading partners, and the dominance of high-value industrial goods such as motor vehicles, petroleum, and pharmaceuticals.

Correlation and cluster analysis revealed deeper structure in the data, including co-movements between product categories and regional trade blocs. These were insightful background for the design and interpretation of predictive models.

Two deep learning models were then developed to forecast UK imports and exports through 2030. While the models could replicate long-term directional trends—such as the widening trade gap—their strengths and limitations of autoregressive forecasting were also emphasized. In particular, their reliance on past values and smooth patterned output limit responsiveness to shocks, policy changes, or external disturbances.

In short, deep learning may serve to produce penetrating high-level estimations of macroeconomic trade patterns, especially supplemented with intense exploratory probing. However, they must be perceived as trend benchmarks rather than precision forecasts. They may advance future improvements, possibly by combining it with outside economic indicators, memory models that operate over time, or priors on structures.

Reference

https://www.cepii.fr/CEPII/en/bdd_modele/bdd_modele_item.asp?id=37

appendix

All code and supporting documents are available on github in the following perma link https://github.com/maxaus2002/ASSESSMENT-2-for-MAST7220/tree/4ee2afca34a870aa6cfba57fd4bb3f3344b36c84

Full version of report available on github Accompanying power point available on github (Mainly contains condesned versions of data ommitted from the condensed report but were present in the full report.)