Liverpool-Humber Optimization for Freight Transportation (LHOFT): Forecasting White Paper

Authors: Peter Neal, Aaron Lowther, Daniel Waller and Idris Eckley (Lancaster University)

Abstract

This paper outlines freight movement forecasts developed during the Liverpool-Humber Optimization for Freight Transportation (LHOFT) project. The very limited quantities of freight flow data that were available prevented the construction of the anticipated bottom-up forecasting framework. This necessitated the development of an alternative approach, producing freight flow forecasts using a top-down approach. Within the paper we outline our approach, starting with publicly available macroeconomic data and using expert knowledge within the LHOFT consortium to develop models and insights into UK freight flows. We also highlight the consideration of `what if' scenarios, which are particularly pertinent given the significant events that have passed since the inception of LHOFT including, inter alia, the effects of Brexit and the COVID-19 pandemic.

1. Introduction

The aim of this paper is to give an overview of UK freight movement forecasts developed within the LHOFT programme. The work presented in this paper is part of the Liverpool-Humber Optimisation of Freight Transportation (LHOFT) project supported by Innovate UK under its End-to-End transportation call.

The LHOFT project was conceived in 2016 with an aim to investigate, and to seek to promote the rebalancing of the movement of UK imports and exports. The majority of UK manufacturing activity is north of the Severn-Wash axis being concentrated on the Midlands and north of England. Conversely, most UK/Europe freight movements enter or leave the UK via ports south of the Severn-Wash axis with the Dover Straits, including Eurotunnel, accounting for a lion's share of this. The transportation of freight to and from the north of England to southern ports is a major source of congestion on the roads, in particular, the M6, M1 and the road infrastructure in the south east of England.

The LHOFT project was conceived before the 2016 EU referendum took place and the commencement of the LHOFT project on 1 August 2017 was during the early stages of the Brexit process. The coinciding of the LHOFT project with Brexit and uncertainties in the UK's trade relations with the EU has brought into focus the significant reliance of the UK on the Dover Straits for trade with the EU. This has been reinforced by movement restrictions necessitated by the Covid-19 pandemic. Thus a key aim of the LHOFT project to make UK freight movements more robust and less susceptible to problems on key shipping routes has become increasingly import.

LHOFT’s goal has been to influence and reshape UK freight movements by seeking to exploit and develop opportunities presented by the Liverpool-Humber (“M62”) corridor, both in terms of the manufacturing base and the connectivity, and in places lack of it, across this belt of northern England. This reshaping has been envisaged at both a tactical and strategic level. At a tactical level, the construction of the LHOFT platform allows companies to explore alternative routes for freight movements and the opportunity for collaborative arrangements to share costs and risks. On a strategic level, the LHOFT project has explored infrastructure requirements, especially in a rail context, to increase the viability of trans-Pennine rail travel to reduce the amount of freight movement on roads to reduce congestion and the carbon footprint.

It is within the above context that we explore the role of forecasting to support strategic planning. Given suitable historic data, forecasting has a pivotal role to play in strategic planning through the identification of trends and seasonality in freight movement. Successful forecasting enables us to identify which commodities are likely to experience increases or decreases in demand and to take appropriate action in the planning of strategic infrastructure projects. This relates not only to governmental road or rail building projects but also to private investment such as port infrastructure and new ships to service increase route demand. To be able to fully utilise the forecasts further information, such as regional and seasonal demand for products, needs to be taken into account.

A stated aim of the LHOFT project was to obtain shipping information from 50 companies in order to have a representative sample of UK imports/exports. The aim was to obtain substantive data on the products that a company imports/exports including origin and destination of product, route taken (inclusive of ports), mode of transportation for each leg of the journey, frequency of movements and time and cost of routes. Further information sought included the reasons for choosing a given route, in particular, the relative importance of time, cost, reliability and other factors in the decision process. This would enable the extrapolation of the data collected to give an informative picture of UK freight movements. Also by gaining an in-depth understanding of current freight movements and future forecasts this enables us to explore the likely consequences of different scenarios on freight movements and help identify infrastructure investment opportunities.

The commercial sensitivity of company shipping information has been a major barrier to obtaining the required information. The widespread utilisation of 3pl’s (third-party logistic) providers means that many companies are only able to provide limited information on their import/export operations. Consequently, we have only been able to obtain extensive shipping information from two project partners Kraft-Heinz and Nestle. This has meant that we have not been able to develop a “bottom up” picture of UK imports/exports and the corresponding picture of freight movements. Thus it has been necessary to seek alternative approaches to answer key questions in terms of current and future freight movements.

The stated aims of the forecast modelling were as follows:- To generate a forecasting model of suitable complexity to predict freight flows for the next 5-10 years.

Macroeconomic and carbon impacts would be included as well as current transport infrastructure barriers and existing utilisation. Specifically, the modelling would seek to:

- *Identify the proportion of freight moving through southern ports but originating from northern industry.*
- *Identify import/export balances.*
- *Compute total economic and environmental cost of baseline and projections for future.*
- *Include and quantify Ireland's impact on England North-South traffic flows.*

The lack of detailed company level data and the uncertainty surrounding the ultimate details of the Brexit process has meant focussing on a subset of the initial modelling goals. In particular, limited analysis of the impact of Ireland on freight movements through the UK has been made as this will be severely affected by Brexit trade arrangements which are still ongoing.

Whilst a "bottom up" approach to forecasting has not been possible during the lifetime of the LHOFT project, progress on forecasting has been made through adopting a "top down" approach. This approach has been based upon obtaining and analysing macroeconomic data available from governmental sources including HMRC (Her Majesty's Revenue and Customs) and DfT (Department for Transport). These data sources provide information on the UK import and export balance and the amount of freight movement through different UK ports. This allows us to obtain a broad-brush picture of UK freight movements. We have sought to delve into these macroeconomic data sets to extract maximum information from these by combining the data sources and utilising additional information. The ultimate aim with the successful adoption of the LHOFT platform to combine the "top down" approach with "bottom up" individual company data to give an extensive map of UK freight movements along with an extensive forecasting model addressing the key questions given above.

The remainder of the paper is structured as follows. In Section 2, we discuss the available macroeconomic data. In Section 3, we discuss what information is available from the macroeconomic data on aggregated freight flows and how these can form the building blocks for a "top down" investigation of freight flows. In Section 4, we build the basic forecast model utilising the macroeconomic data. We outline the time series modelling approaches taken and we highlight the strengths and limitations of these approaches in forecasting freight movement data. In Section 5, we discuss route freight flow allocation and the limitations of using only the aggregated HMRC and DfT data sets, highlighting the importance of additional information and the LHOFT platform in gaining further insights into forecasts. In Section 6, we present a more detailed analysis of Food imports and exports. This analysis combines freight movements of food-based products with expert allocation of current freight movements. This enables us to demonstrate the potential for forecasting in light of more informative data and illustrates how "top down" and "bottom up" data can be combined. In Section 7, we discuss extensions of the forecasting approaches presented in this paper.

2. Macroeconomic Data

The macroeconomic data that we utilised to build the “top down” forecasting models were the HMRC trade data and DfT major port data set.

The HMRC trade data website[1] allows the user to extract a wide range of UK import/export data. This data is updated monthly as new data become available. The following trade information is available although some combinations of data are not available:

- **Time** - Monthly, quarterly or annual data from January 1996.
- **Product** - Products can be broken down by SITC (Standard International Trade Classification) code or HS (Harmonized System) code. For SITC code, data available to level 5.
- **Country** - Trade with foreign countries with aggregation available for EU and regions.
- **Region** – 9 English regions along with Scotland, Wales and Northern Ireland. There are also two categories for unallocated (to region) trade – known and unknown.
- **Tonnage** – The total number of tonnes for any trade combination. Alternatively monetary value in pounds sterling is available.

Given that origin and destination are crucial to the investigation of freight movements in the LHOFT project, we have ensured that the data has the finest granularity possible for region and country. Consequently, on the temporal data we have quarterly data and on product code we have SITC level 2 data. Furthermore, since DfT port movement data are only available since 2000, we have focussed our attention on data since January 2000.

The DfT port and domestic waterborne freight statistics: data tables (PORT)[2] provide an extensive range of maritime and shipping statistics. We have utilised data set PORT0499 which provides annual records of traffic in major UK ports. This data set is updated and released on an annual basis, typically around the beginning of September. The data provided is the annual imports (inward) and exports (outward) flows involving each port. The data provided in each annual release is annual data from 2000 onwards. We have considered the PORT0499 releases in 2017, 2018 and 2019 covering 2000-2016, 2000-2017 and 2000-2018, respectively, in preparing our forecasts. Each release has presented slightly different information but all have included:

[1] https://www.uktradeinfo.com/Statistics/BuildYourOwnTables/Pages/Home.aspx
[2] https://www.gov.uk/government/statistical-data-sets/port-and-domestic-waterborne-freight-statistics-port#major-port-traffic-by-cargo-type

- **Year** – From 2000 onwards
- **Port** – UK port
- **Country** – Country in which the port of arrival and port of despatch is located for UK export and import trade movements.
- **Direction** – Inward (import), outward (export)
- **Cargo Classification** – The type of cargo transportation, both category and description.
- **Tonnage** – The total tonnage on any given combination. Also available, where appropriate, is the total number of units for unitised goods.

3. Forecasting possibilities

We outline how the "top down" forecast modelling using the HMRC and DfT data sets can assist in the overall goal of modelling UK freight movement. We describe the process for exports with a similar process available for modelling imports.

To obtain the full picture of UK export freight movements with Europe we require the origin (**A**) and final destination (**B**) of each freight movement along with the port of exit (**X**) from the UK and the port of entry (**Y**) to Europe. Besides this information we require the product being exported, the mode of transport for each leg of the journey and the frequency and size of each shipment. In order to interpret and understand the decision making process we also require information on the cost of the journey in terms of CO_2 emissions and time as well as monetary. Preferably this information is available for each leg of the journey.

The above description represents the data "holy grail" for mapping out UK export freight movements and for developing a model for how likely a company is to choose a given route/mode of transport for exporting their product. It is only possible to obtain this data from the exporters or their logistic providers and even then the cost per journey leg might not be easily quantified.

The HMRC and DfT data sets described in Section 2 give us aggregated information on freight movements. The ultimate aim would be to disaggregate the data using a combination of individual company data and mathematical modelling to obtain an approximation of UK export freight movements. This picture will become more accurate as more individual company data is available to refine the parameters of the mathematical model and reduce uncertainty in the estimation processes.

The HMRC data provides detailed information on the product being exported (SITC level 5 data available reducing to SITC level 2 data for regional data) and the quantity (weight in tonnes). Aggregation over origin (**A**) is at a UK regional level and final destination (**B**) is aggregated at a foreign country level. Note in some cases where trade levels are low, regional aggregation of countries takes place. This data set provides no information on route taken in freight movement, and so presents aggregation over **X** and **Y** for all routes **A-X-Y-B**. Furthermore, the HMRC data does not provide information on mode of transportation or the cost of journeys.

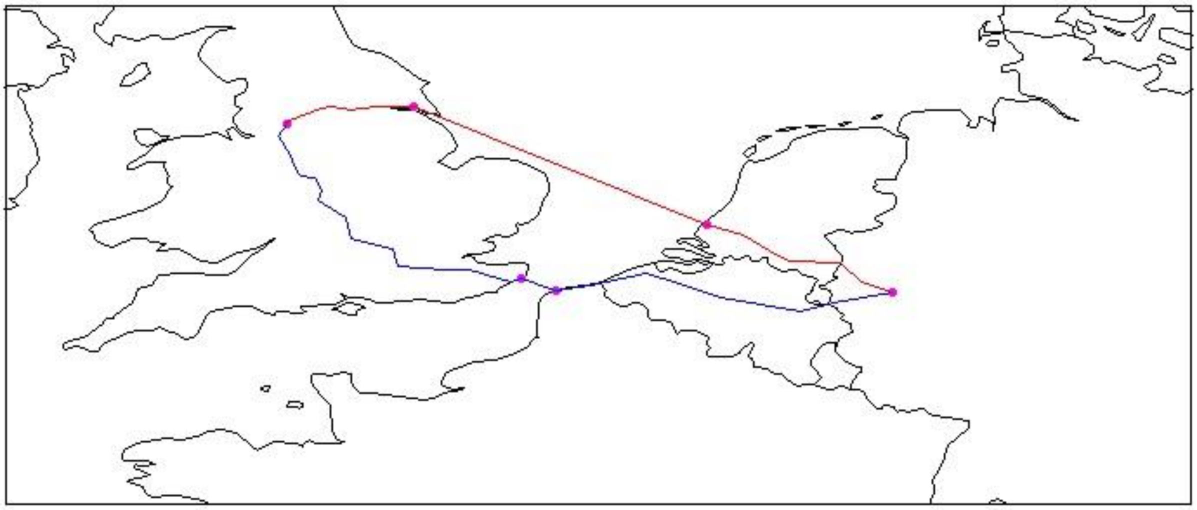

Figure 1: Illustrating routes between Manchester (A) and Cologne (B), via Dover (X) - Calais (Y) (blue) and Humber (X) - Rotterdam (Y) (red).

Figure 1 illustrates two potential route choices between Manchester (North West of England) and Cologne (Germany). These routes are via the Dover straits and the North Sea crossing, Hull to Rotterdam. The HMRC data provides the information at the aggregated level of freight flows per product between the North West of England and Germany. Even with freight flow information at the more refined level between Manchester and Cologne, say, there is further disaggregation required to obtain estimates for what proportion of freight flows on each of the potential routes between the two locations.

The HMRC data tables do allow for the construction of data tables containing information on port of exit, **X**. This information is not available on a regional basis or for trade with the EU. Thus the data provided for a given product presents aggregated data over **A** and **Y** for all routes **A-X-Y-B**, where **B** is a non-EU country. Depending on future trade arrangements between the UK and EU, the set of countries, **B**, may increase. However, since no information is made available on UK origin of exports we do not explore this further in this paper. Given that the LHOFT projects main focus is trade with continental Europe we have not pursued in detail the limited port data available for this trade from the HMRC database.

A limitation of the HMRC data at a regional level is that the regional data is estimated using the proportion of employees in each region. These proportions are then applied to each commodity that a business trades in. This methodology is applied to regional data from the first quarter of 2013 onwards. Prior to this date alternative allocation of trade to regions was used based on the region to which a company's Head Office belonged – not a reliable base. Further information on the construction of regional trade statistics is available from HMRC (2018), Regional Trade Statistics Methodology paper.

The DfT data provide complementary data to the HMRC data in that information is available on the amount of freight/trade shipped between UK ports, **X**, and foreign ports, **Y**, with the latter aggregated at a country level. Given that there is often only one principle shipping route (but possibly several shipping services) between a UK port and a given foreign country, this data is informative in understanding the amount of traffic on a given route. The DfT data does not provide detailed information on the product being exported but on the mode of transportation for the sea-leg of the journey. This enables, with the HMRC data, the inferring of products being moved through a given port. In summary, the DfT data provide by mode of transportation (cargo classification) the total amount in weight (tonnes) aggregated over **A** and **B** for all routes **A-X-Y-B** through a given sea route **X-Y**. For some transportation types, in particular RoRo (Roll-on-roll-off) and LoLo (Lift on, lift off) traffic, the number and type of unit being moved are available.

We are therefore able to use the HMRC data to study the total aggregated movement of freight between **A** and **B** by product type and the DfT data to study the total aggregated movement of freight between **X** and **Y** by transportation type. The temporal nature of the data sets means that we can develop time series models to capture temporal trends, including seasonality, in these data and produce forecasts for freight movements. We present the forecast modelling in Section 4. In Section 6, we discuss approaches to combine the two data sources to estimates, by product, the total amount moving on each route **A-X-Y-B**. We highlight that this is possible for a single product and show that whilst identifiability issues mean we cannot fully apply this with the current data, we note the additional information/assumptions (amount moving, by product, on each **X-Y** combination along with costs) to make this feasible.

4. Forecast modelling

4.1 Data Processing

The initial exploration of the data sets and forecasting considers the HMRC and DfT data sets separately. The first step for both data sets is to process the data into a structure which is amenable for analysis, and in both cases this involved presenting the data in a 4 dimensional array corresponding to the 4 key fields for each data set. Time was a field for each data set. For ease of presentation we construct separate arrays for exports and imports although these can easily be combined into a single 5 dimensional array.

For the HMRC data, the fields and number of categories in each field were as given in Table 1. We have focussed on SITC level 1 rather than the more detailed SITC level 2 to illustrate the methodology more succinctly and also for improved forecasting behaviour which we will discuss in greater detail below. Data are provided for each of the 78 quarters between the first quarter 2000 and the second quarter 2019 and throughout we seek to use the first 72 quarters (18 years) to fit models and the last 6 quarters (first quarter 2018 onwards) to assess performance of the forecasts. There are 15,990 (=13*123*10) combinations (time series) of UK region (unallocated categories combined), Foreign Country and SITC product code to consider.

Field	HMRC Name	Notation	Number in field
1	UK Region	A	13
2	Foreign Country	B	123
3	SITC Code (level 1)	S	10
4	Time	t	78

Table 1: Fields for the HMRC data

For the DfT data, the fields and number of categories in each field were as given in Table 2. As with the HMRC data, we have focussed on the Cargo Category rather than using the more detailed Cargo Group classification. Data are provided for each year 2000 to 2018, inclusive (19 time points). Given that we have fewer time points than for the HMRC data we utilise all the data to fit the models. There are 60,420 (=53*190*6) combinations (time series) of UK Port, Foreign Country port location and Cargo Category to consider.

Field	DfT Name	Notation	Number in field
1	UK Port	X	53
2	Foreign Country	Y	190
3	Cargo Category	C	6
4	Time	t	19

Table 2: Fields for the DfT data

4.2 HMRC Data

The representation of the HMRC export data as a 4 dimensional array enables easy viewing of any time series of interest, whether it be a given UK region, foreign country and SITC product code combination, the whole UK export data or any aggregated combination in between. (As previously we can present and analyse import data in a similar vein.)

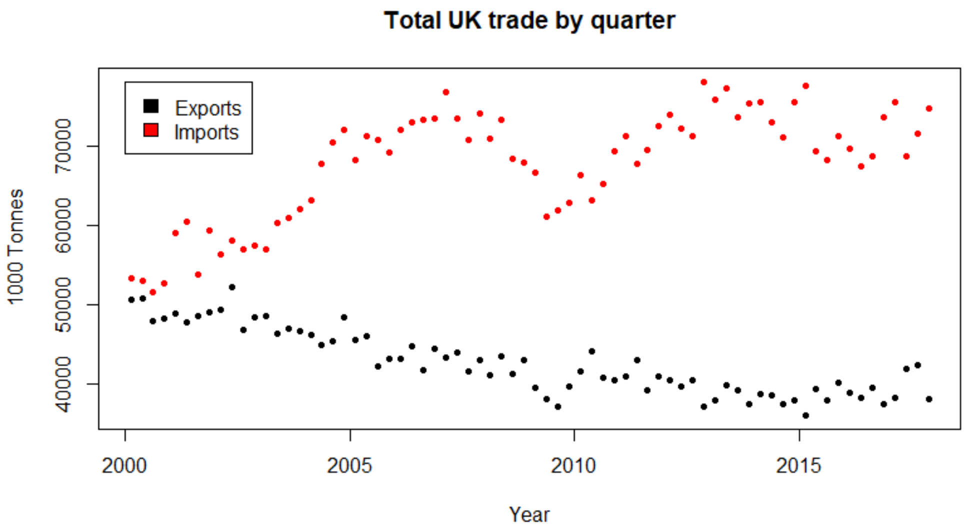

Figure 2: Quarterly imports (red) and exports (black) for the UK, 2000-2017

In Figure 2 we present the total imports and exports in thousand tonnes, by quarter, for the UK from the first quarter of 2000 to the fourth quarter of 2017. We observe that there is a widening trade gap (in tonnage) between total UK imports and UK exports over the period 2000-2018. The effects of the 2008-9 financial crisis are clearly seen in drops in both imports and exports. The total levels of imports and exports look to have stabilised over the past 3-5 years.

There is an observed change in the allocation of trade to regions following the HMRC's change in regional allocation model in 2013. There is a marked increase in some regions, most noticeably Yorkshire and the Humber, with corresponding marked decreases in London and the South East, the principle location of many company Head Offices.

4.3 DfT Data

The DfT data can be presented in a similar manner to the HMRC data in terms of either imports or exports. The availability of data not only in terms of total tonnage but also in terms of RoRo and LoLo units allows for further analysis. It should be noted that only in situations where teu and unit load are applicable is data provided. Therefore in presentation of the forecasting we focus primarily on total tonnage where comparisons with the HMRC data can be made.

The trends in the data mirror those observed in the HMRC data in terms of total exports and imports to and from the UK with a widening trade gap. Given that the DfT data only provides information on the sea crossing between UK and foreign ports we are not able to match up the data on a country level with the HMRC data where the final destination/origin is recorded. The DfT port data consists only of trade through main seaports and thus excludes Eurotunnel, minor seaports, airports and the Irish land border. However, the major seaports account for a large proportion of UK imports and exports.

4.4 Time Series Modelling

The forecasting of the future trends in import and exports either by regions (HMRC) or through ports (DfT) is based on constructing suitable time series models for how these quantities vary over time. A time series modelling approach can be taken regardless of the unit (tonnage, teu or unit load) of measurement.

For both data sets, each time series has three indices alongside the temporal index; region, country and product type for the HMRC data and port, country and transportation type for the DfT data. At the lowest level we are interested in each of these time series; 15,990 for HRMC and 60,420 for DfT. The individual time series for a given combination are often extremely noisy demonstrating considerable variation in levels of trade and offered poor predictive qualities for forecasting. Consequently, we seek to exploit the structure in the data, and in particular, the matrix structure to borrow strength from time series with the same region, country and product type (port, country and transportation type). We will present the approach with the HMRC data but it all translates, with suitable relabelling of type, to the DfT data.

We utilised a grouped time series approached, Hyndman and Athanasopoulos (2019), Chapter 10. This considers the aggregation of time series in a coherent manner without requiring an inherent ordering of the time series. This approach allows us to utilise information from similar time series (for the same region, country and/or product) to build more robust and reliable time series models. Also we consider combinations of the time series from the individual time series for a given region, country and product type to the total aggregated exports (or imports). For example, we could consider imports in manufactured products (SITC 5-8) between the North of England (North West, North East and Yorkshire and the Humber) and the Benelux countries (Belgium, Netherlands and Luxembourg). We have considered two time series modelling approaches for the time series, seasonal ARIMA and exponential smoothing (ETS) models which are applied with the grouping methods. The seasonal ARIMA model is appropriate for time series data with structural temporal dependencies and has good forecasting properties if the model is appropriate. The ETS model makes weaker assumptions about the structure of the time series and is more robust to outlying or unusual observations.

There are a number of methods that ensure that forecasts for a grouped time series are consistent. The simplest reconciliation method is to forecast only the bottom level time series and add the forecasts together to obtain the forecasts for the grouped time series. Alternatively, we could forecast the top level time series and dis-aggregate. This is known as a top-down approach. We focus only on the bottom-up and the optimal combination approaches. The optimal combination approach forecasts each time series in the grouped time series and adjusts the forecasts to make them consistent, see Hyndman *et al.* (2011), Hyndman et al. (2016) and Wickramasuriya *et al.* (2019). We implement the optimal combination approach. It is possible to forecast each time series without reconciling the forecasts. We use a total of three reconciliation approaches namely the bottom-up, optimal combinations and non-reconciled approaches. This led to us implementing six time series modelling combinations.

We utilised the time series models to produce forecasts for future demand. The forecasts consist of point estimates representing the mean (average) prediction for demand at a future time point along with prediction intervals expressing our uncertainty about the forecasts. Wherever possible we constructed the prediction intervals using analytical properties of the time series and we utilised simulation to construct approximate prediction intervals where this was not the case.

4.5 Analysis of data

We focus the data analysis on the HMRC data. Given that we have quarterly data for the HMRC data, we have approximately four times as many observations. In particular, we utilise the first 72 quarters (quarter 1, 2000 to quarter 4, 2017) to fit the time series model and then compare forecasts for the following six quarters (quarter 1, 2018 to quarter 2, 2019) with the observed data.

In order to study both the HMRC and DfT data sets R Shiny[3] apps were written to enable users to explore the data and consider a range of forecasting options. The data were pre-processed with the time series models fitted and the model outputs stored. A detailed technical description of the time series modelling and implementations are presented in Lowther (2020). The model and data were then loaded into the R Shiny apps which can then be used to visualise the observed data along with forecasts.

A total of four R Shiny apps were written, one for the HMRC data and three for the DfT data. The main R Shiny app for the DfT data models freight movement in tonnes, whilst the other two consider subsets of the data, teus and unit loads, respectively. The four R Shiny apps are listed below, along with hyperlinks to the hosting urls:

- HMRC data (tonnes): https://aaronplowther.shinyapps.io/host-hmrc/
- DfT data (tonnes): https://aaronplowther.shinyapps.io/host-dft-tonnage/
- DfT data (teus): https://aaronplowther.shinyapps.io/host-dft-teu/
- DfT data (unit loads): https://aaronplowther.shinyapps.io/host-dft-units/

The R Shiny apps are setup to consider all six forecasting combinations and to produce 90%, 95% and 99% confidence intervals as well as point estimate forecasts. Any combination of the three variables (UK region, foreign country, SITC code for HMRC and UK port, foreign country of port, cargo classification for DfT) can be considered. For example, we can consider the total imports to the North of England (the combined regions of North West, North East and Yorkshire and the Humber) from the Benelux countries (the combination of the countries Belgium, Netherlands and Luxembourg) for SITC product codes 6 & 7 (the combination of Manufactured goods and Machinery & transport equipment). These import freight flow combinations are plotted in Figure 3 along with the observed data and forecasts using the ARIMA model for the six quarters from the first quarter of 2018. The ARIMA forecasts capture the fluctuations in the data well and the 95% confidences intervals demonstrate the uncertainty in forecasted demand. We observe that the forecasts forward are robust to the spikes in imports in the data observed in 2004 and 2006.

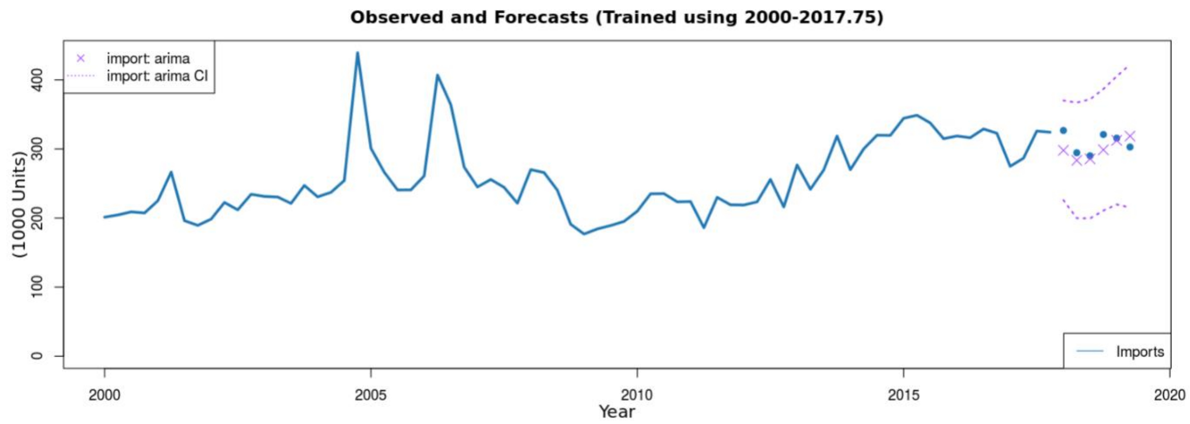

Figure 3: Imports Benelux countries to the North of England for SITC product codes 6 and 7 (Manufactured goods and Machinery & transport equipment). Forecast provided using ARIMA model and 95% confidence interval.

[3] R Shiny: https://shiny.rstudio.com/

In Figure 4, we give the total imports and exports for the UK from 2000 onwards. For these data we provided the reconciled ETS forecasts along with the 99% confidence intervals. We observe that the forecasts give excellent agreement with the observed quarterly data for 2018 and the first half of 2019. The forecasts demonstrate better agreement than those presented in Figure 3 for a subset of the data. This is a general trend that is observed, in that forecasts are more reliable the greater the amount of data aggregation. Whilst the reconciled ETS forecasts are shown in Figure 4, the reconciled ARIMA forecasts are virtually indistinguishable. A general observation is that there is no clear evidence to prefer either the ETS or ARIMA time series model overall in producing the forecasts. We note that the import series shows clear quarterly seasonality whereas the export series shows less variability. These are captured in the forecast predictions and the narrower confidence intervals for the export data.

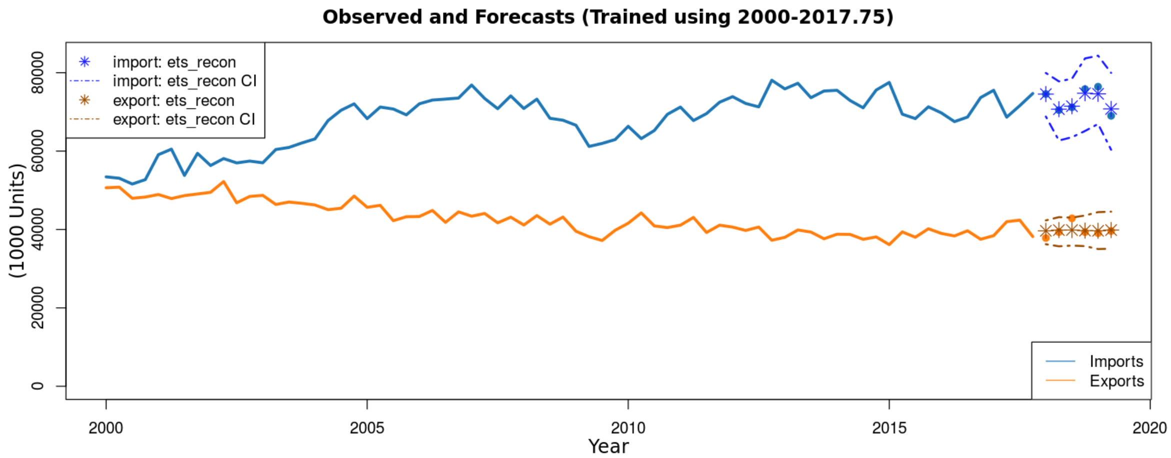

Figure 4: Forecasts for total UK imports and exports with forecasts using reconciled ETS models and 99% confidence intervals. Plot obtained using HMRC forecasting R Shiny app.

5. Route Freight flow forecasting

The HMRC and DfT data sets along with the time series modelling in Section 4 allow us to forecast freight flows at an aggregated level. The aggregation is both spatial (aggregation to region level in the UK and country level in Europe) and route (HMRC data provides information on the level of trade to **A-B**, aggregated over ports **X-Y**, whilst DfT provides information on the level of trade between ports **X-Y**, aggregating over origin and destination). Moreover, the different data sources have different key categorisations, SITC product type for the HMRC data and mode of transportation for DfT data. A question of interest, is can we disaggregate the data to obtain meaningful estimates of the total amount of freight flow on a given route **A-X-Y-B**?

Where a product originates from in a given UK region or European country, and the precise location of its destination for export/import will be, will vary significantly from product-to-product. For many products such as food and household items, the final destination is likely to be approximately proportional to population size of an area within a given region or country. To a lesser extent, population size will be indicative of the breakdown of manufacturing production by area within a given region or country. We can therefore use population demography to approximate the breakdown of imports/exports by area and this can be supplemented by additional information about a given product to refine the breakdown. This can be used for a given SITC product, UK region and European Country to obtain a representative set of locations **A** (UK) and **B** (continental Europe) to allocate routes along.

Given a representative set of locations **A** and **B** for origins and destinations for a product travelling between a UK region and European Country, we can obtain the likely route taken between each pair (**A**,**B**). The LHOFT platform can be utilised to take into account time sensitivity, cost (monetary and CO_2), preferred transportation type and other considerations to propose an optimal route. By averaging the routes over all the representative pairs (**A**,**B**), we can obtain an estimate of the optimal allocation of freight flows to each route. This should provide a reasonable approximation of the routes taken for the given product, region and country combination, but will not include the nuanced requirements of individual companies which will lead to a departure from the proposed allocations. We can then combine the proportion of movement along a given route **A-X-Y-B** with the total amount of trade in the product between **A** and **B**, to obtain an estimate of the total amount of trade on **A-X-Y-B** for the chosen product.

The approach described in the previous paragraph can be applied to all product, UK region and European country combinations. We can then obtain a profile of the total amount of trade on a route **A-X-Y-B** and by aggregating over origin and destination, an estimate of the total amount of trade between a pair of ports **X** and **Y**. We can then compare the estimated total amount of trade on **X**-**Y** with the data provided by the DfT data to test if are estimates are sufficiently close to the observed data.

There will be a discrepancy between the estimates of the total amount of trade through **X**-**Y** and the observed DfT data. These will be due a number of factors including the locations (**A**,**B**) being not fully representative, the inputs for the importance of time sensitivity and cost (monetary and CO_2) not being accurate and alternative choices of routes for various reasons by companies and 3pls. We can refine the analysis based on all these factors as we collect data through the LHOFT platform to better understand the origin and destination locations **A** and **B** and the decision making process for each product in choosing route and mode of transport.

We investigated whether or not we could use statistical approaches to disaggregate the HMRC and DfT data to estimate freight flows along a route **A-X-Y-B**. In order to do this we ran a simulation study with a reduced set of UK regions, European countries, ports, both UK and European, and products. A brief description of the simulation study is provided below with full details provided in a separate report, Neal *et al.* (2019). The simulation study involved a simplified cost structure based upon the distance involved in each leg of the journey and a cost per mile for each leg which was product dependent. This was designed to provide a simple mimic of the LHOFT platform functionality. For each origin, destination and product combination, the HMRC data was used to identify the total amount of freight moving between locations **A** (UK) and **B** (continental Europe). Then for each possible route, **A-X-Y-B**, a proportion of the freight flow was allocated to the route according to the relative cost of the route with a penalisation for more expensive routes. The simulated data were then aggregated over **X-Y**, the total amount of freight flowing through each pair of ports.

The simulated data uses the HMRC data for the total amount of freight flows between origins and destinations **A** and **B** to generate an artificial data set of port-to-port freight flows mimicking the DfT data set. Moreover, for the simulation study we know the total amounts of freight flow on each route **A-X-Y-B**. Given the aggregated data, **A-B** (HMRC) and **X-Y** (pseudo-DfT), it is possible to recover reasonably reliable estimates of **A-X-Y-B** if there is a single product. Specifically, we can use the costs per route and the aggregated data to explore functions for the penalisation of expensive routes to determine estimates of the total amount of the product moving on the route **A-X-Y-B**. Unfortunately this approach does not extend to multiple products (with different penalisation functions) as there is a lack of identifiability with very different product and route combinations leading to similar aggregated port-to-port, **X-Y**, data.

This demonstrates that there is a need for additional data and the LHOFT platform, or similar, to determine accurately the total amount of each product moving along a given route **A-X-Y-B**. However, once an accurate picture of route allocation is obtained this can be combined with the time series forecasts to produce forecasts for each product/route combination. The individual disaggregated forecasts will contain considerable uncertainty. This will not be problematic as our primary interest will be in aggregation of the data, for example, to produce forecasts for the total demand through a port **X**, freight movements along a segment of railway or stretch of motorway.

6. Case Study: Food data analysis

6.1 Introduction

The time series methods detailed in Section 4 and the forecasts produced are appropriate for modelling aggregated data. It is possible to take this approach to construct forecasts for other macroeconomic data scenarios. However, as noted in Section 5, further information and modelling assumptions are required to begin to disaggregate the data to construct specific freight flow for a given route **A-X-Y-B**. In this Section we explore how the forecasts can be extended using a specific example, the movement of Food & Drink products between the UK and members of the EU. We outline the additional information which has been available to the LHOFT consortium to investigate Food & Drink product movements in detail and the additional analysis that this permits.

The key source of data for modelling Food & Drink product movements is the HMRC data tables. We can follow the approaches taken in Section 4 and extract the regional trade data for Food & Drink. We delve into the data in greater detail than in Section 4 by extracting Regional trade with EU countries for Food & Drink products at SITC level 2. The SITC level 2 codes considered were all 0X codes (sub-categories of SITC level 1 code 0 – Food & live animals) with the exception of live animals (00), all 4X codes (sub-categories of SITC level 1 code 4 – Animal & vegetable oils) 11 (Beverages) and 22 (Oil seeds and oleaginous fruit). For these data we used the time series modelling approached outlined in Section 4 to construct models for forecasting at SITC level 2 for UK trade with EU countries in Food & Drink products. This analysis can easily be extended to all SITC level 2 products and all countries, if desired.

The HMRC data allows, at a national level, finer product, import and export, details. The 12 SITC level 2 codes for Food & Drink identified above can be further subdivided into 85 SITC level 3 product codes. The finer identification of products allows for better understanding of the likely mode of transportation for each product. The identification of the likely mode of transportation along with country of origin (import) or destination (export) enables possible routes for the product to be inferred by using port capacity and DfT trade data. Freight flow analysis performed by LHOFT consortium member, PRB Associates has combined expert knowledge of maritime logistics with the HMRC and DfT data sets to estimate, for each product (SITC level 3), the total amount of trade in tonnes, by EU country (**B**), which enters/leaves the UK, via each port (**X**).

The PRB Associates' analysis gives a scientific combination of HMRC data with port of entry/exit information. We have sought to extend this analysis further by combining with data for UK region (**A**). This allows for identification of estimates of the total amount of freight flow on routes **A-X-B**, where **A** and **B** are aggregated at UK region and country level, respectively. This is a major advancement in obtaining detailed route data **A-X-Y-B**, especially as the number of European port, **Y**, options will be limited given the other factors. Specifically, we have the data obtained by PRB Associates which consists of:

- SITC product code at level 3. (85 categories)
- UK port. (42 ports)
- EU countries. (27 countries)

And we have the HMRC Food & Drink data which consists of:

- SITC product code at level 2. (12 categories)
- UK regions (12 regions plus unallocated to give 13 regions).
- EU countries. (27 countries)

We have utilised the 2017 PRB Associates analysis which takes the 2017 HMRC data tables as its source and therefore we match this with the regional HMRC data for 2017. There are very minor discrepancies (typically less than 0.001%) between the total tonnages due to rounding errors in disaggregating the national SITC level 2 data by level 3 SITC code or region. These discrepancies are removed to enable combination of the data.

In order to estimate the total amount of trade **A-X-B** for each SITC level 2 product given the total amounts of trade **A-B** (HMRC regional data) and **X-B** (PRB analysis), we analyse the allocation as a Hitchcock transportation problem, see Hitchcock (1941). The allocation of freight flows to routes **A-X-B** is then optimised via linear programming, see for example, Rao (2009).

Details of the route allocation procedure are provided in a Waller (2020b) and implementation of the route allocation is available as an R shiny app at:

- **Route allocation -** https://danielwaller.shinyapps.io/allocation/

Given that the total imports and exports are known for each product and EU country combination, the route allocation (break down by UK region) takes place separately for each combination. For each of imports and exports there are 270 (10*27) such combinations to consider. The allocation is based upon minimising the distance travelled and can readily be adapted for other metrics such as minimising time and cost. The allocation is based on the UK leg of the journey (**A-X**) as the allocation of trade to the sea and European leg of the journey (**X-B**) is included in the PRB Associates' analysis. (PRB analysis allocates flows across UK ports, determined by sea service capacity – European port allocation estimation is possible but requires further analysis of the allocation model.)

Figure 5: Map of the allocated routes for Imports of Meat & meat preparations (SITC 01) from Austria.

Output from the R shiny app is presented in Figure 5, where there is a map of the allocated routes within the UK for imports of SITC 01 (Meat & meat preparations) from Austria. In 2017, there was a total of 5,201 tonnes of meat & meat preparations imported to the UK from Austria of which it is estimated 4,412 tonnes (84.8%) entered the UK via either the port of Dover or Eurotunnel. The route optimiser shows that goods entering via the Dover straits are shipped throughout the UK whereas other ports generally supply a local region. The R shiny app also allows for exploration of route allocations under changes to port (**X**) usage, *i.e.* changes to the amount of freight flowing on each route (**X-B**). Note that to consider fully the viability of changes to the port of entry, both the UK leg **A-X** and the sea and European leg **X-B** need to be considered.

6.2 Forecasting Food & Drink freight flows

The allocation of Food & Drink data to routes for 2017 gives a baseline model for Food & Drink freight flows. We can combine this baseline allocation with forecasts of product level demands on a regional basis to produce forecasts for the amount of freight flowing through each port. In order to achieve this aim we can apply the forecasting methodology developed in Section 4.

The hierarchical forecasting of trade data in Section 4 was successful at estimating aggregate flows. The highly variable nature of some region, product and country combinations of freight flows meant that lower level forecasts tended to exhibit large amounts of uncertainty. This is particularly pertinent to developing forecasts for Food & Drink freight flows where the data are analysed on a finer (SITC level 2) product level.

We can combine the HMRC forecasts for a given product, UK region (**A**) and country (**B**) combination with the PRB Associates 2017 port (**X**) allocations to produce forecasts for the amount of traffic on a route **A-X-B**. This is done by assuming that the proportion of the traffic for a given product on **A-B** taking a chosen route **A-X-B** will not change over time. This presents a reasonable first-order forecast of freight flow demand but does not take into account shipping capacity and port capacity constraints and other factors which could lead to changes in allocation proportions. Details of the forecasts are provided in Waller (2020a) which outlines the challenges of applying the seasonal ARIMA and ETS models to the finer level product data.

A key component of the forecasting within the LHOFT project is to forecast the total amount of freight flow through UK ports (**X**). By combining the forecasts over products, UK region and country combination, with the current route utilisation rules, we can obtain these forecasts. We observe that, as with the aggregating forecasts produced in Section 4, the aggregation over multiple routes and products produces more reliable estimates of port demand.

The construction of the port level forecasts by combining freight flow forecasts with optimised port allocation enables us to consider alternative forecasts to those based upon current utilisation rules. We can consider how changes to the proportion of freight going through different ports affects the port allocation optimisation along with different metrics (time, money, CO_2 emissions) to optimise over. This leads to different proportions of a product allocated to a given route **A-X-B** which in turn leads to changes in the forecast series for the freight flows through a port **X**. These alternative forecasts can then be viewed either in isolation as presenting a forecast in a given scenario or in combination where an ensemble forecast averaging over proposed changes in freight flow dynamics.

7. Forecasting through the LHOFT Platform

The forecasting analysis which has been possible during the LHOFT project has been limited by an absence of "bottom up", company data. The LHOFT platform offers us the opportunity to rectify this moving forward in two key aspects. Firstly, the collection of suitably anonymised data from companies on their freight flow demands, both import and export, will provide access to this data. Moreover, the LHOFT routing engine will take as inputs company preferences with regards time, money, CO_2, transportation, etcetera in choosing a route. This will enable us to construct estimates that will become increasingly refined for the relative level of importance of each factor for different products.

Secondly, we can utilise the LHOFT platform to estimate for each product the total amount of traffic on a given **A-X-Y-B** route. This can be done by using a more refined UK location (**A**) and foreign country (**B**), for example, by adopting postcode locations. We can estimate the distribution of import/export locations using postcode and other factors such as population and manufacturing density which can be refined in light of data collated from the LHOFT project.

Given that we know from HMRC data the total amount of trade **A-B** for a given product, we can then utilise the LHOFT platform to perform an allocation to routes **A-X-Y-B** as observed in Section 5. We can then continue in a similar manner to the analysis of the Food & Drink data in Section 6 and aggregate the flows over port of entry **X** to the UK. We can present a more refined analysis by shipping route **X-Y** as we will be able to explicitly consider this. We can calibrate the allocations derived using the LHOFT platform by comparing the proposed freight flows through port **X** by cargo category with what is observed in the DfT port statistics. It should however be noted that the LHOFT platform will present the optimal route given the inputs. Therefore discrepancies between the estimated throughput through port **X** and what is observed in the DfT data could be due either to inaccuracies in the input (locations for origin and destination along with preferred constraints on route choice) or the current use of sub-optimal routes or as is likely to be the case combinations of the two.

Acknowledgements

Our grateful thanks to Peter Baker of PRB Associates for providing the analysis of the Food and Drink data and for regular meetings to discuss the data and our analysis.

References

Lowther, A. (2020) LHOFT: Forecasting approaches, comparisons and implementations. (Available on request.)

Hitchcock, F.L. (1941) The distribution of a product from several sources to numerous localities. *MIT Journal of Mathematics and Physics,* **20**, 224–230

HMRC (2018) Regional Trade Statistics Methodology Paper - March 2018. Available at: https://www.uktradeinfo.com/Statistics/OverseasTradeStatistics/AboutOverseastradeStatistics/Pages/PoliciesandMethodologies.aspx

Hyndman, R.J., Ahmed, R.A., Athanasopoulos, G. and Shang, H.L. (2011). Optimal combination forecasts for hierarchical time series. *Computational Statistics and Data Analysis*, **55**, 2579-2589

Hyndman, R.J. and Athanasopoulos, G. (2019). *Forecasting Principles and Practice.* O texts. Available at: https://otexts.com/fpp3/

Hyndman, R.J., Lee, A. and Wang, E. (2016). Fast computation of reconciled forecasts for hierarchical and grouped time series. *Computational Statistics and Data Analysis*, **97**, 16-32.

Lowther, A. (2020) LHOFT: Forecasting approaches, comparisons and implementations. (Available on request.)

Neal, P., Lowther, A. and Eckley, I. (2019) Model specification document – August 2019. WP4.2 Forecasting. (Available on request.)

Rao, S.S. (2009). *Engineering Optimization: Theory and Practice*. (Fourth Edition). John Wiley & Sons.

Waller, D. (2020a). LHOFT – Forecasting for the HMRC data. (Available on request.)

Waller, D. (2020b). LHOFT – Route Allocation. (Available on request.)

Wickramasuriya, S.L., Athanasopoulos, G. and Hyndman, R.J. (2019). Optimal Forecast Reconciliation for Hierarchical and Grouped Time Series Through Trace Minimization. *Journal of the American Statistical Association*, **114**, 804-819.