

# The Global Maritime Container Network: An Application of Conventional Transportation Modelling Techniques

Yanni Huang<sup>1</sup>, Taha Hossein Rashidi<sup>2</sup>, Lauren Gardner<sup>3</sup>

School of Civil & Environmental Engineering, UNSW Australia

<sup>1</sup> yanni.huang@unsw.edu.au, <sup>2</sup> rashidi@unsw.edu.au, <sup>3</sup> l.gardner@unsw.edu.au

## ABSTRACT

Current maritime network models focus primarily on forecasting containerised trade movements between selected groups of countries, but there is a gap in literature regarding worldwide trade prediction at a macroscopic level. With increasing container volumes and evolving economies, there is a need to develop global maritime models in order for ports to remain commercially viable.

This study has developed models, derived from conventional transportation modelling techniques, to understand factors which affect trip generation and trip distribution within the global maritime container network. Two models are proposed, a linear regression model for trip generation, and a gravity model for trip distribution. Container movements for 222 countries in the year 2011 were combined with country level data retrieved from multiple public databases. Analyses of these models focused on assessing accuracy and determining underlying relationships between container trade volumes and explanatory variables.

The results revealed that containerised trade volumes can be significantly represented using trip generation and trip distribution models from the transportation literature. Specifically, socio-economic and demographic indicators that affect import and export containerised trade volumes were identified. It was also found that the perceived impedance between two countries in the maritime container network can be attributed to the distance separating them. The linear regression models captured up to 72% of variation in trade volumes while the gravity model achieved an accuracy of 84%.

These findings support the use of conventional transportation modelling techniques on the global maritime container network. Subsequently, future work can utilise additional years of data to validate and provide a more robust model, enabling the forecast of containerised trade movements on a global scale.

## INTRODUCTION

The world maritime industry has grown eight-fold since 1980 and is said to be the keystone of the international trade network, supporting over 80 per cent of global trade volumes in 2012 (The UNCTAD Secretariat, 2012). Presently, the industry is reported to transport US \$17.7 trillion in commodities (World Trade Organisation, 2013), and the trade competitiveness of all countries depends heavily on effective exploitation of the international port network.

The shipping network is broadly split into three classes of cargo ships, consisting of container vessels, bulk dry carriers and oil tankers. Container vessels are cargo ships that carry all loads in truck-size containers, and have increased in size from 1,000 to 18,000 twenty-foot equivalent units since their introduction in the 1960s. These vessels are now one of the most popular transport modes for trade due to their low cost, high capacity and ease of transshipment (Rodrigue, 2013). However, with increasing vessel size and numbers, the infrastructure of a port (e.g., terminal area, berths, and channel depth) now restricts the capacity to accommodate modern trade. The need to adapt to future trade volumes before they are immediately evident is crucial to a port's commercial viability.

The topic of container movement modelling has been widely approached from economists to traffic forecasters. Millions of dollars are spent each year by ports worldwide to retain a competitive advantage (Elsdon & Burdall, 2004), yet there is a gap in the current literature regarding the complete container network. Moreover, there is a lack of accessible knowledge on quantifying the impact of influential factors. The major consequence of these shortfalls is that some countries do not utilise their full potential to trade, hindering economic development. It is imperative that port authorities and governments are able to accurately adapt port infrastructure and policies to facilitate and benefit from the international trade market, particularly within a growing global economy.

The proposed problem objective is to identify the socioeconomic and demographic factors that significantly influence the volume and pattern of trade in the global maritime container network (GMCN). The network structure currently connects 14,000 ports across 222 countries and territories, transporting manufactures goods, commodities and refrigerated cargo. Using conventional transportation modelling techniques, the attraction, production and distribution of trade can be quantified. Annual worldwide container volumes and travel patterns are required for calibration, and the limited availability of GMCN data is the most significant restriction in the proposed model.

The shipping industry is far less in the public eye than other sectors of the global transport infrastructure, despite current interests in airport, road and train networks (Kaluza, et al., 2010). The main research contribution of this work is the application of transportation techniques on a global maritime network scale. Similar methodology can be extended to develop a forecasting model and to applications beyond transportation networks.

## **LITERATURE REVIEW**

Modelling has long been of interest in the forecasting of trade movement, and spans the fields of macroeconomics, computer science and transportation. A comprehensive review of all such research is beyond the scope of this paper, and the literature review of this section highlights published work directly relevant to the relationships present in the global trade network. This section will also focus specifically on relevant models that have identified significant socioeconomic and demographic factors among containerised trade.

The linear regression model is most prevalent in the field of transportation engineering for the determination of total trips a given zone can generate. In the past ten years, the application of

multiple linear regression has expanded into the maritime industry on a port and regional level (Chou, et al., 2008; Gosasang, et al., 2011). For example, using data on maritime trade volumes, country populations, real final demand and historical gross domestic product figures, Lightfoot, et al. (2009) predicted Australian port activities into 2030. A simplified linear regression model, or linear mixed model, was applied each time the relationship between Australia and a nominated country was to be determined.

This analytical approach is unsurprisingly common for ports in economically active Asian countries, such as Bangkok, South Korea and China, for which there is motivation to forecast trade. Another study revealed that the port of Hong Kong relies on regression analysis to forecast port throughput for its port planning and development. Thirty seven commodity movements are projected separately through the use of explanatory variables such as population, trade values of imports and exports, electricity demand and gross domestic product (Lam, et al., 2004).

Economists tend to focus on general trade relationships, utilising the gravity model of trade. This log-linear equation is popular in the empirical trade literature due to its ability to explain a significant portion of bilateral trade flows, despite the absence of a strong theoretical foundation (Bergstrand, 1984). It traditionally relates bilateral trade to distance between the two trading countries and their respective masses, usually proxied by a measure of wealth (Taplin, 1967). Modern variants of the gravity model of trade incorporate proxies and dummy variables to quantify the effects of natural and artificial trade barriers, such as language, borders and preferential trade agreements (Anderson & Wincoop, 2004). Additional transport cost proxies include effects of infrastructure (Bougheas, et al., 1999), geography (Limao & Venables, 2001), cultural institutions (Sapienza, et al., 2006), and government stability (Marcouiller & Anderson, 2002). In the past ten years, fifty five different gravity models have been published in the field of economics (Anderson, 2011).

Considering the practical value inherent to linear regression models and the extensive research on the gravity model of trade, an intuitive extension of these methods is to the world maritime container trade network as a whole. There is a lack of models on the aggregate level of countries and territories.

A novel analysis of the global cargo shipping network presented by Kaluza, et al. (2010) was the first to incorporate the conventional transportation gravity model in the prediction of global vessel movements. While the model captures broad trends, the results were reportedly too crude for most applications. A similar study by Nuzzolo, et al. (2013) recently employed a partial share approach to simulate production, attraction, distribution, and mode choice for trade amongst ports in Europe. The problem proposed in this paper differs in method from both studies in that the objective is to capture complex trends at the global maritime container network level.

## **PROBLEM DEFINITION**

The objective of this work is to identify the relationship between the volume of containerised trade amongst countries and their socioeconomic and demographic indicators. From this point onwards, the term “trade” will refer exclusively to containerised trade. A trade relation can be

identified as the movement of trade from an origin country to a destination country. The methodology adopts the conventional transportation modelling techniques of trip generation and trip distribution, utilising country attributes and container movement data to infer the relationships present between trading countries. The goal is to accurately calibrate such models, which will aid in the development of port policies, infrastructure investments and government planning. The results may also provide insight into future trade patterns.

## SOLUTION METHODOLOGY

The fundamental nature of the GMCN is one of a transportation system, making it an obvious candidate for the classical transportation model. In this work, the network is aggregated to a national level, with zones defined to be official countries and territories. A trip is defined as a one way interaction between two zones, identified as the movement of a container vessel from an origin zone to a destination zone. The volume or cargo carrying capacity of the vessel, defined as gross tonnage (The International Maritime Organisation, 1970), is assumed to be synonymous with the volume of trade transported. Intrazonal movements are excluded from the models as they are not the focus of this research.

### Trip Generation

The first step of the transportation model is to predict the total number of trips produced by and attracted to each zone. This is achieved through the identification of zone specific indicators that correlate with the variation in the volume of trade generated, presenting a multiple regression problem. The aim is to find a function of  $n$  independent variables,  $X_n$ , that significantly explain the dependent variable,  $Y$ .

Multiple linear regression was implemented on the network of all active trade zones. This model assumes a linear relationship between the dependent variable and chosen predictors, quantified by predictor specific coefficients,  $\beta_n$ . The model also includes a constant term,  $\beta_0$ , and the random error term,  $\varepsilon$ , which captures all remaining variation. The Ordinary Least Squares (OLS) method was implemented to investigate numerous combinations of zonal specific predictors, minimising the sum of squared deviations between observed responses in the dataset and responses predicted by the estimation.

Preliminary analysis of the network revealed that taking the logarithm with base 10 of trade volume allowed for estimation using linear regression. Further, the network displayed different behaviour at different percentiles of trade volume, and was therefore modelled in three trade groups, segregated at the trade volume percentiles which allowed for optimal fit. Attraction trade volumes and production trade volumes segregated at these percentiles produce the same trade groups. Implementing this new model results in an attraction and a production model of the form:

$$Y = \begin{cases} \beta_{g,0} + \sum \beta_{g,n} \cdot X_{g,n} + \varepsilon_g, & 0 < p_z \leq 25 \\ \beta_{g,0} + \sum \beta_{g,n} \cdot X_{g,n} + \varepsilon_g, & 25 < p_z < 80 \\ \beta_{g,0} + \sum \beta_{g,n} \cdot X_{g,n} + \varepsilon_g, & 80 \leq p_z < 100 \end{cases} \quad (1)$$

This formulation introduces the term  $p_z$ , a value between 0 and 100, denoting the percentile of a zone. A zone is grouped by percentile using attraction or production trade volume. Further, the previously introduced terms in the formulation now vary across the trade groups of low, mid and high, denoted by the subscript  $g$ . A significance value of 0.1 was used.

The attraction models aim to predict the total quantity of trade being attracted to each zone, denoted as  $D_j$ , and can be thought of as the zonal demand for trade. The production models aim to predict the total quantity of trade being produced at each zone, denoted by  $O_i$ , and can be thought of as the zonal supply of trade.

### Trip Distribution

The second step of the transportation model is to predict the trip pattern within the network, identifying the volume of trade between each origin and destination pair. This is achieved through the calibration of a model to the observed trip pattern between zones.

The general gravity model used in transportation theory was implemented on the network of all zone pairs. This model assumes that the volume of flow,  $V_{ij}$ , between origin zone  $i$  and destination zone  $j$  is proportional to the volume produced in zone  $i$ ,  $P_i$ , the volume attracted to zone  $j$ ,  $A_j$ , and the distance between the two zones,  $d_{ij}$ . The gravity model form implemented is:

$$V_{ij} = P_i \frac{A_j F_{ij} K_{ij}}{\sum_{j=1}^N A_j F_{ij} K_{ij}} \quad (2)$$

Preliminary analysis revealed an inverse distance provides the best fit. In this work, distance is defined to be the geodesic centroid-to-centroid length between two zones. The friction function adopted,  $F_{ij}$ , represents the impedance for two countries to trade, and took the form:

$$F_{ij} = \frac{1}{d_{ij}^\alpha} \quad (3)$$

Adopting the statistical approach proposed by (Viton, 1994), the optimum value of  $\alpha$  was obtained using an OLS regression of volume against distance. The model also includes a socioeconomic adjustment factor,  $K_{ij}$ , which captures all remaining variation not explained through the friction function. The final model contains nine adjustment factors that cover the nine zone pair categories defined by the trade groups,  $g$ : low to low, low to mid, low to high, mid to low, mid to mid, mid to high, high to low, high to mid, and high to high.

The algorithm for calibration is as follows:

1. Set the adjustment factors,  $K_{ij}$ , to 1 for all zone pairs
2. Calculate the predicted attractions and productions for all zones
3. Calculate the predicted trips for all zone pairs
4. Calculate the ratio of old zone pair trips to new zone pair trips
5. Update all adjustment factors by multiplying the old adjustment factors by the ratio
6. Repeat steps 2 to 5 until all ratios are sufficiently close to 1

7. Group all adjustment factors,  $K_{ij}$ , based on the zone pair categories relevant to the origin zone,  $i$ , and destination zone,  $j$
8. Replace all adjustment factors with the average in each zone pair category
9. Calculate the predicted attractions and productions for all zones
10. Conduct row factoring
11. Conduct column factoring
12. Repeat steps 10 to 11 until all row and column factors are sufficiently close to 1

### **Model Inputs**

The application chosen for analysis is the 2011 GMCN. The data sourced from Lloyd's List include container vessel trips, trip origin, trip destination, and vessel gross tonnage (Lloyd's List Intelligence, 2014).

### **Significant Predictors and Factors**

Significant predictors and factors were identified and incorporated in both models. These variables are listed as follows, with the data source referenced in brackets: coastline length (Central Intelligence Agency, 2013), foreign direct investment inflow (The World Bank, 2014), GDP (The World Bank, 2014), human development index (United Nations, 2014), industry value (Central Intelligence Agency, 2013), labour force (International Labour Organization, 2014), land area (Central Intelligence Agency, 2013), number of airports (Central Intelligence Agency, 2013), number of ports (Lloyd's List Intelligence, 2014), population (Central Intelligence Agency, 2013), population growth (Central Intelligence Agency, 2013), railway length (Central Intelligence Agency, 2013), and geodesic centroid-to-centroid distances between zones calculated by using ArcGIS.

### **Network Structure**

The network analysed in this work was limited to the trade active zones in the calendar year 2011. The network has 222 zones and considers all zone pairs having direct container vessel movements. Only interzonal trips were included in the network.

Container vessel volumes were aggregated to the zonal level to correspond to the zonal-level data set available for predictors. The network was created using the trip data provided by Lloyd's List Intelligence (2014), specifically vessel gross tonnage aggregated across all vessels in the GMCN. Zone-to-zone vessel volumes were calculated by aggregating port-to-port vessel volumes across all ports in a given zone. Where data was missing, the zone or zone pair was removed from the network. The final network for trip generation has 126 non-zero zones, while the trip distribution network has 167 non-zero zones and 3,299 non-zero zone pairs.

If port-level predictor data were available for all ports in a zone, the same methodology could be applied to the disaggregate port-level problem. The port-level model would most likely provide a better tool for tracking container movements across space, useful for port authorities and investment decisions. However, with regulation and profit driven barriers,

data collection on an international level is unrealistic and the model remains presently constrained.

## MEASURE OF PERFORMANCE

The attraction, production and distribution models were computed for the GMCN, resulting in six multiple linear regression models and one gravity model. The complete set of results is not provided, but highlights are discussed in the numerical results and analysis section.

The method of evaluation adopted for the model performances is to measure the proportion of predictions that match the actual pattern that occurred. The following two subsections expand on the measures of performance for the trip generation models and the trip distribution models respectively.

Typically, the calibrated models would then be tested on data from another calendar year, and the performance of the models measured against the corresponding GMCN data set. This would reveal how robust the model is in replicating relationships over time. The models would ideally exhibit the same performances to that measured for 2011, despite different input parameters. Alternatively, using potential discrepancies to adjust the model to the time dependence of trade is a key improvement that has been implemented in current trade models (Lightfoot, et al., 2009). However, due to the lack of data availability, this remains a topic for future research.

### Adjusted Coefficient of Determination

The coefficient of determination, or  $R^2$ , can be interpreted as the percentage of the total variation in the dependent variable that can be explained by a regression model. It is defined as the ratio of explained to total variation, taking a value between 0 (no explanation) and 1 (perfect explanation). However, the addition of any predictors,  $p$ , to a model will inflate the  $R^2$  term. The adjusted  $R^2$  measure accounts for this inflation, and is hence adopted in this work. The measure is calculated as follows:

$$\text{Adjusted } R^2 = 1 - \frac{\sum_{i=1}^n (\hat{Y}_i - Y_i)^2 / (n-p)}{\sum_{i=1}^n (Y_i - \bar{Y})^2 / (n-1)} \quad (4)$$

The formulation requires the number of zones in the total model,  $n$ , the predicted trade volume for origin zone  $i$ ,  $\hat{Y}_i$ , the real trade volume for origin zone  $i$ ,  $Y_i$ , and the mean of the real trade volume,  $\bar{Y}$ .

### Mean Absolute Error Ratio

The mean absolute error ratio (MAER) was chosen to represent the inaccuracy of the trip distribution model, and can be interpreted as the average variation that cannot be explained by the gravity model across the zones calculated. A value of 0 indicates that the model perfectly explains the distribution of trade, while there is no upper bond for the ratio. The mean absolute error ratio is computed as:

$$\text{MAER} = \frac{1}{N^2} \sum_{ij} \frac{|Y_{0,ij} - Y_{ij}|}{Y_{ij}} \quad (5)$$

The formulation requires the number of zone pairs,  $N$ , the original volume of trade from zone  $i$  to zone  $j$ ,  $Y_{0,ij}$ , and the predicted volume of trade from zone  $i$  to zone  $j$ ,  $Y_{ij}$ .

## NUMERICAL RESULTS AND ANALYSIS

### Trip Generation

Interpretation of the final calibrated coefficients from trip generation provides insight into the role of each predictor and the predictive capability of the model. The attraction and production model coefficients are similar in magnitude and possess the same sign, so the results presented have been limited to the attraction model only. The final calibrated coefficients can be found in Table 2 of the Appendix, and the model performance is illustrated in Figure 1.

The results have varying levels of success, with adjusted  $R^2$  values of 0.415 to 0.724. Despite this, the models reveal the multifaceted nature of the WMCN, identifying 11 predictors that significantly contribute to trade attraction and production. Further, the different combinations of predictors across trade groups indicate that at different levels of trade competitiveness, different predictors distinguish zone success in the international market. At different levels of trade, significant zone predictors change. This is in line with previous economic studies where countries of interest are segregated by measures of wealth before modelling (Taplin, 1967).

In the Low trade group, airport numbers have a negative impact on trade volumes, suggesting that airplanes and vessels are competing modes for such countries. The length of coastlines has a negative impact on trade, but this is indicative of a more complex relationship. It could be speculated that zones with smaller coastlines invest more in infrastructure at one port, as opposed to investing that same amount across many smaller ports, allowing a competitive trade advantage. Finally, GDP is observed to have a positive impact with trade volumes, suggesting that countries that trade more correlate with countries that produce more, in line with common literature findings (Barigozzi, et al., 2010).

In the Mid trade group, airport numbers have a positive relationship with trade. This suggests a shift in the role of airports between Low and Mid trade zones. Air travel is an indicator of travel expenditure in this case, or luxury expenses, correlating to a higher potential to trade (World Trade Organisation, 2013). While GDP is also observed to have a positive relationship, the increase in trade associated with a higher GDP is lower than in Low trade zones, indicating a diminishing effect. Foreign direct investment inflows are also associated with higher trade. Conversely, zones with larger industries tend to trade less, suggesting trade demand in such zones is preferentially met locally rather than internationally. Railway track length is similarly associated with lower trade, suggesting that in the Mid trade group, maritime transport competes against railway modes. Interestingly, as the number of ports



increases, trade decreases. Reflecting on the negative effect of coastline length in Low trade zones, this may indicate a similar interaction.

It is worth noting that the production model for the Mid trade group lacks foreign direct investment inflows as a significant indicator. It can be concluded that the investment inflow from other zones only increases demand for trade in this instance, but not production.

For the High trade group, the United Nation's Human Development Index has a positive influence on container trade, while GDP does not. From this, it is evident that economic measurements of development are not sufficient to capture the trade variation in the upper percentile zones. Factors taken into account by the development index, such as life expectancy and standard of living, are also significant (United Nations, 2014). Population density and the labour force size similarly increase with volume of trade, acting as indicators of trade potential. The negative effect of coastline length can again be addressed with the same argument for the Low trade group.

The constant term of the Low trade group is insignificant, with p-value higher than 0.1. However, the constant terms of the Mid and High trade groups are significant. This suggests that the latter two models may be unsuitable for forecasting, as the variation in the trade explained is not being explained entirely through the predictors. If some variation is not being captured, and the constant terms are not constant for different calendar years, the models have unfavourable predictive power. Another year of data is required to assess this possibility. However, as mentioned previously, lack of data availability restricts this assessment for potential future research opportunities.

### Trip Distribution

Interpretation of the final calibrated socioeconomic adjustment factors from trip distribution provides insight into how appropriate the selected friction function is and the relationships between all trade groups. The calibrated socioeconomic adjustment factors, which capture variation unexplained by distance, are provided in Table 1. The gravity model output is illustrated in Figure 2, found in the Appendix.

**Table 1: Calibrated Socioeconomic Adjustment Factors Classified by percentile**

$K_{ij}$ Factors	Destination		
Origin	Low	Mid	High
Low	14.30	0.6752	1.921
Mid	1.706	0.6001	0.9442
High	1.320	1.786	3.549

The model is successful in capturing 84.01% of the 2011 trade pattern, supporting the selection of distance as an indicator of trade impedance. Zone pairs that are further apart do not trade as much as zones pairs that are close together and this is especially evident in Figure 2, with high trade volumes concentrated within the European and Asian zone pairs.

Considering the calibrated socioeconomic adjustment factors, it can be seen that distance is reasonably successful for most trade pairs, and not much variation is required to be explained by the  $K_{ij}$  factors.

For trade originating from the Low trade group, distance grossly underestimates trade to other Low trade zones, underestimates the trade to High trade zones, and overestimates the trade to Mid trade zones.

For trade originating from the Mid trade group, distance underestimates the trade to low trade zones, overestimates the trade to other Mid trade zones, and is reasonably accurate in estimating trade to High trade zones.

For trade originating from the High trade group, distance underestimates the trade to Low and Mid trade zones, and noticeably underestimates the trade to other High trade zones.

The reasons for these adjustment factors are unclear. From the economics literature, underestimations may arise for several reasons. Island nations, countries with the same ethnicity, countries that speak the same language and countries in a trade agreement can be expected to trade more (Sapienza, et al., 2006). Overestimation may similarly arise from the inverse of these circumstances. However, especially with trade agreements, the resulting relationship is often assessable on a case by case basis and requires more detailed data than available (Mansfield & Milner, 2010). More complex interactions, coupled with a better understanding of the GMCN are beyond the scope of this work, and remain for future research.

## **CONCLUSION, CONTRIBUTIONS, CRITICISM, AND FUTURE RESEARCH**

An understanding of the global maritime container network is of increasing importance as infrastructure stresses and international market opportunities continue to rise. The proposed modelling tool is intended to identify significant socioeconomic and demographic relationships within the network and to establish global transport modelling efforts within the maritime network context.

Two models were introduced, generation and distribution, to recreate the trade patterns observed in the 2011 global maritime container network. Multiple predictors were considered, and the estimations were compared with the patterns observed. The modelling results were also assessed intuitively from literature findings.

The accuracy of the generation models was examined by calculating the adjusted coefficient of determination and the distribution model was examined by calculating the mean absolute error ratio, revealing that 42-84% of trade variation can be explained through conventional transportation techniques. Significant predictors were found to include coastline length, GDP, airport numbers, and geodesic distances between zones, among others. The percentile groupings based on zonal trade levels allowed for further network exploration. The segregated generation model revealed that at different levels of trade, different predictors are responsible for trade competitiveness. The distribution model supports the use of distance as

a proxy for impedance to trade, but also captured more complex relationships in the socioeconomic adjustment factors.

The presented work is more importantly a modelling framework, which has the potential to be expanded in the future as more detailed data becomes available. The novelty of the model lies in the use of conventional transportation modelling techniques on the global maritime container network to infer trade patterns, which can aid governments and port authorities in the development of policies and decision making. The major weakness of the proposed methodology is the lack of verifiability due to limited data availability. Without another data set year, the predictive power of the model remains uncertain. In addition, the necessary high level of aggregation resulted in the sacrifice of a more detailed model and analysis.

One extension of the model is to disaggregate the problem to the port-level. Aggregation of some sort is inevitable when modelling networks due to both the available level of data and the desirability of simple measures to complex problems. The level of detail lost can be quantified by establishing another model on the port-to-port level, and a comparison made to reach an acceptable balance between simplicity and detail. Moreover, the port-level model could potentially improve accuracy though capturing complex interactions the current models do not, but implementation is solely dependent on port-level data, which are currently unavailable.

Another planned extension of the model is to obtain an additional data set year to validate the forecasting potential of the model and to develop a model more robust to the fluctuating nature of the shipping industry. Accounting for the time dependence of container trade is an obvious area of improvement for this work, and should significantly increase the predictive capabilities of the models. These areas will be explored in future research.

**REFERENCES**

- Anderson, J. E., 2011. The Gravity Model. *Annual Review of Economics*, Volume 3, pp. 133-160.
- Anderson, J. E. & Wincoop, E. v., 2004. Trade Costs. *Journal of Economic Literature*, 42(3), pp. 691-751.
- Barigozzi, M., Fagiolo, G. & Garlaschelli, D., 2010. Multinetwork of International Trade: A Commodity-Specific Analysis.
- Bergstrand, J. H., 1984. The Gravity Equation in International Trade: Some Microeconomic Foundations and Empirical Evidence.
- Bougheas, S., Demetriades, P. O. & Morgenroth, L. E., 1999. Infrastructure, transport costs and trade. Volume 47.
- Central Intelligence Agency, 2013. *The World Factbook*. [Online] Available at: <https://www.cia.gov/library/publications/the-world-factbook/index.html> [Accessed 26 10 2014].
- Chou, C.-C., Chu, C.-W. & Liang, G.-S., 2008. A modified regression model for forecasting the volumes of Taiwan's import containers. *Mathematical and Computer Modelling*, 47(1), pp. 797-807.
- Elsdon, P. & Burdall, T., 2004. *Container Terminal Planning and Design*, s.l.: Ports 2001.
- Gosasang, V., Chandraprakaikul, W. & Kiattisin, S., 2011. A Comparison of Traditional and Neural Networks Forecasting Techniques for Container Throughput at Bangkok Port. *The Asian Journal of Shipping and Logistics*, 27(3), pp. 463-482.
- International Labour Organization, 2014. *Compilation of statistics*. [Online] Available at: <http://www.ilo.org/stat/Areasofwork/Compilation/lang--en/index.htm> [Accessed 26 10 2014].
- Kaluza, P., Kolzsch, A., Gastner, M. T. & Blasius, B., 2010. The Complex Network of Global Cargo Ship Movements. Volume 7.
- Lam, H. W., Ng, L. P., Seabrooke, W. & Hui, C. E., 2004. Forecasts and Reliability Analysis of Port Cargo Throughput in Hong Kong. Volume 133.
- Lightfoot, A., Bolin, R. & Lubulwa, G., 2009. *Australian Sea Port Activity to 2029–30*, s.l.: Bureau of Infrastructure, Transport and Regional Economics.
- Limao, N. & Venables, A. J., 2001. Infrastructure, Geographical Disadvantage, Transport Costs, and Trade. 15(3).

Lloyd's List Intelligence, 2014. *About Lloyd's List Intelligence*. [Online] Available at: <http://www.lloydslistintelligence.com/llint/about-us.htm> [Accessed 27 10 2014].

Mansfield, E. D. & Milner, H. V., 2010. Regime Type, Veto Points, and Preferential Trading Arrangements. *Stanford Journal of International Law*, Volume 46, p. 219.

Marcouiller, D. & Anderson, J. E., 2002. Insecurity and the Pattern of Trade: An Empirical Investigation. *Review of Economics and statistics*, 84(2), pp. 342-352.

Nuzzolo, A., Crisalli, U. & Comi, A., 2013. An Aggregate Transport Demand Model for Import and Export Flow Simulation.

Rodrigue, J.-P., 2013. *The Geography of Transport Systems*. Third Edition ed. New York: Hofstra University.

Sapienza, P., Zingales, L. & Sapienza, P., 2006. Does Culture Affect Economic Outcomes?. *Journal of Economic Perspectives*, 20(2), pp. 23-48.

Taplin, G. B., 1967. Models of World Trade. 14(3).

The International Maritime Organisation, 1970. *International Conference on Tonnage Measurement of Ships, 1969*. London, The International Maritime Organisation.

The UNCTAD Secretariat, 2012. *Review of Maritime Transport*, Geneva: United Nations Publication.

The World Bank, 2014. *Data Overview*. [Online] Available at: <http://data.worldbank.org/about/data-overview> [Accessed 26 10 2014].

United Nations, 2014. *Frequently Asked Questions - Human Development Index (HDI)*. [Online] Available at: <http://hdr.undp.org/en/faq-page/human-development-index-hdi#t292n42> [Accessed 26 10 2014].

Viton, P. A., 1994. *Calibrating the Gravity Model, Working Paper*. Columbus: Department of City and Regional Planning, The Ohio State University.

World Trade Organisation, 2013. *World Trade Report 2013*, s.l.: World Trade Organisation.

APPENDIX

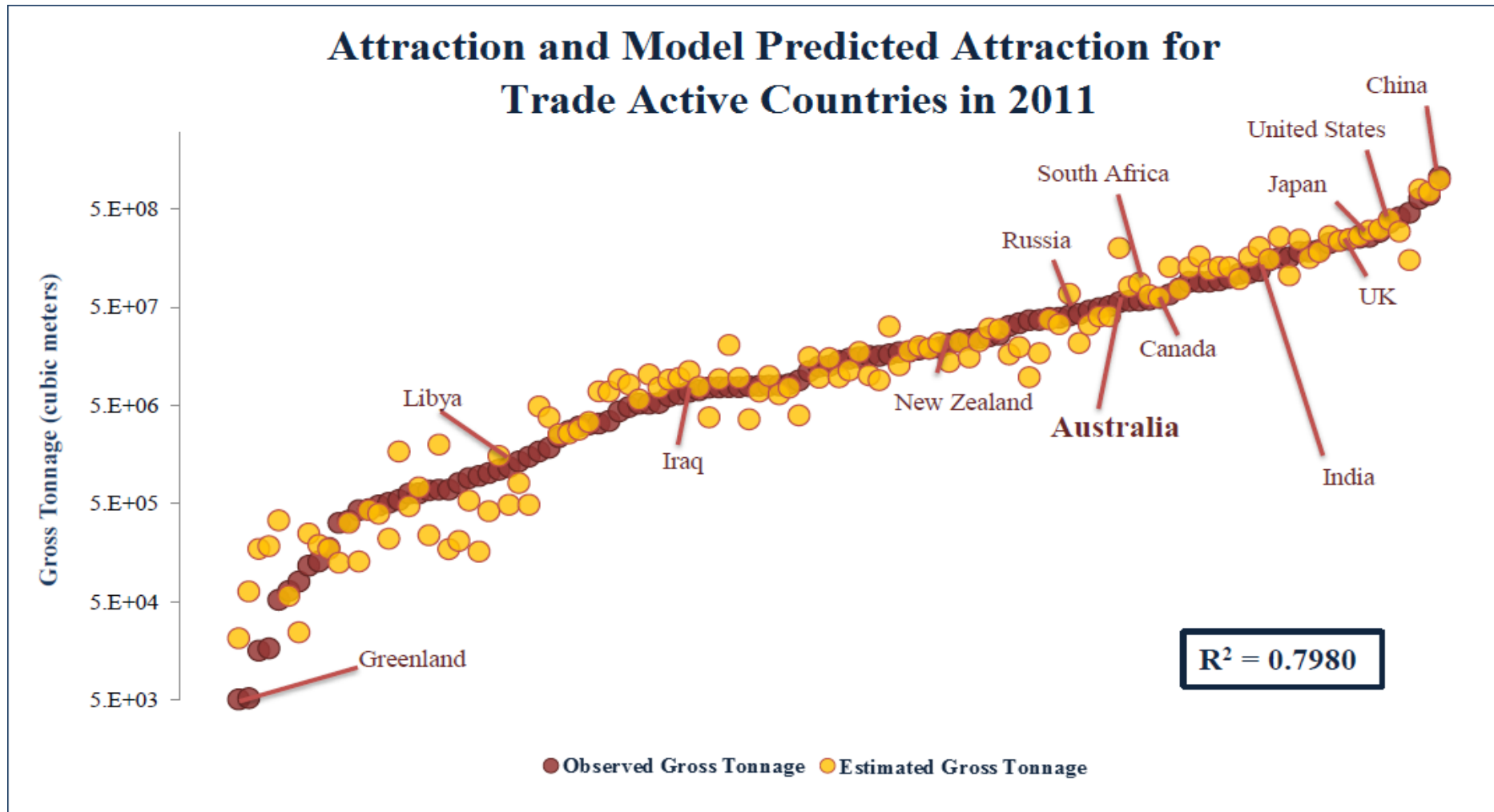


Figure 1: Comparison of Observed Attraction and Model Predicted Attraction

Table 2: OLS Regression Attraction Model

Predictor (unit)	Description	Coefficient	p-value	Adjusted R <sup>2</sup>
<b>Quantiles 0-25</b>				
<b>Constant Term</b>	-	0.059806	0.972	0.415
<b>Airports</b>	The number of airports or airfields	-0.002469	0.004	
<b>logGDP (current USD)</b>	The logarithm with base 10 of GDP	0.597118	0.003	
<b>Coastline (m)</b>	The total length of coastline, including islands	-0.04091	0.003	
<b>Quantiles 25-80</b>				
<b>Constant Term</b>	-	2.182059	0.002	0.665
<b>Airports</b>	The number of airports or airfields	0.000424	0.015	
<b>Industry (% GDP)</b>	The percentage contribution of industry to GDP	-0.011478	0.000	
<b>logGDP (current USD)</b>	The logarithm with base 10 of Gross Domestic Product	0.511820	0.000	
<b>logRailways (km)</b>	The logarithm with base 10 of route length of railway networks	-0.281402	0.001	
<b>logFDI (million USD)*</b>	The logarithm with base 10 of foreign direct investment inflows	0.197236	0.036	
<b>logPorts</b>	The logarithm with base 10 of the number of ports	-0.184178	0.006	
<b>PopulationGrowth (%)</b>	The average annual percent change in population	0.082972	0.015	
<b>Quantiles 80-100</b>				
<b>Constant Term</b>	-	6.752115	0.000	0.724
<b>Coastline (m)</b>	The total length of coastline, including islands	-0.00316	0.002	
<b>HDI (2011)</b>	Human development index as defined by the United Nations	0.017647	0.000	
<b>LabourForce (ppl)</b>	The number of residents available for work	0.00125	0.000	
<b>Population/Land (ppl/km<sup>2</sup>)</b>	Number of residents per square kilometre of land area	0.0681	0.002	

\* Predictor was absent in Production Model

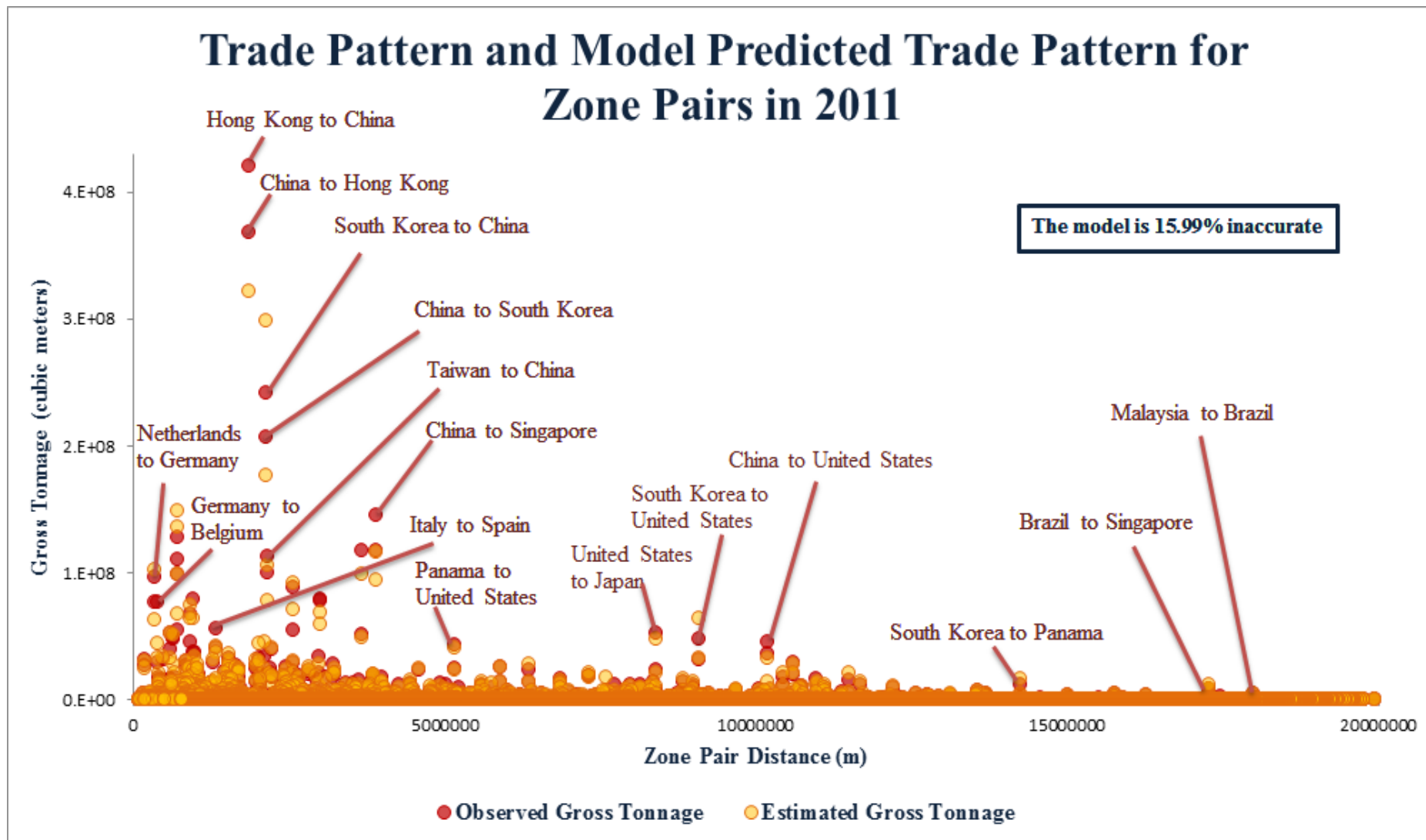


Figure 2: Comparison of Observed Trade Pattern and Model Predicted Trade Pattern