Erasmus University Rotterdam

MSc in Maritime Economics and Logistics

2013/2014

The effects of socio-economic factors on container volume developments

by

Jente Waal LLB BSc

Acknowledgements

In the few weeks that this thesis had to be written, the dazzled gaze in my eyes may have shocked many on the campus for several weeks. The human mind is highly fallible and tunnel vision is luring. Fortunately, I had good support who protected me against this happening.

I would like to thank dr. Peter de Langen for his feedback and assistance during the writing of this thesis. His comments made me reconsider choices and further analyse elements in the paper which constituted a true academic process.

My thanks furthermore go to the MEL Office, in special Ms Veronique Lefel who was able to give me crucial support in the completion of my thesis.

Abstract

In this exploratory research, the development of the container volumes is set in the context of socio-economic factors of influence. The development of the container industry is one of the many remarkable stories in maritime transportation. Nevertheless, research into the macro-economic impact and development of the container volumes is only recently upcoming.

This research contributes to the understanding of the use of the container by looking into the relation between the container volumes in a country and socio-economic factors such as average personal income and urbanisation. Using data from the United Nations Conference on Trade and Development, World Bank and the World Economic Forum, this research will answer the research question: What is the relation between a country's economic development and the number of twenty-foot equivalent units handled per thousand inhabitants? The container volumes are corrected for the population size by regarding them per thousand inhabitants. The research question is answered using five hypotheses regarding the influence of the average income, both individually as opposed to consumption developments, the differences in container volume developments in different income groups, the effect of urbanisation and the trade intensity in a country.

The data is analysed using a three-step approach. First, an analysis of the data of each of the countries in the sample is given using ordinary least squares regression estimation. This analysis reveals the influence of the variable(s) that is/are most explanatory for the development in the container volumes. Subsequently, a cross-country analysis will regard the determinants of differences in container volumes between countries. A panel data analysis will regard the trend development of the container volumes for the whole sample to determine in how far the developments are similar. Using fixed and random effect modelling, the models are tested against the assumptions of the Gauss-Markov Theorem.

It is found that the average personal income is highly explanatory for both changes in the level of container volumes and trend developments. Lower-income countries (with an average income between 10.000-20.000 USD) show significantly higher growth rates than other income groups. The same holds for higher-income countries (with an average income between 30.000-40.000 USD), although the trend growth is much less than the lower-income countries. Differences in container volumes among countries can also be explained by differences in the average personal income. Low-income countries (average income < 30.000 USD) have a significantly higher level of container volumes than high-income countries.

The impact of an increase in consumption or average personal income is amplified by the growth in the degree of urbanisation or trade intensity of a country. An increase in the total consumption, combined with an increase in urbanisation results in a larger increase in container volumes than when the level of urbanisation remained constant.

This research contributes to the understanding of container volume developments in a more macro-economic context, using socio-economic variables and time-series data. Further research should focus on the development of larger datasets.

Table of Contents

List of figures

List of tables

1 Introduction

1.1 Preamble

The content of a ship can be divided into four main categories: liquid and dry bulk, people, general cargo and containers. 60% of the traded goods are transported by container. (World Shipping Council, 2014) Nevertheless, this trade seems to be relatively concentrated in certain regions, as becomes visible from the collage of print screens from Marinetraffic.com entailing all container ships' positions at a point in time (the yellow triangles in Figure 1). The European and Eurasian regions are well-visited by container ships, while Latin America has only got some activity in Brazil. The north of Africa seems well-attended, but does not show that most container ships only pass Africa on their way to Asia or Europe. For the rest, only South Africa seems to have some attendance.

Figure 1 Container ships per selected continent (marinetraffic.com)

These images signal that the role of containerised trade per economy can be different. Where some economies prosper with the use of the container in international trade, others rely more on regional or general cargo shipping. In this paper I investigate the relation between the size of the container volumes corrected for the population size and socio-economic factors describing the economy. I take a sample of countries from regions displayed in Figure 1 that are least influenced by transhipment and of which relatively safely can be assumed that the container trade represents the use of these containers in the national economy. This thesis will try to explain which characteristics of a country's economy determine the amount of containers handled per thousand citizens.

1.2 Aim of the research

The development of containerised trade in a country can depend on a large number of factors due to its broad application. Ranging from fruits to shoes and hardware for computers, most of the products today can be transported in containers. That does not necessarily mean that the container is actually used. For carriers to use the container in short-sea shipping, the use of the container requires substantial infrastructural investments as well of reforms of customs procedures and increased efficiency on the handling locations. (Yang et al., 2014) These types of alterations to the system are not performed equally easy in all countries. More developed countries or countries in which the economic risk is smaller can possibly invest in their port sector more easily than countries that have a more vulnerable economy or state of the nation.

In this thesis, the aim will not lay in identifying the main factors of influence on the adaptation of containerised transport or which changes must be made in the transport systems to promote the use of containerised transport in a country, but rather on the economic conditions under which the use of the container changes. Where many researchers focus on the influence of certain policies or investments on the use of containerised trade, this thesis tries to identify determinants of container volume developments in a broader perspective. Moreover, the thesis will identify common and country-specific influences that affect the level of containerised trade. Therefore, the research question is:

> What is the relation between a country's economic development and the number of twenty-foot equivalent units handled per thousand inhabitants?

This research question identifies the underlying motives for changing to containerised trade as observed in the country's economies for both import and export containers. With economic development, the changes in the wealth of the nation, consumption and foreign direct investments and demographic factors such as population and urbanisation are included in the analysis. The use of the container is measured in twenty-foot equivalent units (TEU) handled, both for exports and imports, per capita in that country. In contrast to the number of containers, TEU measures the volume of the shipped amounts. A container can be even two TEU and therefore such a measure would not relate to the handled volume. Most importantly, the data availability for TEU is better than for containers.

Where the development of the economy can be expressed in mainly financial or production terms, I also take into account social and demographical factors to be able to review the changes in the economic development of the country. These factors might have an effect on the preferences with regard to the use of transport systems or the demand for foreign goods. Therefore, I include population, employment and urbanisation variables to provide on a very general level for the necessary socio-economic indicators.

Using five hypotheses regarding both economic and socio-demographic factors in a country, I will try to give an explorative answer to the formulated research question. All hypotheses regard the effect of economic and socio-demographic factors on the number of containers per capita. The hypotheses are described in paragraph 2.4. Nevertheless, it must be noted that the hypotheses only regard a limited number of factors.

For answering the research question, I will use three statistical approaches. The first regards a country-based analysis of the economic factor(s) that explain(s) the variation in the use of the container per thousand capita the most. The set of variables reflects which independent variable or parameter relates to the containerised trade developments the best. The influence of these factors can be country-based or can have a broader application because more countries share the same dominant driver for container handling growth. The cross-country analysis will further regard why some countries have certain drivers in common and others have alternative drivers. This substantiates the indicated commonalities of the country-based analysis.

Moreover, I will do a panel data analysis using fixed effects to capture both common effects and country-based effects for the variables that turned out to be of major influence in the country-based analysis. Further, the data reveal large variances in levels of average personal income, the use of the container, export and import per country. Therefore, I will do a cross-country analysis of the data to identify significant variances between countries. The countries in the sample are divided into groups to test for differences. Where the cross-country analysis will confirm or reject differences between these groups, the panel data analysis will test whether these groups have existed over time. Moreover, combined with the cross-country analysis, the panel data analysis will provide proof of the relations or will contradict such.

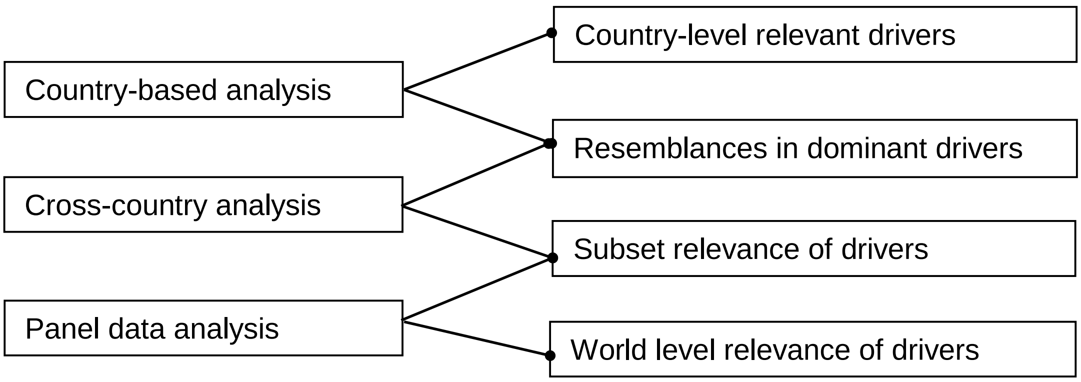


Figure 2 Coherence between analysis techniques and their results

Figure 2 displays how the analysis techniques of this paper are used to add new information to the research and to check the analysis of the previous methods used. Combining these methods will give an inclusive and reliable picture of the effect of the variables on all levels. Where the results of the country-based analysis result in drivers that are of relevance for container trade developments in that country and indicate relevant drivers for common trends, the cross-country analysis will substantiate these common indicators but will also indicate which drivers are important for all countries in the sample. Moreover, the cross-country analysis regards the division of the sample in subsets, i.e. groups based on a division criterion. The panel data analysis subsequently confirms and proves these relations or rejects them and uses the sample for extrapolation to the world population of countries with containerised transportation and sea access.

The common effects of certain variables on containerised trade provided for in the panel data analysis will give an estimate of the worldwide trend in container transport development as the results of the sample are extrapolated to the population of container handling countries. Indeed, the sample will not be random. On the one hand I use certain criteria as the transhipment factor to select the sample of countries, on the other hand the sample was selected aiming for a geographical representation of all regions. Therefore, extrapolation to the whole population should be done with reticence.

The research focuses on a country-based and a cross-country analysis of time series data of several renowned statistic providers such as UNCTAD and the World Bank.

Most research in the field of containerised transportation seems to relate to efficient handling, ship deployment, scheduling and other operational phenomena. Also, the organisation of shippers and the policy regarding price-setting and service development has been research thoroughly. Many scholars have contributed greatly to the knowledge on containerised transport by regarding regional or policy-specific effects and changes. Nevertheless, it seems that the broader picture of the use of the container in the economy and world development of containerised transport still has not been regarded much in detail. Recently, some researchers have started to investigate the development of containerised trade in a more macro-economic context (for example Rua, 2012 and Bernhofen et al., 2013). This thesis will contribute to this new branch of research in containerised trade, but will still be explorative of nature due to the lack of preceding research in this field.

1.3 Setup of this thesis

In Chapter 2, I give a background of the development of the advantages and disadvantages of the use of containers in transportation and the implementation of containerised trade in the transport systems. Moreover, an outline of the theory underlying this thesis will be given.

Chapter 3 describes the data sources and the availability of the data that is used in this paper. It will further give a description of the main tendencies and developments for countries and years visible in the data and possible correlations of the variables.

The three analytical methods deployed in this thesis are further described in Chapter 4. Moreover, this chapter gives an overview of the statistical techniques that are used for testing the hypothesis and doing the analysis.

In paragraph 5.1 I will do the country-based analysis of all countries in the sample and I will give a brief analysis and comparison of the acquired result. In paragraph 5.2,**Error! Reference source not found.** I will elaborate on these results by doing he cross-country analysis of the data.

The results of the panel data analysis are described in paragraph 5.3. Concluding, the main cross-country and country-based results are described in the subsequent chapter, Chapter 6. At last, the method and results are discussed in Chapter 7.

2 Literature review

In this Chapter, I will first consider the basic developments with regard to the introduction of the container in international trade. Therefore, the advantages and disadvantages of the container technology are compared to traditional handling and transport systems in paragraph 2.1. This paragraph will also regard the influence of containerisation on the transport system and the way containerisation has been developing in the economies. Paragraph 2.2 regards how containerisation has had effect on international trade, the competition between countries and how trade in goods in containers has influenced the configuration of the local economy. Paragraph 2.4 introduces the theory on which this thesis will be vested and will be tested, using research publications and the literature review.

2.1 Containerisation

Movement of goods in a container is common today, but was only adopted half a century ago. Where containerisation started with only six countries in the West, nowadays only three economically insignificant countries with access to sea do not have a container port. The pace of development and implementation of containerised transportation therefore has been striking.

Although containerised transportation was already invented in 1956, with the first container ship 'Ideal X' sailing to Houston (Texas, U.S.A.), it was only in 1966 that the first nations started to adopt this new phenomenon in logistics. Nevertheless, containerisation had a profound impact on international trade for the rest of the twentieth century according to many authors (among which Bernhofen et al, 2013; Rua, 2012).

The effects of containerisation on international trade can be subdivided in social, economic and political effects. Although this is not a strict separation –some economic effects have had social side effects or vice versa- this distinction can give a better characterisation of the nature of these effects.

Economically, the container enabled intermodal transportation to become more efficient. Initially, the discharging and loading of the vehicles deployed for transportation could take tremendous effort, involve large costs and long duration, the containerisation allowed for much faster handling. Where at first ships laid in port about two-third of the time, they now often have a turnaround time of only twenty-four hours. The easiness of the handling was also improved by the standardisation of the dimensions of the container which was agreed upon in 1967. The dimensions of the container were set, though in some instances firms still continued their own designs. (Rua, 2012)

The main advantage of the container is not found in the sea transportation, but rather in the handling of the container on land i.e. intermodal handling. Most advantages of intermodal transportation arise from the faster handling of the container and the standardised dimensions (Bernhofen et al, 2013; Rua, 2012; Hummels, 2007). Nevertheless, the cost reductions on the land side might lose their effect on containerisation because of rising prices on the waterside (Hummels, 2007). Hummels though argues that these gains on the landside might even be so large, that they could overcompensate for the rise in container transport prices on the sea lag.

Another economical effect of containerisation was on the reduction of insurance costs for the cargo. Due to pilferage, a lot of the cargo was usually lost during the voyage. Talley (2000) also originates a part of the loss of certain pleasurable or valuable cargo to the long voyage and the needs of the crew on board. The storage in a metal box with a seal on it during the voyage reduced these losses significantly and therefore the costs associated with it. (Rua, 2012) This has also realised cost reductions for transportation of goods.

With the easier handling at intermodal points, such as a port, the labour that was required before for handling the goods became quickly abundant. To illustrate, ECT reports that conventional discharge of cargo could be done with 1,25 tonnes per man per worked hour. This figure is already quite astonishing if one starts thinking of how much the port labourer should bear each time to be able to move 1.250 kg per hour. Nowadays though, ECT is able to move 125 tonnes per man per hour, using automation based on the standardised dimensions and features of the container. (Bottema, 2014) Where carriers thus had to rely on port labourers and were consequently in the grasp of the port labour unions, this dependency decreased significantly after the introduction of the container. (Bernhofen et al., 2013; Rua, 2012)

For carriers, containerisation thus offered new chances in efficiency gains and cost reductions. Whether or not these gains were also shifted to the consignees, is the logical follow-up question. Transportation costs for the movement of a container overseas are determined by the fuel costs, shipping distance and the weight/value ratio of the shipment. The fact that the goods are shipped by container seemed to have a decreasing effect on the trade costs, though this evidence is diluted by the oil price developments. (Hummels, 2007) Moreover, as I will elaborate on in paragraph 2.2, liner conferences have set prices on the long-run average costs level. (Haralambides, 2007) This might have an effect on the potential for container transportation in a country, because the cost advantages will be less.

These oil price developments were of a highly political nature due to the unrest in the seventies in the Middle Eastern world. Other political events also had their effect on the development of containerisation. Rua (2012) argues that the involvement of African countries in containerisation was accelerated by the closure of the Suez Canal. Although this had a negative effect on the development of containerisation for the West-Asian route, the continuation of this trade lane was only possible via the Cape of Good Hope under South Africa. The temporary shift in the trade route incentivised many African countries to invest also in container ports. Other political events of influence on containerisation were the Vietnam war and the Cold War, giving rise to a high demand for containerised transport by the military.

Using the container in transportation has also some disadvantages. As Konings & Thijs (2001) prove, there is a great trade imbalance on certain trade routes. Therefore, empty containers are stacked up in the demand locations and are an economic burden to move back to a point of supply. One on every five containers transported is empty. These reversed logistics may place a high burden for countries that are starting to use the container, because the fixed costs involved in returning the container add up to the fixed costs for the other container infrastructure.

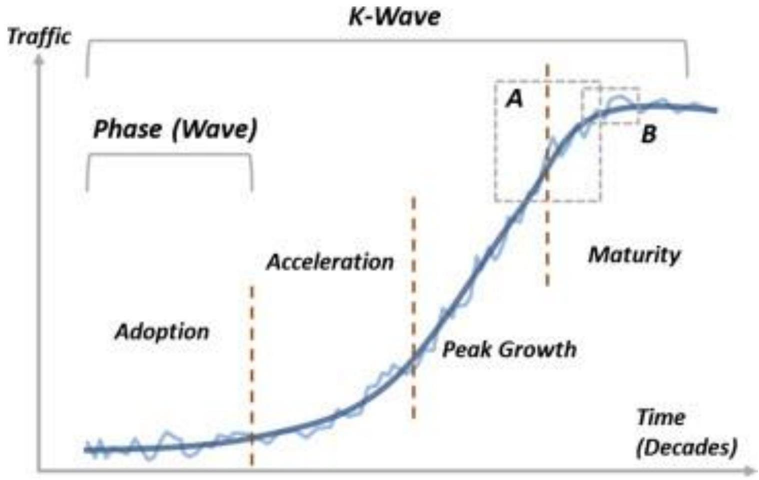


Figure 3 K-wave development of containerisation (Guerrero & Rodrigue, 2014)

The size of the effects though have differed over time and per nation. Guerrero & Rodrigue (2014) prove that the development of containerisation follows a Kondratieff wave, i.e. an s-curve development of four phases as shown in Figure 3, taking about 45-60 years. According to them, in the adaptation phase, the first countries to adopt containerised transportation showed to be able to have more efficient port operations. Its success was confirmed in the acceleration phase, where the availability of containerised transportation rose significantly due to the development of networks. When growth entered the peak growth phase the creation of container transhipment hubs resulted in even rising growth figures. In this phase, the ability to concentrate operations via a hub-and-spoke network enabled shipping companies to have economies of scale in operations. Combined with the high demand instigated in the acceleration phase, these gains were significant. These days, the tremendous growth figures seem to stabilise because the gains from concentration have become less in certain regions. Other economic factors such as the economy's growth rate, are becoming increasingly important for the development of container transport demand. Guerrero & Rodrigue (2014) though stress that this does not account for all regions, since containerised transportation still has large growth perspectives in certain developing regions.

Guerrero & Rodrigue's (2014) publication introduces the research into the development of containerisation in the world trade arena. The research provides evidence for both the network effect that is also described by Rua (2012) as well as the effects of increased efficiency and economies of scale. This research though regards both transhipment and domestic container handling processes. In contrast, this thesis tries to correct for the effect of transhipment by regarding countries in which the container volumes are limitedly affected by transhipment. Where the approach of Guerrero & Rodrigue was feasible for their research, for this research it is required to correct for transhipment to reveal the relation between domestic variables and domestic container volumes. The research has helped to great extend in understanding the opportunities for economies arising from containerisation and how this has influenced world trade through time. This thesis will elaborate on this, by researching which factors determine the success of the container trade given socio-economic circumstances.

Similar to the development of the containerised world trade, Rua (2012) proves that the adoption of container infrastructure in ports follows an s-curve. Adoption is the moment on which countries first allow containers into their ports. Where the use of container transport follows a more linear trend, the adoption of infrastructure is much steeper and quicker. This indicates a diffusion of containerisation since the rise in availability in infrastructure for container handling in ports rises much quicker than the actual use of containers in transportation. She finds that the expected usage of containers, institutions, the size of the country in both physical and economical perspective and trade with Australia and the United Kingdom are of significant influence on the adoption of container infrastructure. This research has provided insight in the main determinants for adaptation of container infrastructure. This thesis will provide additional insight in the determinants for the development of the sector in the country, therefore regarding the phase after adaption.

Rua's (2012) findings seem to contradict the findings of Guerrero & Rodrigue (2014). Where they describe the bundling of trade flows and the development of the hub-and-spoke network, Rua states that diffusion has increased following an S-curve in the last few decades. Nevertheless, the two can be combined. Rua ignores the size of the trade, but regards only the moment of adoption. It is true that container trade has diffused in the last decades. More countries have been included in the container trade network. Nevertheless, the actual size of the trade flows has concentrated on several trade routes and certain hub locations.

Although in recent years with the attempts of Rua, Guerrero and Rodrigue, a more general overview of the nature of the container market has become available, most research in container transport is still done on ship routing and handling efficiency. Economists have thus provided insights in the development of the international trade and globalisation of production processes, which can be described as a demand factor for containerised transportation. Nevertheless, there exists a large research gap in combining these fields with regard to containerised trade. The contribution of this thesis will be that it dissects the main incentives for both developed and developing countries from multiple regions to increase container volumes and transcend from general cargo transportation to containerised transportation.

2.2 International trade and competition

Countries compete on multiple levels on welfare gains, often translated in economic terms to the income per capita. It appears that this competition is induced by a relatively higher average personal income in one country opposed to another, which forms a threat to the latter because it is deprived from its possessions. This competition practically translates in competition on innovation, of which infrastructural innovation can be a part. (Chaudhry & Garner, 2006) For containerisation, countries can increase their comparative advantage by implementing container transport infrastructure and technology in their country. As mentioned in paragraph 2.1, this can not only lower transport prices and improve the competitiveness of the local producers, but can also reduce consumption prices for the population.

Midoro et al. (2005) show that liner shippers focus strategically on mergers and acquisitions, where between ports competition is more on operational efficiency. This results in powerful liner conferences that can put high demands on ports. The market position is one reason for carriers to engage in alliances and conferences.(Haralambides, 2007)

The container shipping market is characterised by high fixed costs due to the investments required in infrastructure and large ships with which the economies of scale the containerised transport offers can be fully used. Market competition would disable investors to recover their initial investments because fixed costs cannot be recovered under perfect market conditions. Therefore, liner conferences were founded in which price setting was the main objective. (Haralambides, 2007) To be able to acquire the cost and efficiency advantages that are involved with container shipping, it is required that conditions for perfect markets are breached by using cartel-like organisations as liner shipping conferences. In a perfect market, the price of the products is determined by the marginal costs of supply. This marginal cost of supply encompasses production and distribution costs, but in the theory of Haralambides (2007) no fixed costs. Both the power of the conferences and the price setting mechanism can decrease the benefits for economies to change traditional transportation into container transportation.

This effect might be even more pronounced for countries of small size both in geographical or economical terms. (Rua, 2012) As Lim (1998) touches upon, there is an abnormal difference in prices for freight shipment per teu per mile between different parts of the world. For Northern Africa, the freight rate is 28.6 pence per teu/mile, where it is only 5.5 pence per teu/mile to the Far East. It is a result from the price setting in conferences of liner shippers, which puts developing regions apparently in a worse position. Because these areas are not the main source of their demand, one could imagine that there are price differences. It could also reflect the lack of power of the small amount of customers since there is so little demand for transport on connections to these economies. Therefore, this paper regards the influence of the size of the domestic consumption on the role of container trade in the economy.

2.3 Economic development and the role of trade

Jacoby & Minten (2009) researched on which households in a developing country the construction of infrastructure to a central market place would have the largest effect. They show that the most remote households gain the most when transport costs of goods are reduced. Therefore, there might be a strong link between the levels of urbanisation and container transport development in this research.

Borensztein et al. (1998) find that foreign direct investments have a positive effect on economic growth via the transferral of technologies. This effect is enhanced if there is a sufficient amount of human capital in a country. Enabling foreign private investors to engage in projects in the country without capital and trade barriers that hamper the profitability of the projects will therefore increase the country’s wealth in the long run. I must though note that this statement is based on the definition of wealth based on a nation’s income, in which no note is taken on how this wealth is spread in society and which externalities or disadvantageous effects foreign investments may have.

The link between the country’s income and the role trade plays in its economy becomes apparent from Erwin & Terviö (2002). They describe that the more countries engage in international trade the higher the real income of this country will be. Therefore, each country has an interest in engaging in international trade to realise welfare growth. Container transportation could help by giving the domestic firms a better position on the international market, and by lowering domestic consumer prices through the competition of local suppliers with their international counterparts.

This paper will therefore also regard the influence of the size of the value of import and export relative to the size of the economy.

2.4 Analytical framework

The subject that is under attention in this thesis is the amount of containers that is handled per capita in nations that have not developed large transhipment operations in their economy. As mentioned before, the benefit of the container lies mainly in the efficient handling in supply chain nodes and the reduction of pilferage. Because of these advantages, one could assume that all suitable trade flows are imported and exported in containers. This, though, appears not to be the case. There are thus factors that restrict the introduction and growth of the container technology for certain goods or economies.

The demand for container transport is based on the underlying demand for goods that are transported. The demand for transport is derived demand and thus container transport is too. So what demand does it serve? From what is this type of transport demand derived? In this paragraph I determine which major factors of influence on the demand of transport of containers have been proven in the literature and how this can be used to construct a theoretical model for the development of containerised trade, apart from country-specific developments.

Verny (2007) states that in general, transport demand is related to the development of the economy. As an indicator of economic development, for the European Union it holds that GDP has a greater influence on transport demand than industrial productivity. On the other hand, internationalisation of production has given rise to transportation because of the movement of semi-finished products. The globalisation nevertheless also caused a concentration of production and distribution. The growth of transport demand in certain places could therefore partly be explained by the decrease in demand in more peripheral places.

As Helpman et al. (2003) have described, multinationals can choose to serve markets by local production or exporting their products to that market from their production site. In the first case, the values of the imports for the country will rise, where in the second case the FDI will rise. Companies may decide to produce locally if transportation costs are high resulting in more local investments. Contrastingly, if transport costs are low, concentration of production can result in economies of scale but subsequently also more export and import of goods. Helpman et al. (2003) describe this as the proximity-concentration paradox. Therefore, economies with a large production capacity are expected to be responsible for a large part in the variance in container transportation via the exports of their products to other markets. With regard to countries, Melitz (2003) shows that increased openness of the economy has a positive impact on the welfare of a country. Increased wealth has a positive effect on consumption and openness on trade. Combined, these factors can have a positive effect on the container volumes because of the economies of scale that can be gained. Therefore, I expect a positive relation between the average personal income (gdppc) and the number of containers that is transported (teupc):

> Hypothesis 1: The average income per capita has a positive influence on the number of handled containers per capita.

One could suggest that imports and exports than would better explain the rise in personal income and therefore in the number of containers that is transported. But even when exports fall, the openness of the economy might increase because more goods are imported and vice versa. For openness, it is not required that both exports and imports have a positive relation with the average personal income. Subsequently, the relationship between exports and teupc and imports and teupc may be ambiguous. Therefore, gdppc is a better indicator. This hypothesis will be tested both on a country-based level, as with cross-country and panel data analysis.

According to Rua (2012) the usage of containers in transportation is determined by the fixed costs involved in the implementation and the number of other countries that have implemented container transport technology. Moreover, exporting products increases 'melt away costs', i.e. costs of theft and pilferage, and transport costs relative to selling it domestically. The melt-away costs for container transport are lower than for traditional cargo handling because the goods are transported in a metal box. Nevertheless, for small countries the fixed costs of infrastructure are higher for container transport than for conventional cargo handling. Although the variable costs, i.e. the handling costs, and melt-away costs, are lower for container transport than for conventional cargo handling, the benefit does not weigh up against the difference in fixed costs that is initially very high because of the low transported volumes. Once the change to container trade can be made, the development goes theoretically quick because of the decrease in unit fixed costs. Another effect of containerisation is that it becomes more profitable to adopt when more countries adopt the technology. Transportation becomes easier and trade flows can be bundled. Therefore, the opportunity costs for conventional cargo handling increase over time.

Does this hold in practice? As Haralambides (2007) argues, liner shipping conferences make that the actual price of container shipping is dependent on the long-run average costs of these companies. These contain both marginal costs (container handling costs) and fixed costs (initial investment costs in ships, for instance). For the first container to be shipped, not only the long-run average costs for the liner service must be paid, but also the initial investment costs for creating the network to allow for containerised transport in the country. As indicated by Rua (2012), the economic size of the country has an effect on the adaptation of container infrastructure. For the shift to container transport to become feasible for transporters, the size of the domestic consumption might be of influence. Therefore, I think that the total domestic consumption will have a positive and significant effect on the development of containerised trade in a country rather than the rise in national income, of which a part is not used for consumption. Thus,

> Hypothesis 2: Consumption is a better explanatory variable for changes in the number of containers per capita in a nation than the income per head of population.

The effect of consumption on teupc might be influenced by exchange rate variability and international demand factors. More demand elsewhere will increase prices and reduce the power purchasing parity of the population, although their consumption has not changed. This though also affects the average personal income. Moreover, not all of the income will be spend on consumption of goods, but also partly on the consumption of services and savings. I will correct for this by using only the value of the consumed goods for determining the domestic consumption. Hypothesis 2 will be tested in the country-based, cross-country and panel data analysis.

There might be another relation between the gross domestic product and the size of the containerised trade in a nation. As Hummels (2009) points out, market power leads to increased transport prices for developing nations and therefore has an effect on trade flows. The number of shipping lines that serve a specific region or country determines the relative higher prices of these lines to other routes. Developing countries are paying higher transport prices than developed nations. I estimate that the relative higher prices will have a dampening effect on the development of the container transport industry in these nations, because less suppliers will be interested in switching from traditional handling to container transportation. Moreover, Rua (2012) finds that the size of the economy is of influence on the adaptation of container transport technology. Thus, in addition to hypothesis 1:

> Hypothesis 3: The positive effect of a change in the average personal income on the number of handled containers per thousand inhabitants is larger for high-income countries than for low-income countries.

The income of a country is measured in terms of the average income per capita. Hummels (2009) says that there is a difference in exports and imports between low- and high-income countries due to price elasticities. According to me, this is not what is expected. As Hummels (2009) already indicates, the nature of the exports and imports is different in developing nations from developed nations, because the demand for the traded goods is more inelastic in developing nations. Where developed nations are thus more responsive to container transport pricing, developing nations are much less. This would show in the relation between the average personal income and the number of containers handled per thousand inhabitants, but not in the values of the imports and exports. The goods that are exported and imported are different between developing and developed nations. The average personal income does not suffer from this divergence.

This hypothesis will be tested using dummy variables of several income groups. It will be tested both with cross-country analysis as well as with panel data analysis.

As Jacoby & Minten (2009) find, households in more remote areas that are more difficult to reach or to supply with goods have higher benefits from cost reductions relative to areas that are easy to reach. Moreover, they find that where the reduction of transport costs indeed increases the wealth of the remote households significantly, the reduction in the good prices does have a much less sizeable effect. Transport prices thus have a significant impact on the wealth of rural countries. A rise in consumption will therefore render more benefits for the development of the container transport infrastructure for rural societies than for urban societies. Given a certain increase in consumption, the more rural a society is, the more incentive there is to reduce transportation costs. To resemble the remoteness of the population within a country, I will use the degree of urbanisation as a proxy. Resulting, the hypothesis is that:

> Hypothesis 4: The increase in consumption has a larger effect on the number of containers handled per thousand capita for more rural countries than for more urban countries.

In the context of this paper, the hypothesis is relevant for determining the effect of container transport on economic development. Containerisation could then be viewed not only as a virtue for developed economies (as Airriess (1989) seems to think), but also for developing economies because it grows according to the economies of scale that can be obtained and the gains in welfare that it will cause.

The hypothesis is tested by regarding the impact of urbanisation on the change in the number of containers handled per thousand capita in the country-based analysis as well as the panel data analysis. Moreover, the combined effect of consumption and changes in urbanisation is tested for in the panel data analysis to see whether increasing urbanisation over time has an amplifying effect on the container per thousand capita ratio, therefore only testing the combined effect.

The container has proven to be advantageous for many countries in engaging in international trade by increasing efficiency and reducing transport costs. (see for example Rua, 2012; Bernhofen et al., 2013) Reduced transport costs enable distributors and sellers to buy products outside their country against comparable or lower prices than domestic producers, which can translate in a price advantage for the customer. Also, domestic producers can compete on the world market by offering their products abroad, without the transport costs having a tremendously disadvantageous effect on their competitive position. It is thus in the interest of domestic producers and consumers to reduce the transportation costs. This might form an extra incentive for containerisation. The size of this interest might relate to the share of the production and consumption that is sold or acquired from abroad, since a larger share implicates that the profits from a reduction in transport costs will be larger. On the other hand, it must be kept in mind that a reduction in transport costs also has a positive effect on the competitive position of domestic producers and on the domestic prices according to trade theories (Van Marrewijk, 2012). Therefore, a reduction in transport costs will possibly incentivise more trade. The fifth hypothesis is:

> Hypothesis 5: The share of imports and exports on the total domestic product has a positive effect on the number of containers handled per thousand capita.

The share of imports and exports on the total domestic product, mathematically expressed as:

$$TIP_i = \frac{imports_i + exports_i}{gdp_i}$$

for a country i, is mostly called the trade intensity parameter (TIP). (Lane Davis, 2007) Not seldom it is referred to as the ‘openness to trade indicator’ (e.g. Department for Business and Innovation, 2013), but it has been argued that this parameter is not a good measure for the openness of trade because it does not take into account the size of the economy and domestic production. (Rodriguez, 2000) Nevertheless, it gives a good indication of the trade intensity of the country and, according to me, indicates well the interests of a country involved in the reduction of transport prices. (Lane David, 2007)

3 Data

3.1 Data sources and availability

Data on container transport and handling are only limitedly available in databases. As becomes apparent from Table 1, the number of data points available for each dataset differs tremendously among datasets and countries. Where for instance there are thirty data points for the U.K. regarding the number of twenty-foot equivalent units handled, scraping data from other databases ultimately results in only six data points for Senegal. Therefore, in the regression it is urgent to take note that the number of data pairs, i.e. combinations in a time series for which all variables in the regression have data, is based on the minimum number of data points in the variable series for a country. The less data pairs that are included in the analysis, the weaker the resulting conclusions should be.

Table 1 N per source per country

N per variable per country country→ variable ↓	Argentina	Australia	Brazil	France	India	Japan	Kenya	Korea	Mexico	Norway	Senegal	South Africa	United Kingdom
Consumption	54	54	54	44	54	43	53	54	54	44	54	54	44
Container transport	9	9	9	9	9	9	6	9	9	6	6	9	9
Exports	34	34	34	34	34	34	33	34	34	34	33	34	34
Imports	34	34	34	34	34	34	33	34	34	34	33	34	34
GDP	33	33	33	33	33	33	33	33	33	33	33	33	33
FDI	42	42	43	43	42	36	43	43	43	40	36	33	42
GCI	8	8	8	8	8	8	8	8	8	8	7	8	8
Population*	65	65	65	65	65	65	65	65	65	65	65	65	65
ITF Containers	0	17	0	24	22	13	0	20	28	10	0	0	30
Unemployment	33	33	27	33	0	33	0	33	23	33	0	19	29
Labour force	41	41	41	41	41	41	41	41	41	41	41	41	41

Transhipment can have an obscure effect to the results given in this research because they are influenced by other factors than domestic factors. If a country has for instance geographically interesting factors that make it feasible for transhipment, these results would show an increase in import and export of containers where this is not caused by increase of the used economic variables. Moreover, if a country imports its goods via another country, any increase in consumption would show in the import and export of that other country. Therefore, I have taken into account the possible transhipment function of ports in that country in the selection of the sample.

Another selection criterion was the geographical spread of the sample. To give a good representation of the development of the world development of container transportation for domestic demand, all regions should be included. Nevertheless, because of its protectionist laws with regard to container ships and sailing the Northern American region is not represented in the sample. For all other regions, i.e.

Latin America, Europe, Australasia and Africa, both northern and southern countries have been included.

The United Nations Conference on Trade and Development (UNCTAD) provided basic data on economic indicators for all countries. Usually, the available data range is between 1980 and 2012. The database contains the country's GDP, export and import in millions of 2005 US dollars at 2005 exchange rates, population in thousands of people, foreign direct investments (FDIs), the percentage of people unemployed and the size of the labour force. Although the latter two are not directly related to the hypotheses formulated in paragraph 2.4, these variables function as control parameters for changes in consumption. It might be that the change in consumption is actually caused by changes in the actual income of inhabitants rather than a change in living standards.

The data for the number of inhabitants of a country originate from the UNCTAD database. All values are provided in thousands of inhabitants. The time series provided for each country are consistent from 1950 onwards. There are no missing values. The UNCTAD database even provides demographic predictions up to 2050. Nevertheless, only a limited amount of this information is used because of the availability of other data sources.

The variable 'Consumption' is retrieved from the World Bank data centre. It describes total expenditures on consumption of goods in a country from 1960 up to 2013 given in 2005 USD and reflects both private and governmental consumption.

The variable 'Container transport' is retrieved from the International Transport Forum, as part of the World Trade Organisation, combined with data from the Organisation for Economic Corporation and Development (OECD). Both the OECD and ITF provide data about the number of TEU that are handled in a country. Unfortunately, the number of data points is limited per country. Where most countries have nine data points, some have only six. Although the database of OECD provides more data points for a number of countries, it does not include all countries in our selection. Therefore, I use the OECD data where possible, but use the ITF data if the OECD does not provide any data. The number of containers could be expressed in real containers, but because the container comes in different sizes (high, low, forty foot, twenty foot, etc.), there might be country-based influences for the container type that is used. One corrects for this by using the number of TEU that is handled. Moreover, up to now most data is recorded only in TEU. Would I use other measures, than the number of data pairs would even more decrease. Adding information from more than two sources could lead to differences in the data due to registration differences. Moreover, the use of national databases to add information is mostly not possible because of language barriers.

The variables Foreign Direct Investments (FDI), Gross Domestic Product (GDP), Imports and Exports are all provided by UNCTAD. The values are given in 2005 US Dollars and exchange rates, to disregard inflation rates and exchange rates effects. All data have a time range of thirty-two years, going from 1980 to 2012.

The data for Urbanisation are given as a percentage of the total population in a country. They are retrieved from the World Bank, who in their turn acquired them from the United Nations. These data range from 1960 to 2013 and include many data points. For each country in the sample, the data series for urbanisation between 1980 and 2012 is complete.

The unemployment rate is given as a percentage of the total labour population and the labour force in number of inhabitants. The data are retrieved from the International Labour Organisation (ILO) and are an estimate of the national governments. The ILO notes that the definition of unemployment for this database is based on the number of people without a job and looking for one, but that this definition differs per country. For Kenya, India and Senegal the ILO does not provide any unemployment rates. Moreover, for the U.K, Mexico, Brazil and South Africa, the data provided between 1980 and 2012 are incomplete. These data series are added to reveal any spurious relations for Consumption with other variables.

The Global Competitive Index (GCI) provides data for the competitiveness of each country in world trade. It is calculated by the World Economic Forum (WEF) taking into account twelve categories of institutional, economical and policy factors related to the production and development of a country. (WEF, 2014) The indexation of the value makes that the score can be compared across countries. Although the data availability is relatively low (we can only get data from 2006 onwards for most countries), the series reflect a relatively stable score for each of the countries in the population as becomes visible in Figure 4. Although we see that there are quite some differences between countries, the change of GCI over time seems to be insignificant. For Senegal, we only have data from 2007 onwards. Moreover, when looking more closely one can distinct two bands of scores with in the upper layer Korea, Norway, Australia, the UK and Japan. In the lower layer, South Africa, Senegal, Mexico, Kenya, India, Brazil and Argentina are situated.

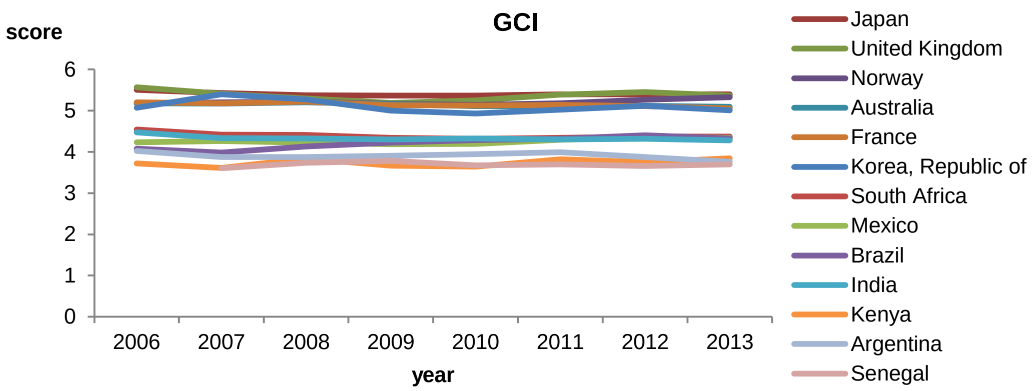


Figure 4 GCI scores for selected countries (own calculations, based on WEF-data)

The dataset contains panel data of a sample of thirteen selected countries over a time period of 33 years (from 1980 to 2012), though not all regressors (variables) are available at all time periods. In the words of Carter Hill et al. (2008), the panel dataset is ‘long’, because the number of regressors is smaller than the number of time periods is observed (r<t). Moreover, the dataset is small, because the number of regressors is limited. There are eleven core regressors, which are used to deduce or calculate other regressors (such as the income per capita or the number of containers per capita) or create dummy variables.

For the analysis, I use a significance level of 5% (α=0,050). With the number of data points provided, a significance level of one per cent would be too strict. Moreover, the intervals for test statistics are usually calculated on the basis of a 5%-alpha. A higher significance level would include too many variables and would allow for too many errors in the data.

Table 2 summarizes the data sources, the data types acquired from it and the description of the measures. The data are collected from a minimal number of sources to reduce possible mismatches (such as between Exports and Imports due to different measurement methods). An exemption is made for containers, which is acquired over two sources for aforementioned reasons.

Table 2 Data, descriptions and sources

Variable name	Description	Source
Containers	Number of containers handled in the country in TEU per year	ITF/OECD
Population	Number of inhabitants in a country per year, in thousand inhabitants.	UNCTAD
GDP	Gross domestic product of a country per year in 2005 USD (constant exchange rates, in millions)	UNCTAD
Imports	Value of the goods imported per year, reported in 2005 USD (constant exchange rates, in millions)	UNCTAD
Exports	Value of the goods exported per year, reported in 2005 USD (constant exchange rates, in millions)	UNCTAD
FDI	Foreign direct investments, reported in 2005 USD (constant exchange rates)	UNCTAD
Consumption	Total private and governmental consumption per year, reported in 2005 USD.	World Bank
GCI	Global Competitiveness Index, scale from 1-7	WEF
Urbanisation	Percentage of the population living in urbanised areas per year	World Bank
Unemployment	Percentage of the labour force looking, but not having a job, varying interpretation	ILO, retrieved via World Bank
Labourf	Number of people that are accounted to the labour population in a country	World Bank

3.2 Cases

The selection of countries in this thesis is done such that the share of transhipment container handling, the part of the container handling where a container is discharged from the ship but later loaded onto another for further travelling to the destination, is of the least influence to the analysis. This selection is based on its geographical position (distance to main trade routes) and the trade with surrounding nations. The criterion has some disadvantages. One of the greatest is that most of the countries for which large data sets are available about container handling are transhipment countries. For example, the Netherlands and Belgium offer relatively long time series. Implicitly, for origin/destination countries, in general less data are available.

Moreover, I have included the United Kingdom and Australia not only because they mostly function as a origin or destination in the container trade, but also because the trade with these countries has been highlighted by Rua (2012) as a major determinant of adaptation of container technology for other countries.

Another determinant for choosing the countries for which the analysis is performed, is their regional representation. Some publications only regard trends in containerisation for certain regions, but this would make extrapolation to other countries and regions impossible. (See for example Airries, 1989 and Beresford et al, 2012) The Middle East is not represented, as well as Northern America. Nevertheless, three countries from Africa, Latin America, Asia and Europe are selected. Australia will be included in the Australasian region.

Only countries with containerised transportation facilities and access to sea are selected. Although containers can be transported via land, this would involve other factors such as the distance to the sea to have an effect on the results, for which other countries do not have these effects.

Argentina, Australia, Brazil, France, India, Japan, Kenya, the Republic of Korea, Mexico, Norway, Senegal, South Africa and the United Kingdom are included in the analysis. These countries show a relatively acceptable availability of data. Moreover, with three countries in Latin America, three in Asia, three in Europe, one in Australia and three in Africa, the dataset gives a good overview over different world regions.

3.3 Describing the data

In this paragraph I will elaborate on trends that are visible in the data comparing the countries.

First, I will regard the dependent variable. The development of container handling per country over time, corrected for its population size is reflected in Figure 5. The number of TEU per thousand capita increases for all countries over time, although for some significantly more than for others. As it seems, the growth in handled TEU for Korea grows with about twenty TEU per thousand inhabitants per year. The runner-up, Australia, has only half of that annual growth. Japan and the United Kingdom follow a similar trend as well as Norway, which gives rise to thoughts that the categorization of countries into income categories might be of significant influence on their container handling development. Although the rest of the countries seem to follow also a slow growth in teupc, the use of the container seems to be minor compared to the UK, Norway and other countries. Nevertheless, container handling still seems to be a developing business, as revealed by the steady growth in all countries.

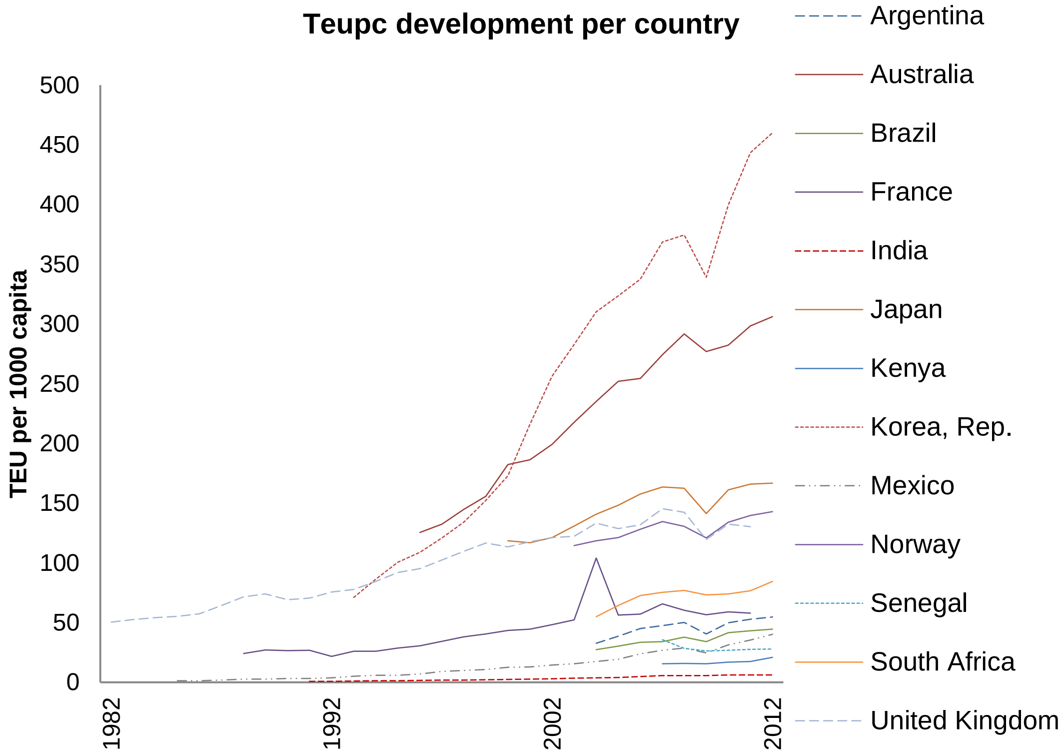


Figure 5 teupc developments per country (own calculations)

The tremendous Korean growth can only partly be explained by transhipment. For Korea, the transhipment factor was about 9% in 2013, which hardly explains the trend visible in Figure 5. (JIFFA, 2013) An awkward point is the value for teupc in France in 2004. Suddenly, an unexplainable spike appears in its time series. I redo the calculation for this year with the original data, but the same value keeps returning. It seems to be an outlier, which will be regarded further in the country-based analysis.

There are thus apparently two components in the development, or rather, two questions to be asked when regarding a country. The first question regards the initial level of teupc. Has the country already had a relatively large teupc factor? The second question regards the growth. Given the teupc level, how quickly does the teupc rate develop over time? To divide the country, I separate the countries in three groups giving all possible combinations of these factors. Given the linear regression model:

$$teupc_i = c_i + b_i YEAR + c$$

The coefficient c_i reflects the baseline level, the number of containers handled in the first year of the dataset. This coefficient can be either large or small. Subsequently, the year-to-year growth b_i can be large (b_i>0, showing that there is significant growth of the factor), or relatively constant (where I assume b_i to be null). The countries can therefore be grouped as in Table 3.

Table 3 Groups in teupc development

Group	Model description	Countries
1	$B_1>0$, C is large	Korea, Australia
2	$B_1=0$, C is large	Japan, United Kingdom, Norway
3	$B_1=0$, C is small	Argentina, Brazil, France, India, Kenya, Mexico, Senegal, South Africa

Concluding, it seems that both Korea and Australia have high growth rates in teupc and a high level of handled containers per thousand capitain their economy, relative to the size of the population. The rest of the countries seem to follow a similar trend, but The United Kingdom, Norway and Japan seem to have a higher level than the rest of the countries in the sample.

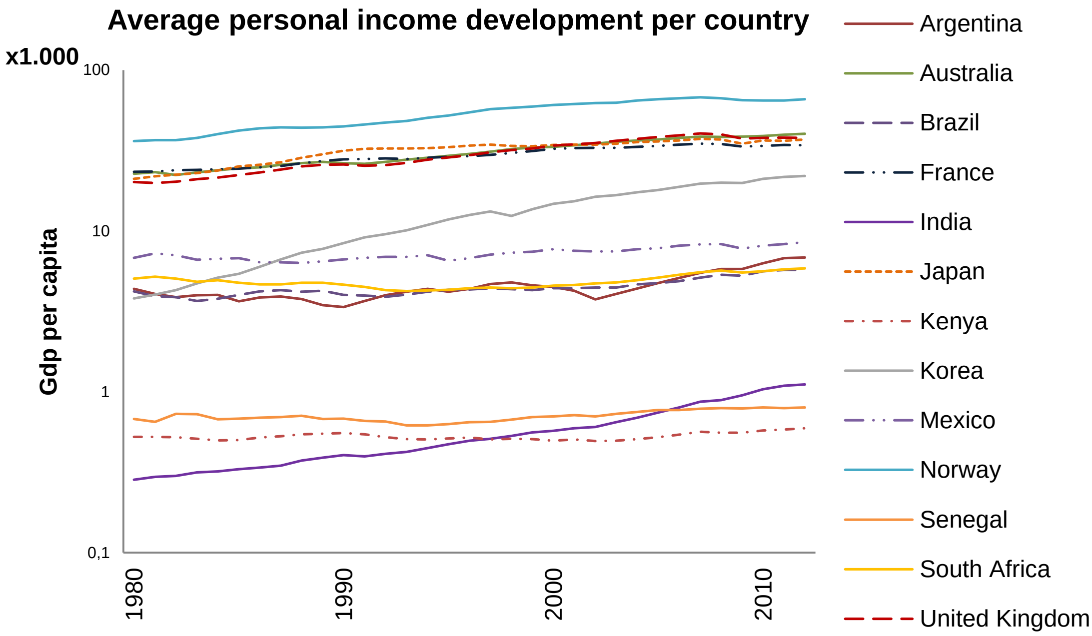


Figure 6 Average personal income developments per country (own calculations)

From Figure 6, it seems that there are three strata of personal income. The first stratum, containing Norway, Japan, France, Australia and the United Kingdom, have an average income per capita of between 34.000 and 65.000 dollars. The second stratum, in which Mexico, South Africa, Brazil and Argentina are located, has an average income of between five and nine thousand USD. The outcast in this regard is Korea, which seems to move from stratum 2 to stratum 1 during this period. Lastly, India, Kenya and Senegal are located in the third stratum with an average personal income less than 1.000 USD, although India describes a quite steep and positive trend. Concluding, the strata could be defined as follows:

Table 4 teupc and gdppc groups/strata

↓group (n) strata→	1 (5)	2 (5)	3 (3)
1 (2)	(1) Australia	(1) Korea	
2 (3)	(3) Japan, UK, Norway		
3 (8)	(1) France	(4) Mexico, South Africa, Brazil, Argentina	(3) India, Kenya, Senegal

Combining the deductions from Figure 5 and Figure 7 I make a table of all combinations of groups and strata with between brackets the corresponding count (Table 4), reflecting the possible relations between countries in terms of teupc growth and level, and gdppc developments. One could say that low-income countries have smaller chances of either high teupc rates or high teupc development than high-income countries.

With regard to the urbanisation rates, Figure 7 shows that all countries have seen increasing urbanisation rates over the years in the data set. The gap between the data for 1980 (arrow) and 2012 (cross) are the largest in the case of the Asian countries, as it seems. African nations still have the lowest urbanisation rates. In the light of Jacoby & Minten (2009), the largest benefits for containerisation would be expected to be found there. On the contrary, compared to

Table 4 all countries in group 3, strata 3 (i.e. low development, low average income, low teupc levels) also have seemingly significantly lower urbanisation rates. To be concrete, India, Kenya and Senegal have all an urbanisation rate lower than 50%.

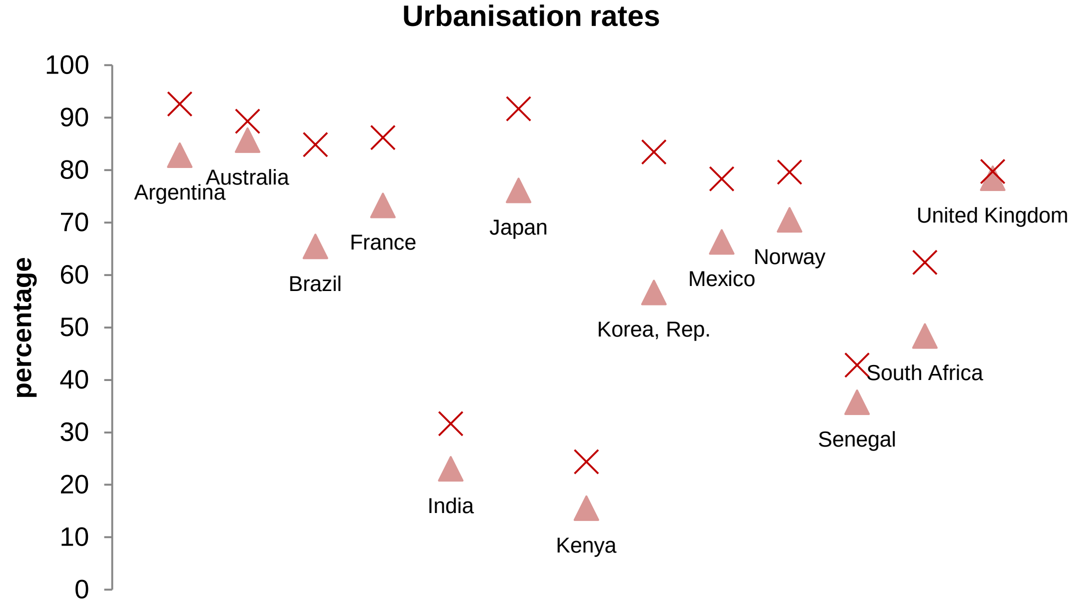


Figure 7 Urbanisation rates (own calculations)

As Figure 8 shows, the aforementioned stratification is also visible in the consumption per capita rate, but such can be deemed logical because consumption is a part of the total personal income of a person. Nevertheless, compared to Figure 6, it appears that the variation is larger in Consumption between countries in the same stratum. The shift from the second stratum to the first of Korea is less apparent than in Figure 6. Where for most countries the average personal income remains relatively stable over the regarded period, it is slightly increasing for the personal consumption. Therefore, I expect that there is multicollinearity between gdppc and Consumption, but that the variation in gdppc cannot fully explain the variation in Consumption.

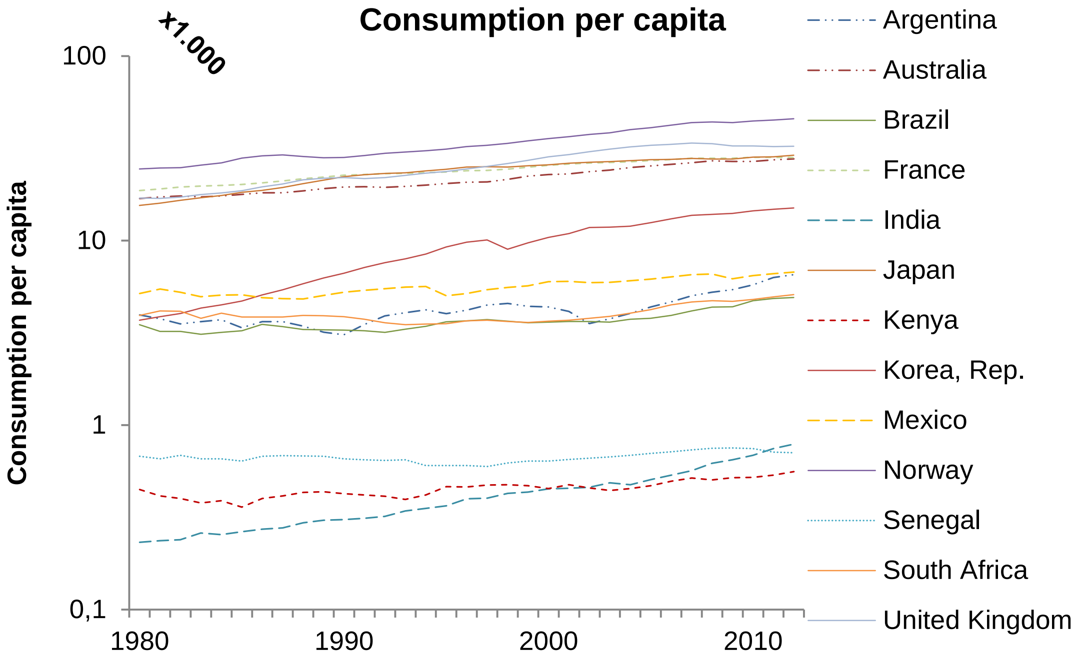


Figure 8 Consumption per capita per country (own calculations, UNCTAD data)

Figure 9

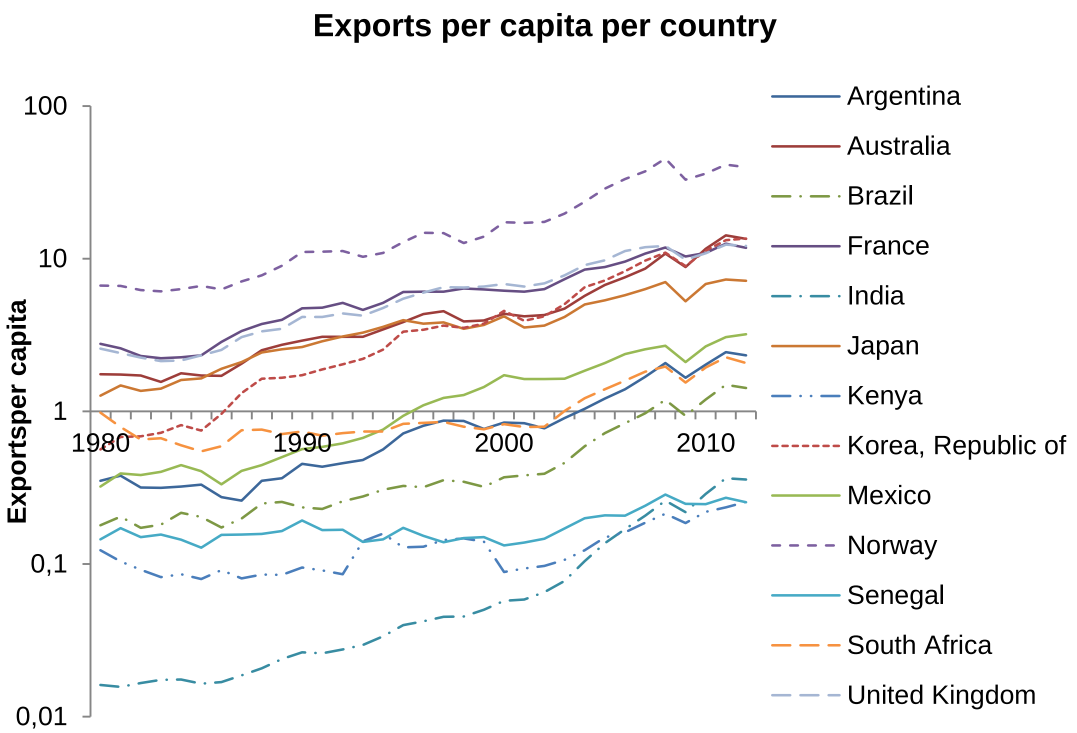


Figure 9 and Figure 10 show a similar pattern for each country in imports and exports per country between 1980 and 2012. Therefore, I think that the multicollinearity will be critical between these factors (with a tolerance level smaller than .2). When severe multicollinearity is the matter, than only one of the two can be used, because the cause of the variation in imports is the same as the variation in exports though its effect may be larger on one. There will though be some variation and possibly the variation in imports has a bigger size than exports. It will be no surprise that the ranking between the countries is the same as with the average personal income, but the differentiation is larger with regard to imports and exports. There are no clear strata, as is the case for the average personal income and consumption.

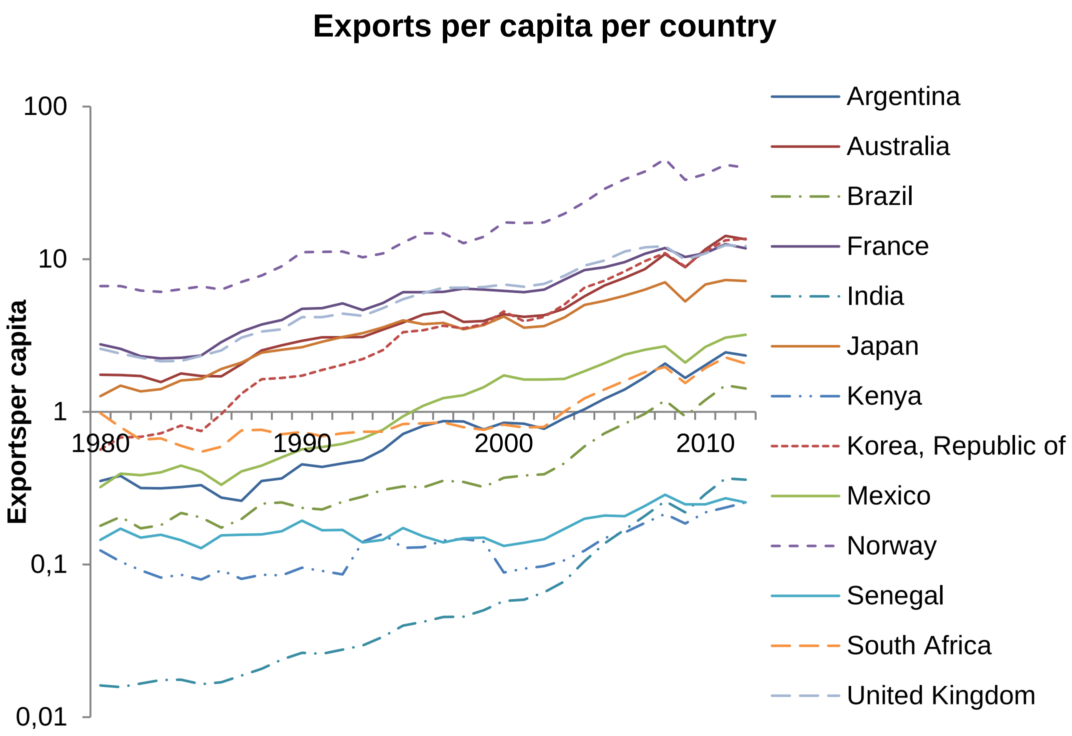


Figure 9 Exports per capita per country (own calculations, UNCTAD data)

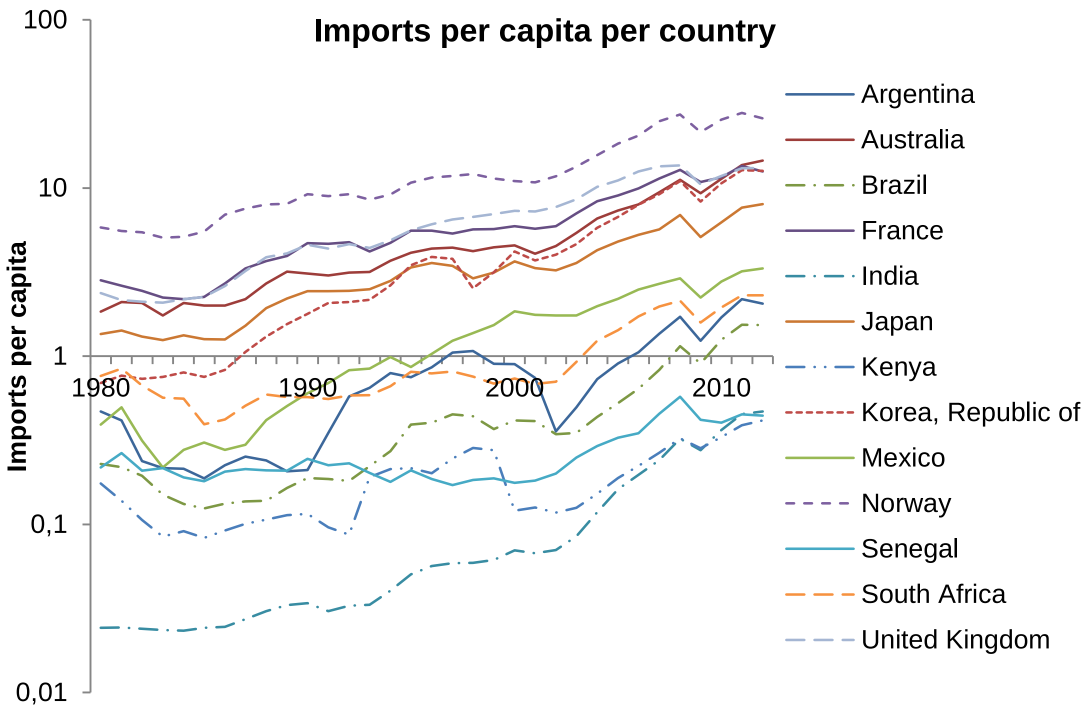


Figure 10 Imports per capita per country (own calculations, UNCTAD data)

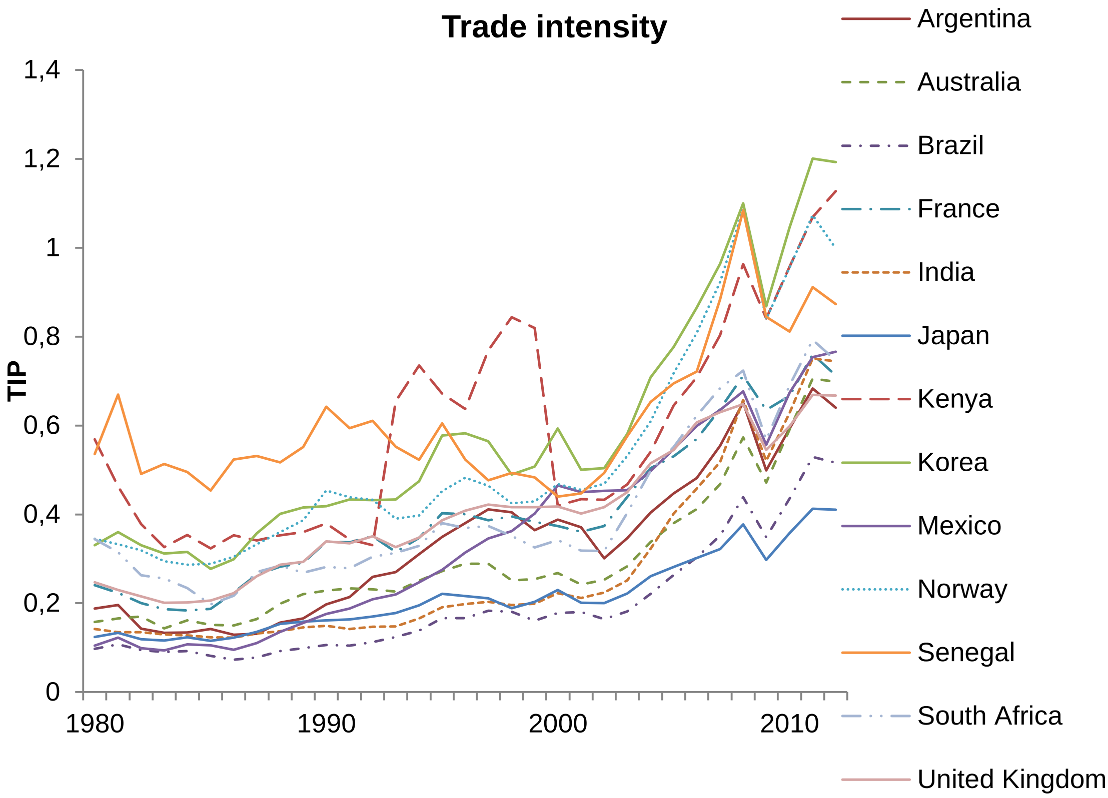


Figure 11 Trade intensity parameters for all countries (own calculations, UNCTAD data)

With regard to the combination of export, import and Gross Domestic Product (GDP) in the trade intensity parameter (TIP), this ratio shows a quite similar development of the teupc development for the last ten years as appears from Figure 11. Strangely enough, for Kenya the trade intensity is very high between 1992 and 1999. Indeed, scholars indicate that export and import have increased severely mid-nineties due to import tariff liberalisations and the increased cooperation between countries in the region. (Ikiara & Mutua, 2006) This was though not persistent due to political domestic conflicts in the country, although in recent years it seems that the trade development has picked up the trend.

All countries show a steep but short drop in 2009 due to the economic countries. Also between 2011 and 2012 most countries show a decline, although Mexico and Kenya still have increasing TIP. Where all rates remain relatively stable from 1980 to 2001, the intensity increases significantly afterwards. Taking into account the limited number of observations for multiple countries in the sample, I expect that the TIP has a positive but weaker correlation with teupc because the trends in teupc seem to be less pronounced than the trend in TIP in the last few years.

4 Method

This chapter describes the three analysis methods used to analyse the data as described in the previous chapter and reflect on the results from different angles. I will highlight the way the model is built up as well as how the model is tested. At the end of the chapter a map for the generation of models is given including the consequences if one of the conditions is not met. As became apparent from the previous chapter, the number of data pairs in the country samples are very limited due to the lack of data on containers. The models that we will produce should therefore be interpreted not in too much detail and should be tested thoroughly.

Using the data on containers, GDP and the population size, I create two new ratio's: gdppc and teupc. These are the average income per capita (gdppc), calculated using the formula:

$$gdppc_{t,i} = \frac{GDP_{t,i}}{Population_{t,i}} \; where\; N = 33$$

Where t is the year and i indicates the country for which the value is calculated. The average number of containers per thousand persons (teupc), is calculated using the formula:

$$teupc_{t,i} = \frac{Containers_{t,i}}{Population_{t,i}} \; where\; N = N_{Containers(t,i)}$$

The number of observations in the dataset is all years between the data limits (1980 upto 2012) for gdppc and dependent on the number of container data for teupc.

Similarly, I calculate the average consumption per capita using the variables Consumption and Population:

$$Cpop_{i,t} = \frac{Consumption_{i,t}}{Population_{i,t}}$$

As with gdppc, for Cpop the data are available for all years included in the database.

The statistical program that I will use is IBM SPSS version 20. The feasibility of this program exceeds others in that it can easily change and import data from external sources and provides clear answers that are ready for reporting. Therefore, the cross-sectional and country-based analysis of the data would be easier to do in this program. The panel data analysis is performed with Eviews version 8, which is more suitable for panel data because it automatically identifies individuals (i.e. countries) and repeated values (i.e. years). Moreover, the fixed and random effect modelling performed in the panel data analysis is easier in Eviews.

4.1 Country-based time series analysis

In the analysis of the time series of each of the countries separately, I will give a country profile of the relation between their socio-economic indicators and teupc. The analysis regards the impact of the variables on container volumes per thousand capita per country to distil the main driver for containerised transport growth or declines. Moreover, the size of the effect can be measured in a beta weight with regression analysis. This analysis assists mainly to answer hypotheses 1, 2 and 5.

I use a multiple regression analysis where possible, but may need to fall back on simple regression models in case of multicollinearity. The multiple regression analysis is initially built up stepwise with all variables. This means that the dependent variable is first tested on the independent variable that has the greatest bivariate and significant correlation with the dependent variable. Subsequently, the ‘second best’ correlating independent variable is entered unless there is multicollinearity with the previous variable and so on, until no other significantly correlated independent variables can be found. The advantage of stepwise regression therefore is that it automatically excludes independent variables that are not relevant with regard to the dependent variable or show very nearly perfect multicollinearity. The resulting included variables are again regressed in a multiple regression model, or, when it is only one variable, in a simple regression model using Ordinary Least Squares (OLS) where the model is calculated on the basis of the total value for the residuals.

The regression is performed with ‘teupc’ as dependent variable and all other variables (except for Containers and Population) as dependent variable. The mentioned variables are required to be ignored, because they form the basis for the ratio. Further analysis is done with the independent variables that have significant influence over teupc.
Before testing the model on the underlying assumptions, the goodness of fit is tested using the adjusted R^2-value. This is needed, since the model will only be relevant when the explanatory value of the model is sufficient. The variables in the model must explain for a large part the variance in the dependent variable. I will only regard models where the adjusted R^2-value is higher than or equal to .5. Instead of using the r^2-value, which is done most often, the limited number of data points requires me to use the adjusted R^2-value, which accounts for this limitation. Therefore, the adjusted R^2-value will always be lower than the regular r^2-value.

$$DV = c + b_1 IV_1 \ldots + b_i IV_i + e$$

A regression model using ordinary least squares techniques must be the best linear unbiased estimator for complying to the Gauss-Markov Theorem. These are that the expected value of the dependent variable (DV) depends on the value of the independent variables (IV). Moreover, the errors in the estimation need to be homoskedastic and the dependent variable is uncorrelated. The errors must be normally distributed. (Carter Hill et al., 2008)

The independent variables should not have too much effect on each other. This could affect the effect of the variables on the dependent variable in the model and therefore the model might not be the best estimate of the dependent variable. It could even result in opposite effects of the independent variables than it actually has. This effect, multicollinearity, can be tested for by the tolerance levels. (Mansfield & Helms, 1982) The closer the tolerance level is to 1, the less multicollinearity affects the beta weights and significance levels. Field & Miles (2010) say that the critical tolerance level is 0.1, but that one should be cautious for multicollinearity at 0.2. I will uphold the latter as the critical level, also because the some of the datasets are limited. Multicollinearity, as the definition already indicates, only has effect on multiple regression models. In case of a single regression model, the tolerance test will be left out.

Using the Durbin-Watson test statistic, the model will be tested for autocorrelation of the errors. Autocorrelation would suggest that the value of the model’s errors can be explained by the previous values of the errors. (Bhargava et al., 1982) This would affect the ability of the model to accurately estimate the dependent variable, given an independent variable value. The critical lower boundary (d_L) and upper boundary (d_u) for the Durbin-Watson test between which the test statistic must be situated to accept the null hypothesis that there is no autocorrelation at the significance level of α=0.05, are given in Durbin & Watson (1951). For the number of regressors we use the number of significant independent variables in the model. For N, we use the number of data pairs available. In case a reason is found to suspect autocorrelation, the significant lag is included in the original model to test further on the validity of the model. The use of the Durbin-Watson test statistic is than no longer possible, since it cannot cope with lagged values of the dependent variable in the model. (Carter Hill et al., 2008)

The obtained regression analysis will be tested for the compliance with the requirement of homoscedasticity of the residuals. If the residuals would not be homoskedastic, than there might be another correlation that is highly explanatory for the change in the dependent variable. (Carter Hill et al., 2008) This would significantly damage the explanatory value of the model. Heteroskedasticity is that the difference between the actual and predicted value changes significantly over the data series on which the model is based. Therefore, the model is more accurate in certain stages of the dataset than in others, indicating that a part of the variation in accuracy is due to variances in another variable that is not considered or included. The estimate can therefore be not the best. A White test statistic will be used to indicate whether there is any reason to suspect heteroskedasticity.

The White test offsets the square root of the residuals against the independent variables, their interaction term and their square root. The virtue of the White test, in contrast to the Breusch-Pagan-Godfrey test, is that it aggregates all these terms by taking the variables and the square root of the predicted value. (Gupta, 2000) The predicted value is either displayed as $\widehat{teupc}$ or teupcpred. This way, one loses less degrees of freedom than with the Breusch-Pagan-Godfrey test and is much more accurate. Thus:

$$resid^2 = b_0 + b_1IV_1 + \cdots b_kIV_k + b_{k+1}IV_1^2 + b_{2k}IV_k^2 + IV_1 * IV_2 + \cdots + IV_{k-1} * IV_k + e$$
$$=$$
$$resid^2 = b_0 + b_1IV_1 + \cdots b_kIV_k + b_{\mathrm{k+1}}\widehat{teupc}^2 + e$$

The fit of the model is multiplied by the number of data pairs available, which is consequently compared with the Chi-squared value of that number of data pairs. In contrast to the models, not the adjusted R^2, but the normal R^2 is used because the adjusted R^2 can take values below zero, which is not possible for the White test statistic. For the X^2-squared values, I use the table and formula provided in Carter Hill et al. (2008). Mathematically:

$$W \leq \mathrm{X}_n^2$$
$$W = nR^2$$

For the model to be homoskedastic, none of the weights in the abovementioned equation should have a significant influence on the squared value of the residuals:

$$H_0: b_1 = b_2 = 0$$

The null hypothesis is rejected if the White test statistic is larger than the critical X^2-value. This critical X^2 value depends on the alpha (chosen to be .05) and the number of degrees of freedom. (White, 1980)

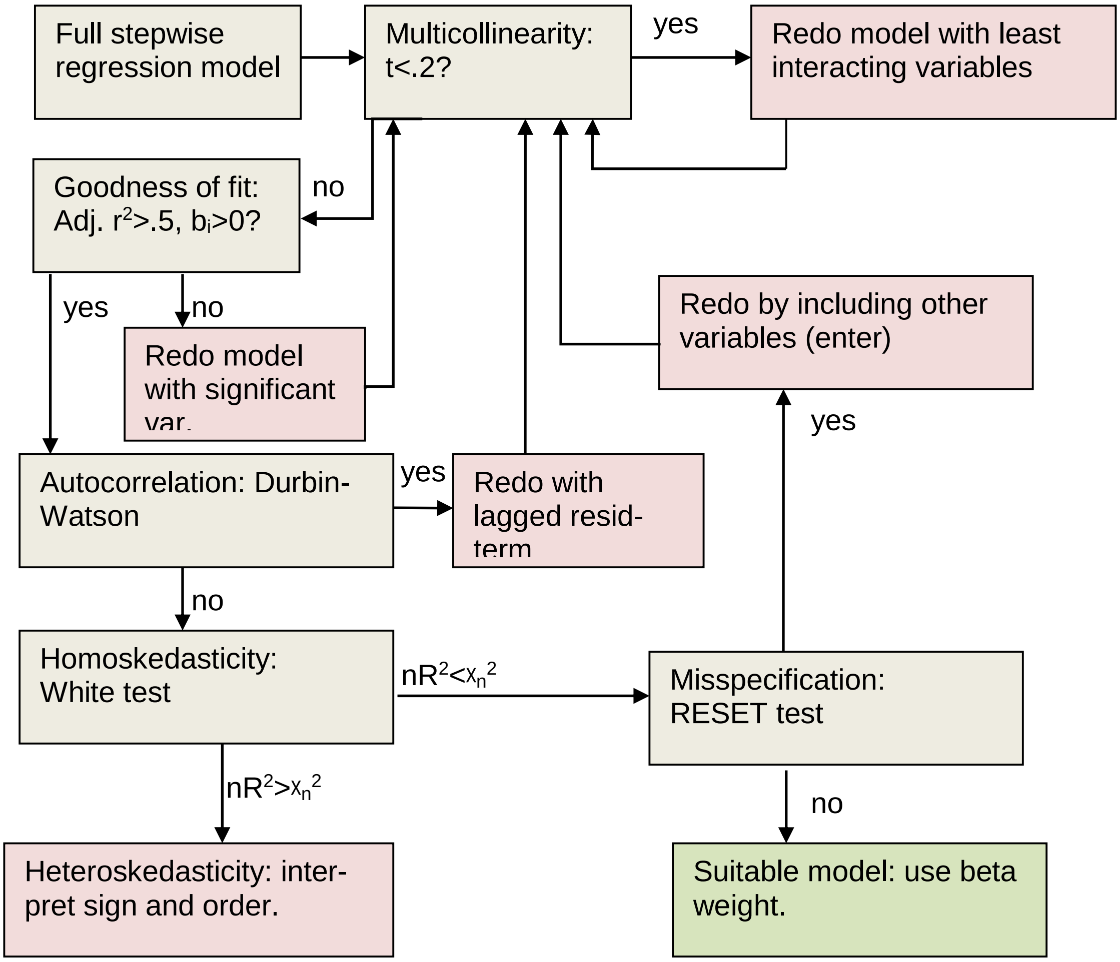
Using a Ramsey RESET-test, I will check for functional misspecification of the model. Basically, the model is tested for a linear relationship between the independent variables and the dependent variable. Would the relationship be different, for example second order or third order relationships, than this would be indicated by an improved explanatory effect of the added second and third power predicted values of the dependent variable. Such would significantly reduce the sum of the squared errors (SSE). Using an F-test, the Ramsey RESET test statistic will indicate whether the null hypothesis, i.e. that the model is not misspecified, holds. If this is the case, than there can reasonably be assumed that the relation is linear. If the null hypothesis is rejected, than there is an indication that the relationship is quadratic or even third power.

The Ramsey RESET test statistic can be calculated using the following formula:

$$F = \frac{(SSE_m - SSE_r)/t}{SSE_r/(n-r)}$$

Where SSE_M is the sum of squared errors in the original model that is tested for misspecification, SSE_R is the model where the second and third power predicted values for teupc are added to the original model (the Ramsey model), t is the number of Ramsey terms that is added relative to the original model, n is the number of observations and r is the number of regressors, i.e. independent variables and Ramsey terms that is added. In this thesis, I test against possibilities of second or third power relationships (thus, t is always two). More Ramsey terms can be added, but will have a small and insignificant contribution. Such will already be visible in the increased explanatory power when two Ramsey terms are added.

Concluding, the following steps are being taken for data analysis with regard to country-based analysis:



4.2 Cross-country analysis

From the data analysis in paragraph 3.2 it becomes apparent that there is a tremendous difference in some of the variables between countries. The teupc for India is for instance still much lower than the teupc for Australia. In the cross-country analysis, I will try to dissect why these variances exist between countries, which tendencies can be found and how these can be explained. This analysis puts the results gained in the country-based analysis in a broader context and validates them. Moreover, it will add new information by analysing regional and cross-country effects on the levels of the dependent variable and independent variables.

I select 2012 as the year of interest for the cross-country analysis. For this data point, France and the United Kingdom have not reported the number of containers handled this year. Therefore, solely for the cross-country analysis I add the data from the ITF to the sample. The use of these data in the time series analysis would reduce the number of data points that are available for these countries significantly.

The results of the cross-country analysis can give preliminary proof for all hypotheses, although it must be offset against the evidence from the time series analyses.

A correlogram will show the resemblance between the variables for all countries. Using a Pearson correlation test, the multicollinearity between the variables can be revealed and the influence of the independent variables on the dependent variables becomes apparent. The correlation is bivariate, meaning that only the relation between variable x and y are measured, such that multicollinearity does not have any influence over the Pearson value. The test is two tailed, because relations in this case can be both positive and negative.

A scatter plot of the variables that are corrected for population differences can reveal differences in characteristics between countries with regard to these variables. First, I will produce a scatter plot of the economic variables and subsequently a scatter plot of the teupc rate compared to socio-demographical variables.

To classify the countries into income groups, I create dummy variables for the income class per country. Each class regards a range of 10.000 USD, as indicated in Table 5, where countries that belong to neither of the classes are in the reference category of an income of USD 40.000,- or more per capita. Insignificant dummies will be combined with the next dummy variable, unless this dummy becomes insignificant thereafter. Then, it will be added to the reference category by excluding it from the equation.

Table 5 Dummy for gdppc classes

Dummy	=1	=0
G1	0-10.000	>10.000
G2	10.001-20.000	<=10.000, >20.000
G3	20.001-30.000	<=20.000, >30.000
G4	30.001-40.000	<=30.000, >40.000
Reference category	40.000+	

The effect of the stratification of countries and income groups can be twofold: on the one hand, it might turn out that certain groups have a different teupc level than the reference group, resulting in a deviating coefficient in the model. Also, it might be that a certain group has a different trend from the reference group, which will cause a significant interaction term of the independent variable and the dummy variable. Mathematically:

$$DV = b_0 + b_1 D_1 + \cdots + b_\delta D_\delta + b_{\delta+1} IV + b_{\partial+2} D_1 IV + \cdots + b_{2\partial+1} D_\partial IV$$

With δ being the number of dummies minus one for the reference category, DV being the dependent variable and IV being the independent variable. Practically, b_0 is the coefficient for the dummy variable of the reference category. Likewise, $b_{\delta+1}$ indicates the trend variable for the reference category. With a different teupc level,b_1-b_δ will be significant. With a different trend, $b_{\delta+2}$-$b_{2\delta+1}$ will be significant. It depends from the hypothesis whether the difference in trend or the difference in the level is under attention.

In case of hypothesis 3, I am interested in differences in the beta weight for gdppc among income groups. Using dummy variables for indicating income groups and regions, I regress δ-1 dummy groups (where δ is the number of dummy variables) on teupc in a combined term of the group variable. For example, for the stratification of the average personal income in dummy variables, I regress the interaction term of the dummy variable with gdppc, add gdppc as independent variable and regress these terms on teupc:

$$teupc = b_0 + b_1 G1 * gdppc + \cdots + b_4 G4 * gdppc + b_5 gdppc + e$$

If b_1 and b_3 are not significant:

$$teupc = b_0 + b_1(G1 + G2) * gdppc + b_2(G3 + G4) * gdppc + b_3 gdppc + e$$

The excluded group, in the example G5, is the reference variable of which the effect is captured in b_0 and b_3. If a beta weight of a dummy group is not significant, I combine this group with the following group. The dummy variables are combined by creating an interaction variable of the independent variable times the maximum of dummy 1 and dummy 2. Mathematically:

$$IV_i * (D_1 + \cdots + D_k) = IV_i * (\max(D_1; \ldots; D_k))\ with\ k = \partial - 1$$

If all dummy variables are insignificant, than there is no effect of the selection on the outcome for teupc.

Significant interaction terms indicate that the trend of this group is different from other groups. The beta weight reflects how large the deviation is from the trend of the reference group.

As with the country-based analysis, the models will be tested for multicollinearity using tolerance levels, for autocorrelation by regressing the one-period lagged errors on the errors, for heteroskedasticity using the White test statistic and for misspecification using the Ramsey RESET test. All procedures for the calculation of the test statistics are the same as in the previous paragraph.

4.3 Panel data analysis

The explanatory power of the variables indicating the main driver per country for the development of containerisation might have a common denominator for all countries, or for the countries in this subset. Moreover, the cross-country differences that are found as a result of the method under paragraph 4.2 can have a dampening or more pronounced impact per country. The analysis should be extended to the determination of common effects to the development of containerisation. The panel data analysis will be used with regard to all hypotheses, but will regard only a variant of hypothesis 4.

Using the data from the sample as described in paragraph 3.2, I create a panel data series of thirteen countries and the six basic regressors that appear to be relevant according to the country-based analysis (apart from the time identifier (year) and individual identifier (country)). With these basic regressors, I can create added regressors such as teupc, gdppc and Cpop as well TIP, of which the calculation methods have been described in paragraph 3.2. With regard to Cpop I must note that its use must be delicately, since in Consumption also governmental consumption is included. Nevertheless, this can be regarded indirect consumption of the citizens, who at the end make use of these goods.

The relation between the dependent variable (DV) teupc and the independent variables (IV) in a panel data set are both influenced by variances over time and over the individuals, in this case the countries, in the data set. These variances are deviations from the common trend among the countries and can therefore be described as errors. Because each country in the data set starts on a different level, for instance because of country-based characteristics that hopefully will be revealed in paragraph 4.2, the intercept of the trend line of this individual in the data set may differ significantly from other individuals. This is statistically regarded as a fixed effect. The trend in the data is than a common denominator among individuals, but individual characteristics make that the individuals have a different level of the variable. In contrast, if both the intercept and the trend respond to individual characteristics, both are significantly different from both the intercept and the trend from other individuals. Nevertheless, part of the variation in the data of the individual can be common with other individuals, but this effect can be more pronounced or dampened with regard to the other countries. If this is the case, than there are random effects influencing the trend estimate for which has to be accounted. Also it is possible that countries have the same intercept, but nevertheless have different trend estimates for which still is accounted for in random effect modelling. (Carter Hill et al., 2008)

A fixed effect model contains a trend estimate beta weight that is valid for each country. The country-based effects are captured in a dummy variable for each of the individuals in the dataset indicating the difference from the trend intercept. Moreover, as usual the model contains an idiosyncratic error term that accounts for deviations that can neither be explained by the dummy variables, nor the independent variables. Therefore, a fixed effect model can be described as:

$$\hat{y} = b_1 IV_{1,i} + \cdots + b_p IV_{p,i} + c_1 DV_1 + \cdots + c_i DV_i + e$$

Where b_p is the common trend effect of the independent variables on the dependent variable and c_i is the individual-specific intercept. $\hat{y}$ is the estimation of the dependent variable, given the described model. DV_i is the dummy variable that is one for country i and null if it is not. 'e' contains the residuals from the trend line, which have to be uncorrelated and with a mean of null for $\hat{y}$ to be the best estimate. If e is auto correlated, the one-period value of e will be included to correct for the correlation. Fixed effects will not be reported, because they entail country-specific characteristics which already will be regarded in the country-based analysis. Comments on the fixed effects will therefore not contribute further.

Random effect models contain an individual-specific trend estimate incorporating the difference from the common trend estimate its beta weight. They can also contain an individual-specific dummy variables to account for differences in the intercept as with fixed effect models, but this is not required if the intercepts do not differ significantly. Moreover, the model contains an idiosyncratic error term as well. The model can be described by:

$$\hat{y} = b_{1,i} DV_1 + b_{p,i} DV_i + b_1 IV_1 + b_p IV_p + e$$

Where $b_{p,i}$ is the country-based estimate of the variance from the trend b_p, DV_i is the dummy variable for each country i and IV_p is the value for independent variable p.

A Hausman test will indicate whether in a certain model random effects modelling or fixed effect modelling is justified. The null hypothesis of this test is that a random effect model is justified, where the alternative hypothesis is that a fixed effect model is justified. The Hausman test is performed on the basis of a random effects model in Eviews. The first step in the analysis is therefore to create a random effects model. The model tests whether the coefficient weights are different for the subset of countries in a random or fixed effect model. If the p-value of the test statistic is above the critical level of alpha, the coefficient weights are indeed significantly different and this model can be used for further testing. Otherwise, the model has to be redone with only fixed effects. (Carter Hill et al., 2008)

Using the same criteria as with the country-based time series analysis, the goodness of fit is tested against a critical adjusted R^2 of .50. Moreover, all beta weights of the independent variables need to be significant. The intercept for the countries does not necessarily be significantly different, but will be reported to see which country is used as default country, thus which country is used for the common coefficient.

Both Eviews and SPSS do not provide suitable autocorrelation tests such as a Breusch-Pagan-Godfrey test for fixed and random effect models. Therefore, testing for autocorrelation is done by taking the one-period lagged values of the residuals of the model and regressing them on the residual values of that model. Mathematically:

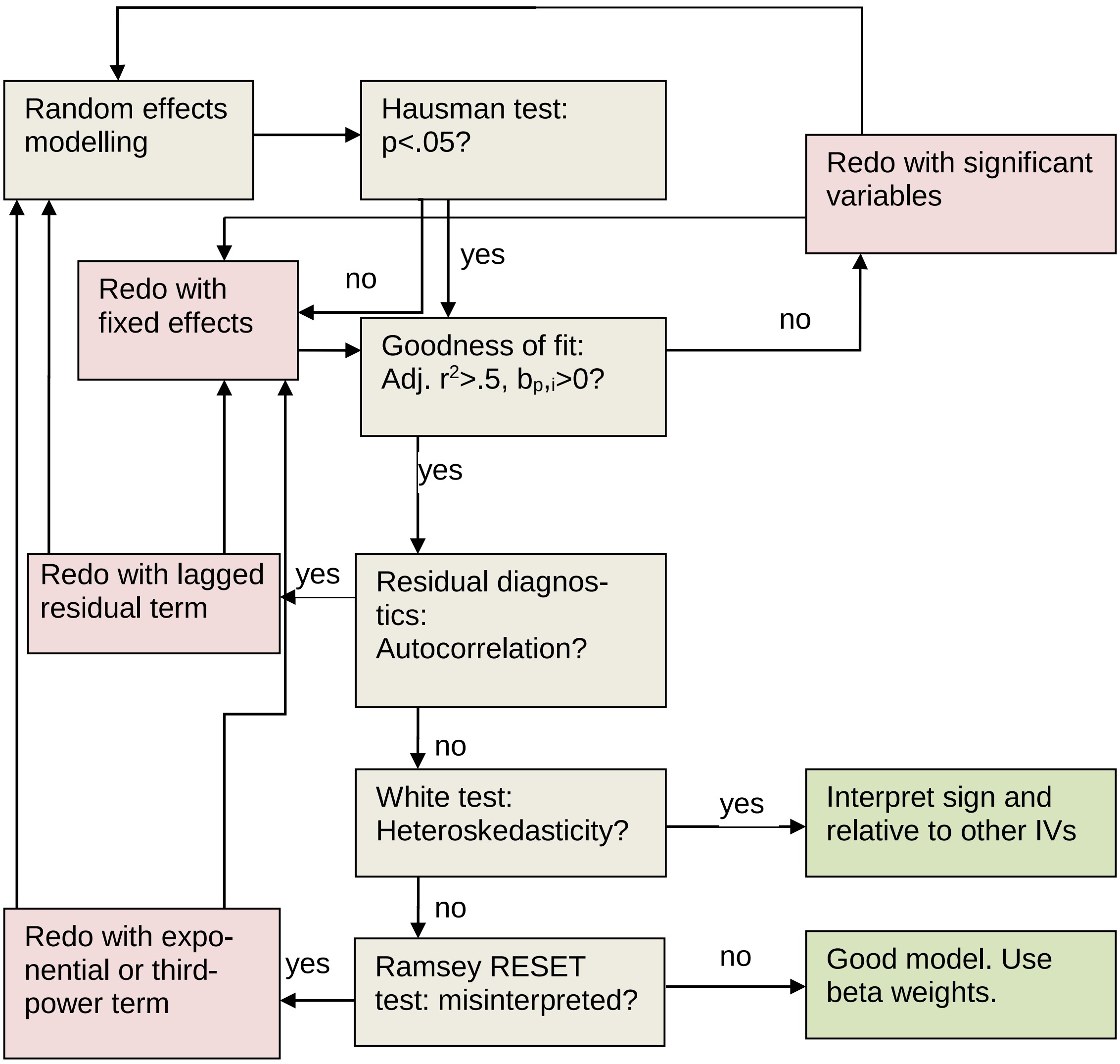
$$resid_t = b_0 + b_1 resid_{t-1} + e$$

The p-value of b_1 will indicate whether or not there is autocorrelation. The fit of the model indicates the influence of the residuals on the initial model. Is there autocorrelation indeed, than I will correct for this by including a one-period lagged residual term in the model. The corrected model reveals a more accurate beta weight for the independent variable and reveals whether the initial significance is correct.

Again, a White test will be used to test for heteroskedasticity in the residuals. In both SPSS and Eviews the White test statistic needs to be calculated manually on the basis of data provided in the results. Again, heteroskedasticity will affect the reliability of the linear model because over time in the data set, it might not be the best fitting model. Nevertheless, it has explanatory value as is provided for by the R^2 value. Thus, a model that is infected with heteroskedasticity is suitable for cautious interpretation.

The relation between the dependent variable and independent variables might be exponential or otherwise polynomial. A Ramsey RESET test will reveal whether the explanatory power of such a high-power model including the same set of independent variables will indeed be significantly larger. This test can also not automatically be performed in the statistical software for fixed effect models, but has to be calculated manually.

Do none of the abovementioned test provide significant problems, than it can be concluded that the model is suitable for forecasting and interpreting. Summarizing, the method for the panel-based data analysis can be described by the following structure:


Random effects modelling
Hausman test: p<.05?
Redo with significant variables
yes
no
Redo with fixed effects
Goodness of fit: Adj. r²>.5, b_p,i>0?
no
yes
Redo with lagged residual term
yes
Residual diagnostics: Autocorrelation?
no
White test: Heteroskedasticity?
yes
Interpret sign and relative to other IVs
no
Redo with exponential or third-power term
yes
Ramsey RESET test: misinterpreted?
no
Good model. Use beta weights.

5 Results

5.1 Country-based analysis

In this chapter, I will execute the method on the collected data described in chapter 4. The analysis is displayed per country to report the methodological steps and conclusions that have been taken. A summary of the models with the relevant data can be found in Appendix A.

This analysis tries to give a more thorough understanding of the main drivers of changes in the degree of container handling relative to the population of a country. It tries to provide insights and indications for evidence for hypotheses 1, 2, 4 and 5. Because only one country is regarded, hypothesis 3 cannot be validated using a country-based analysis.

5.1.1 Argentina

The number of data pairs is with nine all-inclusive observations very small for Argentina. Doing the stepwise multiple regression with teupc as dependent variable and all others as independent variables, only FDI appears to be significant (p=.001). Since there is only one significant variable, there is no need to test for multicollinearity. The fit of the simple regression model is very high with an adjusted R^2 of .894. The model with only FDI shows that the independent variable is still significant (p=.002), but the beta weight seems unfeasible because it is very small (b_1=.002). Therefore, I adopt FDI by taking the natural logarithm to correct for scale effects. This indeed proves to be helpful since the beta weight increases to 16.558, but the significance remains about the same (p=.001).

Durbin and Watson (1951) do not give critical values in their paper for this number of data pairs. Nevertheless, the test statistic of 1.268 is sufficiently below the critical upper value they give for the lowest number of data points, i.e. d_u=1.36 and for n=15. There is thus no indication for autocorrelation of the error terms.

For the test on homoskedasticity, I do the White test with the dependent variable FDI and the squared predicted values on the residuals of the model. This results in the following equation:

$$resid^2 = c + b_1 logfdi + b_2 teupcpred^2 + e$$

The resulting R^2 is .760, and with six data points, W=4.560. Since the null hypothesis is:

$$H_o: b_1 = b_2 = 0$$

the number of degrees of freedom is 2-1=1. Correspondingly, the X^2 value is 3.841. Since $W>X^2$, I reject the null hypothesis and assume there is proof for heteroskedasticity. With heteroskedasticity, the fit of the regression line might not be the best over time, because the variation in the residuals varies.

For the Ramset RESET test statistic, the SSE of the initial model and the SSE of a regression of the significant independent variable with the squared and third power predicted value are required. Therefore, I regress teupc on these parameters. The resulting Ramsey model has an SSE of 60.301, where the initial model had an SSE of 67.589. Therefore:

$$F = \frac{\frac{67.589 - 60.301}{2}}{\frac{60.301}{9-3}} = \frac{3.644}{10.050} = .363$$

The critical F-value for the Ramsey RESET test statistic, given that there are nine data pairs and three terms in the Ramsey model, is $F_{9,3}$=27.35. Because $F<F_{9,3}$, the null hypothesis of no misspecification is not rejected. Therefore, there is no indication of misspecification.

Concluding, it appears that the logarithmic value of the FDI has the most explanatory effect on teupc for Argentina. The linear regression model:

$$\widehat{TEUpc} = -100.176 + 16{,}558 * logFDI$$

is complying to the Gauss-Markov Theorem, except that there is proof of heteroskedasticity in the residuals.

The nature of this relation might be laying in the developing nature of the Argentinean economy. As Figure 12 shows, Argentina has the highest average rate of the Foreign Direct Investments over the Gross National Income. This means that in the period 1980-2012, on each dollar earned by its population, the most investments are done from abroad in this nation. It reflects how the international market is involved in the Argentinean economy. The increase in FDI in countries like Argentina can possibly be explained by the increased investment of production facilities in low-wage countries and consequently the globalisation of production. (Molnar et al., 2007) Therefore, the relation between log(FDI), but also FDI itself, and the rate of the number of containers per head of the population can be explained by the need of the international market to export the products that are produced in Argentina to their selling markets. The use of the containers is apparently a logical choice for the exportation of a part of these products.

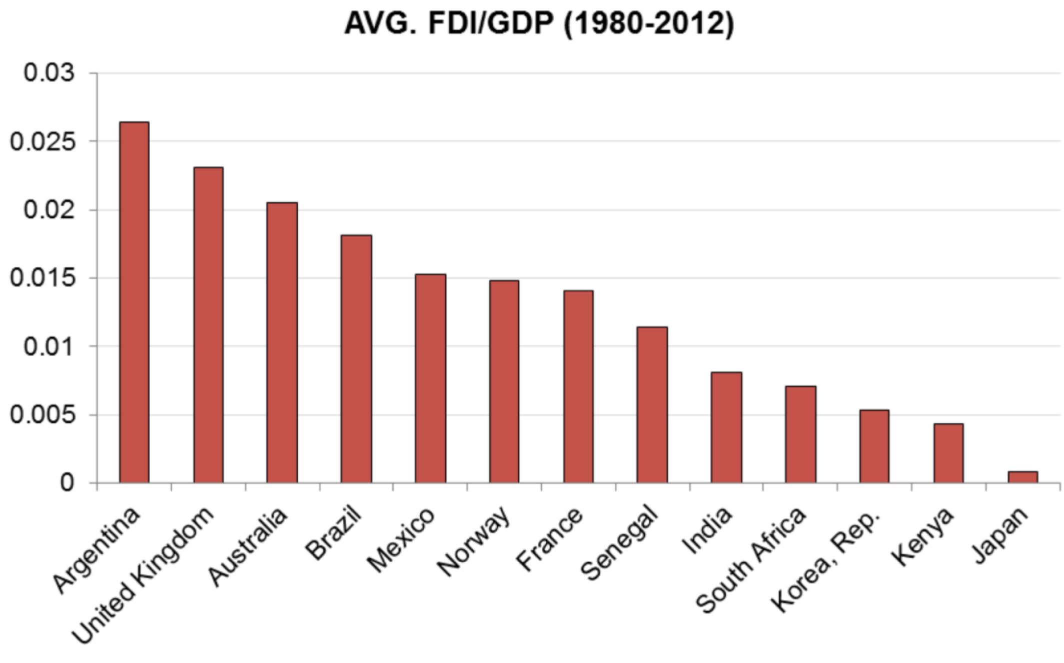


Figure 12 Average FDI/GDP rate (own calculations based on UNCTAD data)

5.1.2 Australia

With seventeen data pairs, the N-value is not too large, but also not the lowest I will use in this research. I start with a stepwise regression of all independent variables on teupc.

The full stepwise model shows that in the most inclusive model, the tolerance levels are very high (t=.997), indicating barely any multicollinearity. Moreover, the explanatory value of the model is high with an adjusted R^2 of .983. The two variables included, gdppc and Consumption, are significant (p=.000 and .005 respectively), but where the beta weight of gdppc is feasible with a value of b_1=9.028, the beta of Consumption is very low with a beta weight of $b_2=1.316*10^{-6}$. This might be due to the scale effect between Consumption and teupc. This scale effect might even hide some of the multicollinearity between gdppc and Consumption. Therefore, I redo the model with the natural logarithmic value of Consumption and gdppc.

Indeed, the corrected model shows that though it is highly explanatory (adjusted R^2=.994), the independent variables are collinear, appearing from the low tolerance level of t=.088. Therefore, I have to exclude one of the variables. First, I redo the analysis with gdppc, because its significance on the model was the highest from the beginning. The adjusted R^2 of .992 is very high. The independent variable is highly significant (p=.000) and its beta value is feasible though has slightly increased compared with the bivariate model previously discussed, to b_1=9.485. Unfortunately, the Durbin Watson test statistic of .589 is lower than the critical lower level of d_L=1.13, considering we only have one regressor and seventeen data pairs. Therefore, I regress the residuals of the model on their one-period lagged values to see whether there is any autocorrelation. Indeed, it appears there is, since the beta weight of $resid_{t-1}$ is significant (p=.005). It accounts for nearly 44% of the variation in the error terms (r^2=.437). (A2.3) Therefore, the one-period lagged value of gdppc and teupc are included in the model.

The resulting model doesn't lose explanatory value, as the adjusted R^2 of .997 indicates. Gdppc is still significant (p=.000), as are the lagged values of teupc and gdppc (respectively p=.005 and .081). The new beta weight of gdppc has slightly decreased to b_1=9.009, but still quite similar. Testing for homoskedasticity, I take the residuals of the given model and regress it on the variables and lagged values:

$$resid^2 = c + b_1 gdppc + \mathrm{b_2 resid_{t-1}} + \mathrm{b_3}\widehat{\mathrm{teupc}}^2 + e$$

The resulting r^2 is .095. With an N=16, the W=16*.145=1.52. The null hypothesis is:

$$H_0: b_1 = b_2 = b_3 = 0$$

Therefore, our degrees of freedom are 3-1=2. The corresponding critical X^2 value is 5.991. Since $W<X^2_{max}$, there is no proof of heteroskedasticity.

To test for misspecification, I create a Ramsey model by regressing the parameters of the initial model with the squared and third power predicted values on teupc. The resulting model has an SSE of $7.5*10^{10}$, where the initial model had an SSE of $9.1*10^{10}$. Thus:

$$F = \frac{\frac{9.1*10^{10} - 7.5*10^{10}}{2}}{\frac{7.5*10^{10}}{17-4}} = \frac{7.8*10^9}{5.8*10^9} = 1.359$$

Since with seventeen data pairs and four regressors, the critical F-value is approached by the tabulated value of $F_{20,4}$=5.80, I can conclude that $F<F_{20,4}$, thus that there is no indication of misspecification.

Concluding, the income per capita has a significant influence on the number of containers per teupc. The relation can be interpreted in the sense that the income of an Australian has a positive effect on the use of the container for import. As we saw already in the beginning of the paragraph, the effect of the change in income is comparable with the change in consumption, though has a slightly better explanatory effect. A reason might be that Australia is dependent for their consumption of (luxury) goods for imports via container, rather than national production of goods.

5.1.3 Brazil

For Brazil, unfortunately there are no ITF data available for containers. The analysis will therefore be limited to only nine data pairs. For the unemployment rate, six data points are missing for the period 1980-2012.

The comprehensive model of teupc has a high explanatory value (adjusted R^2 is .973). Only one independent variable, Exports, is found to have significant effect on teupc (p=.001). Therefore, no testing for multicollinearity needs to be done. The beta weight for Exports is though not feasible because of its small value (b_1=7.7*10^{-5}). The model needs to be redone with the natural logarithmic value of Exports, log(Exports).

This new model has a good fit (adjusted R^2 = .943) and the beta weight of log(Exports) is feasible (b_1=17.306) and significant (p=.000). Because for an N =9 and only one regressor, the Durbin Watson test statistic must be between 1.08 an 1.36, where it is 1.777 in this model. Consequently, the residuals must be regressed on the current and one-lag logarithmic value of the Exports. This though appears not to be of significant influence, since the beta weight of $resid_{t-1}$ is insignificant with a p=.980. Thus, there is no autocorrelation between the residuals.

The White test is performed by regressing the independent variable and the square root of the independent variable on the square root of the residuals:

$$resid^2 = b_0 + b_1 logE + b_2 logE^2$$

The resulting r^2, .317, renders a White test statistic of W=9*.317=2.853. Since the null hypothesis is:

$$H_o: b_1 = b_2 = 0$$

the number of degrees of freedom is 2-1=1. Correspondingly, the X^2 value is 3.841. Since $W<X^2$, there is no proof for heteroskedasticity.

The model that we have obtained can be described by the simple linear function:

As follows, the predicted value for teupc is:

$$\widehat{teupc} = -174.139 + 17.306 logE$$

To test for misspecification, I take the squared and third power value of the predicted values for teupc and regress them with log(Exports) on teupc. The resulting regression does not succeed, because there is no correlation between these variables and the initial model. Therefore, the F-test is null, leading to the conclusion that there is no misspecification.

Concluding, we could say that the number of containers per capita handled is positively influenced by the size of the exports in the case of Brazil over the limited time period I have been able to investigate. This relation, described by the model:

$$teupc = -174.139 + 17.306 logE + e$$

is robust and complying to the Gauss-Markov assumptions.

The influence of exports for Brazil might be because of its growing industrial capacity due to the movement of many production processes to Brazil as a consequence of production globalisation. Moreover, some of the Brazilian exports such as fruits might have changed to transportation in reefer containers, boosting the influence of exports further.

5.1.4 France

For France, the dataset is relatively large with 24 data pairs. Notably, for 2004 I have excluded the value for Containers, because it was twice as high as in neighbouring years. In other factors, this was not seen for this year. Moreover, even exports only showed a moderate growth for this year. The eventual inclusion of such an extreme and unexplainable figure might cause problems for the model, because the fit is not the best. The first assumption of the Gauss-Markov Theorem would then be violated. It is to prevent this from happening, that I exclude 2004 from the dataset. Nevertheless, the dataset remains relatively large with 23 data pairs.

The initial stepwise regression did not render any result, because none of the variables are entered into the model. To see what is wrong, I enter all variables standard in the model. Of all independent variables, the tolerance levels are low, of which only Unemployment and gdppc not critically. This multicollinearity has affected the beta weights and significance of all variables. FDI, Imports, Labourf and Urbanisation are even excluded from the model due to problems with the tolerance levels.

The correlogram of all variables indicates that nearly all variables are correlated with each other. I use the variables with a Pearson correlation value larger than .900 in the model on teupc, because these have the largest effect on the dependent variable. These are Consumption, Exports, Labourf, Urbanisation, GDP and gdppc. GDP is excluded from the model, because of the low tolerance levels. Urbanisation and Exports are also insignificant. The tolerance level of the rest of the values is troubling. Therefore, I select the two variables with the highest tolerance levels. These are Exports and gdppc.

Although still low, the tolerance levels have come to an acceptable level of .225. Both variables are significant (Exports at p=.024 and gdppc at p=.000). The beta weight for gdppc is feasible (b_2=3.993), but for Exports it is negligible (b_1=$1.8*10^{-5}$), therefore, I take the natural logarithm of the values of Exports. This unfortunately makes log(Exports) insignificant (p=.123), perhaps because of the tolerance level (t=.145) that indicates severe multicollinearity. Therefore, I exclude Exports and log(Exports) from the model.

This does not reduce the explanatory value severely (from adjusted r^2=.950 to adjusted- r^2=.946). Gdppc is still significant (p=.000) with a beta weight of 5.037. The Durbin-Watson test statistic is with .938 lower than the lower critical value at N=23 and r=1 of d_L=1.26. Therefore, autocorrelation of the residuals is indicated. This is supported by the regression of the one-period lagged residuals on the current value, because the significance is p=.021. Although the adjusted r^2=.200, therefore indicating that the impact is little, I must include the lagged term in the model. The model therefore becomes:

$$teupc = b_0 + b_1 gdppc + b_2 resid_{\mathrm{t}-1} + e$$

Inclusion of the lagged residual term increases the adjusted R^2 slightly to .964. Also the beta weight of gdppc has slightly improved to 5.239. Both terms are still significant (p=.000 and p=.024 for $resid_{\mathrm{t}-1}$).

I calculate the squared values of predicted value for teupc (teupcpredsq), using the model:

$$\widehat{teupc} = -121.398 + 5.239 * gdppc + .434 * resid_{t-1}$$

Inserting the teupcpredsq (squared values of the predicted values for teupc) term in the model and regressing the model against the squared residuals renders the White test statistic. Because there are 22 data pairs and r^2=.008, W=22*.008=.176. The corresponding critical X^2 value, given the number of data pairs, is 3.841. Since $W<X^2$, there is no indication of heteroskedasticity in the results.

Subsequently, I test the obtained model for misspecification using the Ramsey RESET test. I add the squared and third power predicted values for teupc to the model. The SSE declines from 138.976 in the model to 128.377 in the model with the added terms. Therefore, the Ramsey RESET test statistic is:

$$F = \frac{(138.976 - 128.377)/2}{128.377/(22-4)} = \frac{5.2995}{7.1321} = .7431$$

Given N=22 and r=4, the critical F-value is comparable to the tabulated F-value of $F_{20,4}$=5.80. Because $F<F_{max}$, there is no indication of misspecification.

Concluding, the number of containers that is handled in France per capita is positively influenced by the income per capita according to the model:

$$\widehat{teupc} = -121.398 + 5.239 * gdppc + .434 * resid_{t-1}$$

This relationship is robust. Although, I must note that the high multicollinearity which I encountered in the beginning of this paragraph indicates that many factors are of significant influence on teupc, this factor has the most pronounced influence.

5.1.5 India

Using twenty-two data points for India, the availability of data for India is relatively high compared to the other countries tested for. There are though data missing for Unemployment.

The comprehensive stepwise linear model only includes gdppc as independent variable. Where nearly all other variables are excluded because of their low tolerance levels, Labourf is excluded because its impact on teupc is insignificant, as is the case with GCI and FDI. Solely regressing gdppc on teupc increases the explanatory value to .963 (adjusted R^2). The independent variable is significant (p=.000) and feasible, since its value is 8.403. It indicates that a rise in the income per capita will offset an even larger rise in the number of containers handled in India. With .359, the Durbin- Watson test statistic is far below the lower critical value of 1.24, given the number of observations and the fact that there is only one regressor.

As we see in the regression of the residuals on their one-period lagged values, there is indeed autocorrelation. The beta weight for the one-period lagged values is significant with p=.000 and positive, with a weight of .902. The adjusted R^2 of .625 suggests that the autocorrelation explains more than 60% of the variance in the initial model. Subsequently, I include the one-period lagged value for the dependent and independent variable.

Both lagged variables are significant with a p-value of .000 for teupc_1 and .002 for gdppc_1. Gdppc itself is still significant with p=.003. The explanatory value has increased to adjusted R^2=.988. It results in a beta weight of 10.2 for gdppc, which is quite some higher than 8.403.

Subsequently, I test for heteroskedasticity using the White test:

$$resid^2 = c + \mathrm{b}_1\widehat{\mathrm{teupc}} + \mathrm{b}_2\widehat{\mathrm{teupc}}^2 + e$$

The resulting r^2 is .408. With N=21, the W= 21*.408=8.568. The null hypothesis is:

$$H_0: b_1 = b_2 = 0$$

Therefore, our degrees of freedom is 2-1=1. The corresponding critical X^2 value is 3.841. Since $W>X^2_{max}$, there is proof of heteroskedasticity.

Using the model I obtained earlier, I create the predicted values for teupc and its squared and third power values. With these values, I can execute a Ramsey RESET test for assessing whether the model is misspecified. Using the abovementioned formula:

$$F = \frac{.749 - .319/2}{.319/(21-5)} = \frac{.215}{.020} = 10.784$$

In this case, with 21 observations and five regressors, the critical F-value is again comparable with $F_{20,5}$=4.56. Since $F>F_{max}$, there is reason to believe the model is misspecified. The model that includes both gdppc as well as the squared values of gdppc reveals significant multicollinearity and can therefore not be used for interpretation. Neither has it any improved explanatory power. Therefore, the model with gdppc and $resid_{t-1}$ explains the relationship well, but the beta weights of the model should not be interpreted apart from their sign. This is both due to the misspecification and the heteroskedasticity in the model.

Concluding, it appears that the growth in India's container transportation is explained by the growth in the income of the population. More containers come in as the income of the country rises. There is though heteroskedasticity that can affect the accuracy of the model.

5.1.6 Japan

With only thirteen data points, the availability of information is rather limited for Japan. Further, the information for Japan is complete.

The initial comprehensive regression model includes only two variables: Exports and gdppc. Their tolerance level is high (t=.887), so there will be little effect of multicollinearity. Because the beta weight of Exports is again very small, I will use the natural logarithm of Exports (log(Exports)).

Regressing only these two independent variables on teupc results in a model that has a good explanatory value (adjusted R^2=.810). Both variables have a significance of p=.000. Where the beta weight of gdp is 3.868, the weight for log(Exports) is 188.017. Although the tolerance levels are lower (t=.371), they do not approach a critical level. Multicollinearity thus has not too much effect.

With a Durbin-Watson test statistic of .874, it is not impossible that there is autocorrelation of the errors. Although there is no lower limit given for N=13 and two regressors, the lower limit for the corresponding N=15 is .95. It is on the edge, and either way reason to investigate it further. The autocorrelation model of the errors shows that the one-period lagged errors do not have a significant effect on the current value for the errors. Therefore, it is reasonable to conclude there is no autocorrelation.

For the White test, I take the squared values of the residuals, gdppc and loge. The regression:

$$resid^2 = b_0 + \mathrm{b}_1\widehat{\mathrm{teupc}} + \mathrm{b}_2\widehat{\mathrm{teupc}}^2 + e$$

Has an R^2 of .557. Consequently, the White test statistic is $W=n*R^2=13*.557=7.241$. Since the null hypothesis is:

$$H_0: b_1 = b_2 = \cdots = b_5 = 0$$

We have 5-1=4 degrees of freedom. The corresponding X^2 critical value is 3.841. Since $W>X^2$, there is suggestion of heteroskedasticity.

The model can thus being described as:

$$\widehat{teupc} = -2,.503.027 + 3.868 * gdppc + 188.017 * loge + e$$

For the Ramsey RESET test, I take the second and third power values of the predicted value for teupc. These values are added to the regression model that is tested. The SSE for the Ramsey RESET model is 1917.441, where it is 2032.088 for the initial model. The corresponding F-value is:

$$F = \frac{\frac{2032.088 - 1917.441}{2}}{\frac{1917.441}{13-4}} = \frac{114.647}{213.049} = .538$$

Since the critical F-value with N=13 and r=4 is approached by $F_{12,4}=14.37$ and $F<F_{12,4}$, there is no indication of misspecification.

It can be concluded that teupc in Japan is influenced by the income per capita and Exports, where the natural logarithm of this value gives the best estimations for the value of the number of containers handled per thousand capita due to scale effects. The model is robust and conform the assumptions of the Gauss-Markov theorem.

An explanation might be that on one hand, the Japanese society is a wealthy society, which is largely dependent on imports for their consumption goods. On the other hand, the economic power of their society may lay in the production and export of high-valued goods, which are often transported in containers.

5.1.7 Kenya

The number of data pairs for Kenya is very limited. In the initial comprehensive model, only Consumption has a significant influence on teupc (p=.002), but its beta weight is very small ($1.2*10^{-9}$). Therefore, it is useful to convert the value of Consumption into its natural logarithm, to correct for scale effects. The resulting simple linear regression model of log(Consumption) on teupc has a high explanatory value (adjusted r^2=.890). The beta of log(Consumption) is significant with a p-value of .003 and a weight of 26.0. Therefore the model has an appropriate fit. Nevertheless, the Durbin-Watson test statistic indicates autocorrelation of the errors. Although with only six data points, we are far outside the reach of the indicated critical values of one regressor and 15 data pairs, the upper critical level of 1.36 is not even close to the value of 2.525 that was acquired. Nevertheless, the small number of data pairs might have an influence on this value. This is shown by the regression of the one-period lagged residuals on the residuals of the model. With a p-value of .403, it is shown that there is no autocorrelation in the error terms.

In this case, to test for heteroskedasticity, I regress the squared residuals on log(Consumption) and its squared values:

$$resid^2 = c + b_1 logC + b_2 logC^2 + e$$

The resulting regression analysis has an r^2 of .177. Correspondingly, the White test statistic is W=6*.177=1.062. In this test, the null hypothesis is:

$$H_0: b_1 = b_2 = 0$$

Thus, we have 2-1=1 degrees of freedom. The X^2 critical value is correspondingly 3.841. Since $W<X^2$, there is no heteroskedasticity.

Subsequently, the model is tested for misspecification using the Ramsey RESET test. Upto now, the model describes the following relation:

$$teupc = -603.93 + 26.0\ logC + e$$

Consequently, the predicted value for teupc (teupcpred) is:

$$\widehat{teupc} = -603.93 + 26.0\ logC$$

For testing whether there is any evidence of misspecification, I compute this value, its squared and its third power value (teupcpredsq and teupcpredthrd respectively). The residuals are small (.688), where in the model they are 2.275. Therefore, the F-value is:

$$F = \frac{(2.275 - .688)/2}{.688/(6-3)} = \frac{.7935}{.2293} = 3.46$$

With only six data points and three regressors, the critical F-value is 8.94. Therefore, since $F<F_{max}$, there is no indication of misspecification.

Concluding, it is shown that there is a robust relation between the number of containers per capita and the country's total consumption. The natural logarithm gives a more feasible beta weight of 26.0.

This relation might be because of the nature of the Kenyan economy. Where Kenya has a relatively rural society and thus produces primary goods themselves, it has to rely on imports via containers to satisfy their demand for luxury goods. Kenya has known a great development in the last few years, giving rise to the national income. Subsequently this has resulted in a growth of imports of luxury goods. Nevertheless, it is striking that the degree of urbanisation than did not have any significant effect. This might be because the different levels of urbanisation only explain different levels of container handling rather than different trends.

5.1.8 Republic of Korea

Since the data for containers are available from 1993 onwards in the case of Korea, we have a relatively large number of data pairs. The N=20.

The comprehensive model excludes only Exports and Imports. There is large multicollinearity in the model with regard to gdppc, Consumption and Labourf. The correlogram of the relations shows that actually only GDP, gdppc, Labourf, FDI and Consumption have a significant effect on teupc. Moreover, all factors have a significant correlation between them of p=.000. Therefore, multicollinearity is expected.

Indeed, the model including these variables has large multicorrelation, as shown by their tolerance levels. Apart from FDI (t=.226), all tolerance levels are below .100). Because GDP and gdppc are based on the same data, except that gdppc is the ratio of GDP over the Population, I exclude the latter and continue only with gdppc. Moreover, I include FDI in the model, Consumption and Labourf. Still, there is multicollinearity which leads me to exclude Labourf. Still, between gdppc and Consumption there is multicollinearity. Only when I exclude either gdppc or Consumption, the tolerance levels have become acceptable.

So which model to continue with? I make a selection of models based on the multicollinearity statistics. Model 1 describes the relation between FDI and gdppc. This model has only one significant variable, i.e. gdppc (p_{FDI}=.295, p_{gdppc}=.000. Model 2 describes the relation between FDI and Consumption. The tolerance levels are acceptable (t=.411), but FDI is again not significant (p=.645). Consumption is significant (p=.000), but not feasible ($b_2=1.0*10^{-9}$). Model 3 describes the relation between FDI and the natural logarithm of the values of Consumption. Still, tolerance levels are fine (t=.409), FDI is insignificant (p=.637) but log(Consumption) is (p=.000). The beta weight of log(Consumption) has though improved to 528.0. Model 4, a simple linear regression of log(Consumption) on teupc, has an explanatory value of .956 (adjusted r^2), where model 5, a simple linear regression of gdppc on teupc, has an explanatory value of .980 (adjusted r^2). Concluding, the best model we have when correcting for insignificant independent variables and multicollinearity is model 5, describing the relationship between gdppc and teupc. Nevertheless, the difference with model 4 is small, which needs to be accounted for in the interpretation of the final results.

The beta weight of gdppc in model 5 is significant (p=.000) and feasible (b_1=33.446). Testing for autocorrelation, the Durbin Watson test statistic is 1.134, which is slightly outside the limits of 1.20 and 1.41 considering we have only one regressor and twenty data pairs. The regression of the one-period lag of the residuals on the residuals itself though shows that there is no autocorrelation (p=.106).

To test for heteroskedasticity, I square both the residuals and the values of gdppc, use the latter in the regression on the first to acquire the r^2 for the White statistic. The regression

$$resid^2 = b_0 + b_1 gdppc + b_2 gdppc^2 + e$$

has an r^2=.018, thus the White statistic is W=20*.018=.36. Since the null hypothesis is:

$$H_0: b_1 = b_2 = 0$$

The number of degrees of freedom is 2-1=1. Therefore, the critical X^2-value is 3.841. Since W is well below this value, there is no reason to suggest heteroskedasticity.

The model describes a positive relation between gdppc and teupc:

$$teupc = -291.920 + 33.446 * gdppc + e$$

Therefore, the estimate of teupc is:

$$\widehat{teupc} = -291.920 + 33.446 * gdppc$$

By computing this value, its squared and third power value, I am able to do the Ramsey RESET test. Including these values in the regression of gdppc on teupc, I find a SSR of 4,774.471. This is 5,817.354 for model 5. Using these values, I can calculate F:

$$F = \frac{(5817.354 - 4774.471)/2}{4774.471/(20-3)} = \frac{521.4415}{280.85} = 1.857$$

With three regressors, but twenty data pairs, the critical F-value is 8.66. Since $F<F_{max}$, there is no indication of misspecification.

Concluding, the model describing the relation of gdppc on teupc is highly explanatory and robust. Gdppc has a significant and positive effect on teupc for Korea.

In the light of the multicollinearity with amongst others Consumption, this relation should be interpreted also in relation to these variables. Multicollinearity indicates that the change in one variable can be explained for a large part by the variation in another. The in- or decrease of gdppc is thus related to changes in Consumption. Apparently, the Korean public does mostly consume the increase in their income, instead saving a larger amount. This would have reduced the multicollinearity. It also indicates that this rise in income is mostly spent on goods that are imported in containers.

5.1.9 Mexico

For Mexico, there is a promising data set since there are 28 data points. The comprehensive model only includes Exports as explanatory variable. This is mainly caused by the low tolerance levels of the other independent variables. In the case of Unemployment, the tolerance levels are very high, but nevertheless it is of no significant influence to the dependent variable teupc. Export is significant (p=.000), but nearly non-existent (b_1=.000). Therefore I correct the model by regressing the natural logarithm of Exports (loge) on teupc. Because all other variables are significantly correlated with Exports, I cannot enter more variables into the model.

The fit of the model is good, since adjusted r^2=.826. The beta weight of loge is still significant (p=.000) and the weight is feasible (b_1=12.354). With a Durbin Watson test statistic of .139, there is reason to suspect autocorrelation. The test statistic is far below the lower critical value of 1.33 at N=28 and with one regressor. Therefore, I regress the residuals of the model on their one-period lagged values. It appears indeed that the beta weight is significant (p=.000) and has a weight of 1.063. Moreover, the one-period lagged residuals explain for a large part the variance in the residuals of the current period (adjusted r^2=.858). Therefore, I have to redo the model, including the one-period lagged residuals to see the real beta weight of log(Exports).

Both terms are significant (p=.000), but have sufficiently high multicollinearity terms (t=.966). The beta weight of log(Exports) has increased from 12.354 to 13.705. The model too has gained explanatory value: where the adjusted R^2 was .826, it is now .985. The lagged value of the residuals is added to the model.

Subsequently, the model must be tested for heteroskedasticity. Therefore, I square the residuals of the abovementioned model, the independent variable and the one-period lagged value of the residuals that are included in the model:

$$resid_2^2 = b_0 + b_1 logE + b_2 logE^2 + b_3 resid_{t-1} + b_4 resid_{t-1}^2 + b_5 logE * resid_{t-1} + e$$

The resulting regression has an r^2 of .180. Thus, since we have 27 data pairs, the W=27*.180=4.86. Since the null hypothesis is:

$$H_0: b_1 = b_2 = \cdots = b_5 = 0$$

We have 5-1=4 degrees of freedom. The corresponding critical X^2-value is 9.488. Since $W<X^2$, there is no indication of heteroskedasticity.

I calculate the predicted value for teupc, its squared and third power value using the model that was tested so far:

$$teupc = c + b_1 logE + b_2 resid_{t-1} + e$$

Therefore, the predicted value for teupc is:

$$\widehat{teupc} = -146.028 + 13.705\, logE + 1.113\, resid_{t-1}$$

By entering the squared and third power values to the model, a model with a SSE of 20.865 is obtained. Where the initial value for SSE was 46.328, the Ramsey RESET test statistic is:

$$F = \frac{(SSE_1 - SSE_2)/J}{SSE_2/(N-r)} = \frac{(46.328 - 20.865)/2}{20.865/(27-4)} = \frac{12.7315}{.9072} = 14.034$$

Since the critical F-value is 5.80, the model is convincingly misspecified. This suggest that the actual relation between the independent and the dependent variables is not linear. It could be that the estimation would improve when we would a squared or a third-power term.

Knowing this, I regress the log(Exports), squared log(Exports) on teupc. It is indeed significant (p=.000), though with a negative beta weight for log(Exports) (b_1=-163.932, but b_2=7.632). The adjusted R^2-value has become very high with ..975. Misspecification is logically high, because we have used the same value. Nevertheless, since the Ramsey RESET test revealed misspecification, I ignore the multicollinearity at hand.

The Durbin-Watson test statistic indicates autocorrelation in the error term (d=.521, which is lower than d_L=1,26., given N=28 and r=2). Therefore, I will regress the one-period lagged residuals of the second order model on their current values. It proves that the lagged values indeed have severe and considerable influence over the current errors, since the significance level of the beta weight is p=.000, and the explanatory value is adjusted r^2=.513.

By inclusion of these error terms, the model's fit increases to adjusted r^2=.988. All variables are significant at p=.000. The beta weight for log(Exports) has even decreased to -177.376 and log(Exports)^2 has increased to 8.214.

The second order model therefore can be described as:

$$teupc = b_0 + b_1 logE + b_2 logE^2 + b_3 resid_{t-1} + e$$

Where:

$$\widehat{teupc} = 959.580 - 177.376 * logE + 8.214 * logE^2 + .827 * resid_{t-1}$$

To test for heteroskedasticity in the error terms, I regress the squared residuals on the predicted value of teupc and its the squared value:

$$resid^2 = b_0 + b_1 \widehat{teupc} + b_2 \widehat{teupc}^2 + e$$

With an r^2 of .077, the W=27*.077=2.079. Considering that the null hypothesis for the White test is:

$$H_0: b_1 = b_2 = 0$$

The model thus has 1 degree of freedom. The critical X^2-value for one degrees of freedom is 3.841. Since $W<X^2$, there is no reason to suspect heteroskedasticity in the model.

With the new terms in the model, it is relevant to do another Ramsey RESET test to see whether this relationship is actually fitting the data. Therefore, I also submit a fourth power predicted value of teupc. Where the SSE was 35.022, it has decreased to 22.740. Therefore, the Ramsey RESET test statistic is:

$$F = \frac{(35.022 - 22.740)/2}{22.740/(27-4)} = \frac{6.141}{.989} = 6.211$$

Considering we have 27 data pairs and 4 regressors, the critical F-value of the relatively close $F_{30,4}$ is 5.75. Because $F>F_{max}$, there is still misspecification. Nevertheless, adding another term to the model based on the same independent variable wouldn't increase the adjusted R^2 any further. Nor will it give a more feasible relationship.

Concluding, the number of containers handled in Mexico is significantly dependent from the logarithmic and squared logarithmic values of exports. Container handling in Mexico is apparently predominantly influenced by the exports. An explanation for this might be that Mexico is situated very close to the United States of America. Because production costs in Mexico are lower than in the U.S.A., it might be that many companies have decided to allocate their production for the U.S.A. market in Mexico and export these goods from there to the U.S.A. The number of containers handled is highly affected by changes in Exports because this is mainly driven by increased container trade to the U.S.A.

5.1.10 Norway

In the case of Norway, the dataset is limited to ten data pairs. The initial comprehensive stepwise linear regression reveals that only GDP is of great influence to teupc. Although the tolerance levels of all other variables are fine, they are all far from significant. Independently, the correlogram indicates that only gdppc and GCI are not significant.

The simple linear regression of GDP on teupc has a higher explanatory value (adjusted r^2=.775), although one of the lowest up to now. The beta weight for GDP is significant (p=.000), but small (b_1=.001). Therefore, I redo the model with the natural logarithm of GDP: loggdp. This beta weight is much higher (b_1=.001) and still significant (p=.000), but the explanatory value of the model has decreased, though only marginally (adjusted r^2=.766).

The Durbin-Watson test statistic of 1.466 is indicating that there is no autocorrelation. A regression of the residuals on their one-lagged values is confirming that (p=.576). Therefore the model is:

$$teupc = b_0 + b_1 loggdp + e$$
$$teupc = -2398.667 + 199.813 * loggdp + e$$

A regression of the squared value of the residuals on loggdp and the squared values of loggdp:

$$resid^2 = b_0 + b_1 loggdp + b_2 loggdp^2 + e$$

results in an r^2 of .020. With only 10 data pairs, W=10*.020=.20. Since the hypothesis is:

$$H_0: b_1 = b_2 = 0$$

There is only one degree of freedom. The corresponding X^2-value is 3.841, which is much higher than W. Therefore, there is no significant heteroskedasticity in the model.

To test whether there is any misspecification, I calculate the predicted value of teupc (teupcpred) and its squared value and third power value (respectively teupcpredsq and teupcpredthrd). Where the SSE was 170.592, it has decreased to 129.239 by adding the two adapted predicted values to the model. Therefore, the Ramsey RESET test statistic is:

$$F = \frac{SSE_1 - SSE_2/J}{SSE_2/(N-r)} = \frac{170.592 - 129.239/2}{129.239/(10-3)} = \frac{20.6765}{18.463} = 1.12$$

With ten data pairs and three regressors, the critical F-value is 8.79. Since $F<F_{max}$, there is no indication of misspecification. Therefore, the relationship is linear.

Concluding, the number of containers per capita in Norway is positively influenced by the size of the national income. It relates linearly.

As a high-wage country, Norway is for a large part relying on import for its consumption goods. These products are generally not produced in Norway, which is to my knowledge typically more a service-oriented economy. Many independent variables did not have any effect on the number of containers imported. Even the income per capita was not of significant effect. A reason might be that gdppc and GDP are nearly perfectly correlated. In the few years in the sample, Norway barely accrued a higher income per capita.

5.1.11 Senegal

For Senegal, there are only six data points available. The initial stepwise comprehensive regression only includes Consumption and GCI as significant variables. The adjusted R^2 of this model is .969, which is very high. Both are significant at respectively p=.003 and p=.008. There is little multicollinearity, because the tolerance levels are high with t=.904. Because the beta weight for Consumption is very small (b_1=-6.34*10^{-9}), I correct the model by taking the natural logarithm of Consumption instead. The tolerance level decreases slightly, but is still very acceptable (t=.896). The explanatory value has increased insignificantly (R^2=.971).s

Unfortunately, the Durbin Watson test statistic is very high (d=3.178). The critical upper value is 1.54 for two regressors at N=15. A regression of the one period-lagged values of the residuals though shows no correlation on the current value of the residuals (p=.133). There is thus no sign of autocorrelation in the residuals.

Therefore, I can test the following model:

$$teupc = b_0 + b_1 logC + b_2 GCI + e$$

For testing the homoskedasticity, I calculate the predicted values for teupc based on Consumption and GCI:

$$\widehat{teupc} = 1484.154 - 58.794 logC - 28.539 GCI$$

These values are regressed on the squared value of the residuals. In this model, the r^2 is low with .179. Therefore, the White test statistic is W=6*.179=1.074. Since the assumption than is that in the model

$$resid^2 = b_0 + b_1\widehat{teupc} + b_2\widehat{teupc}^2 + e$$

The value of both b_1 and b_2 is nil, we have only one degree of freedom. The critical X^2-value is therefore 3.841, which is much higher than W. Therefore, there is no heteroskedasticity in this model.

Using the predicted value, I calculate the third power values. Adding teupcpredsq and teupcpredthrd to the model, I find that this regression has an SSE of .378, where it is 1.031 in the initial model. Therefore, the Ramsey RESET test statistic is:

$$F = \frac{\frac{1.031 - .378}{2}}{\frac{.378}{6-4}} = \frac{.3265}{.189} = 1.728$$

Since the critical F-value is 6.16, given that N=6 and r=4, $F<F_{max}$, thus there is no evidence for misspecification. The linear model seems to have the best fit.

Strikingly, the effect of a rise in Consumption has a negative effect on the number of containers that is handled for Senegal. This is against my expectations, because a rise in consumption of a population will give an absolute rise in imports, of which a part is containerised. An explanation may lay in the very limited number of data pairs. Since there are only six data pairs for Senegal, it is hard to do a reliable estimation of the relation. As can be observed in Figure 13 the relationship may actually be positive if the data pair at the left is ignored. Probably, this had the obscure effect on the relationship.

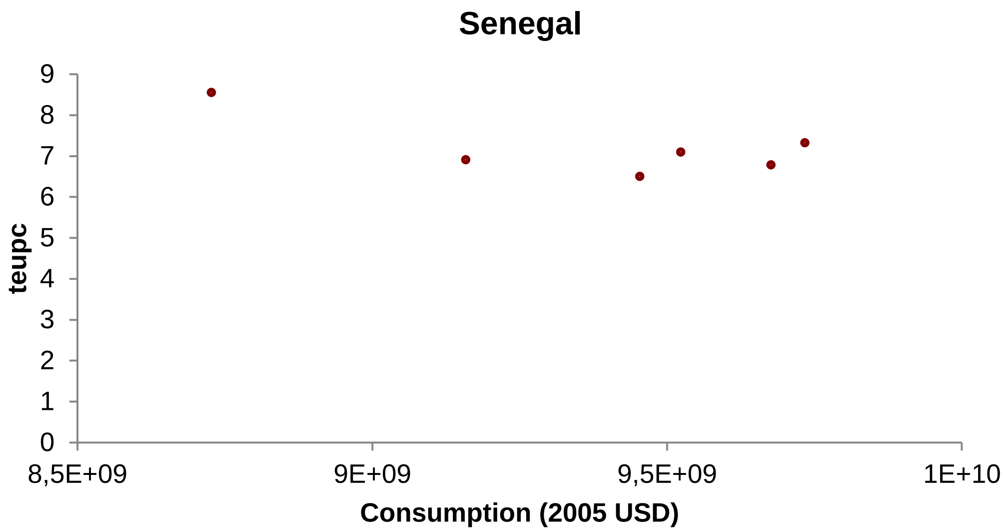


Figure 13 Data pairs for Senegal (own calculations, based on UNCTAD data)

Concluding, the Global Competitive Index and the natural logarithm of the Consumption both have a robust but negative effect on the number of containers handled in Senegal. This is a rather surprising result, since in most countries there would be a positive relation. An explanation might be the effect of a possible outlier or the small amount of data available for Senegal.

5.1.12 South Africa

For South Africa, there are only nine data pairs available. The initial comprehensive linear regression indicates that only gdppc has significant influence on teupc. The rest of the independent variables show either tolerance problems (e.g. Imports and Exports), or are not significant (Consumption).

In the corrected model, gdppc is highly significant (p=.000) and the model is highly explanatory (adjusted r^2=.901). The beta weight is feasible (b_1=26.267). The Durbin-Watson test statistic is with 1.436 though too high, since the upper critical level at N=15 with one regressor is 1.36. The regression of the one-period lagged residuals on the residuals though shows no autocorrelation (p=.646). Therefore, the model is:

$$teupc = b_0 + b_1 gdppc + e$$

And the estimation of teupc:

$$\widehat{teupc} = -71.637 + 26.267 gdppc$$

With the estimation, I calculate the squared values of the estimate. Furthermore, I take the square of the residuals. The r^2 of the regression of teupcpred and teupcpredsq on residsq is .074. Thus, the White test statistics is W=9*.074=.666. The critical X^2-value, given that we have nine data pairs and one degree of freedom, is 3.841. Since $W<X^2_{max}$, there is no indication of heteroskedasticity in the errors.

At last, I calculate the third power values of the predicted teupc. Adding both this value and teupcpredsq to the simple regression model gives an SSE of 42.703, where it was 48.878 initially. Therefore, the Ramsey TEST statistic is:

$$F = \frac{48.878 - 42.703/2}{42.703/(9-3)} = \frac{3.0875}{7.1172} = .434$$

Considering that N=9 and r=3, the critical F-value is 8.81. Because $F<F_{max}$, there is no indication of misspecification. Therefore, the relation is seemingly linear.

Concluding, the number of containers per capita in South Africa is driven by the income per capita in South Africa. This relationship is positive and robust.

An explanation might be that South Africa is increasingly consuming goods from abroad to satisfy their consumption, rather than producing them themselves. Another explanation might be that South Africans are earning more over time, because they are able to participate in international trade and export more products in containers.

5.1.13 United Kingdom

With thirty data pairs, the United Kingdom has the largest available amount of data on containers.

The initial comprehensive stepwise regression indicates that GDP is solely of significant influence on teupc. Most variables are excluded because they are insignificant, except for Consumption and gdppc, who have very low tolerance levels. Because the beta weight of GDP is too low, I will use the natural logarithm of this value.

The corrected model indeed shows a more feasible beta weight (b_1=123.042) in which loggdp is still significant (p=.000). The explanatory value of the model is high, as the adjusted r^2=.979 indicates. The Durbin-Watson test statistic indicates that there is no autocorrelation (d=1.496). This is confirmed by the insignificant effect of the one-period lagged residuals on the current residuals (p=.268).

I calculate the predicted value for teupc (teupcpred) based on the following model:

$$teupc = c + b_1 loggdp + e$$

$$\widehat{teupc} = -1672.087 + 123.042 * loggdp$$

By regressing the squared residuals of the model on teupcpred and its squared values, I obtain the r^2 that is required for the White test statistic:

$$resid^2 = b_0 + b_1 * \widehat{teupc} + b_2 * \widehat{teupc}^2 + e$$

R^2 is .078. With thirty data pairs, W=30*.078=2.34. Since the null hypothesis for this White test is:

$$H_0: b_1 = b_2 = 0$$

Because the White test is performed on two beta weights, there is one degree of freedom. The corresponding critical X^2-value is 3.841. Because $W<X^2$, the null hypothesis is not rejected. Thus, there is no indication of heteroskedasticity in the model.

For the Ramsey RESET test, I compute the third power values of teupcpred (teupcpredthrd). Regressing both loggdp and teupcpredsq and teupcpredthrd on teupc, the SSE drops from 546.618 to 436.812 compared to the simple linear model with only loggdp. Therefore, the Ramsey RESET test statistic is:

$$F = \frac{(546.618 - 436.812)/2}{436.812/(30-3)} = \frac{54.903}{16.18} = 3.394$$

Since N=30 and r=3, the critical F-value is 8.62. Because $F<F_{max}$, the model is not misspecified.

Concluding, the change in GDP has a robust and significantly positive effect on the number of containers per capita in the United Kingdom. Using the logarithmic values of GDP gives a more feasible beta weight.

Over the last thirty years, the industrial activity in the U.K. might have decreased because of the movement of production sites to low-wage countries. This has benefitted the U.K. on one hand, because many multinationals and related firms are based in the U.K. Therefore, the national income has risen. On the other hand, the U.K. became more dependent on the import of goods via containers rather than that they were produced in the country. This explains the relationship between the rise in income and the rise in imports per container per capita.

5.1.14 Conclusion and interpretation

Table 6 Summary of significant beta weights (sign if heteroskedastic model)

Dependent variable: teupc						
	Independent variable →					
↓Country	Gdppc	GDP	FDI	Cons	Exports	GCI
Argentina			+*			
Australia	9.769					
Brazil					17.306*	
France	5.239					
India	+**					
Japan	+				+*	
Kenya				26.000*		
Korea	33.446					
Mexico					(-177.376*, 8.214*)**	
Norway		199.813*				
Senegal				-58.794*		-28.539
South Africa	26.267					
United Kingdom		123.042*				
* logarithmic value of independent variable						
** polynomial relation (quadratic)						

The country's profiles reveal differences in the driving force of the containerised transportation booms that is observed. In Table 6, the results of the previous analysis are summarised.

For Argentina, the natural logarithmic value of FDI has the most significant and robust impact on the number of containers per capita that is handled in the country.

For Brazil, Japan, Mexico the natural logarithmic value of the exports have the most significant and robust impact on the number of containers per capita that is handled in the country. The beta weight for Brazil is 17.306. For Japan, also the size of the income per capita has a significant effect. The relation between the logarithmic value for the exports and teupc is convex for Mexico.

For Kenya and Senegal the natural logarithmic value of the total consumption has the most significant and robust impact on the number of containers per capita that is handled in the country. For Senegal, also the GCI has a significant and robust effect. The beta weight for Kenya is 26.000, where it is -58.794 for Senegal. The negative sign for Senegal can be explained by the limited amount of data available and the effect of a possible outlier on the model.

For Australia, France, India, Korea and South Africa the income per capita has the most significant and robust impact on the number of containers per capita that is handled in the country. The beta weight for Australia is 9.769, for France 5.239, for India and Japan positive, for Korea 33.446 and for South Africa 26.267. The size of the latter two is much larger than the others. I expect that these countries are developing both in their domestic economy and on the international market. The rise in the income for these countries thus is related to increased import and export of goods via containers. Thus, this income will be earned or spent for a quite significant part abroad. The established economies of France and Australia do not further develop in the same terms apparently. India will have a low beta weight because of the closed character of its economy.

In the case of Norway and the United Kingdom, it is not the income per capita, but the total national income that has the most significant effect. The beta weight for Norway is 199.813, where it is 123.042 for the United Kingdom. The difference is considerable. An explanation might be that the domestic production of goods in the United Kingdom is larger than in Norway. Therefore, Norway has to import more goods it consumes, which explains partly a greater influence of GDP on teupc.

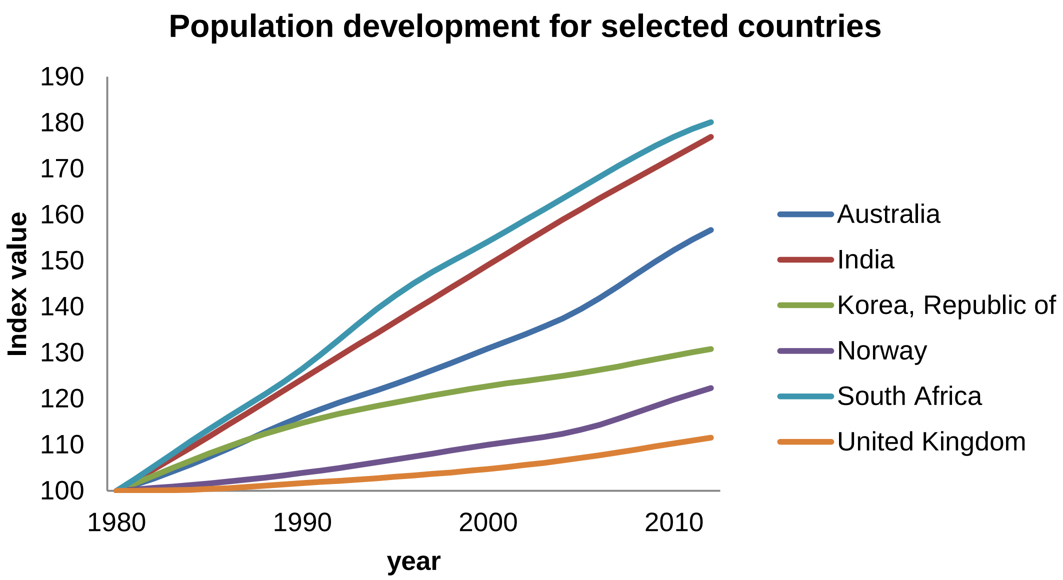


Figure 14 Population development (index) for selected countries (own calculations, based on UNCTAD data)

An explanation for this dichotomy of the average income-driven countries and GDP-driven countries might lay in the development of the population. As Figure 14 shows, the development of the population in the United Kingdom and Norway has developed slower than in the countries where the income per capita has a significant impact on the number of containers per capita handled in that country. In the stepwise regression, the independent variable that explains most of the changes in teupc is preferred over the one with which the SSE is larger. As was observed, the explanatory value of income-related variables is very large, mostly more than 90%. There is thus a very strong relationship between Containers and GDP. By correcting Containers for the number of people in the nation, for countries with large changes in their population over time, it is thus required to do the same for GDP. Otherwise also the variation in GDP that is explained by changes in the population is regarded.

The nature of the driver differs in this dataset. Where the size of the Exports and FDI is theoretically determined on the international trade markets, i.e. an exogenous factor of influence, Consumption and GDP or gdppc are influenced by the countries themselves because they are more influenced by domestic production and consumption. They are thus more endogenous. To investigate whether there is a relationship between the size of the beta and the fact that the driver is exogenous for a country (which I reflect in a binary variable for which 1=exogenous and 0=endogenous), I plot the two against each other. As indicated in Figure 15 there appears to be no relationship between the two. Neither does a relationship appear in Figure 16, reflecting the beta size relative to the nature of the driver.

Only the models of India, Japan and Argentina are affected by heteroskedasticity of the residuals. Therefore, in predicting future values the country-based models for these countries should not be used.

So what might determine the driver of adaptation of container transport technology in their trade systems?

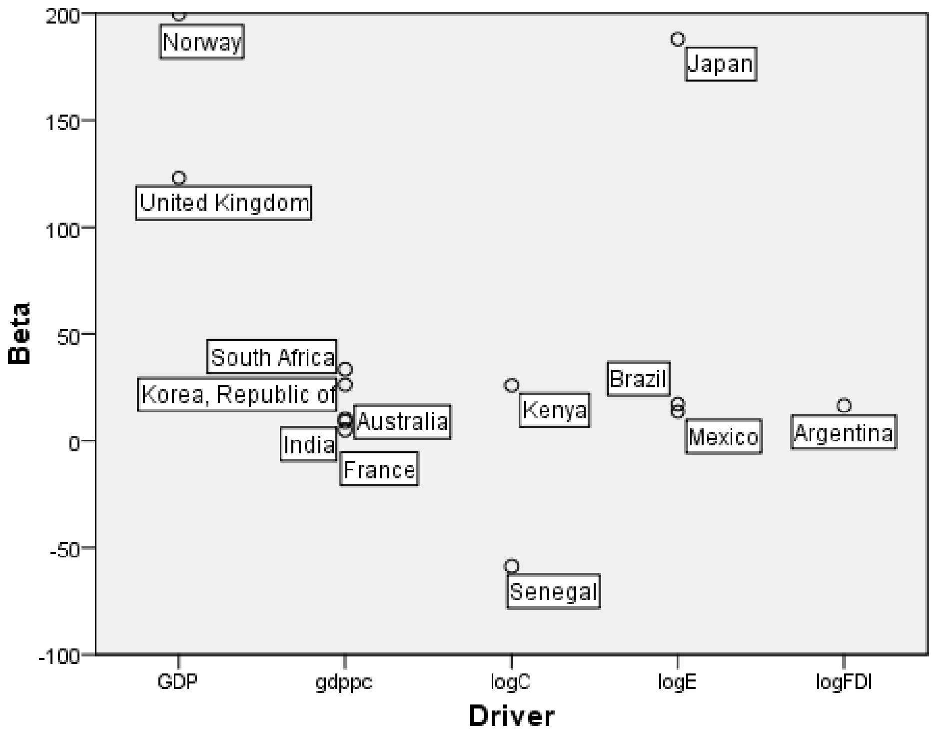


Figure 15 Scatterplot of driver and beta size (own calculations)

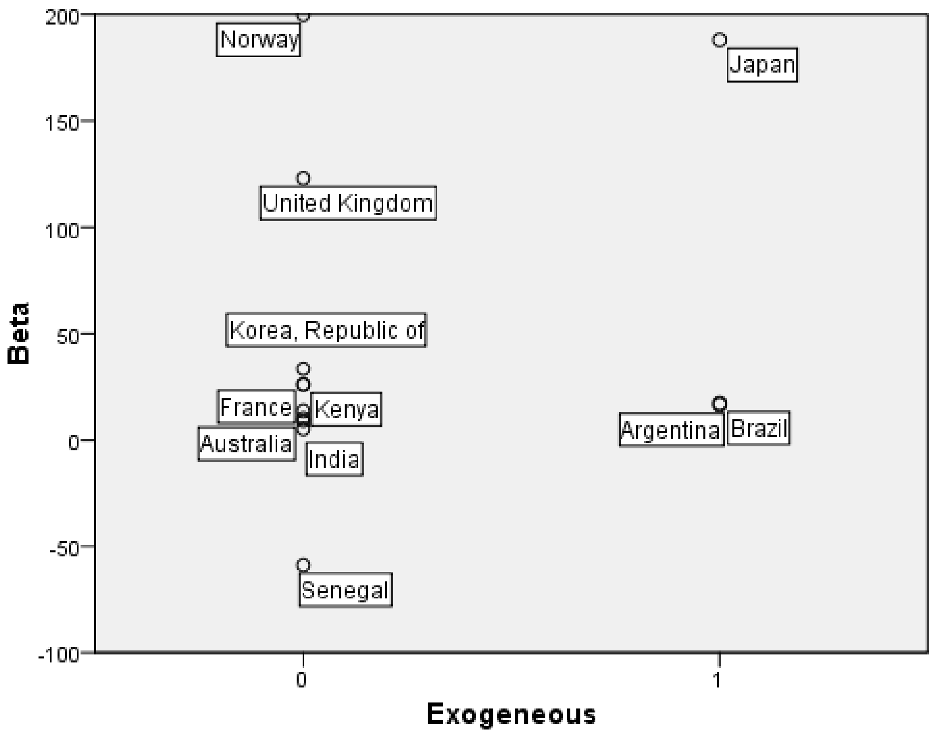


Figure 16 Beta size and nature of beta (own calculations)

5.2 Cross-section analysis

This section regards the analysis on a country-to-country basis of differences in levels of independent and dependent variables. All results and plots are included in Appendix B This analysis is used for validating hypotheses 1-5.

5.2.1 Correlation between variables

The correlogram of the dependent variable and independent variables shows that GDP is highly correlated with Exports, Imports, Consumption (Cons) and Containers (Cont). It is also significantly correlated with TIP, but in contrast to the other correlations is this negative of nature, which is logical since GDP is in the denominator of the TIP-ratio. Export and import are nearly perfectly correlated, appearing from the Pearson correlation coefficient of nearly one. The same holds for the relation between Consumption and GDP as well as their values corrected for the population sizes. The size of the labour force appears to be highly correlated with the size of the total population with a Pearson value of .998. Surprisingly, also import is highly explanatory for the consumption (or vice versa), because the Pearson value is .909. Exports and Imports have both a significant and similar effect on the number of handled containers, although the explanatory power is about .75.

On the basis of the correlogram, I created two graphic representations of the significant relations between the factors (Figure 17 and Figure 18). Variables that are very close next to each other have a Pearson correlation approaching one. If these variables are seen as one, the picture in Figure 18 emerges, showing a simplified representation of relations between variables. GDP and Consumption, as well as Cpop & gdppc, Pop & Labourf and Imports & Exports are expected to be critically multicollinear. Therefore, no models should be accepted in which both of the pairs are expressed.

The scatter plot of the economic variables with teupc is displayed in Report 2 of Appendix B. It shows the relative position of each of the countries with regard to a set of two variables. It appears that between the average annual personal consumption (Cpop) and the average annual income per person (gdppc), there is virtually a straight line, indicating that as the income rises, the consumption rises with about the same amount. As shown in many of the country-based analyses, if a model includes both gdppc and Cpop there is a critical degree of multicollinearity. The relation with the two economic indicators and teupc is therefore also highly similar. There is thus no evidence that one of the two (Cpop or Consumption or gdppc or GDP) explains the changes in teupc better than the other variable.

From Figure 18 we can distil there is also a significant relation between Unemployment and Urbanisation, which appears to be negative (Pearson=-.742). Moreover, Urbanisation is positively correlated with Cpop and gdppc as well as with GCI. Also teupc is correlated with GCI. A regression of GCI on teupc nevertheless reveals that indeed the correlation is significant (p=.024), but that the explanatory power of the model is very limited (adjusted R^2=.330). Even when I include an interaction dummy for competitive economies or a coefficient dummy for accounting for a different level in teupc, the explanatory power of the model does not rise above an adjusted R^2 of .413 which is too low to have a good fit.

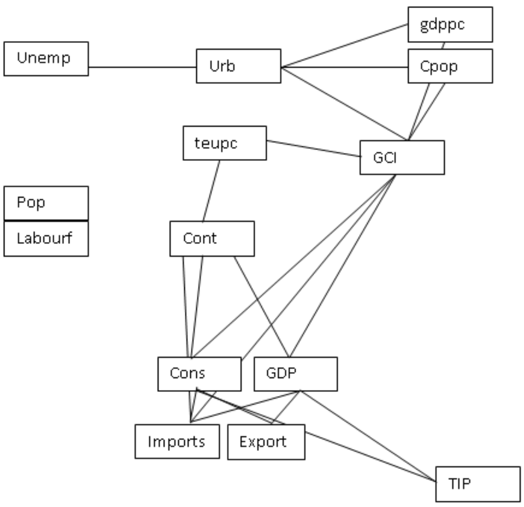


Figure 17 Graphical representation of correlations

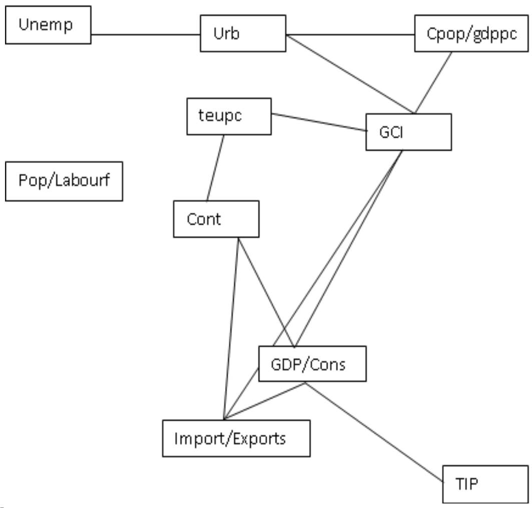


Figure 18 Simplified graphical representation of correlations

5.2.2 The effects of income and income groups

The regression of the dummy variables for gdppc as described in Table 5 (p. 32) reveals that G2 is automatically excluded because it is monotonic. Therefore, I will use G2 instead of G5 as reference group. Only G3 and G5 are significant. I combine G1 and G3, and G4 and G5, but the combined term for G4 and G5 is excluded automatically because of multicollinearity. Therefore, I combine groups G1, G2 and G3 and use the combined groups G4 and G5 as reference groups. Both the interaction variable and the independent variable gdppc are significant (p=.003 respectively p=.005) and both beta weights are positive. The adjusted R^2 of the model is sufficiently high with .638. The tolerance levels are high with t=.939, showing only marginal multicollinearity. The regression of the one-period lagged residuals on the residuals itself renders no significant relation. Therefore, there is no indication of autocorrelation.

For investigating whether the model is homoskedastic, I use the White test statistic. I regress the squared residuals on the predicted value of teupc according to the model and the square root of the predicted value:

$$resid^2 = b_0 + b_1\widehat{teupc} + b_2\widehat{teupc^2} + e$$

The r^2 value of this model is .197. With an N=13, the W=13*.197=2.561. The X^2 critical value is correspondingly 3.841. Since $W<X^2$, there is no indication of heteroskedasticity.

By adding the squared and third power predicted value to the model, I test whether the model might be misspecified. Without these terms, the SSE is 61,452.091 where it is 31,650.806 when these terms are added. Therefore, the Ramsey RESET test statistic is:

$$F = \frac{61452.091 - 31650.806/2}{31650.806/(13-4)} = \frac{14900.6425}{3516.756} = 4.237$$

Because N=13 and r=4, the critical F-value of $F_{13,4}$ can be approached by $F_{12,4}$, which is 5.91. Because $F<F_{12,4}$, there is no indication of misspecification.

Concluding, the model:

$$teupc = -20.430 + 4.106gdppc + 14.746(G1 + G2 + G3) * gdppc + e$$

could be understood such that an increase in personal income offsets a larger growth in the number of containers per capita in low-income countries (gdppc≤30.000) than in high-income countries (gdppc>30.000). Therefore, the cross-country analysis provides evidence that the opposite of hypothesis 3 is true.

5.2.3 The effect of an urbanised population

The correlogram (Report 1) indicated that there was no significant correlation between the degree a country is urbanised and the number of TEU handled per thousand inhabitants. It does thus indicate that more urbanised countries do not necessarily make greater use of the container.

A regression of the independent variable Urbanisation on teupc indeed indicates that changes in teupc can only marginally be explained by changes in Urbanisation, because adjusted R^2=.156. Moreover, the model does not give a reliable beta weight for Urbanisation since it is not significant (p=.100). Therefore, this model is not to be interpreted further. (Report 3)

There is no evidence from the cross-country analysis that the degree of urbanisation of a country does have a significant effect on the number of containers handled per capita. The cross-country analysis does not provide any evidence that hypothesis 4 is true.

5.2.4 The effect of a higher trade intensity

As with urbanisation, the correlogram already indicated that there is no significant relation between the Trade Intensity Parameter (TIP) and the number of containers handled in a country per thousand inhabitants (teupc). Such is confirmed by the regression of TIP on teupc. The explanatory value of the model is negligible with an adjusted R^2 of .007. Also, the parameter is not significant (p=.320). Therefore, there is no indication that TIP can explain any variation in the level of container handling per country. The cross-country analysis does not provide any evidence to confirm that hypothesis 5 is true. (Report 3)

5.2.5 Conclusion and interpretation

There is only correlation between the GCI of a country and the number of containers handled in a year with regard to teupc. The latter is rather logical, because it forms the basis for calculation of the ratio.

The variables Consumption and GDP, Imports and Exports, Cpop and gdppc and Pop and Labourf will cause multicollinearity when occurring together in a model. Their correlation is nearly perfect. Therefore, only one of the two should be regarded in a model. The variation in Consumption therefore is nearly fully the same as the variation in GDP. Both have a similar effect on teupc. Therefore, the cross-country analysis indicates that Consumption is not necessarily a better indication for changes in teupc than GDP either on an individual level (Cpop and gdppc) or national level (Consumption and GDP). Therefore, I find no proof for hypothesis 2 in the cross-country analysis.

Low- and middle-income countries (gdppc≤30.000 USD) develop teupc quicker than high-income countries given a certain rise in welfare. Possibly, there are more economies of scale that can be gained in upcoming economies which container transport creates, than in developed economies. This effect is contrary to what I expected in hypothesis 3. The model indicates that the effect of a change in the personal income is positive for all countries and income groups, but stronger for low- and middle-income countries. Therefore, there is reason to believe hypothesis 3 should be rejected, but hypothesis 1 should be accepted according to the cross-country analysis.

In this sample, the level of urbanisation of a country did not have to seem a significant effect on the level of containerised transportation in that country. Neither was TIP of significant influence on the level of teupc. The cross-country analysis did thus not give proof for hypotheses 4 and 5.

5.3 Panel data analysis

This paragraph regards the analysis of hypotheses 1-3 and 5, and hypothesis 4 regarding the effect of an increase in urbanisation. All data results are reported in Appendix C.

5.3.1 Income and consumption effects

Hypothesis 2 supposes that consumption is a better explanatory variable than the average personal income. Statistically, this could be proven if, when both a model with Consumption and a model with gdppc have significant beta weights for the independent variables, the explanatory power for the model of Consumption is higher than the model of gdppc. Both models must in this case be complying with the Gauss-Markov Theorem. (Carter Hill et al., 2008) Because of the correlation that was noticed in paragraph 5.2.1 and in many instances in the country-based analysis, a model with both Consumption and gdppc is not reliable and can therefore not be interpreted.

The regression of Consumption on teupc in a random effect model renders a very low adjusted r^2 of .197. A Hausman test though indicates that the model should contain fixed effects rather than random effects (p=.001). Indeed, the explanatory value of the fixed effects model of Consumption increases significantly to .815 (adjusted R^2). Also the beta weight of Consumption is significant (p=.000), but is very small and therefore not feasible ($.121 \cdot 10^{-10}$). I redo the analysis with the logarithmic value of Consumption, to account for scale effects. The explanatory value of the model increases slightly to .835 with this adaption, but the beta weight becomes much more feasible (b_1=126.748) and still significant (p=.000). A regression of the one period lagged residuals on the residuals reveals that there is autocorrelation affecting the beta weight of log(Consumption). A corrected model has a nearly perfect fit (Adj. r^2=.990) and two significant beta weights. The beta weight of log(Consumption) has slightly increased to 127.453.

Because the r^2 of the regression of the predicted and squared predicted values on the squared residuals is .022, the White test statistic is W=.022*13=.286, since there are thirteen countries in the panel data analysis. The critical X^2-value is 3.841, indicating that there is no heteroskedasticity in the residuals.

The SSE in the Ramsey RESET-model is only slightly lower with 15620.87 than the initial 16606.13 in the model. Therefore:

$$F = \frac{\frac{16606.13 - 15620.87}{2}}{\frac{15620.87}{13-4}} = \frac{492.63}{1735.652} = .284$$

With N=13 and r=4, the critical F-value approaches $F_{12,4}$=5.19. Since $F<F_{12,4}$, there is no misspecification of the model. Therefore, the model:

$$teupc = -3365.005 + 127.453 * \log(Consumption) + .986 * resid_{t-1} + e$$

Is robust and has a high explanatory power with adjusted r^2=.990.

Subsequently, I regress gdppc on teupc using a random effects model. Also in this case, the Hausman test is significant, indicating that a fixed-effects model should be used (p=.000). The corrected model has a high fit (adjusted R^2=.873), but shows autocorrelation (p=.000). The inclusion of the residual term renders the following model:

$$teupc = -62.363 + 8.048gdppc + .979resid_{t-1} + e$$

The fit of the model is very high with an adjusted R^2 of .991. The regression of the predicted and squared predicted values on the squared residuals has a fit of .185. Subsequently, the White test statistic is W=13*.185=2.405. Because the critical X^2-value is 3.841, $W<X^2$ and there is no indication of heteroskedasticity. For testing for misspecification, I regress the squared and third power predicted values of the model with gdppc and the one-period lagged residuals on teupc. The SSE for the initial model was 14.305,19, where in the Ramsey RESET model it is 14.086,52. Therefore:

$$F = \frac{\frac{14305.19 - 14086.52}{2}}{\frac{14086.2}{13-4}} = \frac{218.67}{1565.133} = .140$$

Because $F<F_{12,4}$, there is no indication of misspecification. This model is thus robust and confirming to the Gauss-Markov Theorem. Apart from the question whether or not Consumption is a better explanatory variable, the model with gdppc is thus significant, has a good fit and is robust. The effect is positive. The increase of gdppc with a thousand USD thus offsets an increase of the number of containers per thousand people of eight.

Since both models are complying with the Gauss-Markov Theorem, both beta weights can be used to estimate teupc. The fit of the model with gdppc is only marginally better than the model with log(Consumption). An F-test can indicate whether the fit of the model with log(Consumption) is better than the model with gdppc:

$$F = \frac{(SSE_{gdppc} - SSE_{\log(Consumption)})/J}{SSE_{\log(Consumption)}/(N-r)}$$

Where J is the number of variables in the hypothesis, N is the number of individuals in the panel dataset and r is the number of parameters in the model of log(Consumption). The hypothesis is that the model of Consumption is significantly better than the model of gdppc. Formally:

$$H_0: {r^2}_{\log(Consumption)} = {r^2}_{gdppc}$$

$$H_1: {r^2}_{\log(Consumption)} \neq {r^2}_{gdppc}$$

Using the values for SSE obtained from the model renders the following F-test statistic:

$$F = \frac{14305.19 - 16606.13/2}{16606.13/13 - 2} = -\frac{1150.47}{1509.65} = -.762$$

Since N=13 and r=2, the critical F-value is approached by $F_{12,2}$=19.41. Because $F<F_{12,2}$, there is no proof for rejection of the null hypothesis. Therefore, there is no evidence that the predictive value of gdppc is different from log(Consumption). Hypothesis 2 can thus not be proven.

Moreover, the panel data prove that there is a positive relation between gdppc and teupc. The model corrected for autocorrelation of the residuals reveals that an increase of the average personal income with 1,000 USD realises a growth in the number of containers handled of eight per thousand inhabitants. Although there are fixed effects for the level of teupc, the trend is common to all countries in the sample as indicated by the Hausman test. Hypothesis 1 is thus confirmed.

5.3.2 Income group effects

Hypothesis 3 assumes a difference in the influence of gdppc on teupc for low- and high-income countries. Using the dummy variable as described in Table 5 on page 32, I will use the panel data to test for different levels of teupc for different groups and different trend estimates.

The panel data have already presented evidence for a positive relation between gdppc and teupc. As discussed in the cross-section analysis of paragraph 5.2.2, there are coefficient and trend effects.

First I will investigate whether there is a difference in coefficients. The dummy variable for countries with an average personal income of 40.000 USD or higher will be used as a reference category. All dummy variables are significant, but the explanatory value of the model is slightly lower than the critical value (adjusted R^2=.471). Thus, it is not useful to include income-group specific coefficients in a model because apparently different levels of gdppc do not sufficiently explain differences in teupc.

The inclusion of interaction terms between gdppc and the dummy variable might indicate differences in trends in growth for several countries. Indeed, all beta weights are significant except for G1. Combining G1 and G2 though reduces the explanatory value such, that it is below an adjusted R^2 of .5 since adjusted R^2=.438. Consequently, I incorporate the effect of G1 in G5 by excluding its interaction term. The resulting model reveals that the interaction term gdppc*G3 is not significant. Therefore, I combine G2 with G3. The same problem as with G1 occurs. G3 is therefore excluded from the model. The resulting model both has sufficient explanatory value (adjusted R^2=.530) and significant beta weights for the interaction term of G2 and G4. I test for autocorrelation by taking the one-period lagged residuals and regressing it on the residuals. The beta weight is significant (p=.000) and seems highly explanatory for the variation in the errors because the adjusted R^2=.860. The one-period lagged residuals term is included in the model and increases the explanatory value of the model significantly (adjusted R^2=.935). Still, all beta weights are significant but differ slightly from the model without $resid_{t-1}$.

The regression of the predicted and squared predicted values of teupc on the squared residuals results in a fit of .538 (R^2). With an N=13, the White test statistic is W=13*.538=6.994, which is more than the critical X^2 value with one degree of freedom of 3.841. Therefore, there is indication of heteroskedasticity. The interpretation of the results should be done with care and the beta weights should only be limitedly interpreted because the model is not the best linear unbiased estimator.

The SSE for the Ramsey RESET model is 108018.7, where it was initially 110491.2 for the model with gdppc, gdppc*G2, gdppc*G4 and the one-period lagged residuals variable. Therefore, the F-value is:

$$F = \frac{\frac{110491.2 - 108018.7}{2}}{\frac{108018.7}{13-4}} = \frac{1236.25}{12002.08} = .103$$

The critical F-value is approached by the value for $F_{12,4}$=5.91. Because $F<F_{12,4}$, there is no indication of misspecification.

To be able to draw any conclusions on differences of growth rates, I do Wald tests to compare the beta weights of gdppc with gdppc*G2 and gdppc*G4. The null hypothesis $b_{gdppc} = b_{gdppc*G2}$ is convincingly rejected with a p=.000. $b_{gdppc*G2}$ is thus larger than b_{gdppc}. The Wald test also confirms that $b_{gdppc} \neq b_{gdppc*G4}$ with p=.000. Using the same method, it appears that also both interaction terms have a significantly different beta weight. The growth rates for countries with an average personal income between 10.000-20.000 USD, is thus significantly larger than all other country groups. Also the growth rate for countries with an average personal income between 30.000-40.000 USD is significantly larger than other income groups, but smaller than for countries with an average personal income between 10.000-20.000 USD. Moreover, it must be concluded that there is no evidence for different growth rates amongst the other income groups. Hypothesis 3 should thus be rejected, although there is a difference between growth rates of different income groups.

5.3.3 Urbanisation effects

Hypothesis 4 compares the effect of an increase in consumption in a relatively more urbanised population with a more rural population. Although panel data is not suitable for measuring cross-country variation in absolute terms, I can test whether a change in urbanisation combined with a change in consumption has a positive effect on the number of containers handled. There might be other indicators influencing the relative size of containerised trade. Therefore, I have to correct for this country-specific variation with either a random effects model or a fixed effects model.

I regress the interaction term of urbanisation and consumption (urb*cons) on teupc in a random effects model. The Hausman test statistic indicates that a fixed effects model should be used (p=.001).

Furthermore, I test whether the effect of the interaction term is solely caused by the relation between consumption and teupc. The question is whether there is a spurious relation. Therefore, I include both the interaction term and Consumption. Would there indeed be a spurious relation, than the independent variable Consumption would be much more suitable to explain variances in teupc than the interaction term. Therefore, the inclusion of the independent variable would make the interaction term insignificant. As the regression shows, this is though not the case. Consumption has a p-value of .578 where the interaction term has a p-value of .035. There is thus no evidence of a spurious relation with Consumption.

The resulting fixed effects model has a good explanatory power (adjusted R^2=.817) and the beta weight for the interaction term is significant (p=.000). The beta weight is positive, but not feasible ($b_1=1.41*10^{-12}$). Therefore, I alter the interaction term by taking the natural logarithm of Consumption. Indeed, this beta weight is more feasible with a value of .323.

The regression of the one-period lagged residuals on the residuals though gives proof of autocorrelation of the errors (p=.000) which is highly explanatory for the variation in the errors (adjusted R^2=.932). The inclusion of the one-period lagged residual term significantly increases the explanatory value of the model (adjusted R^2=.989), but does not have an effect on the significance of the interaction term. The beta weight of the interaction term remains the same.

The regression of the model with the squared predicted values has an R^2 of .776. The White heteroskedasticity test statistic is therefore 10.088, which is much larger than the critical X^2 value of 3.841, given there is only one degree of freedom. Therefore, there is heteroskedasticity in the model.

The regression of the model with the squared and third power predicted value on teupc results in the Ramsey RESET model. Its SSE is 15651.97, which is naturally less compared to 16735.83 in the original model. The F-value of the Ramsey RESET test statistic is:

$$F = \frac{\frac{16735.83 - 15651.97}{2}}{\frac{15651.97}{13-4}} = \frac{541.93}{1739.11} = .312$$

Since $F_{12,4}$=5.61 approaches the critical F-value of $F_{13,4}$ and $F<F_{12,4}$, there is no indication of misspecification of the model.

Although the model is not robust, the explanatory power of the model is high. There is a relation between the interaction term of urbanisation and the natural logarithm of consumption with teupc. This relation is positive. Therefore, if both urbanisation and consumption rises, this will have a positive effect on teupc. What the size of the effect is, cannot be told because of the heteroskedasticity in the model. This model provides proof for a variant of hypothesis 4.

5.3.4 Trade intensity effects

Hypothesis 5 regards the effect of more intense trade activity relative to the size of the economy on the number of containers handled in a country per thousand people. Using panel data, I can regard the question whether an increase in TIP has a positive effect on teupc.

The regression of TIP on teupc reveals that the use of random effect models is justified (p=.104). Nevertheless, the explanatory value of the model is very limited and too low for consideration (adjusted R^2=.389). The fixed effects model contrastingly shows a much better and sufficient fit (adjusted R^2=.855) and a significant beta weight (p=.000). I thus have to continue with the fixed effects model since the random effect model does not have sufficient power. An explanation might be that size of the effect of a change in TIP on teupc is to a great extent country-specific.

The beta weight of the fixed effects model is feasible with a value of 172.260. The regression of the one-period lagged errors on the errors reveals that the model is coping with autocorrelation. Therefore, the one-period lagged values of the errors are included in the model. The resulting beta weight for TIP is 152.855 and still significant. The model fit has increased to adjusted R^2=.985.

Regressing TIP and $resid_{t-1}$ on the squared residuals including the squared predicted values, results in a model with an R^2 of .120. The White test statistic is therefore W=13*.120=1.56, which is lower than the critical X^2 value of 3.841. Therefore, there is no indication of heteroskedasticity.

Concluding, I regress teupc on the predicted values of the model, the squared predicted values and the third power predicted values. The resulting model has an SSE of 18906.01, where the initial model had an SSE of 23937.38. Therefore, the Ramsey RESET test statistic is:

$$F = \frac{\frac{23937.38 - 18906.01}{2}}{\frac{18906.01}{13-4}} = \frac{2515.685}{2100.67} = 1.198$$

Since F is smaller than $F_{12, 4}$=5.61, the value that approaches the critical F-value of $F_{13, 4}$, there is no indication of misspecification.

Lastly, I would like to point out that the trade intensity parameter is a bit ambiguous: it does not regard differences in effects of increases in imports or exports, nor in changes of the gross domestic product. One could say a country's container handling is too: container handling regards both export and import containers. Nevertheless, in paragraph 3.2 it was discussed that exports and imports follow a similar trend and are highly collinear. It might therefore be of little added value to include import or export, since the same trend will be visible in the trade intensity parameter as when only imports or exports is included.

Still, the model discussed is robust and shows a significant and positive relation between the trade intensity parameter and the number of containers handled per thousand inhabitants. An increase in trade intensity of .1 will increase container handling with about 150 containers per thousand inhabitants. The Hausman test statistic indicates that countries do vary from this trend, but that this variation is limited. The model gives support to hypothesis 5.

5.3.5 Conclusion and interpretation

Concluding, the panel data analysis provides evidence in support of hypothesis 1. The growth of the average personal income with 1.000 USD will result in an increase of teupc in about eight containers per thousand inhabitants. This effect is robust and positive.

Hypothesis 2 should be rejected according to the panel data analysis. The comparison of the model with gdppc as independent variable and the model with consumption does not lead to the conclusion that consumption is a better explaining variable than gdppc.

Hypothesis 3 should formally be rejected according to the panel data analysis. Countries in lower-income groups do not have in common that there growth rates are significantly lower than the growth rates in high-income groups. Nevertheless, a robust model is found that indicates that countries in the income group between 10.000-20.000 and 30.000-40.000 USD have significantly higher growth rates at the same increase in average personal income of the population. Moreover, the increase in the first group is significantly larger than in the second. Because of heteroskedasticity in the model, it is not possible to interpret the beta weights further.

The effects of a change in consumption in a rural and more urban country do differ significantly. The model with the interaction term of urbanisation and the natural logarithm of consumption supports hypothesis 4, although this model is heteroskedastic. On the basis of the model, one could say that at constant consumption levels or similar increase in consumption, more urbanised countries show a larger (increase in) teupc than more rural countries.

The change in the trade intensity of the country has a positive and significant effect on the number of containers handled in that country according to hypothesis 5. The model describing the effect of TIP on teupc does confirm this statement. The model is robust.

6 Conclusions

This chapter will summarize the main results obtained in this research using the hypotheses that were formulated in paragraph 2.4.

In Argentina, FDI is the most significant driver for containerized trade. Exports is most significant for Brazil, Japan and Mexico where the effect of a change in exports on teupc is significantly larger in Japan than in Brazil. The amount that is consumed in an economy is the most significant factor of influence for Senegal and Kenya. Nevertheless, it must be noted that for both countries the number of data pairs is very limited, which might be the reason for the counterintuitive beta weight of Senegal. A representation of its data pairs shows that its negative sign might be caused by one outlier, where the actual trend is more stable which might explain the significance of GCI.

Most countries in the sample (Australia, France, India, Korea and South Africa) have the average personal income as the main driver for the container trade developments in their economy. For the United Kingdom and Norway, this is the national income, but this might be because Population has little influence on the value of both teupc and gdppc. I.e. gdppc and GDP follow the same trend where in other countries the growth of the population has an effect on the trend of gdppc compared to GDP.

Concluding, the country-based analysis suggests that hypothesis 1 is true for most of the countries in the sample. The cross-country analysis shows that indeed the average personal income explains for a large part the differences in levels of teupc. The panel data analysis confirms the latter conclusion, but adds that the increase in gdppc has a monotonic effect on teupc for the whole sample, regardless of the country's teupc level. If the average income rises with a thousand USD, in general the teupc increases with eight containers per thousand inhabitants. There might be differences in trend growth rates between countries, but these are not significant. Therefore, this research provides evidence confirming hypothesis 1.

Hypothesis 2 regards whether the consumption in an economy has a better explanatory effect of changes in the number of containers handled per thousand inhabitants than the average personal income in that country. In the country-based analysis, the regression revealed large multicollinearity between Consumption and GDP in many instances, but not between Consumption and gdppc. Only for Senegal and Kenya, the explanatory power of Consumption was larger than other included independent variables. The cross-country analysis reveals nearly perfect multicollinearity between Consumption and GDP and between Cpop and gdppc. Therefore, the cross-country analysis proves that the factors can be used alternatively. Moreover, the model with Consumption as explanatory variable has got equal power as a model with gdppc in the panel data analysis. Therefore, no evidence is found in support of hypothesis 2.

Hypothesis 3 regards the divergence in the effect of the average personal income for low- and high-income countries. As an indication, the country-based analysis shows a large beta weight for South Africa (b_1=26.267), but a small beta weight for higher-income countries like Australia (b_1=9.769) or France (b_1=5.239). These beta weights suggest that there indeed might be a difference amongst income groups. In the cross-country analysis, there can indeed be found evidence for a difference in effect between low- and high-income countries. Countries with a gdppc≤30.000 USD are found to have significantly higher teupc levels than countries with a higher average personal income. This is contrary to what the hypothesis suggests. The panel data analysis indeed reveals that for countries with an average personal income between 10.000-20.000 and 30.000-40.000 USD the growth rates are significantly larger and that this difference is significantly larger in the first group than in the second, but that the lowest and intermediary income group do not have significantly different growth rates. Therefore, hypothesis 3 should be rejected. The analysis provides evidence that suggests that lower income countries are picking up quickly in the development of the container volumes, but that countries with a very low income are still lacking behind.

Hypothesis 4 regards the difference in effect of consumption changes for rural and urbanized countries on the container volumes in a country per thousand inhabitants. The cross-country analysis finds no proof in support of this hypothesis. The panel data analysis though shows that for a country the increase in urbanization combined with an increase in consumption will have a positive effect on the container volumes. Therefore, where urbanization might not explain for differences between countries in the use of container transport in their economies, it might explain for differences over time. Hypothesis 4 should therefore be accepted with regard to time-series variations in teupc, but not cross-country variances in teupc levels.

Hypothesis 5 regards the effect of the trade intensity parameter, measuring the size of the values for export and import on the total national income, on the container volumes. Where for a few countries in the country-based analysis exports appears to be of significant influence on teupc, imports is for neither of them. For Senegal, the competitiveness of the economy is although of significant influence. The cross-country analysis finds no significant correlation between teupc and TIP. Therefore, also this analysis does not provide any evidence in support of hypothesis 5. Nevertheless, the panel data analysis finds a positive and significant relation between TIP and teupc. The analysis provides proof in support of hypothesis 5. An increase in TIP with .1 will increase the number of handled containers with 17 per thousand inhabitants, but the Hausman test indicates that there is country-specific variance from this trend. Hypothesis 5 should therefore be accepted, although TIP explains developments in teupc for a country over time and not the different container volume levels compared to other countries.

Overall, it appears that none of the countries deviate from the common trend significantly. The only trend deviations can be found with regard to the influence of gdppc on teupc for different income groups and there is a suggestion of small trend deviations for TIP. Nevertheless, growth factors in all countries seem to be comparable and have a strong relation with the development of the container handling in the country. The urbanisation can cause relative higher growth rates for some countries. The cross-country analysis finds that socio-economic variables can explain for differences in the level of teupc.

As an answer to the research question, the development of the average income per capita has the most pronounced effect on the development of the number of twenty-foot equivalent units handled in a country. The growth rate is significantly higher for countries with an average income between 10.000 and 20.000 USD and slightly higher for countries with an average income between 30.000 and 40.000 USD. The consumption of the population has a similar and positive effect as the national income. The effect of consumption on the container volumes is amplified for countries by urbanisation developments. If countries engage more in international trade, i.e. the relative size of the values for export and import to the national income grows, the amount of containers handled per capita in that country increases.

7 Discussion

Although evidence has been found for most of the hypotheses or small variances on these hypotheses, the consequences of this confirmation should be put into the perspective of the advantages and disadvantages of this research.

A disadvantage of influence on the results might have been the differences in data availability for the countries in the sample. Where for some countries only six data pairs are available (e.g. for Senegal), for other countries there are longer time series available (e.g. the United Kingdom). Therefore, for countries with small time series, the panel data analysis provides less evidence of a relation than countries with longer time series. The validity of the conclusions for the latter is stronger than for the first. Improvement of the data set was not possible by using other sources such as the national databases within this research due to the lack of access to certain promising databases or language barriers. In most instances, the data availability in the databases used in this thesis is the same as the national or other databases that I have been able to access. Nevertheless, macro-economic research into the development of containerized trade has only recently gained attention of only a few researchers (e.g. Rua, 2012; Bernhofen et al., 2013). This research is highly explorative. It provides a first analysis of key indicators for containerized trade developments in different economies across the world.

The results of this research indicate that policy makers, investors and companies must consider country-specific circumstances for estimating the feasibility of local investments in containerized trade facilities. Not only economic but also socio-demographic variables are of influence on the development of the local container industry. Founding the decision to invest in a container terminal solely on the expectation of increased exports in a country has proven to be an insufficient argument. Rather, expectations of welfare and consumption and the ability of a country to urbanize should be regarded, as well as the level of average personal income in the country. These factors seem to become of increased importance in countries where investments bare more risk due to political factors or lack of existing infrastructure, which is the case for most low-income countries.

A remarkable conclusion is that lower-income countries have exceedingly high growth rates compared to other income groups, even compared to low-income countries (average income <10.000 USD), indicating that there is a transition phase where the increase in income has an extraordinary effect on the container volume developments. The same holds for higher-income countries (20.000-30.000 USD), although these growth rates do not exceed normal growth rates as excessive as the lower-income groups. An explanation could be that this is a ‘catching-up development’, but this is left to further research to substantiate. Another reason could be that these economies could be more focussed on the production of actual goods for supply to the international market, where other economies are more locally or service oriented.

Using three methods for regarding the data of the sample makes that all hypotheses have been abundantly checked for both country-specific and time-specific effects. Some of the results reveal that where there are country-specific effects, there are no time-specific effects and vice versa. Therefore, the different analysis methods have both a complementary and correcting function that secure the results further.

It must be noted that the sample of countries is not taken random. The countries included in the analysis have been selected by me on the basis of their geographical position, apparent transhipment function and data availability. Of each of the continents, several countries are selected. Moreover, countries along the main trade routes are excluded. The goal of these criteria is that the container data must reflect reactions on variations in economic factors which are not diluted by changes in factors that influence transhipment.

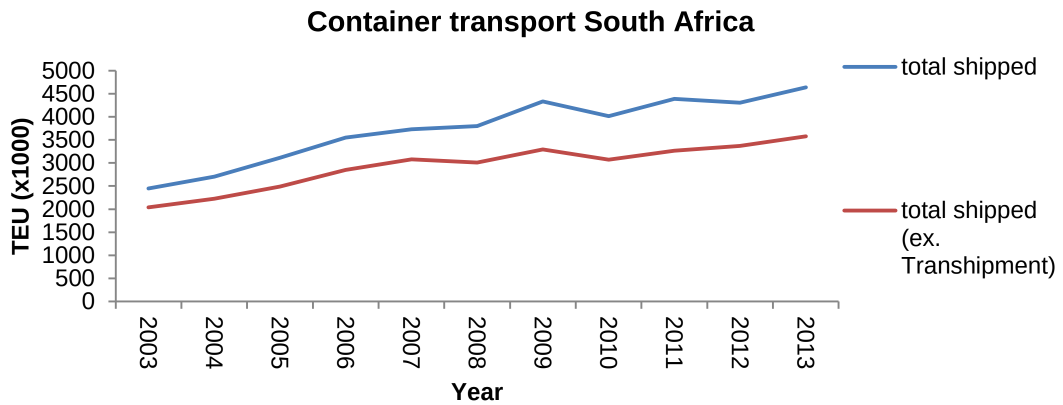


Figure 19 Container data for South Africa (own calculations, based on Transnet (2004-2013) data)

It can be argued that South Africa is not a suitable country for the sample because of its geographic position along the Southern trade route via Cape of Good Hope. An analysis of the container transport data of Transnet (2004-2013) shows that the trend in total shipped containers shows a slightly steeper but similar development as the shipped containers where transhipped containers are excluded. (Figure 19) Therefore, I concluded that the transhipment containers had a minor effect on container trade development in South Africa. On the other hand, it might have had some effect on the level of container handling compared to other countries in the sample. The data of Transnet (2004-2013) have been used as reference data to assess the feasibility of South Africa in the sample. For the analyses, the same data as for the other sources have been used.

The data availability for African countries is limited. Most countries have only six data pairs. Nevertheless, it is preferred to include these countries for geographical representation than to exclude them from the analysis. Containerised transport has not yet caught the loop there. The inclusion of these countries in the analysis can help to distil factors that explain for the lack of containerised transportation in these regions.

The analyses revealed significant importance of the effect of the level of average personal income on the number of containers that are handled per thousand inhabitants. Neither this level, nor the impact of the change in the variable, is better explained by changes in the consumption of the population. Nonetheless, the growth of containerised trade is larger in certain income groups than in others. A reason for this might be that where people with an average income lower than 10.000 USD tend to spend an increase of income mostly on primary goods such as food and drinks that is either locally produced or not imported ‘in boxes’, i.e. with containers, people with higher income levels spend more of their income on luxury goods, non-food products or processed foods that are imported in boxes. The differences on expenditures on food for different income groups have been found by scientists for different societies, among others Putnam & Allshouse (1994) for the United States. It does though not explain why the third income group of people with an income between 20.000-30.000 USD show the same growth patterns as the lowest and highest income groups. There might be an explication in the role fixed costs play in deciding to expand containerised transport facilities. Intermediate economies might experience higher income risks for container terminal operators or shippers since demand might drop below a breakeven point, but I lack evidence for this statement.

The urbanisation of population only has an invigorative effect on the growth of container handling due to consumption or income growth. Urbanisation itself does not explain changes in container transportation, though it seems to spur gains in efficiency. Such seems logical since the container is often associated with large production and demand poles, but is less feasible for delivering goods to small societies with little demand of that product. Figuratively speaking, a container full of soap bars will be sold much easier in a large city as Amsterdam than in a small village such as Twello. This seems nevertheless to contradict the findings of Jacoby & Minten (2009), who found that rural populations have the largest gains in wealth with container trade.

A country's involvement in international trade seems surprisingly not to have any significant correlation with the size of the container transportation in that country. There might be several reasons for these findings. The first and most sobering reason could be that the ratio used to measure the intensity of international trade in a country is fallible. As Lane David (2007) discusses, this measure for openness of a country results that do not fit expectations. Rodriguez (2000) indicates that there is a negative correlation between the parameter and the size of the economy, regardless of its relative openness. Therefore, only very robust and consistent results could be used for drawing conclusions in this regard. Lane David (2007) contrastingly indicates that the use of this measure is well-accepted by the scientific society and can be used in some instances where trade policy is not regarded, such as in this thesis. Therefore, the initial decision to use this parameter was in itself justified, but did not result in solid and significant results.

The research offers new insights in the development of the container volumes for low-income countries and the growth rates that accompany this development. New is also the amplifying effect of urbanisation on consumption growth with regard to container volume developments. It would surprise only few that the average income has great impact on the development of container volumes. Nevertheless, the research provides new insights in the strength of this relationship, regarded in combination with other independent variables such as Consumption.

At last, although there have been found some significant, interesting and robust relations between the variables that contribute to today's view on both macro-economic developments and investment planning policies, it must be noted that these results are exploratory of nature. As has been discussed before, little research has been on the impact of container trade on economies and welfare developments. Therefore, further research should focus on finding more reliable and substantiated explanations for the revealed relations to overcome the sometimes necessarily superficial discussion of the nature of relations between the variables. Moreover, attention should be paid to the expansion of the dataset.

References

Airries, C.A. (1989), The spatial spread of container transport in a developing regional economy: North Sumatra, Indonesia, *Transportation Research part A*, 23:6, p. 453-461.

Bell, M.G.H., X. Liu, P. Angeloudis, A. Fonzone & S.H. Hosseinloo (2011), A frequency-based maritime container assignment model, *Transportation Research part B: Methodological*, 45:8, p. 1152-1161.

Beresford, A., S. Pettit, Q. Xu & S. Williams (2012), A study of dry port development in China, *Maritime Economics & Logistics*, 14:1, p. 73-98.

Bernhofen, D.M., Z. El-Sahli & R. Kneller (2013), *Estimating the effects of container revolution on world trade*, Munich: CESifo Working Paper Trade policy, No. 4136.

Bhargava, A., L. Franzini & W. Narendranathan (1982), Serial Correlation and the Fixed Effects Model, *The Review of Economic Studies*, 49:4, p. 533-549.

Borensztein, E., J. de Gregorio & J. Lee (1998), How does foreign direct investment affect economic growth?, *Journal of International Economics,* 45, p. 115-135.

Bottema, U.H. (2014), *Terminal operations & Logistics* (presentation), Rotterdam: ECT.

Button, K. (2010), *Transport Economics,* Cheltenham: Edward Elgar, 3rd edition.

Carter Hill, R., W.E. Griffiths & G.C. Lim (2008), *Principles of Econometrics*, New York: John Wiley & Sons, Third edition.

Chaudhry, A. & P. Garner (2006), Political Competition Between Countries and Economic Growth, *Review of Development Economics,* 10(4), p. 666-682.

Department for Business Innovation & Skills (2013), *Openness to trade: exports plus imports as a share of GDP, ranked against major competitors* (Business Performance Indicator), London: Government of the United Kingdom.

Durbin, J. & G.S. Watson (1951), Testing for Serial Correlation in Least Square Regression. II, *Biometrika*, 38:1-2, p. 159-177.

Field, A. & J. Miles (2010), *Discovering Statistics using SAS,* London: Sage Publications Ltd.

Guerrero, D. & J-P Rodrigue (2014), The waves of containerization: shifts in global maritime transportation, *Journal of Transport Geography*, 34, p. 151-164.

Gupta, V. (2000), *SPSS for beginners*, VJBooks Inc.

Haralambides, H.E., ‘Structure and operations in the liner shipping industry’, in: Hensher, D.A. & K.J. Button (2007) *Handbook of Transport Modelling*, Oxford: Elsevier, p. 607-622.

Helpman, E., M.J. Melitz & S.R. Yeaple (2004). Export versus FDI with Heterogeneous Firms, *The American Economic Review*, 94(1), p. 300-316.

Hummels, D., J. Ishii & K. Yi (2001), The Nature and growth of vertical specialization in world trade, *Journal of International Economics*, 54, p. 75-96.

Hummels, D. (2007), Transportation Costs and International Trade in the Second Era of Globalization, *The Journal of Economic Perspectives,* 21:3, p. 131-154.

Hummels, D. (2009), The trade reducing effects of market power in international shipping, *Journal of Development Economics,* 89, p. 84-97.

Ikiara, M.M. & J.M. Mutua (2006), Identifying Complementary Measures to Ensure the Maximum Realisation of Benefits from the Liberalisation of Trade in Environmental Goods and Services. Case Study: Kenya (Trade and Environment Working Paper No. 2004-02), France: OECD.

Irwin, D.A. & M. Terviö (2002), Does trade raise income? Evidence from the twentieth century, *Journal of International Economics,* 58:1, p. 1-18.

Jacoby, H.G. & B. Minten (2009), On measuring the benefits of lower transport costs, *Journal of Development Economics*, 89, p. 28-38.

Japan International Freight Forwarders Association, *Transhipment Containers via Korean Ports Climb 8.9% in Jan-Sept,* JIFFA 1st of November 2013. Accessed online via http://www.jiffa.or.jp/en/news/entry-2531.html on the 6th of August 2014.

Konings, R. & R. Thijs (2001), Foldable Containers: a New Perspective on Reducing container-Repositioning Costs, *EJTIR*, 4, p. 333-352.

Lane David, H. (2007), *A Guide to measures of Trade Openness and Policy*, Indiana University South Bend.

Lim, S. (1998), Economies of scale in container shipping, *Maritime Policy & Management*, 25:4, p. 361-373.

Mansfield, E.R. & B.P. Helms (1982), Detecting multicollinearity, *The American Staticstician,* 36:3, p. 158-160.

Marine Traffic (2014), *Owned Fleet of Largest container lines,* Marinetraffic.com. Accessed online via http://www.marinevesseltraffic.com/2013/07/owned-fleet-of-largest-container-lines.html on the 18th of August 2014.

Marrewijk van, C. (2012), *International Economics*, Oxford University Press: Oxford, 2nd edition.

Midoro, R., E. Musso & F. Parola (2005) Maritime liner shipping and the stevedoring industry: market structure and competition strategies, *Maritime Policy & Management*, 32:2, p. 89-106.

Molnar, M., N. Pain & D. Taglioni (2007), *The internationalisation of Production, International Outsourcing and Employment in the OECD*, OECD Working Papers, No. 561.

Putnam, J.J. & J.E. Allshouse (1994), *Food consumption, Prices and Expenditures, 1970-93*, Washington: U.S. Department of Commerce.

Rodriguez, C.A. (2000), *On the Degree of Openness of an Open Economy*, Buenos Aires: Universidad del CEMA.

Rua, G. (2012), Fixed costs, Network Effects and the International Diffusion of Containerisation.

Russo, F. & G. Musolino (2013), Geographic Factors Affecting the Presence of Transhipment Services in Regional Maritime Container Markets, *Geographical Analysis,* 45, p. 90-102.

Talley W.K. (2000), Ocean Container Shipping: Impacts of a Technological Improvement, *Journal of Economic Issues*, 34:4, p. 933-948.

Transnet (2004-2013), *Summary of cargo handled at ports of South Africa*, Pretoria: Transnet. Accessed online via http://www.transnetnationalportsauthority.net/DoingBusinesswithUs/Pages/Port-Statistics.aspx on the 20th of July 2014.

Venables, A.J. (2005), Spatial disparities in developing countries: cities regions and international trade, *Journal of Economic Geography*, 5, p. 3-21.

Verny, J. (2007), The importance of decoupling between freight transport and economic growth, *EJTIR*, 7:2, p. 113-128.

White, H. (1980), A heteroskedasticity-consistent covariance matrix estimator and a direct test for heteroskedasticity, *Econometrica*, 48:4, p. 817-838.

World Economic Forum (2014), *The Global Competitiveness Index data forum*, weforum.org. Accessed via http://www.weforum.org/issues/competitiveness-0/gci2012-data-platform/ on the 19th of August 2014.

World Shipping Council (2014), *About the industry*, worldshipping.org. Accessed via http://www.worldshipping.org/about-the-industry on the 19th of August 2014.

Yang, C.C., Tai, H.H. & Chiu, W.H. (2014), Factors influencing container carrier's use of coastal shipping, *Maritime Policy & Management,* 41:2, p. 192-208.

Appendices

Appendix A Country-based regression results

Model description				Beta weights (if significant) and p-values			Model characteristics	
Parameters	IV	DV	Test statistic	B_1 (p)	B_2 (p)	B_3 (p)	(Adj.) R^2	SSE
Argentina								
FDI	FDI	Teupc	Full stepwise	.002 (.001)			.894	
FDI	FDI	Teupc	Simple linear	.002 (.001)			.751	
Log(FDI)	FDI	Teupc	Simple linear	16.558 (.001)			.818	67.589
log(FDI), $teupred^2$	FDI	$resid^2$	White heteroske-dasticity				.760	
log(FDI), teup-$cpred^2$, $teupcpred^3$	FDI	Teupc	Ramsey RESET					60.301
Australia								
Gdppc, Consump-tion	Gdppc, Consump-tion	teupc	Full stepwise	9.028 (.000)	$1.32*10^{-6}$ (.005)		.983	
Gdppc, log(Consumption)	Gdppc, Consump-tion	teupc	Multiple regression	7.920 (.000)	417233.476 (.000)		.994	
Gdppc	Gdppc	Teupc	Simple linear	9.485 (.000)			.992	
$Resid_{t-1}$	Resid	Resid	Autocorrelation	.690 (.005)			.369	
Gdppc, $resid_{t-1}$	Gdppc	Teupc		9.769 (.000)	.775 (.002)		.996	$9,1*10^{10}$
Gdppc, $resid_{t-1,}$	gdppc	$Resid^2$	White heteroske-				.095	

Model description				Beta weights (if significant) and p-values			Model characteristics	
Parameters	IV	DV	Test statistic	B_1 (p)	B_2 (p)	B_3 (p)	(Adj.) R^2	SSE
teupcpred^2			dasticity					
Gdppc, resid$_{t-1}$, teupcpred^2, teupcpred^3	gdppc	Teupc	Ramsey RESET					$7,5*10^{10}$
Brazil								
Exports	Exports	Teupc	Full stepwise	$7.654*10^{-5}$ (.001)			.973	
log(Exports)	Exports	Teupc	Simple linear	17.306 (.000)			.943	
Resid$_{t-1}$	Resid	Resid	Autocorrelation	(.980)				
log(Exports), teupcpred^2	Exports	Resid^2	White Heteroskedasticity				.317	
France								
Consumption, Labourf, gdppc	Consumption, Labourf, gdppc	teupc	Full model	$-6.901*10^{-11}$ (.005)	.006 (.015)	5.531 (.020)		
Exports, gdppc	Exports, gdppc	Teupc	Multiple regression	$1.799*10^{-5}$ (.024)	3.993 (.000)			
Log(Exports), gdppc	Exports, gdppc	Teupc	Multiple regression	4.068 (.000)	(.123)			
Gdppc	Gdppc	Teupc	Simple linear	5.037	(.000)		.946	
Resid$_{t-1}$	Resid	Resid	Autocorrelation	.439 (.021)			.238	
Gdppc, resid$_{t-1}$	Gdppc	Teupc		5.239 (.000)	.434 (.024)		.967	138.976

Model description				Beta weights (if significant) and p-values			Model characteristics	
Parameters	IV	DV	Test statistic	B_1 (p)	B_2 (p)	B_3 (p)	(Adj.) R^2	SSE
Gdppc, $resid_{t-1}$, $teupcpred^2$	Gdppc	$Resid^2$	White heteroske-dasticity				.008	
Gdppc, $resid_{t-1}$, $teupcpred^2$	Gdppc	Teupc	Ramset RESET					128.377
India								
Gdppc	Gdppc	Teupc	Full stepwise	3.897 (.005)			.780	
Gdppc	Gdppc	Teupc	Simple regression	8.403 (.000)			.963	
$Resid_{t-1}$	Resid	Resid	Autocorrelation	.902 (.000)			.625	
Gdppc, $resid_{t-1}$	Gdppc	Teupc		7.909 (.000)	.999 (.000)		.989	.749
Gdppc, $resid_{t-1}$, $teupcpred^2$	Gdppc	$Resid^2$	White heteroske-dasticity				.408	
Gdppc, $resid_{t-1}$, $teupc^2$, $teupc^3$	Gdppc	Teupc	Ramsey RESET					.319
Gdppc, $gdppc^2$	gdppc	Teupc	Misspecification	18.441 (.000)	-6.747 (.000)		.986	
Japan								
Gdppc, Exports	Gdppc, Exports	Teupc	Full stepwise	4.781 (.000)	.000 (.004)			
Gdppc, log(exports)	Gdppc, exports	Teupc	Multiple regression	3.868 (.000)	188.017 (.000)		.810	2032.088
$Resid_{t-1}$	Resid	Resid	Autocorrelation	.308 (.158)				
Gdppc, loge, $gdppc^2$, $loge^2$,	Gdppc, loge	$Resid^2$	White heteroske-dasticity				.557	

Model description				Beta weights (if significant) and p-values			Model characteristics	
Parameters	IV	DV	Test statistic	B_1 (p)	B_2 (p)	B_3 (p)	(Adj.) R^2	SSE
loge*gdppc Gdppc, loge, $teupc^2$, $teupc^3$	Gdppc, loge	Teupc	Ramsey RESET					1917.441
Kenya								
Consumption	Consumption	Teupc	Full stepwise	$1.211*10^{-9}$ (.002)			.908	
Log(Consumption)	Consumption	Teupc	Simple linear	26.000 (.003)			.890	2.275
$Resid_{t-1}$	Resid	Resid	Autocorrelation	(.403)				
Consumption, $teupcpred^2$	Consumption	$Resid^2$	White heteroskedasticity				.177	
Consumption, $teupcpred^2$, $teupcpred^3$	Consumption	Teupc	Ramsey RESET					.688
Korea								
FDI, gdppc	FDI, gdppc	Teupc	Multiple regression	(.295)	34.999 (.000)		.986	
FDI, Consumption	FDI, Consumption	Teupc	Multiple regression	(.645)	$1.004*10^{-9}$		.964	
FDI, log(Consumption)	FDI, log(Consumption)	Teupc	Multiple regression	.001 (.637)	528.042 (.000)		.954	
Log(Consumption)	Consumption	Teupc	Simple linear				.956	
Gdppc	Gdppc	Teupc	Simple linear	33.446 (.000)			.980	5817.354
Gdppc, $gdppc^2$	gdppc	$Resid^2$	White heteroskedasticity				.018	
Gdppc, $teupcpred^2$, $teupcpred^3$	Gdppc	Teupc	Ramsey RESET					4774.471

Model description				Beta weights (if significant) and p-values			Model characteristics	
Parameters	IV	DV	Test statistic	B_1 (p)	B_2 (p)	B_3 (p)	(Adj.) R^2	SSE
Mexico								
Exports	Exports	Teupc	Full stepwise	.000 (.000)			.981	
Log(Exports)	Exports	Teupc	Simple linear	12.354 (.000)			.829	
$Resid_{t-1}$	Resid	Resid	Autocorrelation	1.063 (.000)			.858	
Log(Exports), $resid_{t-1}$	Exports	Teupc		13.705 (.000)	1.113 (.000)		.985	46.328
Log(exports), $resid_{t-1}$, $teupcpred^2$	Consumption	$Resid^2$	White heteroske-dasticity				.230	
Log(exports), $resid_{t-1}$, teup-$cpred^2$, $teupcpred^3$	Exports	Teupc	Ramsey RESET					20.865
Log(Exports), $Log(Exports)^2$	Exports	Teupc	Simple exponential	-163.932 (.000)	7.632 (.000)		.494	35.022
$Resid_{t-1}$	Resid	Resid	Autocorrelation	.789 (.000)			.989	
Log(Exports), $Log(Exports)^2$, $Resid_{t-1}$	Exports	Teupc		-177.376 (.001)	8.214 (.001)	.827 (.958)	.988	
Log(Exports), $Log(Exports)^2$, $Resid_{t-1}$, teup-$cpred^2$	Exports	$Resid^2$	White heteroske-dasticity				.077	
Log(Exports), $Log(Exports)^2$, $teupcpred^2$, teup-	Exports	Teupc	Ramsey RESET					22.740

Model description				Beta weights (if significant) and p-values			Model characteristics	
Parameters	IV	DV	Test statistic	B_1 (p)	B_2 (p)	B_3 (p)	(Adj.) R^2	SSE
cpred³								
Norway								
GDP	GDP	Teupc	Full stepwise	.001 (.039)			.528	
Log(GDP)	GDP	Teupc	Simple linear	199.813 (.001)			.766	170.592
$Resid_{t-1}$	Resid	Resid	Autocorrelation	(.576)				
Log(GDP), teupcpred²	GDP	$Resid^2$	White heteroskedasticity				.020	
Log(GDP), teupcpred², teupcpred³	GDP	Teupc	Ramsey RESET					129.239
Senegal								
Consumption, GCI	Consumption, GCI	Teupc	Full stepwise	$-6.34*10^{-9}$ (.003)	-29.137 (.008)		.969	
Consumption, GCI	Consumption, GCI	Teupc	Multiple regression	$-6.340*10^{-9}$ (.003)	-29.137 (.008)		.969	
Log(Consumption), GCI	Consumption, GCI	Teupc		-58.794 (.003)	-28.539 (.008)		.971	1.031
$Resid_{t-1}$	Resid	Resid	Autocorrelation	(.133)				
Log(Consumption), GCI, teupcpred²	Consumption, GCI	$Resid^2$	White heteroskedasticity				.179	
Log(Consumption), GCI, teupcpred², teupcpred³	Consumption, GCI	Teupc	Ramsey RESET					.378
South Africa								

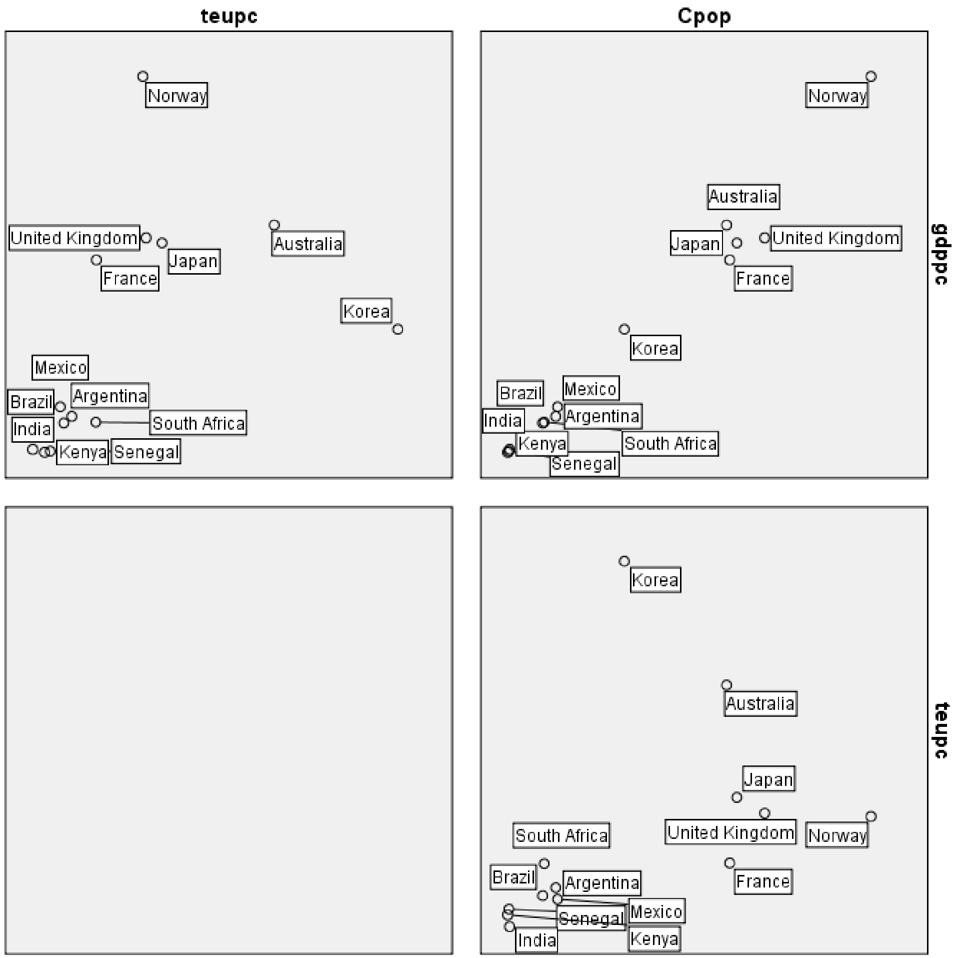
Model description				Beta weights (if significant) and p-values			Model characteristics	
Parameters	IV	DV	Test statistic	B_1 (p)	B_2 (p)	B_3 (p)	(Adj.) R^2	SSE
Gdppc	Gdppc	Teupc	Full stepwise	19.367 (.018)			.649	
Gdppc	Gdppc	Teupc	Simple linear	26.267 (.000)			.901	48.878
$Resid_{t-1}$	Resid	Resid	Autocorrelation	(.646)				
Gdppc, $gdppc^2$	Gdppc	$Resid^2$	White heteroskedasticity				.074	
Gdppc, $teupcpred^2$, $teupcpred^3$	Gdppc	Teupc	Ramsey RESET					42.703
United Kingdom								
GDP	GDP	Teupc	Full stepwise	.000 (.002)			.919	
Log(GDP)	GDP	Teupc	Simple linear	123.042 (.000)			.979	546.618
$Resid_{t-1}$	Resid	Resid	Autocorrelation	(.268)				
Log(GDP), $log(GDP)^2$	GDP	$Resid^2$	White heteroskedasticity				.078	
Log(GDP), $teupcpred^2$, $teupcpred^3$	GDP	Teupc	Ramsey RESET					436.812

Appendix B ***Cross-country analysis results***

Report 1 ***Correlogram of variables***

correlogram		E	I	Pop	FDI	Cons	Cpop	Cont	GCI	Urb	Lf	TIP	Unemp	teupc	gdppc
GDP	Pearson	.889**	.914**	.111	.148	.997**	.442	.693**	.655*	.416	.140	-.594*	-.238	.179	.372
	P-value	.000	.000	.719	.629	.000	.131	.009	.015	.157	.648	.032	.508	.559	.211
I	Pearson	.990**		.211	.254	.909**	.448	.735**	.735**	.400	.233	-.399	-.273	.325	.377
	P-value	.000		.489	.403	.000	.125	.004	.004	.176	.444	.177	.445	.278	.204
Pop	Pearson	.113			.096	.080	-.326	.076	-.137	-.458	.998**	-.137	-.079	-.298	-.318
	P-value	.712			.755	.794	.277	.804	.656	.116	.000	.655	.829	.322	.289
FDI	Pearson	.240				.172	.252	.068	.366	.350	.126	-.410	-.154	.126	.221
	P-value	.430				.573	.406	.825	.219	.241	.682	.164	.670	.683	.469
Cons	Pearson	.884**					.444	.662*	.647*	.428	.111	-.616*	-.211	.143	.367
	P-value	.000					.129	.014	.017	.145	.719	.025	.558	.641	.217
Cpop	Pearson	.513						.205	.876**	.583*	-.322	-.096	-.388	.443	.992**
	P-value	.073						.502	.000	.036	.283	.755	.268	.130	.000
Cont	Pearson	.759**							.578*	.402	.105	-.148	-.306	.672*	.189
	P-value	.003							.038	.174	.733	.629	.390	.012	.537
GCI	Pearson	.787**								.619*	-.122	-.181	-.347	.621*	.854**
	P-value	.001								.024	.692	.554	.326	.024	.000
Urb	Pearson	.474									-.431	-.446	-.742*	.476	.559*
	P-value	.102									.142	.127	.014	.100	.047
Lf	Pearson	.135										-.171	-.147	-.295	-.316
	P-value	.660										.576	.685	.328	.293
TIP	Pearson	-.330											-.135	.300	-.029
	P-value	.271											.711	.320	.924
Unemp	Pearson	-.324												-.340	-.417
	P-value	.362												.337	.231
teupc	Pearson	.413													.478
	P-value	.161													.099
gdppc	Pearson	.448													
	P-value	.125													

Report 2 Scatter plots of economic indicators



Report 3 Regression analyses

Model description				Beta weights (if significant) and p-values				Model characteristics	
Parameters	IV	DV	Test statistic	B_1(p)	B_2 (p)	B_3(p)	B_4 (p)	(Adj.) r^2	SSE
GCI	GCI	teupc		127.023 (.024)				.330	
Gdppc, PI123	gdppc	teupc		4.106 (.005)	14.746 (.003)			.638	61452.091
$resid_{t-1}$	gdppc	resid	Autocorrelation	-.110 (.727)					
P1, P3, P4, P5	gdppc	teupc	Dummy significance	(.847)	19.062 (.001)	(.143)	2.918 (.035)		
$\widehat{teupc}^3$, gdppc, PI123, $\widehat{teupc}^2$	gdppc	teupc	Ramsey RESET						31650.806
Urbanisation	Urbanisation	teupc		(.100)				.156	
TIP	TIP	teupc		(.320)				.007	

Appendix C Panel data regression results

Model description					Beta weights (if significant) and p-values				Model characteristics		
Parameters	IV	DV	Test statistic	B_1 (p)	B_2 (p)	B_3 (p)	B_4 (p)	B_5 (p)	(Adj.) r^2	SSE	Type (Hausman p)
Consumption	Consumption	teupc							.197		Random (.001)
Consumption	Consumption	teupc		$.121E^{-10}$ (.000)					.815		Fixed
Log(consumption)	Consumption	teupc		126.748 (.000)					.835		Fixed
$Resid_{t-1}$	resid	resid	Autocorrelation	.986 (.000)					.927		Fixed
Log(consumption), $Resid_{t-1}$	Consumption	teupc		127.453 (.000)	.986 (.000)				.990	16606.13	Fixed
Teupcpred, $teupcpred^2$	Consumption	$Resid^2$	White heteroskedasticity						.022		
Log(consumption), resid(t-1), $teupcpred^2$, $teupcpred^3$	Consumption	teupc	Ramsey RESET							15620.87	
Gdppc	Gdppc	Teupc							.390		Random (.000)
Gdppc	Gdppc	Teupc		10.372 (.000)					.873		Fixed
$Resid_{t-1}$	resid	resid	Autocorrelation	.983 (.000)					.995		Fixed
Gdppc, $Resid_{t-1}$	gdppc	teupc		8.048 (.000)	.979 (.000)				.991	14.305.19	Fixed
Teupcpred, $teupcpred^2$	gdppc	$Resid^2$	White heteroskedasticity						.185		Fixed

Model description					Beta weights (if significant) and p-values				Model characteristics		
Parameters	IV	DV	Test statistic	B_1 (p)	B_2 (p)	B_3 (p)	B_4 (p)	B_5 (p)	(Adj.) r^2	SSE	Type (Hausman p)
Gdppc, $resid_{t-1}$, teupcpred^2, teupcpred^3	gdppc	teupc	Ramsey RESET							14086.52	Fixed
G1, G2, G3, G4, gdppc	gdppc	teupc		13.970 (.060)	185.149 (.000)	33.653 (.028)	56.794 (.000)	2.335 (.000)	.471		OLS
Gdppc, gdppc*G1, gdppc*G2, gdppc*G3, gdppc*G4	gdppc	teupc		2.031 (.000)	.708 (.745)	12.356 (.000)	.839 (.176)	1.647 (.000)	.530		OLS
Gdppc, gdppc*(G1+G2), gdppc*G3, gdppc*G4	gdppc	teupc		2.640 (.000)	11.287 (.000)	1.697 (.009)	2.120 (.000)		.453		OLS
Gdppc, gdppc*G2, gdppc*G3, gdppc*G4	gdppc	teupc		1.997 (.000)	12.204 (.000)	.759 (.181)	1.597 (.000)		.532		OLS
Gdppc, gdppc*(G2+G3), gdppc*G4	gdppc	teupc		1.723 (.000)	2.613 (.000)	1.533 (.001)			.286		OLS
Gdppc, gdppc*G2, gdppc*G4	gdppc	teupc		2.155 (.000)	11.923 (.000)	1.383 (.000)			.530		OLS
$Resid_{t-1}$	Resid	Resid	Autocorrelation	1.066 (.000)					.860		OLS
Gdppc, gdppc*G2, gdppc*G4, $Resid_{t-1}$	Gdppc	Teupc		2.329 (.000)	12.050 (.000)	1.147 (.000)	1.068 (.000)		.935	110491.2	OLS
Teupcpred, teupcpred^2	gdppc	Resid^2	White heteroskedasticity						.538		OLS
Gdppc, gdppc*G2, gdppc*G4, $Resid_{t-1}$teupcpred^2, teup-	gdppc	teupc	Ramsey RESET							108018.7	OLS

Model description					Beta weights (if significant) and p-values				Model characteristics		
Parameters	IV	DV	Test statistic	B_1 (p)	B_2 (p)	B_3 (p)	B_4 (p)	B_5 (p)	(Adj.) r^2	SSE	Type (Hausman p)
cpred^3											
Urb*cons	Urbanisation, Consumption	teupc	Hausman								Random (.001)
Urb*cons	Urbanisation, Consumption	Teupc		$1.41e^{-12}$ (p=.000)					.817		Fixed
Urb*log(Consumption)	Urbanisation, Consumption	Teupc		.323 (.000)					.818		Fixed
$Resid_{t-1}$	Resid	Resid	Autocorrelation	.986 (.000)					.932		Fixed
Urb*log(Consumption), $resid_{t-1}$	Urbanisation, Consumption	Teupc		.323 (.000)	.986 (.000)				.989		Fixed
Teupcpred, teupcpred^2	Urbanisation, Consumption	$Resid^2$	White heteroskedasticity						.776		Fixed
Urb*log(Consumption), $resid_{t-1}$, teupcpred^2, teupcpred^3	Urbanisation, Consumption	Teupc	Ramsey RESET							15651.97	Fixed
Tip	Tip	Teupc	Hausman	172.260 (.000)					.389		Random (.104)
Tip	Tip	Teupc		175.677 (.000)					.855		Fixed
$Resid_{t-1}$	Resid	Resid	Autocorrelation	.927 (.000)					.982		Fixed

Model description				Beta weights (if significant) and p-values					Model characteristics		
Parameters	IV	DV	Test statistic	B_1 (p)	B_2 (p)	B_3 (p)	B_4 (p)	B_5 (p)	(Adj.) r^2	SSE	Type (Hausman p)
Tip, $resid_{t-1}$	Tip	Teupc		152.855 (.000)	.938 (.000)				.985	23937.38	Fixed
Tip, $resid_{t-1}$, $tippred^2$	Tip	$Resid^2$	White heteroskedasticity						.120		Fixed
Tip, $resid_{t-1}$, $tippred^2$, $tippred^3$	Tip	Teupc	Ramsey RESET							18906.01	Fixed