ERASMUS UNIVERSITY ROTTERDAM

Erasmus School of Economics

Master Thesis Economics and Business: International Economics

The Emergence and Growth of Developing Countries' Exports – a Dynamic Factor Proportions Approach.

Name student: Maaike Bekkers

Student ID number: 504185

Supervisor: Prof. Dr. Maarten Bosker

Second assessor: Dr. Julian Emami Namini

Date final version: 14-01-2020

Abstract

How developing countries should design their export diversification strategies to provoke stable growth patterns has remained an incomplete answered question in the economic trade literature. This research adopts a factor proportions approach in an attempt to contribute to this question, by analyzing how changing production capabilities affect both the emergence and growth of exports from developing countries. A key contribution of this paper to the existing literature on factor proportions models is the separation of the impact of two distinct mechanisms that determine the evolution of these capabilities: changes in factor endowments and changes in production structures. The global evidence provided by this paper is consistent with the view that factor proportions are a precondition for export sustainability and growth, rather than strict determinants of what a country should produce. In particular, this paper demonstrates that changes in endowments have a substantial impact on export growth. On the contrary, production structure changes affect the emergence of new exports, but the size of this effect is almost negligible.

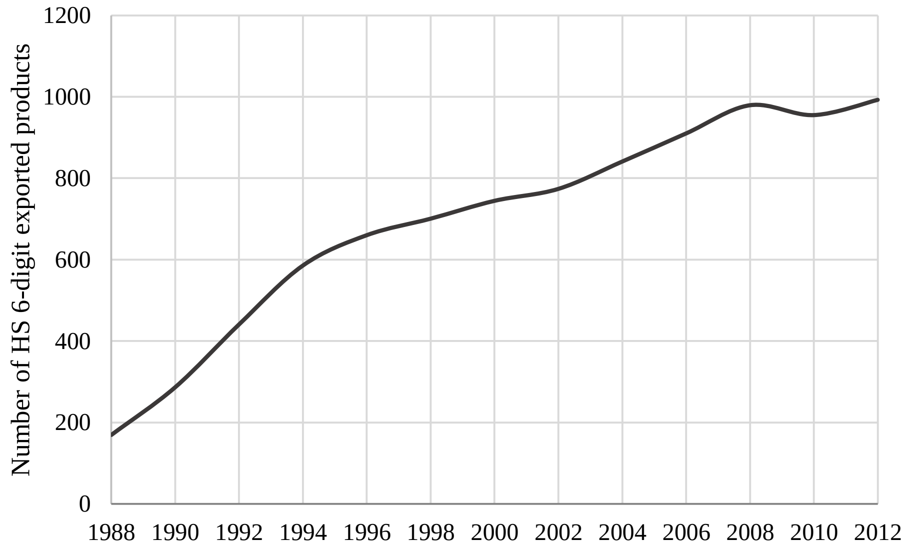
Keywords: Factor Proportions, Comparative Advantage, Export Emergence, Export Growth

Table of Contents

1. Introduction

Changing export patterns over time and between countries have become popular research topics over the last years. In particular the export baskets of developing countries have gained much attention. The diversity and quality of products within these baskets are often seen as a prediction for economic growth as they reflect production capabilities (Basu & Das, 2011; Schott, 2003; Sutton & Trefler, 2016). Developing countries' exports are frequently concentrated on a few products with very volatile demand, so that high growth volatility is provoked. To ensure a more stable growth pattern, it is often pointed out that diversification of their export basket should be high on the agenda (Agosin, 2007; Hesse, 2009). That developing countries adopt diversification strategies is evident from the increasing number of exported products by those countries, as is shown in Figure 1. However, for such 'new' exports to be successful, both survival and growth are required (Besedeš & Prusa, 2006).

Figure 1. Average Number of Exported Products for a Selection of Developing Countries.



Notes: Averages are based on a selection of 51 developing countries, of which an overview is provided in Appendix 9.1. Source: World Integrated Trade Solution and author's own calculations.

This has led to a much debated question on this topic: what exports tick those two boxes and how should developing countries diversify their exports? Literature on export diversification establishes two main diversification strategies. The first option is path-following diversification. This option is based on standard trade theories according to which accumulation of production capabilities determines comparative advantage and in turn determines what a country should export. Alternatively, path-defying industrial policies are increasingly being designed to imitate export activities of more developed countries that are not necessarily in line with their current production capabilities (Agosin, 2007). The basis for such policies is that the type of exported products matters for economic growth and that the expectation is that products that are exported by developed countries have a higher potential for creating economic growth than e.g. primary commodities (Hausmann, Hwang, & Rodrik, 2007). However, empirical evidence suggests that such exports are more difficult to be sustained. Despite survival rates of exports from developing countries being low in general, they are even lower for path-defying exports (Nicita, Shirotori, & Klok, 2013). Low export survival rates may imply that production capabilities differ in their impact on the decision to start exporting versus the sustainability of exports. In order for developmental policies to encourage sustainable export diversification, it is necessary to understand some underlying forces affecting production capabilities and their impact on export emergence and growth/sustainability in more depth.

This paper tries to provide more insights in these forces, by building on an empirical application of neoclassical trade theories in which factor proportions play an important role in determining a country's production capabilities. Empirical models based on such theories are often restricted to predict exclusively path-following exports. According to these models, a country should exclusively export products that use factor proportions similar to its factor endowments in the production process. The approach adopted in this research is less restrictive as it does not take a stand on the categorization of exported products to be either path-following or path-defying. Factor endowments are seen as a precondition for the establishment of sustainable exports rather than being strict determinants of what a country should produce. The further away the factor endowments are from the factor intensities used in the production of a product, the higher the opportunity costs and the less likely it is that these endowments can support competitive exports (Redding, 1999). This paper therefore predicts that a country can start exporting a product for which not all necessary production capabilities are accumulated yet. For such exports to be sustained over time, the gap between its factor endowments and the comparative advantageous factor inputs used in production should be reduced. Specifically, this research exploits changes in the magnitudes of these 'gaps' over time to identify their impact on changing export baskets of developing countries. The magnitude of this gap will be referred to as the distance to comparative advantage (distance to CA hereafter) as this gap reflects how far a country's production capabilities (factor endowments) are from the revealed comparative production structure of a particular product. In contrast to previous studies, this method provides the opportunity to separate the effect of two mechanisms that determine the evolution of this distance: changes in factor endowments and changes in production structures. This allows for an even more detailed comparison of the underlying forces and their impact on the emergence and growth of exports.

Overall, the findings of this research are in line with the expectation that production capabilities affect the emergence and growth of developing countries' exports differently. Whereas their effect on the likelihood of exporting is almost negligible, yet depends on the changing distance to CA due to changes in production structures, the growth of exports is revealed to be affected exclusively by the change in a country's own factor endowments. The signs of these effects are as expected: the change in distance to CA, due to either of the two defined effects, has a negative impact on the emergence and growth of developing countries' exported products. These findings should be taken into account when designing trade policies, to ensure sustainable growth patterns for these countries, and to increase export survival rates.

The rest of the paper is structured as follows. Section 2 reviews related literature that is relevant to this research. The data and methodology are introduced in Section 3 and 4 respectively. Section 5 provides and describes the results of this research and checks their robustness when controlling for productivity differences between countries. In Section 6 these results are discussed and Section 7 concludes.

2. Literature review

This paper is related to different strands of literature on the empirical determinants of changing export baskets, regarding the intensive margin as well as the extensive margin.[1] Whereas the first strand of literature has put their focus on factor-proportions motives for trade, and changing export baskets to be a response to changing comparative advantage, the second and more recent strand concentrates on exports to follow a pre-determined path based on accumulated production capabilities.

There exists an extensive body of literature in which changes in relative abundance of factor endowments are focused on in explaining dynamic export patterns. The interaction between a country's relative factor endowments and production characteristics of traded products determines what a country exports according to this area of research (Vanek, 1968). Thus, this theory establishes relative factor endowments to be the relevant source of a country's comparative advantage (Bernstein & Weinstein, 2002). This strand of literature has gained

[1] Throughout this paper, the growth of exports of existing country-product exports is referred to as the intensive margin of exports. When referring to the extensive margin of exports, the emergence of a new exported country-product pair is meant, without taking into account new export destinations of a product that has already been exported by a country.

much attention from researchers and has prompted intense empirical scrutiny. When testing predictions of the factor proportions theorem, a large amount of researchers has focused on very restrictive models, assuming factor price equalization and/or the use of identical technologies between countries (Bernstein & Weinstein, 2002; Davis & Weinstein, 2001; Harrigan, 1995). The general outcome of those tests is that they result in poor predictions of the observed trade patterns and that they sometimes even result in paradoxical findings (Leontief, 1953). Researchers have responded to this by relaxing some of the restrictive assumptions. Romalis (2004) deviates from the assumption of factor price equalization and Harrigan and Zakrasjek (2000) in addition allow for unmeasurable technological differences between countries. Both provide evidence in favor of changing factor endowments to predict observed trade specialization patterns. Fitzgerald and Hallak (2004) adjust the baseline framework by taking productivity differences into account and find that even among OECD countries with relatively small differences in factor proportions, those differences are strongly correlated with differences in what countries produce.

Even though the factor proportions models with relaxed assumptions can predict trade patterns at a highly aggregated level well, the necessary detailed trade data for such research was exclusively available at the industry level. This made empirical tests for predictions of these models at the product-level non-existent (Schott, 2004). Nevertheless, the observation of similarly endowed countries not holding similar export baskets suggests that products are at some distance from each other in terms of their required production capabilities, even when the factors used in the production process might be similar (Hausmann & Rodrik, 2006). Hausmann and Klinger (2006) and Hidalgo, Klinger, Barabasi and Hausmann (2007) have formalized this idea in their seminal papers on the Product-Space framework. Within this framework it is hypothesized that when countries add products to their export basket, these products should be related to the products they already export. This theorem is distinct from previous approaches to describe changes in revealed comparative advantage, in that it abstracts from identifying which particular sources of product relatedness or similarity are important in explaining these changes. Rather, it is argued that the relatedness between products should be reflected by the probability of other countries holding a comparative advantage in both products (Hausmann & Klinger, 2006). The intuition for this being that the ability of a country to produce and export a product is based on being able to produce and export products with similar requirements regarding e.g. physical factors, infrastructure and/or technology (Hidalgo et al., 2007).

Both seminal papers provide evidence in favor of positive correlations between evolving comparative advantage in products and those products being ‘related’ to products in the current export basket. However, such simple correlations cannot discriminate between this correlation reflecting the relationship as suggested by the framework and spurious relatedness. Coniglio, Vurchio, Cantore and Clara (2018) argue that their methodology, on the other hand, can make this distinction by comparing a randomly generated process of distributing new products to a country's export basket to the actual distribution. They find that the path-dependence hypothesis is confirmed by their results for small, least advanced and resource abundant countries. However, they do point out that deviations are not rare (Coniglio et al., 2018)**.** An additional test of the Product-Space theorem at the sub-national level, based on a similar methodology, leads to the same findings; an overall confirmation of the tendency of path-dependence, with exceptions holding a sizable share (Coniglio, Lagravinese, Vurchio, & Armenise, 2018). The literature on export diversification refers to these exceptions as being path-defying changes of the export basket, which can be thought of as being the result of combining knowledge and capabilities in a new way to result in breakthroughs **(**Coniglio et al., 2018). The concept of product relatedness is too agnostic to characterize particular factors or channels that give rise to such breakthroughs.

Recently, a large group of multi-disciplinary scholars started questioning if this concept in itself teaches us about economic development and if it matters for economic policy analysis because of its rather static and descriptive character (Hidalgo et al., 2018). Some papers have responded to this criticism by attempting to identify the importance of various channels that contribute to the observed relatedness between products. Bahar, Rosenow, Stein and Wagner (2018) study both demand and supply channels and find that supply channels and in particular technological similarity within the supply channel, matter for the emergence of exports. Additionally, they find that workforce similarity explains much of the growth of exports. Such studies provide interesting insights in how and why products and/or exports are related. Especially their focus on growth in addition to the emergence of exports contributes to the understanding of changing export baskets. Nonetheless, their methods still cannot fully incorporate the dynamics of international trade patterns such as the possible 'breakthroughs' mentioned before.

This research tries to bridge the gap between the factor-proportions models that were only able to predict aggregate trade flows and the more detailed and disaggregated but agnostic approaches of describing trade patterns based on product relatedness. The availability of a dataset that provides revealed factor intensities at the disaggregated product level as well as factor endowments at the country level, allows for a microeconomic factor based comparison of production capabilities to production requirements. This is exactly the essence of the factor-proportions models. This research is related to that of Lectard and Rougier (2018) and Nicita, Shirotori and Klok (2013) in that it uses a similar comparison of these factor proportions to proxy for production capabilities. However, whereas previous research on these models focused on country endowments to be the dynamic and changing factor, this research adds another dynamic aspect, namely changing production structures to reflect possible 'breakthroughs'. These two dynamic forces are compared on their impact on the emergence and growth of exports. Furthermore, from research on the Product-Space framework it has become evident that more channels at the product level may influence export baskets, which is allowed for by the inclusion of a careful set of fixed effects.

3. Data

This research investigates the impact of changes in the magnitude of the distance to CA based on factor intensities on the emergence and growth of developing countries' exported products. The analysis uses panel data covering 51 developing countries and 810 different products, from 1984-2006.[2] The sample size is determined by the data coverage in the relevant databases that will be described in the following subsections.

This research classifies products according to the SITC 5-digit Rev.1 classification which includes 1,035 different products. All necessary data could also be obtained at the more detailed HS 6-digit classification, that distinguishes between approximately 5,300 products. The choice for not using the latter one is based on the lower aggregation to enable a more relevant discussion on changing export baskets. As this study is mainly based on production side capabilities to affect the emergence of new products, at a certain level the difference between products may not be very meaningful. Table 1 demonstrates this by showing that at the higher level of disaggregation, products are distinguished by such small differences that not production capabilities, but rather demand side characteristics are expected to influence the presence of particular products in an export basket. The higher aggregation of the products is therefore expected to only identify changes from an accounting viewpoint, rather than from a capabilities viewpoint. Similar arguments against the use of overly disaggregated product data are common in the literature on new production or export discoveries (Klinger & Lederman, 2004).

[2] Within this research, a country is classified as a developing country when at the start of the timespan, hence 1984, its GNI per capita in USD was below the threshold of 'middle low income countries' set by the World Bank at that time. See Appendix 9.1 for an overview of the 51 countries included in the sample.

The timespan 1984-2006 is chosen based on the availability of data on the one hand and an arbitrary length of five years to construct the change variables, so that four time intervals could be constructed.[3] The base years for those time intervals are t=1985, 1990, 1995 and 2000, and for each interval T - t = 5.[4]

Table 1

Selected Part of a Concordance Table for the HS 6-digit and SITC 5-digit Product Classification

HS 6-digit Classification		**SITC 5-digit Classification**	
481310	Cigarette paper in booklets or tubes	64291	Cigarette paper cut to size
481320	Cigarette paper in rolls of width not exceeding 5 cm		
481390	Other cigarette paper		
961310	Pocket lighters, gas fueled, nonrefillable	89934	Mechanical, etc. lighters
961320	Pocket lighters, gas fueled, refillable		
961380	Other lighters		
961390	Lighter Parts		

Source: World Integrated Trade Solution (WITS).

3.1 Data Sources

3.1.1 Trade Data

The database providing the necessary trade data is the UN Comtrade database (United Nations Statistics Division, 2019). This database covers both agricultural and manufacturing exports, while it excludes services. Important for this analysis is that it does not always report non-exported products as zero trade flows. The creation of a harmonized sample size for all countries and years is therefore necessary to ensure comparability between export baskets over time, and in particular to identify new exported products.[5] Especially for developing countries, imports are usually reported with more accuracy than exports, so that in order to limit potential errors in reported trade flows, mirror data is used. Furthermore, if there are still reporting errors

[3] Three-year averages are used in the construction of some of the variables used in this research. Therefore, the years based on which the variables are constructed are 1984-2006 instead of 1985-2005.
[4] The adoption of the five-year difference for constructing the change variables is arbitrarily chosen and is based on the data availability and the intention of constructing at least four different base years to allow for some within variation in the panel data.
[5] The sample size is harmonized by adding non-reported products per country and year and assigning them a trade value of zero.

left in the data, those are no longer related to countries' income levels which limits measurement error issues.

3.1.2 Revealed Factor Intensities and Country Endowments

A database compiled by Shirotori, Tumurchudur and Cadot (2010) provides the necessary variables for the calculation of the distance to CA at the country-product-year level. It contains factor endowments as well as revealed factor intensities, which are calculated based on those factor endowments. This database is created to provide a tool to empirically test traditional trade theories of factor-content trade at a highly disaggregated level, as its creators observed that such tests were lacking due to non-availability of data on production structures at the product level. Their collection of raw data on factor endowments at the national level of human, natural and physical capital led to the firs part of their database. It includes endowment data on 92 countries in the balanced panel, which together accounted for 87.3% of world exports during the sample period.[6] This research draws only on the use of human and physical capital, in line with Cadot, Carrère and Strauss-Kahn (2011) as natural capital is considered to be more stable over time. This does not fit into the adopted identification strategy of this research, which is based on changes of these factor endowments.

The collected factor endowments are used in the calculation of the revealed factor intensities at the SITC 5-digit product classification, by combining them with trade data. Factor endowments of all countries exporting a particular product are weighted and combined into factor intensities of that product. The idea behind this variable construction method is that a product exported predominantly by countries that are richly endowed with a particular factor of production is revealed to be intensive in the use of that factor. That is, the revealed factor intensity of factor f of product j for country c at year t is:

$$\hat{f}_{jt}^{c} = \sum_{i} \frac{X_{jt}^{i}/X_{t}^{i}}{\sum_{i}(X_{jt}^{i}/X_{t}^{i})} \cdot f_{it}, \quad \text{with } i \neq c \tag{1}$$

Where $f = k, h$ refers to either the physical or human capital endowment of country i respectively. Physical capital is measured as real capital stock per worker, and human capital

[6] The database distinguishes between two panels of revealed factor intensities: a balanced and an unbalanced panel. The balanced panel of data on country endowments includes only those countries for which endowment data for all years is available, to ensure a consistent comparison between indices over several years. The drawback of this method is that if there is systematic bias in the selection of the countries in the balanced panel (e.g. underreporting) then this might lead to a bias of the composed factor intensities. The unbalanced panel includes all countries for which endowment data was available in a particular year. This research is based on a comparison of the factor intensities over time, so that in the trade-off between 'width' and 'consistency' the latter one is most important.

as the average number of years of schooling of the labor force.[7] These endowments are weighted by a slightly modified version of Balassa's revealed comparative advantage index where the value-share of product *j* of the total exports of country *i* is divided by the sum of product *j*'s share across all countries exporting that product. These weights per product sum up to one, where countries that compared to other countries export a larger share of product *j* are assigned a larger weight to their factor endowment.[8] Country *c* is excluded from the calculation to avoid endogeneity. Due to the availability of the necessary trade data for constructing these indices, the revealed factor intensities for physical and human capital are provided for 810 out of the 1,035 classified products at the SITC 5-digit level.

3.2 Variable Construction

3.2.1 Export Emergence

The creation of the dependent variable for the extensive margin of export growth in this study is based on the adopted definition of 'new exports'. The straightforward method of defining a product with a positive reported trade value at a particular year, and zero trade values before to be 'new' might lead to misidentification as it could include sample sales or coding mistakes. In the existing literature on new exports it is common to control for this by restricting new products to include only products with a trade value above a certain threshold (Hausmann & Klinger, 2006; Klinger & Lederman, 2004). In particular, the threshold of 10,000 USD is often used as a benchmark, as this coincides with the adopted definition of 'new exports' by the World bank in constructing their trade indicators (World Integrated Trade Solution). This research will follow previous studies and therefore adopt the 10,000 USD threshold for classifying new exports.

Interruptions in reported trade values are another analytical challenge faced in the empirical literature on export emergence. Some researchers have decided to ignore those missing values as they argue that they are likely to be reporting errors (Besedeš & Prusa, 2006; Blyde, 2008). Others mention that such missing values reflect low survival rates of new exports, so that these are reflecting real trade patterns (Brenton, Saborowski, & Von Uexkull, 2010). Following Wagner and Zahler (2015), misclassification due to interruptions in reported trade values are limited without taking a stand on the cause of those gaps by taking averages of the

[7] A more detailed description of these measures and the collection of the factor endowments is provided in Shirotori, Tumurchudur and Cadot (2010).
[8] The methodology of weighting factor endowments in this way is adopted from Hausmann, Hwang and Rodrik (2007).

reported export values at the cutoffs between periods *t* and *T*, where $T > t$. In particular, the product is defined as 'new' when it has not been exported between [t-1, t+1] with an average export value above 10,000 USD and when it is exported with an average above the threshold of 10,000 USD between [T-1, T+1]. The dependent variable reflecting export emergence in this research is a dummy variable constructed based on the above definition:

$$\text{New Export}_{ijt\rightarrow T} = \begin{cases} 1, & \text{if } \left[\overline{x_{ijT}} > 10{,}000 \middle| \overline{x_{ijt}} < 10{,}000\right] \\ 0, & \text{if } \left[\overline{x_{ijT}} < 10{,}000 \middle| \overline{x_{ijt}} < 10{,}000\right] \end{cases} \qquad (2)$$

Exports with a value above 10,000USD in period *t* are therefore excluded from the sample for the extensive margin of export growth. Using an arbitrary cutoff in defining such variables is unavoidable, however the results from this research are found to be robust to small changes of the cutoff level.[9]

3.2.2 Export Growth

The construction of the second dependent variable, which reflects the intensive margin of export growth, is more straightforward. The only condition to be met for a country-product pair to be included in the sample used in this case, is that the average export value for that pair should be at least 10,000 USD between [t-1, t+1]. The argument for this is the same as before, namely minimization of the possible bias due to reporting errors. For the same reason, the export values are again calculated as an average over [t-1, t+1] and [T-1, T+1]. The growth rate of the export value of a country-product pair between the years *t* and *T* is the measurement for the intensive margin of exports.[10] That is:

$$\text{Export growth}_{ijt\rightarrow T} = (\overline{x_{ijT}} - \overline{x_{ijt}}) / \overline{x_{ijt}}) \quad \text{if} \quad \overline{x_{ijt}} > 10{,}000 \qquad (3)$$

3.2.3 Changing Distance to Comparative Advantage

The main independent variable in this research is the distance to CA, which is a measure of the difference between the factor endowments of a country and the revealed factor intensities of a particular product. Following Cadot, Carrère and Strauss-Kahn (2011) the Euclidean distance function is used for this. This function can only be used with variables that are on the

[9] See Appendix 9.2 for the results when different cutoff levels are used.

[10] The calculation of the growth rate in this way allows for the inclusion of zero trade values in period T, for which by construction the growth rate is equal to -1. This happens in 4% of the observed trade flows. These trade flows and in particular their in- or exclusion, do not affect our results. Appendix 9.3 provides the baseline results with these ending trade flows excluded.

same scale, which is not the case here. Therefore, standardized variables are used in the construction of the distance variable:

$$\text{Distance to CA}_{ijt} = \left[\left(h_{it} - \hat{h}^{i}_{jt}\right)^2 + \left(k_{it} - \hat{k}^{i}_{jt}\right)^2\right]^{1/2} \quad (4)$$

Where h_{it} and k_{it} represent the standardized human capital and physical capital endowments of country *i* at year *t* respectively, and $\hat{h}^{c}_{jt}$ and $\hat{k}^{c}_{jt}$ are expressions for the standardized revealed human and physical capital intensities of product *j* in that year.[11]

This research focusses on how changes in distance to CA affect the decision to start exporting a product as well as the growth of exports. The construction of the change of the Euclidean distance to CA between *t* and *T* allows for the total change to be split up in two parts that reflect two separate effects; the endowment effect and the revealed production structure effect. By construction, these two effects add up to the total change in distance to CA: [12]

$$\Delta\text{Distance to CA}_{ijt\to T}= \left[\left(h_{iT}-\hat{h}^{i}_{jT}\right)^2 + \left(k_{iT}-\hat{k}^{i}_{jT}\right)^2\right]^{1/2} - \left[\left(h_{it}-\hat{h}^{i}_{jt}\right)^2 + \left(k_{it}-\hat{k}^{i}_{jt}\right)^2\right]^{1/2} \quad (5)$$

$$\Delta\text{Endowment distance}_{ijt\to T} = \left[\left(h_{iT}-\hat{h}^{i}_{jt}\right)^2 + \left(k_{iT}-\hat{k}^{i}_{jt}\right)^2\right]^{1/2} - \left[\left(h_{it}-\hat{h}^{i}_{jt}\right)^2 + \left(k_{it}-\hat{k}^{i}_{jt}\right)^2\right]^{1/2} \quad (6)$$

$$\Delta\text{Production structure distance}_{ijt\to T} = \left[\left(h_{iT}-\hat{h}^{i}_{jT}\right)^2 + \left(k_{iT}-\hat{k}^{i}_{jT}\right)^2\right]^{1/2} - \left[\left(h_{iT}-\hat{h}^{i}_{jt}\right)^2 + \left(k_{iT}-\hat{k}^{i}_{jt}\right)^2\right]^{1/2} \quad (7)$$

The total change in distance to CA is simply the difference between the distance to CA at the end and the start of the five-year time interval. The first effect is based on the changing distance to CA to become smaller or bigger based on the change of the country's own endowment structure. An increase in this distance means that the endowments of a country are even further away from the revealed factor intensities calculated at time *t* than before its endowments changed. The latter effect incorporates possible changes of the factor intensities over time.[13] It essentially reflects the difference between the total change in distance to CA and the endowment distance. An increase in the production structure distance means that, given the new endowments, the changing production structure increases the distance between production

---

[11] Standardized factor intensities are calculated according to the following formula: standardized value = (original value – mean) / standard deviation.
[12] Please note that similar to specification (4) standardized variables are used in the construction of the distances.
[13] Changes in the factor intensity of a product can be the results of either changes in the factor endowments of the mix of countries that export that product, and/or of changes in the mix of countries that export that product.

capabilities and requirements as compared to the distance between both when only the endowments would have changed.

3.3 Descriptive Statistics

Table 2

Summary Statistics

Variable	*N*	*Mean*	*SD*	*Min*	*Mdn*	*Max*
Panel A $\overline{x_{ijt}} < 10{,}000$						
New $Export_{ijt\rightarrow T}$	105,970	0.165	0.371	0	0	1
ΔDistance to $CA_{ijt\rightarrow T}$	105,970	-0.002	0.688	-4.957	0.003	6.135
Distance to CA_{ijt}	105,970	1.754	1.007	0.007	1.599	7.233
ΔEndowment $distance_{ijt\rightarrow T}$	105,970	0.006	0.160	-0.879	0.015	0.964
ΔProduction structure $distance_{ijt\rightarrow T}$	105,970	-0.006	0.672	-4.838	-0.005	5.803
Panel B $\overline{x_{ijt}} > 10{,}000$						
Export $growth_{ijt\rightarrow T}$	59,270	3.715	27.712	-1	0.394	1366.033
ln Export $value_{ijt}$	59,270	5.566	2.456	2.303	5.056	16.659
ΔDistance to $CA_{ijt\rightarrow T}$	59,270	0.003	0.604	-4.736	-0.004	5.735
Distance to CA_{ijt}	59,270	1.472	1.022	0.007	1.467	7.271
ΔEndowment $distance_{ijt\rightarrow T}$	59,270	-0.005	0.190	-0.953	-0.014	0.965
ΔProduction structure $distance_{ijt\rightarrow T}$	59,270	0.009	0.578	-4.844	0.003	5.545

Notes: Variables are constructed as described in Section 3.2. ln Export $value_{ijt}$ reflects the natural logarithm of the export value in 1,000USD.

Table 2 presents some summary statistics for the variables constructed for this research. The table is divided in two panels to display the statistics for the samples regarding the extensive and intensive margin of export growth separately. Panel A shows that the introduction of a new exports happens quite frequently, as it has a probability of 16.5%, which is in line with what Figure 1 shows. From the bottom half of the table it should be noted the median and the mean of the export growth variable differ substantially, being 371.5% and 39.4% respectively. This indicates that the data for this variable is skewed and that the calculated mean is probably larger than the actual distribution of the data would suggest. As the average world export growth of the four time intervals used in this research was 56.1% (World bank), the median indeed seems to provide a more realistic picture of the data. Another interesting thing to point out is that the mean of the initial distance to CA is higher in the sample for the extensive margin (panel A) than in the sample for the intensive margin (panel B), which is what would be predicted by traditional factor proportions models.

4. Methodology

This research aims to find support for the hypothesis that the likelihood of starting to export a particular product as well as the growth of exports increases when the distance to producing with revealed comparative advantage becomes smaller. Additionally, it tries to find evidence for the suggestion made in the introduction, that export emergence and growth are affected differently by the change in distance to CA. Therefore, this paper explores the effects of the two distinct effects that cause changes in the distance to CA on explaining both the extensive as intensive margin of export growth separately.

4.1 The Extensive Margin of Export Growth

With respect to the regression regarding the extensive margin of export growth, hence the emergence of a new product in the export basket of a country, the dependent variable is the binary variable New Export_{ijt}. When designing an empirical model with a binary dependent variable, the main models available are the linear probability model on the one hand, and the binary response models probit and logit on the other hand (Wooldridge, 2016). It is important to keep each model's limitations into consideration, when deciding on which model to use. For linear probabilities models these are that the probabilities are not necessarily distributed on the (0,1) interval, and constant partial effects of all explanatory variables. Nevertheless, for this research, those models are still preferred over probit models. Not being able to include fixed effects into the regression – a limitation of probit models – is expected to affects this research even more. Not controlling for any country-product related fixed effects would ignore all other heterogeneous factors across countries, which are proven to be important in explaining changing export baskets by the Product-Space framework (Hausmann & Rodrik, 2006). Logit models do allow for the inclusion of fixed effects and have a probability distribution on the (0,1) interval, but have the drawback that they are difficult to interpret. In many situations the linear and logistic models give practically similar results (Hellevik, 2009). Additionally, Angrist and Pischke (2008) argue that average effects from linear probability models resemble marginal effects of non-linear models. However, the high-dimension of the country-product fixed effects makes the estimation with logit models harder than OLS regression. Stata's memory does not allow for the explicit introduction of dummy variables to reflect fixed effects when the number of groups is considerably large (Guimarães & Portugal, 2009). Whereas there exists a build-in syntax in Stata for estimating high dimensional fixed effects specifications with OLS models, for logit regressions such syntax does not exist. Therefore, the baseline specification model for the extensive margin of exports is estimated with OLS regression and is described by:

$$\text{New Export}_{ijt\rightarrow T}=\beta_0+\beta_1\Delta\text{Distance to CA}_{ijt\rightarrow T}+\beta_2\text{Distance to CA}_{ijt}+\gamma_{ij}+\tau_{it}+\varepsilon_{ijt} \qquad (8)$$

where the variables are constructed as described in Section 3.2, γ_{ij} reflects country-product fixed effects and τ_{it} represents the country specific time fixed effects. The country-product fixed effect is included to allow for other forces than factor endowments to play a role in determining production capabilities of a country, as described above. Such forces might change over time, but as the variables included in this regression only vary at three dimensions (country-product-year), fixed effects can maximally vary at two. This means that including the country-product fixed effects is the best possible option here. Nevertheless, the restricted within country-product variation might make the use of country-product fixed effects inappropriate because too many variables have to be dropped from the regression.[14] The second-best option in this case would be to control for a higher aggregation of the product-classification. In addition to using the product SITC 5-digit classification which is used throughout the rest of this paper, results when including a fixed effect based on the SITC 4-digit classification, referred to as a country-industry fixed effect, are provided as well. The country specific time fixed effect is added to capture macro-economic shocks that might influence the overall export and production capabilities of a country. The initial distance to CA is included in the specification to allow for country-product pairs at different initial distances to CA to respond differently to changes in this distance.[15] If it is the case that country-product pairs at different distances actually respond different to the changing distance, excluding this independent variable would lead to omitted variable bias.

[14] The dependent variable New Export_{ijt} is defined in such a way that once a country starts exporting a product, the country-product pair is not considered in the next periods sample for the extensive margin, as for each period only non-exported products can become new exports. This restricts the within country-product variation. When including country-product fixed effects, all singleton observations have to be dropped from the regression.

[15] The initial distance is the distance to CA in the base year, so in 1985, 1990, 1995 or 2000.

The most important independent variable measures the relationship between the changing distance to CA of country *i* and product *j* between *t* and *T* and the decision of country *i* to start exporting that product. A negative and significant β_1 would reflect the expected negative relationship, such that a decreasing distance to CA makes it more likely that a country starts exporting product *j*. In the first specification, the two different channels through which distance to CA can be influenced are considered jointly, but in addition they will be considered separately. The adjusted regression specification for the extensive margin of export growth will then look as described below:

$$\begin{aligned}\text{New Export}_{\text{ijt}\rightarrow\text{T}} = \beta_0 + \beta_{1a}\Delta\text{Endowment distance}_{\text{ijt}\rightarrow\text{T}} + \\ \beta_{1b}\Delta\text{Production structure distance}_{\text{ijt}\rightarrow\text{T}} + \\ \beta_2\text{Distance to CA}_{\text{ijt}} + \gamma_{\text{ij}} + \tau_{\text{it}} + \varepsilon_{\text{ijt}}\end{aligned} \quad (9)$$

In this case, the effect of the change in total distance to CA, which was reflected by β_1 in specification (8), is split up in the change in endowment distance, and the change in the production structure distance. The estimated coefficients of the impact of these two changing distances on the emergence of a new exported product are reflected by β_{1a} and β_{1b} respectively.

4.2 The Intensive Margin of Export Growth

Regarding the intensive margin, hence the growth of the exported value of an already existing product in the export basket of a country, the dependent variable is the continuous variable Export $\text{growth}_{\text{ijt}}$. In this case, specification (8) should be adjusted by replacing the binary dependent variable by the continuous variable Export $\text{growth}_{\text{ijt}}$.[16] Another adjustment to this specification is made by including the natural logarithm of the initial level of export value. This allows for products that differ in the magnitude of their export value to differently respond to changes in the other independent variables. Nevertheless, as pointed out by Knight, Loayza and Villanueva (1993) the inclusion of this variable might cause endogeneity problems as by construction, the initial level of a variable is correlated to the growth variable. Interpretation of

[16] While being aware that this simple adjustment of the dependent variable in the specifications for the extensive margin and the intensive margin overlooks the possible problem of data censoring, this paper does rely on two distinct OLS regressions instead of using for example a two-step Heckman selection model to estimate the intensive margin. The main arguments for this choice being that the probit selection models do not allow for the inclusion of fixed effects, and the difficulty of fulfilling the requirement of including a selection instrument variable in the selection model. The exclusion of the various fixed effects from the specifications would almost certainly lead to omitted variable bias because of unobserved heterogeneity at the country-product level. On the other hand, the data available for this research does not allow for the inclusion of a valid instrument that can distinct the selection from the outcome model. Nevertheless, an attempt is made to overcome the problem of data censoring by estimating a two-step Heckman selection model in the best possible way, for which a description of the method and the results are provided in Appendix 9.4. No evidence in favor of data censoring was found.

the regression outcomes when this variable is included should hence be done with care. This results in the following specification in its most extensive form:

$$\text{Export growth}_{\text{ijt}\to\text{T}} = \delta_0 + \delta_1 \Delta\text{Distance to CA}_{\text{ijt}\to\text{T}} + \delta_2 \text{Distance to CA}_{\text{ijt}} + \delta_3 \ln \text{Export value}_{\text{ijt}} + \varphi_{\text{ij}} + \varsigma_{\text{it}} + \upsilon_{\text{ijt}} \quad (10)$$

Again, variables are constructed as described in Section 3.2, and ς_{it} represents the country specific time fixed effect. However, as in this case, almost all country-product pairs are included in each time interval, the inclusion of country-product fixed effects is justified and reflected by φ_{ij}. The specification will again be adjusted to include the distinction between the endowment effect and the production structure effect:

$$\begin{aligned}\text{Export growth}_{\text{ijt}\to\text{T}} = \delta_0 &+ \delta_{1\text{a}} \Delta\text{Endowment distance}_{\text{ijt}\to\text{T}} + \\ &\delta_{1\text{b}} \Delta\text{Production structure distance}_{\text{ijt}\to\text{T}} + \\ &\delta_2 \text{Distance to CA}_{\text{ijt}} + \delta_3 \ln \text{Export value}_{\text{ijt}} + \\ &\varphi_{\text{ij}} + \varsigma_{\text{it}} + \upsilon_{\text{ijt}}\end{aligned} \quad (11)$$

The interpretation of the coefficients on the main independent variables is quite similar to the interpretation when estimating the extensive margin of growth. The difference being that now these coefficients measure the effect of the changing distances on export growth. In particular, δ_1 represents the effect of the change in total distance, $\delta_{1\text{a}}$ the effect of the change in endowment distance and $\delta_{1\text{b}}$ the effect of the change in production structure distance.

5. Results

This section outlines the empirical results of this study. The results on the impact of changes in distance to CA on the extensive and intensive margin of export growth will be described separately. The results are also provided and described when the change in total distance to CA is split up in the two underlying effects. In addition, the models will be re-estimated with productivity adjusted CA distances to check for the robustness of the main findings.

5.1 The Extensive Margin

Table 3 presents the results from the baseline estimation of the extensive margin as described by specification (8)(8). The regression outcomes show that an increase in the distance to CA appears to reduce the likelihood that a country starts exporting a product, as the

Table 3

The Extensive Margin of Export Growth

Dependent Variable: New Export$_{ijt\rightarrow T}$				
Variables	(1)	(2)	(3)	(4)
ΔDistance to CA$_{ijt\rightarrow T}$	-0.012***	-0.006***	-0.013***	-0.006***
	(0.002)	(0.002)	(0.002)	(0.002)
Distance to CA$_{ijt}$	-0.015***	-0.003	-0.018***	-0.002
	(0.002)	(0.003)	(0.002)	(0.003)
Country-Product FE	No	Yes	No	Yes
Country-Industry FE	Yes	No	Yes	No
Country-Year FE	No	No	Yes	Yes
Observations	105,345	102,014	105,344	102,013
R^2	0.253	0.407	0.277	0.472

Notes: The number of observations differs per column as the number of dropped singletons differs per specification due to the inclusion of fixed effects. Regression is performed in Stata with the command reghdfe. Standard errors are robust and presented in parenthesis. (***), (**) and (*) denote significance at 1%, 5% and 10%, respectively.

coefficient on the change in distance to CA is highly significant and negative throughout all four columns. However, there is a jump in the magnitude of this coefficient when industry fixed-effects (columns 1 and 3) rather than product fixed-effects (columns 2 and 4) are included in the regression. This indicates that there is some heterogeneity at the product level even within an industry. The almost doubled magnitude of the coefficients might reflect the existence of unobserved factors that affect both the emergence of an exported product and the changing distance to CA, which cannot be picked up by the country-industry fixed effects. The inclusion of country-year fixed effects in the third and fourth column hardly affects the magnitude and significance of the coefficients of the independent variables. Nevertheless, including these fixed effects is more reliable because it controls for possible macro-economic shocks, and is validated by a significant joint F-test.[17] Henceforth, the focus will be on interpreting the coefficients in the fourth column, in which only the effect of the main independent variable is found to be statistically significantly different from zero. The value-estimate of -0.006 implies that an increase of one standard deviation of the change in distance to CA, holding all other factors

[17] A joint F-test on the included fixed effects in column four was performed and was found to be significant at the 1% significance level ($F_{29259,72751}$= 2.215, p = 0.000).

fixed, results in a decrease of 0.4 percentage points in the likelihood of starting to export a 'new' product. This is over forty times smaller than the unconditional probability of export emergence, which was 16.5% (see Table 2).

Table 4

The Extensive Margin of Export Growth

Dependent Variable: New Export$_{ijt\rightarrow T}$				
Variables	(1)	(2)	(3)	(4)
ΔEndowment distance$_{ijt\rightarrow T}$	-0.001	-0.016*	-0.021	-0.022
	(0.010)	(0.010)	(0.014)	(0.014)
ΔProduction structure	-0.012***	-0.006***	-0.014***	-0.006***
distance$_{ijt\rightarrow T}$	(0.002)	(0.002)	(0.002)	(0.002)
Distance to CA$_{ijt}$	-0.015***	-0.003	-0.018***	-0.002
	(0.002)	(0.003)	(0.002)	(0.003)
Country-Product FE	No	Yes	No	Yes
Country-Industry FE	Yes	No	Yes	No
Country-Year FE	No	No	Yes	Yes
Observations	105,345	102,014	105,344	102,013
R^2	0.253	0.407	0.277	0.472

Notes: The number of observations differs per column as the number of dropped singletons differs per specification due to the inclusion of fixed effects. Regression is performed in Stata with the command reghdfe. Standard errors are robust and presented in parenthesis. (***), (**) and (*) denote significance at 1%, 5% and 10%, respectively.

The results in table 4 allow for a comparison between changes in factor endowments and changes in production structures and their impact on export emergence, as these regression outcomes are based on specification (9). Whereas the coefficient on the changing production structure distance is significant at the 1% level throughout all columns, the coefficient on the changing endowment distance is only significant in the second column at the 10% level. The reported coefficients respond in a similar way to the inclusion of the various fixed effects, so that for the same reasons as described above the fourth column is the preferred one. The signs in front of both the coefficients on the changing distances to CA reflect that an increase in these distances decreases the likelihood of exporting. However, only the effect of the changing production structure is found to be significant and equal to the effect of the change in total distance in Table 3. An increase of one standard deviation of the change in the production structure distance, ceteris paribus, again results in a decrease of 0.4 percentage points in the likelihood of starting to export a 'new' product.

5.2 The Intensive Margin

The results of the baseline regression on the intensive margin of export growth, as described in specification (1010) are illustrated in Table 5. Whereas the country-product fixed effects are included in all columns, the four columns differ in their inclusion of country-year fixed effects and the natural logarithm of the initial export value. Including the country-year fixed effects does seem to affect the coefficients of the dependent variable in this case. Comparing the output when the natural logarithm is not included (columns 1 and 3), both the sign and the magnitude of the coefficients on the change in distance to CA, and the initial distance to CA have considerably changed. The coefficients in the first column, hence when country-year fixed effects are excluded, are in line with expectations, whereas the coefficients in the third column are not. In particular, the positive correlation between the change in distance to CA and export growth is the opposite of what this paper hypothesized. Nevertheless, as the coefficients of all independent variables were highly insignificant at all defined levels in both column 1 and 3, not much meaningful information can be subtracted from this comparison. The theoretical justification together with the significant joint F-statistics again makes specifications including this fixed effect preferred over the ones that do not include this.[18]

Regarding the inclusion of the natural logarithm of the initial export value, it is more difficult to argue what is best. Including this variable might cause endogeneity problems by construction, while excluding this variable is expected to cause omitted variable bias which is another source of endogeneity. The results in the third and fourth column show that the inclusion of this variable lowers the magnitude of both the distance variables. This indeed provides evidence in favor of omitted variable bias. The possible endogeneity problems from including this variable are expected to be less severe, as the main argument for this problem is 'conditional convergence' of export growth which merely plays a role at the country-level and not at the product-level.[19] The country-year fixed effects should therefore make up for this. The coefficients on the change in distance to CA are not only insignificant, but their indication of a

[18] Joint F-test on the included fixed effects in columns three and four were performed and were found to be significant at the 1% significance level ($F_{16420,36866}$= 1.097, p = 0.000 and $F_{16420,36865}$= 1.201, p = 0.000).

[19] 'Conditional convergence' is referred to in the growth literature to indicate that countries with an initially lower income level grow faster than richer countries and converge to similar income levels.

positive correlation with export growth is opposite of what was hypothesized within this paper. The single significant coefficient of this regression shows that a one-unit increase in the natural logarithm of the initial export value decreases the growth of the export value with 8.4 percentage points.

Table 5

The Intensive Margin of Export Growth

Dependent Variable: Export growth$_{ijt \rightarrow T}$				
Variables	(1)	(2)	(3)	(4)
ΔDistance to CA$_{ijt \rightarrow T}$	-0.293	0.247	0.045	0.072
	(0.390)	(0.371)	(0.405)	(0.387)
Distance to CA$_{ijt}$	-0.339	0.365	0.075	-0.026
	(0.510)	(0.488)	(0.559)	(0.534)
ln Export value$_{ijt}$		-6.836***		-8.398***
		(0.377)		(0.448)
Country-Product FE	Yes	Yes	Yes	Yes
Country-Year FE	No	No	Yes	Yes
Observations	53,290	53,290	53,290	53,290
R^2	0.285	0.353	0.303	0.368

Notes: Regression is performed in Stata with the command reghdfe. Standard errors are robust and presented in parenthesis. (***), (**) and (*) denote significance at 1%, 5% and 10%, respectively.

Even though the total change in distance to CA did not have any significant effect on export growth, the change in a country's own factor endowments, or the change in production structures might. To test for this, specification (11) is regressed and the results of this are provided in Table 6. Interestingly, the change in the production structure distance is never found to have a significant effect on export growth. The same holds for the initial distance to CA. To the contrary, the coefficients on the change in the endowment distance are found to be significant in three of the four columns. The insignificant coefficient on the change in endowment distance in the second column seems to be the result of the inclusion of the initial export value and the exclusion of country-year fixed effects. This indicates that there is some unobserved heterogeneity at the country-year level, that is correlated to both the initial level of export value and the change in endowment distance. Adding country-year fixed effects controls

Table 6

The Intensive Margin of Export Growth

Dependent Variable: Export growth$_{ijt \to T}$				
Variables	(1)	(2)	(3)	(4)
ΔEndowment distance$_{ijt \to T}$	-5.650***	-0.998	-5.379***	-3.606***
	(1.687)	(1.541)	(2.000)	(1.377)
ΔProduction structure	-0.023	0.309	0.180	0.165
distance$_{ijt \to T}$	(0.395)	(0.378)	(0.417)	(0.399)
Distance to CA$_{ijt}$	-0.341	-0.363	0.062	-0.034
	(0.509)	(0.488)	(0.588)	(0.532)
ln Export value$_{ijt}$		-6.826***		-8.390***
		(0.375)		(0.009)
Country-Product FE	Yes	Yes	Yes	Yes
Country-Year FE	No	No	Yes	Yes
Observations	53,290	53,290	53,290	53,290
R^2	0.286	0.353	0.303	0.368

Notes: Regression is performed in Stata with the command reghdfe. Standard errors are robust and presented in parenthesis. (***), (**) and (*) denote significance at 1%, 5% and 10%, respectively.

for this, which is done in columns three and four, in which the coefficient on the change in endowment distance is significant. The magnitude of this coefficient varies substantially between columns three and four, which is the result of the inclusion of the initial export value. This variable seems to take away some of the explanatory power of the change in endowment distance, so that not including this variable would probably cause omitted variable bias. For the same reason as before, the results in the last column are therefore favored over the ones in the third column. The value-estimate of -3.606 implies that an increase of one standard deviation in the change of endowment distance, ceteris paribus, decreases the export growth of a country-product pair with 68.5 percentage points, which is a quite substantial decreases as compared to the mean unconditional export growth of 371.5% over the five-year time interval. Compared to the median value of 39.4%, which is expected to reflect the unconditional export growth more realistically, this decrease is even more substantial.[20] No evidence is found to indicate that the

[20] See Section 3.3 for an explanation of why the median is expected to provide a more realistic picture of the data.

effect of the change in the production structure distance on the intensive margin of export growth is statistically different from zero.

5.3 Robustness Check – Productivity Differences

One of the main conclusions from the discussed literature in Section 1 is that productivity differences should be accounted for when using factor intensities and factor endowments in describing and predicting trade patterns. Within the models designed for this research, the country-product fixed-effects pick up the differences in the productivity levels of the countries included in the sample, but do not allow for these levels to change over time. Furthermore, these possible productivity differences between countries are not accounted for in the construction of the revealed factor intensities. Technology frontiers of most developed countries lie above those of developing countries, indicating that developed countries tend to produce more efficiently (Caselli, Coleman, & John, 2006). The revealed factor intensities for products exported mainly by high income countries are therefore expected to be underestimated, and overestimated for products exported mainly by low income countries. Replacing country endowments with productivity adjusted country endowments solves this problem.

The literature on tests of the hypothesis of factor price equalization has been faced with similar challenges of incorporating factor efficiency differences. One of the most influential methods was proposed by Trefler (1993) who derives capital and labor efficiency indices based on capital and labor endowment data, rent/wage ratios and the US technology matrix. Production factors were multiplied by the efficiency indices to explain differences in factor prices. Whereas the assumptions made in this research regarding the construction of the efficiency indices have been criticized (Repetto & Ventura, 1997), Trefler's (1993) main idea of constructing productivity indices has been widely adopted (Caselli et al., 2006)**.** Whereas the previously mentioned papers had to construct those indices themselves, which lead to the inclusion of mainly developed countries due to data availability, this research takes advantage of recent databases that include productivity parameters. The only database containing a big share of the countries of the sample used in this research is the Penn World Tables database, which unfortunately only includes data on total factor productivity per country instead of capital- and labor productivity separately (Feenstra, Robert C., Robert Inklaar and Marcel P. Timmer, 2015). Even though the data in this database is the best suitable data available for incorporating productivity differences into this research, the implicit assumption of equal (changing) productivity of human and capital is quite strong, especially because productivity levels are measured relative to the United States (US hereafter). Developed countries such as the US generally have a high stock of physical capital that complements their stock of human capital, so that these production factors can be combined efficiently to generate output. This efficiency is reflected by a high total factor productivity. Developing countries on the other hand, often hold high stocks of human capital but lack complementary assets. Therefore, it is expected that the unavailability of physical capital, more so than productivity of the available physical capital leads to lower total factor productivity levels. This unavailability does negatively impact the productivity of human capital. Consequently, for developing countries, the productivity of human capital is expected to be lower than the productivity of physical capital compared to the US. For countries with capital-labor ratios that differ extensively from the US, the assumption of equal physical and human capital productivity compared to the US is thus quite strong.

The incorporation of these productivity parameters in the baseline models requires a re-construction of the revealed factor intensities from scratch. The productivity adjusted revealed factor intensities are now constructed as:

$$\text{Productivity adjusted } \hat{f}_{jt}^{c} = \sum_{i} \omega_{ijt} \cdot A_{it} \cdot f_{it}, \quad \text{with } i \neq c \tag{12}$$

Where A_{it} is the total factor productivity index of country *i* at year *t* with the US being the reference country (US=1), and $f = k, h$ again refers to physical or human capital respectively. Indices below one thus indicate that the country is less productive than the US, and more productive countries are indicated by a value of this index above one. Countries with less productive factors need to use more factor endowments in their production than more productive countries to produce the same. Multiplying the factor endowments with these indices thus results in the creation of 'effective' factor endowments. For example, a country A with a relatively low factor productivity, say A_{At}=0.40, that is endowed with a high amount of capital k_{At}=400 and a country B with a high factor productivity, say A_{Bt}=1.60 but a low capital endowment k_{Bt}=100 thus have similar 'effective' endowments (0.40·400=1.60·100=1600). Incorporating this parameter into the formula for constructing the distance to CA results in:

$$\text{Productivity adjusted Distance to } CA_{ijt} = \begin{bmatrix} \left(A_{it} \cdot h_{it} - \text{Productivity adjusted } \hat{h}_{jt}^{i}\right)^2 + \\ \left(A_{it} \cdot k_{it} - \text{Productivity adjusted } \hat{k}_{jt}^{i}\right)^2 \end{bmatrix}^{1/2} \tag{13}$$

The total factor productivity indices are only available for a limited sample of countries. Of the 92 countries included in the calculations of the indices for the balanced panel, only 78 can be included in the calculation of the adjusted indices. The countries for which the productivity parameters were not available were all low income countries, which may bias the revealed factor intensities if production structures are correlated with income levels.[21] Additionally, the unavailability of those parameters for these countries, reduces the sample of developing countries in this research to 37 countries.[22] There is thus a tradeoff between the incorporation of productivity differences and the 'width' of countries on which the revealed factor intensities are based.

The results of the regressions with the productivity adjusted variables are provided in table 7 for the extensive, and table 8 for the intensive margin of export growth.[23] Data from Table 7 can be compared with data in Tables 3 and 4, which shows that most striking difference is in the magnitude of the coefficients. The coefficients on the change in total distance to CA are again found to be significant at the 1% significance level. However, the magnitude of these coefficients has decreased in both columns. For the same reasons as before, the specification in which country-product fixed effects are included is the preferred one. As compared to the regression with the baseline variables, the magnitude of the coefficient of this variable is three times smaller than for the baseline results, namely -0.002 instead of -0.006. This means that a one standard deviation increase in the change in distance to CA, decreases the likelihood of exporting with 0.2 instead of 0.4 percentage points.[23] The same difference holds for the regression in which the total change in distance to CA is split up in two effects. Similar to the results before, the change in production structure distance is the only of the two for which evidence is found that its effect on the likelihood of exporting is significantly different from zero. The value of the coefficient is the same as for the coefficient on the change in total distance to CA. However, a one standard deviation increase in the change in the production structure distance, decreases the likelihood of exporting with 0.1 instead of 0.4 percentage points.[23] The unconditional likelihood of exporting in this adjusted sample is 19.0%. Whereas the size-effects of these two variables in explaining the emergence of a 'new exported product' were already quite small, the effects when using productivity adjusted variables become almost negligible.

[21] The countries that were included in the calculations of the revealed factor intensities in the original database, but for which no data on total factor productivity is available from the Penn World Tables are: Algeria, Congo, El Salvador, Gambia, Ghana, Liberia, Malawi, Mali, Nepal, Pakistan, Papua New Guinea, Syrian Arab Republic, Uganda and Zambia.

[22] See Appendix 9.5 for an overview of the countries included in the sample for the robustness check for productivity differences.

[23] See Appendix 9.5 Table 13, for a summary of the descriptive statistics of the variables based on the samples for the productivity adjusted distances.

Table 7

The Extensive Margin of Export Growth – Productivity Adjusted Variables

Dependent Variable: New Export$_{ijt \rightarrow T}$				
Variables	(1)	(2)	(3)	(4)
Productivity adjusted	-0.006***		-0.002	
ΔDistance to CA$_{ijt \rightarrow T}$	(0.002)		(0.002)	
Productivity adjusted		-0.002		0.020
ΔEndowment distance$_{ijt \rightarrow T}$		(0.011)		(0.024)
Productivity adjusted		-0.007***		-0.002**
ΔProduction structure		(0.002)		(0.001)
distance$_{ijt \rightarrow T}$				
Productivity adjusted	-0.012***	-0.012***	0.002	0.002
Distance to CA$_{ijt}$	(0.002)	(0.002)	(0.004)	(0.003)
Country-Product FE	No	No	Yes	Yes
Country-Industry FE	Yes	Yes	No	No
Country-Year FE	Yes	Yes	Yes	Yes
Observations	68,948	68,948	66,142	66,142
R^2	0.289	0.289	0.478	0.478

Notes: The number of observations differs per column as the number of dropped singletons differs per specification due to the inclusion of fixed effects. Regression is performed in Stata with the command reghdfe. Standard errors are robust and presented in parenthesis. (***), (**) and (*) denote significance at 1%, 5% and 10%, respectively.

The results of the productivity adjusted regressions for the intensive margin of export growth are provided in Table 8, which should be compared to Tables 5 and 6. In line with the baseline results, the change in the total distance to CA is again no significant predictor of export growth (columns 1 and 3). Another finding similar to the baseline results is that the change in endowment distance is the only significant variable. The value-estimates of those significant coefficients are also not that different from the baseline estimates. Whereas the magnitude of the estimate of the coefficient of the change in the endowment distance is slightly higher when the initial export value is excluded (column 2), it is slightly smaller when this variable is included (column 4). The drop in the magnitude of this coefficient, when including the natural logarithm of the initial export value, is therefore even bigger with the productivity adjusted variables, then before. Therefore, the preferred column in table 8 is again the fourth one, which reveals that an increase of one standard deviation in the change of endowment structure distance decreases export growth with 45.7 percentage points over the five-year time interval.[21] Stated otherwise, this decreases the median of the unconditional export growth of 46.8% by almost its entire value. Compared to the mean of the export growth variable in this adjusted sample, which is 392.0%, the decrease of unconditional export growth is approximately 12%.

Table 8

The Intensive Margin of Export Growth – Productivity Adjusted Variables

Dependent Variable: Export growth$_{ijt \rightarrow T}$				
Variables	(1)	(2)	(3)	(4)
Productivity adjusted	0.047		-0.020	
ΔDistance to CA$_{ijt \rightarrow T}$	(0.308)		(0.296)	
Productivity adjusted		-5.439***		-2.929**
ΔEndowment distance$_{ijt \rightarrow T}$		(1.894)		(1.566)
Productivity adjusted ΔProduction structure distance$_{ijt \rightarrow T}$		0.120 (0.308)		0.019 (0.295)
Productivity adjusted	-0.047	-0.075	-0.192	−0.206
Distance to CA$_{ijt}$	(0.458)	(0.459)	(0.439)	(0.440)
ln Export value$_{ijt}$			-8.679***	-8.672***
			(0.493)	(0.492)
Country-Product FE	Yes	Yes	Yes	Yes
Country-Year FE	Yes	Yes	Yes	Yes
Observations	46,158	46,158	46,158	46,158
R^2	0.302	0.302	0.369	0.369

Notes: Regression is performed in Stata with the command reghdfe. Standard errors are robust and presented in parenthesis. (***), (**) and (*) denote significance at 1%, 5% and 10%, respectively.

6. Discussion

The results obtained in this research show that the view proposed in this research, namely that factor endowments should be seen as a precondition rather than strict determinants of the establishment of sustainable exports, is appropriate when looking at export baskets of developing countries. However, to some extent, these findings do clash with predictions of how factor proportions can predict the establishment and growth of exports according to traditional trade theories. Before discussing some of the confirmed predictions and contradictions of this research, the preferred models, with either the constructed baseline variables or the productivity adjusted variables, have to be chosen. The main difference in the estimates between those two options lies in the magnitudes of the estimated coefficients of the changing distances, which are overall smaller for the productivity adjusted regressions. Based on the smaller sample size and the strong assumption of equal human and physical capital productivity levels compared to the US of the productivity adjusted models, the models with the baseline variables are the preferred ones. As the results on the change in total distance to CA were not statistically significant, or equal to one of the two distinct effects, the focus will be on the specifications in which the change in endowment distance and production structure distance are treated separately (Tables 4 and 6).

Starting with the confirmation of the predictions of this research, the growth and hence sustainability of exported products is revealed to be affected by changes in endowment distance. Export growth decreases with 68.5 percentage points when the change in endowment distance increases with one standard deviation. This effect was smaller, but still of substantial size when incorporating the productivity differences. This finding is in line with the notion of the existence of high opportunity costs for sustaining the production when products are produced by countries that do not produce with comparative advantage (e.g. Fitzgerald & Hallak, 2004; Redding, 1999; Schott, 2004). The insignificant impact of changing endowment distance on export emergence, on the other hand, is inconsistent with the standard factor proportions theory. According to this theory, accumulation of factor endowments is exactly what motivates countries to change their export baskets to conform to their current factor endowments (Schott, 2003). This might imply that diversification strategies are indeed becoming more focused on path-defying exports, without developing countries considering changes of their own factor endowments.

Despite the insignificant effect of changes in endowment distance on export emergence, this research does show that the likelihood of starting to export a new product is affected by changes in the production structure distance. This could possibly be explained by more regular path-defying changes that are based on new breakthroughs (Coniglio et al., 2018). Diversification of developing countries' export baskets would then be provoked by structural change in production methods. Another explanation for this could be that the mix of countries that produce a particular product has changed, due to which the revealed factor intensities change as well if the new mix of countries is differently endowed. This could be the result of relocation of production processes from more advanced to developing countries (Lin, 2011). Nevertheless, the size of this effect is extremely small and was even smaller when using the productivity adjusted variables. A decrease of one standard deviation in the change in the production distance only increases the unconditional probability of exporting of 16.5% with 0.4 percentage points. However, the changes in the production structure distance do not seem to affect growth of already existing exports at all. An intuition for this could be that those countries have established a production structure that requires a predefined input of production factors and stick to this structure because reorganization of the production process might be too costly to gain from possible new production methods.

Although evidence is provided that export emergence and export growth are differently affected by changes in distances to CA, when interpreting these findings, the main drawbacks of this research should be taken into account. Even though the possible problem of data censoring is addressed in the best possible way with the available data, the absence of evidence supporting data censoring does not mean that this is not a problem in this research.[24] A more detailed analysis with a better selection instrument should be performed to assess whether data censoring is really not an issue here. Additionally, not enough data on productivity parameters was available to increase the reliability of this research by including these productivity differences in the construction of the distances to CA.

7. Conclusion

This research tries to contribute to the understanding of how production capabilities affect the emergence and the growth of exported products. In particular, changes in the magnitude of the distance to producing with comparative advantage are exploited over time to identify their impact on changing export baskets of developing countries. A novelty of this research lies in the separation of the total change in distance to producing with comparative advantage into two parts: the change in the distance to CA from changes in a country's own endowments and from changes in the revealed comparative production structure. Overall the prediction of an increase in the distance to CA to negatively affect growth of export baskets, without taking a stand on this growth to happen at the extensive or intensive margin, is supported by the results of this research. Interestingly, the proposition made in the introduction,

[24] See Appendix 9.4 for a description of the performed two-step Heckman Selection Model to address the possible problem of data censoring.

that the emergence and growth of exports are differently affected by the changes in production capabilities is confirmed by the results of this research.

Based on the findings of this research, it can be concluded that the emergence of new products in the export baskets of developing countries is negatively affected by the change in the distance to CA. In particular, changes in revealed comparative production structures rather than changes in their own endowments matter. Nevertheless, this effect is found to be quite small, if not negligible. On the other hand, the growth and hence sustainability of developing countries' exports are revealed to be related to changes in their own factor endowments, not to changes in the total distance to CA nor to changes in revealed comparative production structures. An increase of one standard deviation in the change of the endowment distance decreases the growth of the export value of that product with 68.5 percentage points over the five-year time interval.

The key insight from this research, that indeed factor proportions affect the emergence and growth of exports of developing countries differently, provides interesting insights for economic policy. Promoting new exports for which a developing country does not yet possess the necessary capabilities will only make sense if in the near future the relevant production factors are acquired. The ability of a country to increase its factor endowments towards the use of these factors in revealed comparative production structure should thus be considered in the formulation of economic trade policies. As the effect of the change of changing distances to CA on export emergence was found to be considerably small, future research on other determinants is necessary to be able to understand and predict changes in the extensive margin of developing countries export baskets.

8. References

Agosin, M. R. (2007). Export diversification and growth in emerging economies. *Santiago, Chile, University of Chile, Working paper* N233.

Angrist, J. D., & Pischke, J. (2008). Mostly harmless econometrics: An empiricist's companion, Princeton university press.

Bahar, D., Stein, E., Wagner, R. A., & Rosenow, S. (2018). The birth and growth of new export clusters: Which mechanisms drive diversification?

Basu, S. R., & Das, M. (2011). Export structure and economic performance in developing countries: Evidence from nonparametric methodology.

Bernard, A. B., Eaton, J., Jensen, J. B., & Kortum, S. (2003). Plants and productivity in international trade. *American Economic Review, 93*(4), 1268-1290.

Bernstein, J. R., & Weinstein, D. E. (2002). Do endowments predict the location of production? Evidence from national and international data. *Journal of International Economics, 56*(1), 55-76.

Besedeš, T., & Prusa, T. J. (2006). Product differentiation and duration of US import trade. *Journal of International Economics, 70*(2), 339-358.

Blyde, J. (2008). What drives export survival? An analysis on export duration in Latin America. *Inter-American Development Bank, Washington, DC.*

Brenton, P., Saborowski, C., & Von Uexkull, E. (2010). What explains the low survival rate of developing country export flows? *The World Bank Economic Review, 24*(3), 474-499.

Cadot, O., Carrère, C., & Strauss-Kahn, V. (2011). Export diversification: What's behind the hump? *Review of Economics and Statistics, 93*(2), 590-605.

Caselli, F., Coleman, I., & John, W. (2006). The world technology frontier. *American Economic Review, 96*(3), 499-522.

Coniglio, N. D., Lagravinese, R., Vurchio, D., & Armenise, M. (2018). The pattern of structural change: Testing the product space framework. *Industrial and Corporate Change, 27*(4), 763-785.

Coniglio, N. D., Vurchio, D., Cantore, N., & Clara, M. (2018). On the evolution of comparative advantage: Path-dependent versus path-defying changes.

Davis, D. R., & Weinstein, D. E. (2001). An account of global factor trade. *American Economic Review*, *91*(5), 1423-1453.

Feenstra, Robert C., Robert Inklaar and Marcel P. Timmer (2015). The next generation of the Penn world table. *American Economic Review, 105(10)*(3150-3182).

Fitzgerald, D., & Hallak, J. C. (2004). Specialization, factor accumulation and development. *Journal of International Economics, 64*(2), 277-302.

Greenaway, D., Kneller, R., & Zhang, X. (2010). Exchange rate uncertainty and export decisions in the UK. *Journal of Economic Integration*, 734-753.

Greene, W. H. (2003). Econometric analysis, Pearson Education India.

Guimarães, P., & Portugal, P. (2009). A simple feasible alternative procedure to estimate models with high-dimensional fixed effects.

Harrigan, J. (1995). Factor endowments and the international location of production: Econometric evidence for the OECD, 1970–1985. *Journal of International Economics, 39*(1-2), 123-141.

Harrigan, J., & Zakrajsek, E. (2000). Factor Supplies and Specialization in the World Economy.

Hausmann, R., Hwang, J., & Rodrik, D. (2007). What you export matters. *Journal of Economic Growth, 12*(1), 1-25.

Hausmann, R., & Klinger, B. (2006). Structural transformation and patterns of comparative advantage in the product space.

Hausmann, R., & Rodrik, D. (2006). Doomed to choose: Industrial policy as predicament. *John F.Kennedy School of Government, Harvard University, 9.*

Heckman, J. J. (1976). The common structure of statistical models of truncation, sample selection and limited dependent variables and a simple estimator for such models. *Annals of economic and social measurement, volume 5, number 4* (pp. 475-492) NBER.

Hellevik, O. (2009). Linear versus logistic regression when the dependent variable is a dichotomy. *Quality & Quantity, 43*(1), 59-74.

Hesse, H. (2009). Export diversification and economic growth. *Breaking into New Markets: Emerging Lessons for Export Diversification,* 55-80.

Hidalgo, C. A., Balland, P., Boschma, R., Delgado, M., Feldman, M., Frenken, K., Morrison, A. (2018). The principle of relatedness. Paper presented at the *International Conference on Complex Systems,* 451-457.

Hidalgo, C. A., Klinger, B., Barabasi, A. L., & Hausmann, R. (2007). The product space conditions the development of nations. *Science (New York, N.Y.), 317*(5837), 482-487. doi:317/5837/482.

Klinger, B., & Lederman, D. (2004). Discovery and development: An empirical exploration of “new” products, *World bank*.

Kneller, R., & Pisu, M. (2004). Export-oriented FDI in the UK. *Oxford Review of Economic Policy, 20*(3), 424-439.

Knight, M., Loayza, N., & Villanueva, D. (1993). Testing the neoclassical theory of economic growth: A panel data approach. *Staff Papers, 40*(3), 512-541.

Lectard, P., & Rougier, E. (2018). Can developing countries gain from defying comparative advantage? Distance to comparative advantage, export diversification and sophistication, and the dynamics of specialization. *World Development, 102,* 90-110.

Leontief, W. (1953). Domestic production and foreign trade; the American capital position re-examined.

Lin, J. Y. (2011). From flying geese to leading dragons: New opportunities and strategies for structural transformation in developing countries.

Melitz, M. J. (2003). The impact of trade on intra-industry reallocations and aggregate industry productivity. *Econometrica, 71*(6), 1695-1725.

Nicita, A., Shirotori, M., & Klok, B. T. (2013). Survival analysis of the exports of least developed countries: The role of comparative advantage. *Policy Issues in International Trade and Commodities Study Series, No. 54.*

Redding, S. (1999). Dynamic comparative advantage and the welfare effects of trade. *Oxford Economic Papers, 51*(1), 15-39.

Repetto, A., & Ventura, J. (1997). The Leontief-Trefler hypothesis and factor price insensitivity.

Romalis, J. (2004). Factor proportions and the structure of commodity trade. *American Economic Review, 94*(1), 67-97.

Schott, P. K. (2003). One size fits all? Heckscher-Ohlin specialization in global production. *American Economic Review, 93*(3), 686-708.

Schott, P. K. (2004). Across-product versus within-product specialization in international trade. *The Quarterly Journal of Economics, 119*(2), 647-678.

Shirotori, M., Tumurchudur, B., & Cadot, O. (2010). *Revealed factor intensity indices at the product level* UN.

Sutton, J., & Trefler, D. (2016). Capabilities, wealth, and trade. *Journal of Political Economy, 124*(3), 826-878.

Tobin, J. (1958). Estimation of relationships for limited dependent variables. *Econometrica: Journal of the Econometric Society*, 24-36.

Trefler, D. (1993). International factor price differences: Leontief was right! *Journal of Political Economy, 101*(6), 961-987.

United Nations Statistics Division. (2019). Trade statistics. Retrieved from http://comtrade.un.org.

Vanek, J. (1968). The factor proportions theory: The n-factor case. *Kyklos, 21*(4), 749-756.

Wagner, R., & Zahler, A. (2015). New exports from emerging markets: Do followers benefit from pioneers? *Journal of Development Economics, 114*, 203-223.

WITS. World integrated trade solution product concordance. Retrieved from https://wits.worldbank.org/product_concordance.html.

Wooldridge, J. M. (2016). *Introductory econometrics: A modern approach* Nelson Education.

World Integrated Trade Solution. Trade outcomes indicator. Retrieved from https://wits.worldbank.org/trade_outcomes.html.

9. Appendices

9.1 Overview of the Countries Included in the Sample

1. Algeria
2. Benin
3. Bolivia
4. Venezuela
5. Cameroon
6. Central African Republic
7. China
8. Colombia
9. Congo
10. Costa Rica
11. Ecuador
12. Egypt
13. El Salvador
14. Fiji
15. Gambia
16. Ghana
17. Guatemala
18. Honduras
19. India
20. Indonesia
21. Iran
22. Jamaica
23. Jordan
24. Kenya
25. Liberia
26. Malawi
27. Mali
28. Mozambique
29. Nepal
30. Nicaragua
31. Niger
32. Pakistan
33. Panama
34. Papua New Guinea
35. Paraguay
36. Peru
37. Philippines
38. Poland
39. Rwanda
40. Senegal
41. Sierra Leone
42. Sri Lanka
43. Sudan
44. Syrian Arab Republic
45. Thailand
46. Togo
47. Tunisia
48. Turkey
49. Uganda
50. Zambia
51. Zimbabwe

9.2 Baseline Results with Different Cut-Off Levels

Table 9

The Extensive Margin of Export Growth – Different Cut-Off Levels

Dependent Variable: New Export$_{ijt\rightarrow T}$				
Variables	Cut-off level 8,000 USD	Cut-off level 8,000 USD	Cut-off level 12,000 USD	Cut-off level 12,000 USD
ΔDistance to CA$_{ijt\rightarrow T}$	-0.006***		-0.007***	
	(0.002)		(0.002)	
ΔEndowment distance$_{ijt\rightarrow T}$		-0.024		-0.017
		(0.015)		(0.014)
ΔProduction structure		-0.006***		-0.006**
distance$_{ijt\rightarrow T}$		(0.002)		(0.002)
Distance to CA$_{ijt}$	-0.004	-0.004	-0.003	-0.003
	(0.003)	(0.002)	(0.003)	(0.003)
Country-Product FE	Yes	Yes	Yes	Yes
Country-Industry FE	No	No	No	No
Country-Year FE	Yes	Yes	Yes	Yes
Observations	99,126	99,126	104,249	104,249
R^2	0.489	0.489	0.486	0.486

Notes: The number of observations differs per column as the number of dropped singletons differs per specification due to the inclusion of fixed effects. Regression is performed in Stata with the command reghdfe. Standard errors are robust and presented in parenthesis. (***), (**) and (*) denote significance at 1%, 5% and 10%, respectively.

Table 10

The Intensive Margin of Export Growth – Different Cut-Off Levels

Dependent Variable: Export growth$_{ijt\rightarrow T}$

Variables	Cut-off level 8,000 USD	Cut-off level 8,000 USD	Cut-off level 12,000 USD	Cut-off level 12,000 USD
ΔDistance to CA$_{ijt\rightarrow T}$	-0.179		-0.225	
	(0.486)		(0.374)	
ΔEndowment distance$_{ijt\rightarrow T}$		-3.835***		-3.128**
		(1.587)		(1.515)
ΔProduction structure		-0.082		-0.031
distance$_{ijt\rightarrow T}$		(0.484)		(0.386)
Distance to CA$_{ijt}$	-0.417	-0.425	-0.302	−0.091
	(0.758)	(0.759)	(0.537)	(0.535)
ln Export value$_{ijt}$	-8.881***	-8.872***	-8.162***	-8.155***
	(0.463)	(0.461)	(0.456)	(0.453)
Country-Product FE	Yes	Yes	Yes	Yes
Country-Year FE	Yes	Yes	Yes	Yes
Observations	55,815	55,815	51,218	51,218
R^2	0.394	0.394	0.370	0.370

Notes: Regression is performed in Stata with the command reghdfe. Standard errors are robust and presented in parenthesis. (***), (**) and (*) denote significance at 1%, 5% and 10%, respectively.

9.3 Baseline Results with Ending Exports Excluded from the Sample

Table 11

The Intensive Margin of Export Growth – Ending Exports Excluded

Dependent Variable: Export growth$_{ijt \rightarrow T}$

Variables	(1)	(2)	(3)	(4)
ΔDistance to CA$_{ijt \rightarrow T}$	0.083	0.1042		
	(0.419)	(0.399)		
ΔEndowment distance$_{ijt \rightarrow T}$			-5.399^{**}	-3.460^{**}
			(2.010)	(1.894)
ΔProduction structure distance$_{ijt \rightarrow T}$			0.221	0.194
			(0.431)	(0.412)
Distance to CA$_{ijt}$	0.166	-0.031	0.147	0.049
	(0.574)	(0.497)	(0.572)	(0.545)
ln Export value$_{ijt}$		-8.622^{***}		-8.613^{***}
		(0.461)		(0.460)
Country-Product FE	Yes	Yes	Yes	Yes
Country-Year FE	Yes	Yes	Yes	Yes
Observations	51,831	51,831	51,831	51,831
R^2	0.307	0.373	0.307	0.373

Notes: Regression is performed in Stata with the command reghdfe. Standard errors are robust and presented in parenthesis. (***), (**) and (*) denote significance at 1%, 5% and 10%, respectively.

9.4 The two-step Heckman Selection Model

The estimation of the impact on changing distance to CA on export growth may be subject to estimation bias due to the non-random selection of the sample. The sample only includes export observations for the products that are exported. This is a common problem in the economic literature, which is referred to as a censored data.[25] A variety of techniques has been developed to correct for this selection problem, of which the two most prominent techniques are the Tobit model (Tobin, 1958) and the Heckman selection model (Heckman, 1976). The basic idea of those sample selection models is that the outcome variable is only observed if some criteria is met. When there is a relation between meeting the criteria, and the effect of the independent variable on the outcome variable, this will lead to biased estimates of the effect of the independent variable. Both models are based on a different set of assumptions and therefore their shortcomings differ as well. One of the most important limitations of the Tobit model for is that both the probability that an observation will be censored and the value of the dependent variable are based on the same set of variables and coefficients (Greene, 2003). This limitation stems from the computational convenience of only having to estimate one equation for this model. The Heckman selection model on the other hand, compromises on this convenience as it is based on a two-step control function approach, which allows for the selection stage (which results in the outcome to either be observed or not) to be described by different variables and coefficients than the outcome stage (the value of the original dependent variable). As this research is focused on identifying possible differences in the determinants of export diversification and export growth, the latter less restrictive model is the preferred one.

Incorporating the main idea of the Heckman selection model to this research means that the outcome variable export growth is only observed if some criterion, defined in terms of the likelihood of exporting is met. In first of the two stages of this selection model a regression for the probability of observing the dependent variable (here, the probability of exporting) has to be formulated specified by a probit model (Heckman, 1976). The second stage models the expected value of the dependent variable, conditional on that it is observed. Combined, this results in a dummy variable indicating the probability of exporting and a realization of a continuous variable, which in this case is the growth of exports. For the first stage to work, this model requires that all variables of the second stage regression should be included in the first stage. In addition, this regression should include at least one variable that is not included in the

[25] Censored data is data for which not all outcomes are observed. For some outcomes it is only indicated if they lie below or above a certain threshold (Wooldridge, 2016).

original regression (Wooldridge, 2016).[26] For this reason, a dummy variable which indicates if the country was already exporting the product in year t is included in the selection equation. This variable is often incorporated within the Heckman model when the dependent variable is export growth (e.g. Greenaway, Kneller, & Zhang, 2010; Kneller & Pisu, 2004). The argument for the inclusion of this variable is that this is consistent with export models that take into account the sunk costs of exporting (Bernard, Eaton, Jensen, & Kortum, 2003; Melitz, 2003).[27] As a probit model does not allow for the inclusion of fixed effects, controlling for other relevant heterogeneous production capabilities in addition to factor endowments is not possible here as relevant data for this is unavailable. This is the downside of taking into account possible data censoring. Using the core variable specified before and adding the additional variable $\text{Exporter}_{\text{ijt}}$, the selections model become:

$$\text{Export decision}_{\text{ijt}\rightarrow\text{T}} = \alpha_0 + \alpha_1\Delta\text{Distance to CA}_{\text{ijt}\rightarrow\text{T}} + \alpha_2\text{Distance to CA}_{\text{ijt}} + \alpha_3\text{Exporter}_{\text{ijt}} + u_{\text{ijt}} \quad (14)$$

$$\text{Export decision}_{\text{ijt}\rightarrow\text{T}} = \alpha_0 + \alpha_{1a}\Delta\text{Endowment distance}_{\text{ijt}\rightarrow\text{T}} + \alpha_{1b}\Delta\text{Production structure distance}_{\text{ijt}\rightarrow\text{T}} + \alpha_2\text{Distance to CA}_{\text{ijt}} + \alpha_3\text{Exporter}_{\text{ijt}} + u_{\text{ijt}} \quad (15)$$

Where $\text{Exporter}_{\text{ijt}} = 1$ if the country already exported product j with an export value of at least 10,000 USD at year t and 0 otherwise. The outcome equations are given by:

$$\text{Export growth}^{*}_{\text{ijt}\rightarrow\text{T}} = \pi_0 + \pi_1\Delta\text{Distance to CA}_{\text{ijt}\rightarrow\text{T}} + \pi_2\text{Distance to CA}_{\text{ijt}} + \epsilon_{\text{ijt}} \quad (16)$$

$$\text{Export growth}^{*}_{\text{ijt}\rightarrow\text{T}} = \pi_0 + \pi_{1a}\Delta\text{Endowment distance}_{\text{ijt}\rightarrow\text{T}} + \pi_{1b}\Delta\text{Production structure distance}_{\text{ijt}\rightarrow\text{T}} + \pi_2\text{Distance to CA}_{\text{ijt}} + \epsilon_{\text{ijt}} \quad (17)$$

Where $\text{Export growth}_{\text{ijt}\rightarrow\text{T}} = \text{Export growth}^{*}_{\text{ijt}\rightarrow\text{T}}$ if $\text{Export decision}_{\text{ijt}\rightarrow\text{T}} = 1$ and it is not observed when $\text{Export decision}_{\text{ijt}\rightarrow\text{T}} = 0$.[28] Stata has a built-in syntax to estimate such models.

[26] This criterion is not absolutely necessary, however when the variables included are exactly the same in the two stages it is extremely difficult to distinguish sample selection from a wrongly specified functional form of the original regression.

[27] This variable might affect export growth as well because of for example export experience or learning by doing. Nevertheless, with the available data, this was the best available option, even though the interpretation of the findings of the model should be interpreted with caution because of this ‘imperfect’ instrument.

[28] Please note that the natural logarithm of the initial export value is excluded from the outcome equation, whereas it is included in the baseline specifications on the intensive margin of export growth. The reason for this is that this variable is essentially a continuous version of the selection variable $\text{Exporter}_{\text{ijt}}$. Including this variable would thus invalidate the requirement of including an additional selection variable in the selection model, that is not included in the outcome model.

The results from the regression of the Heckman selection model are provided in Table 12. The most striking finding from this table is that for both the specifications (the total change in distance to CA and the distinct distance changes), lambda is not significantly different from zero. This implies that no evidence is found for sample selection due to data censoring. The other coefficients in this table will not be interpreted, as there now is no reason to believe anymore that these coefficients are reflecting the true effects of the variables better than the coefficients in the baseline intensive margin model. Additionally, because of the exclusion of the various fixed effects included in the baseline models, the estimates in Table 12 are even expected to be further from the true effects.

Table 12

Two-Step Heckman Selection Model

Dependent Variable:	Export decision$_{ijt\rightarrow T}$		Export growth$_{ijt\rightarrow T}$	
Variables	(1)	(2)	(3)	(4)
ΔDistance to CA$_{ijt\rightarrow T}$	0.005		-0.449***	
	(0.010)		(0.197)	
ΔEndowment distance$_{ijt\rightarrow T}$		-0.007		-2.236***
		(0.032)		(0.604)
ΔProduction structure distance$_{ijt\rightarrow T}$		0.007		-0.271
		(0.011)		(0.205)
Distance to CA$_{ijt}$	-0.043***	-0.044***	-0.551***	-0.553***
	(0.006)	(0.006)	(0.120)	(0.120)
Exporter$_{ijt}$	4.674***	4.674***		
	(0.062)	(0.062)		
Lambda	-0.432	-0.517	-0.432	-0.517
	(1.819)	(1.820)	(1.819)	(1.820)
Observations	165,240	165,240	165,240	165,240

Notes. Standard errors are in parentheses. Regression is performed in Stata with the command heckman. The inverse mills ratio is reflected by lambda. If lambda is significantly different from zero, this suggests there is evidence in favor of sample selection. (***), (**) and (*) denote significance at 1%, 5% and 10%, respectively.

9.5 Robustness Check – Productivity Differences

Overview of the Countries Included in the Sample for the Robustness Check

1. Benin
2. Bolivia
3. Venezuela
4. Cameroon
5. Central African Republic
6. China
7. Colombia
8. Costa Rica
9. Ecuador
10. Egypt
11. Fiji
12. Guatemala
13. Honduras
14. India
15. Indonesia
16. Iran
17. Jamaica
18. Jordan
19. Kenya
20. Mozambique
21. Nicaragua
22. Niger
23. Panama
24. Paraguay
25. Peru
26. Philippines
27. Poland
28. Rwanda
29. Senegal
30. Sierra Leone
31. Sri Lanka
32. Sudan
33. Thailand
34. Togo
35. Tunisia
36. Turkey
37. Zimbabwe

Table 13

Summary Statistics – Robustness Check Productivity Differences

Variable	*N*	*Mean*	*SD*	*Min*	*Mdn*	*Max*
Panel A $\overline{x_{ijt}} < 10{,}000$						
New Export$_{ijt\rightarrow T}$	69,511	0.190	0.393	0	0	1
Productivity adjusted ΔDistance to CA$_{ijt\rightarrow T}$	69,511	-0.031	0.750	-6.113	-0.034	8.063
Productivity adjusted Distance to CA$_{ijt}$	69,511	1.418	0.907	0.001	1.249	9.537
Productivity adjusted ΔEndowment distance$_{ijt\rightarrow T}$	69,511	-0.024	0.168	-0.569	-0.012	0.562
Productivity adjusted ΔProduction structure distance$_{ijt\rightarrow T}$	69,511	-0.007	0.734	-5.878	-0.014	8.004
Panel B $\overline{x_{ijt}} > 10{,}000$						
Export growth$_{ijt\rightarrow T}$	50,369	3.920	28.679	-1	0.468	2063.610
ln Export value$_{ijt}$	50,369	5.745	2.493	2.303	5.285	16.659
Productivity adjusted ΔDistance to CA$_{ijt\rightarrow T}$	50,369	0.008	0.653	-6.159	0.003	5.809
Productivity adjusted Distance to CA$_{ijt}$	50,369	1.369	0.878	0.002	1.177	8.003
Productivity adjusted ΔEndowment distance$_{ijt\rightarrow T}$	50,369	0.003	0.156	-0.561	0.000	0.473
Productivity adjusted ΔProduction structure distance$_{ijt\rightarrow T}$	50,369	0.005	0.638	-5.971	0.004	5.651

Notes: Variables are constructed as described in Section 3.2. ln Export value$_{ijt}$ reflects the natural logarithm of the export value in 1,000USD.