The Impact of Market Potential on Income Distribution

Thesis presented by
Jonathan Teixeira Dos Reis

Supervisor
Florian Mayneris (UCL/UNamur)

Reader
Eric Toulemonde (UCL/UNamur)

Academic year: August 25, 2017

In order to obtain the Joint Degree
Master 120 en Sciences économiques, Orientation générale, Finalité spécialisée (UCL/UNamur)
and
Master en sciences économiques (UNIGE)

Ecole d'économie de Louvain/UCL • Place Montesquieu 3 • 1348 Louvain-la-Neuve / Belgium
Département des Sciences économiques/UNamur • Rempart de la Vierge 8 • 5000 Namur / Belgium
University of Geneva/SES – Uni Mail, Bd du Pont d'Arve 40 – CH-1211 Genève / Suisse

Université Catholique de Louvain

Abstract

Faculté des sciences économiques, sociales, politiques et de communication
Master en sciences économiques (UNIGE)

Master in Science of Economics

The Impact of Market Potential on Income Distribution

by Jonathan TEIXEIRA DOS REIS

This paper analyses income inequalities through the perspective of spatial geography. Constructing for 2004 and 2012 the Market Potential thanks to the gravity equation, we attempt in linking spatial geography with inequalities within countries, measured by the Gini index. We propose that the direction of the relationship depends on a specific combination between the wage ratio skilled-unskilled workers and the share of the skilled workers in total population. The results show that the relation is non-linear: globally negative for the entire sample and slightly negative for advanced economies. Disentangling the effect, this difference might come from a higher share of skilled workers and a higher degree of mobility in advanced economies.

JEL classifications: J31, J61, O10, 019, O57, R10
Keywords: Economic Geography, Market Potential, inequalities, Gini index

Acknowledgements

I especially thank my supervisor the Professor Florian Mayneris for its beneficial advices and support in this adventure. I also would like to give particular thanks to Kevin Antunes, Stéphane Berberat, Ian Bill, Othman Gacem, Cristina Iselin and Ka-fat Leung for their precious help.

Table of content

List of Tables

List of Figures

Chapter 1

Introduction

Thomas Piketty's book on the dynamics of inequalities[1] has had an important effect on the research literature. Depicting the internal causes of the evolution of inequalities, one can also be interested in estimating the effect of external factors on the dynamics of this phenomenon. In particular, the effect of the huge increase, since the 90's, of the exchange of goods and services between countries and the international evolution of the factors of production might be worth exploring.

Various inequality evolutions among countries are published in the literature. De Oliveira Cruz and Naticchioni (2012) find a decrease of inequalities for Brazil during 2002 and 2009. Ho, Wei, and Wong (2005) analyze the increase in inequalities between 1991 and 2002 in Honk Hong. Hering and Poncet (2010) attempt in explaining the increase in inequalities in China since 1980. Charnoz, Coudin, and Gaini (2011) find a decrease in inequalities between 1995 and 2004 for France.

Moreover, different arguments are used to explain the internal development of countries. GDP growth, differences in endowments, institution differences, etc. More recently, the geography of location also emerged as a significant constituent for the determination of wage differences across entities. Indeed, New Economic Geography (NEG) environment provides interesting insights to study the welfare impact on individuals due to the location choice of different agents. Even if its primary goal is to analyse and explain why economic activity agglomerates in certain regions, links between returns of production factors and the location of various agents can be derived from models in this field. These links depend on different hypothesis. The mobility of factors is one of them and plays a crucial role in the determination of the evolution of factor inputs returns.

Through the interaction between transport costs and increasing returns to scale, NEG models attempt to explain the concentration or the dispersion of the economic activity. One implication that can be derived from this field is that the firm's access to markets counts in determining the profits it makes and the returns a firm can afford to pay to its factors of production.

Indeed, to provide goods to individuals, a firm must sell its products to consumers that are spread around the world. If a firm sells its products to consumers that are close, it will have less transport costs when shipping its products. As a result, this firm will make more profits. However, firms should not be to close to consumers, otherwise they will not benefit from increasing returns to scale. Consequently, they do not follow exactly the demand of their products. The combination of transport costs and increasing returns to scale results in the agglomeration of the production in specific places.

With free entry and if a firm makes profits (due to an decrease of its transport costs for example), these benefits will immediately shrink to zero because more firms will be tempted to enter the market to also make profits. How do they decrease to

[1] Piketty (2014).

zero? *New* firms increase the competition inside this market and lower the demand addressed to our firm. This drives the profits to zero. On the other hand, if firms and workers are not mobile, our firm can apply a higher price on its goods. Therefore, it has the capacity to remunerate higher its factors of production. With this example, we observe that depending on the mobility or not of firms and workers, higher accessibility to consumers implies an increase of the number of firms and workers or an increase in the returns to factor inputs.

When it comes to the labor factor of production, many papers underline the positive link between access to markets and differences in wages across countries or regions. Indeed, a difference in market access between countries induces a difference in returns to workers across those entities. A consistent number of papers have found a positive link between the two[2].

Redding and Venables (2004) look at the role of geographical location in explaining differences in wages for 101 countries. They use a model of economic geography from Fujita, Krugman, and Venables (1999) where they derive two mechanisms through which geographical location affects wages: distance of demand and distance of supply. They test two relationships among others: the link between wages and the distance of the demand and the link between wages, distance of the demand and distance of the supply in a single equation. In the NEG field, this equation is called the "wage equation". Redding and Venables find that distance of the demand and distance of the supply explain most of the differences in wages *across* countries.

Head and Mayer (2011) extend their work, but formalize the impact of economic geography differently. They use only a measure of proximity to markets (which is the same as the distance of the demand) that they call Market Potential. They find that the Market Potential drives differences in wages and thus economic development. This means that large (lot of consumers) and centrally located countries (low transport costs) are richer than those with small markets[3].

Finally, Hanson (2005) find the same link between Market Potential and differences in wages. One of the divergences in his study stands in its level of observation. The two previous authors consider the country level, while he chooses the regional level.

The three previous papers take different measures to capture the impact of economic geography, but they have in common the goal of analyzing the impact of geographical location on wages compared across countries. They find that the Market Potential is a powerful driver of wages. Less investigated, predictions on the evolution of the income distribution *in* a region could be formulated.

Still, does the Market Potential do better than other elements in explaining the income distribution? This question goes beyond our scope of analysis and requires to compare Market Potential with others variables. Instead, we will focus on the capacity of the Market Potential in providing interesting insights about the income distribution. Does it increase the wage of a specific group? If yes, does it also benefit the entire society? Or, in other words, do we manage to obtain a specific shape for the income distribution depending on where a country is located compared to foreign markets?

When Market Potential (from now on MP) is low, for example when there is very high trade costs as stated in Krugman and Venables (1995), the equilibrium of the two regions occurs when all the sectors remunerate their factor of production the same return (here, workers across regions have the same characteristics). This

[2]In its literature review, p.20-23, Redding (2013) mentioned some papers that find this "association".
[3]Head and Mayer (2011), p.288.

means that each individual in a given region will have the same wage. Thus, each individual will have the same share of the total income. Taking the Gini index has a measure of inequalities, the Lorenz curve will perfectly match the 45° line in figure 2.1. This is true under the assumptions that each worker can move without frictions across sectors. If we believe that the Krugman and Venables (1995) model describes the organization of the world economy, high level of trade costs between regions (so, low level of MP) should be observed with very low level of inequality within the region.

Furthermore, different papers investigate the consequences in wage disparities within the unit of observation[4]. In the international economic field, Verhoogen (2008) find that the effect of Trade between Mexico and United States has a positive return on skilled Mexican workers, since the goods exported by the former were oriented toward high-quality products. They then test a decrease in the price of these products (through devaluation of the *peso*) and find that this effect benefits relatively more to productive firms and their skilled workers.

Ho, Wei, and Wong (2005) investigate the impact of the opening process of Hong Kong with China on the wage differences between university, secondary and primary graduates. Their goal is to identify the direction of the causality and they use thus co-integration techniques. They obtain that the causality runs from the outsourcing Trade to wage differential and that it raises the wage of the skilled workers, while decreasing those who are less educated (negative impact for primary graduates).

Helpman, Itskhoki, Muendler, and Redding (2017) determine that two-third of the transmission mechanism between Trade and wage inequalities is driven by within "sector-occupation", where between firms wage differences accounted for a large part in this effect. They then derive a model where Trade affects wage inequalities through two mechanisms: market access[5] and selection effect. They find a non-linear relationship between Trade and inequalities, where the former first increase inequalities and then decrease it.

They do not use the Market Potential measure to identify the effect of the opening process of Trade, but links can be established with the economic geography field. Indeed, an improvement in Trade process increases the accessibility of the demand that consequently increases the MP of a country.

In the NEG literature, Redding and Schott (2003) and Fallah, Partridge, and Olfert (2011) gave two solid foundations for the causal link between Market Potential and wage disparities. The former analyses the impact of distance on wages and incentives to invest in education. They find that remote countries have lower wages and tend to invest less in human capital. A lower MP is associated with lower educational attainment. They use the theoretical framework from Fujita, Krugman, and Venables (1999) where two types of factors of production are presented: skilled and unskilled labor. The intuition is as follows: if the MP increases, there will be an increase in the demand for skilled workers if the sector that benefits from increasing returns to scale uses intensively skilled workers. In the presence of immobile workers across countries, the wage of the skilled workers must increase, which will encourage more unskilled workers to invest in education to become skilled. The nominal wage of the skilled workers increases and the nominal wage of the unskilled workers decreases[6].

[4] Behrens and Robert-Nicoud (2014) and Behrens, Pokrovsky, and Zhelobodko (forthcoming) provide insights linking market size, sorting of workers and inequalities in urban economics field.

[5] In this particular paper, the market access is not exactly the same as we understand it in our paper.

[6] Redding and Schott (2003), p.15-16 and Fallah, Partridge, and Olfert (2011), p.870.

Looking at the wage disparities at the regional level in the United States, Fallah, Partridge, and Olfert (2011) derived a "spatial skill demand equation" (from the same theoretical model as Redding and Schott (2003)), where an increase in Market Access (synonym of Market Potential) increases wages' disparities. Dividing the labor force into three types of workers, they found out that an increase in Market Access increases the gap between high and low skilled labor and between medium and low skilled labor.

Fally, Paillacar, and Terra (2010) find a positive impact of Market Access on the wage differentials across states and industries in Brazil. They implement robustness check with the inclusion of firm productivity, endowments, etc. to carry out this result. The relationship was still positive and significant.

Hering and Poncet (2010) obtain a higher positive impact for skilled Chinese workers than unskilled ones after an increase in Market Access. They also mention a second adjustment mechanism through which the economy respond to a change in Market Access: workers move across sectors to meet changes in demand. Therefore, they suggest that the wage mechanism is weaker in the case where the labor market is highly competitive[7].

Indeed, one issue concerning the four previous papers is that they supposed that the wage differential between specific population groups is enough to determine the income distribution. Only considering the wage disparities might not be sufficient. It actually depends on the exact share of these groups in the overall population. If there are only two types of people (skilled and unskilled workers) and skilled workers are the larger group, the MP increases the wage differential between the two groups, as we saw earlier. However, an increase in MP will increase the demand for skilled workers and thus increase the proportion of people with the same wage. Workers meet changes in demand as stated in Hering and Poncet (2010). The region will be composed with a larger share of individuals with the higher wage. The increase or decrease in inequality depends thus on these two forces, that might go in opposite directions. The first one is related to the wage disparities and the second to the share of each group in overall population, which is influenced by workers mobility.

In the NEG approach two mobility can be studied[8]. The mobility of firms and the mobility of workers. Firms decide to locate where there is a higher demand for their product and where they can be supplied the easiest (Venables (1996) and Strauss-Kahn (2001)). This is called the *backward linkage*. Workers choose their location considering the wage they receive and the supply of goods they can consume. This is the *forward linkage*. Workers will be employed here, since we are interested in modeling inequalities across people. In this framework, Redding and Schott (2003) again analyze also the impact of Market Potential on mobility of workers, assuming that unskilled workers can become skilled through education. They find that a higher MP is consistent with a higher educational attainment.

In this paper, implications on the mobility of workers and implications on wage disparities within the unit of observation will be combine in a same environment to determine a broad pattern for inequalities evolution. Using the NEG approach is helpful since mobility and wage disparities are influenced by the same element, the Market Potential. They are key ingredients to measure the inequality dynamics.

A second question that emerges is whether the link between MP and inequalities actually holds also for countries and not only on a regional level. This might

[7] Hering and Poncet (2010), p.156.

[8] Head and Mayer (2004b), p.15.

not be so simple since it requests another level of estimation and modeling. Taking international interactions to explain within country disparities is not without any sense since Fally, Paillacar, and Terra (2010) find that international spatial interaction have a higher impact than local and national interactions on regional differences for Brazil[9].

On the inequality side, an information that can combine wage distribution and the share of a group in total population at the country level is the Gini index. This is an originality of our paper since no authors used it to estimate the dynamics of inequalities. We will employ it as our measurement of income distribution.

Our strategy increases the understanding of the consequences on income inequality. First, economy geography approach as the advantage of providing implications on the two dimensions that we suppose do be determinant in shaping income distribution. Second, the Gini index, that measures the income distribution, is constructed using these two dimensions.

For this end, we will consider the simplest case with only two groups forming the population. This simplicity puts in evidence some links, but inevitably lacks realistic backdrop that an explanation needs. Finally, this paper does not aim at determining if the Market Potential is actually the primary driver of the inequalities[10].

The contribution of this paper will be twofold. First, a structure to model the overall impact of the Market Potential on the income distribution will be provided. Second, we will disentangle and attempt in testing the multiple effects of the Market Potential on the Gini index. This requires five chapters: chapter 2 presents the theoretical framework where we explain the impact of spatial economy on wages. The construction of the Market Potential will then be formulated. Finally, the Gini index and its dynamics will be evaluated. In chapter 3, we will present the data and our estimation strategy. Chapter 4 will be dedicated to explain the empirical estimation. In chapter 5, we present the results. Finally a discussion and a conclusion stand in chapter 6.

[9] Fally, Paillacar, and Terra (2010), p.162.

[10] Bosker and Garretsen (2009) provide evidence for the role of institutions and neighboring institutions in shaping the cross-countries income differences.

Chapter 2

Theoretical environment

This theoretical environment will be divided in three parts. Firstly, the theoretical framework presents the model that can, at best, incorporate the two propositions of New Economy Geography. Accordingly, we choose the model developed by Fujita, Krugman, and Venables (1999), then add the mobility between population groups formulated by Redding and Schott (2003) and use inequalities' dynamics presented by Fallah, Partridge, and Olfert (2011). Herein, only the relevant steps of the model are showed.

The second part depicts the Market Potential, where its origins are explained. Challenges to derive the MP are also mentioned. One of them relates to obtain a correct structure to estimate it. This structural equation is called the gravity equation and requires specific elements in itself as multilateral resistance terms, measurement of the demand, etc... These are issues that needed to be considered to find the MP.

Finally, the Gini index is presented as the measure to capture the income distribution of the different countries. This index has a favorable outcome to gather at the same time the cumulative share of the individual's income from the lowest to the highest and the cumulative share of the income that is earned. These two scales are connected to our two propositions where the MP can be hold accountable. The Gini index represents thus a suitable measure. This last part begins with a short description of it and ends with its implications when we consider only two groups composing the overall population.

2.1 Theoretical framework

The model we present originates from Fujita, Krugman, and Venables (1999)[1]. It was first elaborated by Krugman (1991) and from the work of previous authors in the 90's and 80's. Here, it is not described entirely, since we are only interested on the implications of the MP. The main purpose of this paper remains the empirical analysis. We also indicate the adjustments that we perform regarding the original model for the purpose of applying our ambitions. Therefore, we comply with the two previous articles if the reader searches the full development. The model described by them is resumed as follows.

Supposedly, the world is made of two countries, where each of them produces two types of goods in two different economic sectors. The agricultural sector produces an homogeneous agricultural good subjected to perfect competition and the manufacture sector produces differentiated goods (each firm produces one variety) under monopolistic competition. The customers consume both the agricultural good and all the varieties of the manufacturing sector giving them a specific utility. They

[1] Fujita, Krugman, and Venables (1999), p.45-55.

maximize their utility under their budget constraint, which yields the demand for each manufacturing variety:

$$q_i(v) = \mu \cdot Y \cdot p_i(v)^{-\sigma} \cdot P_i^{\sigma-1} \tag{2.1}$$

$q_i(v)$ is the demand in country i of the variety v. The total income is divided between the agricultural good and the manufacture variety expenditure. The former is represented by $\mu \cdot Y$, where μ is the expenditure share of manufactured varieties and Y is the total income. p_i is the price of the variety v. P_i is the price index in country i and measures "the minimum cost of purchasing a unit of the composite index"[2] of manufacturing varieties and its value is different in each country. The price index is an important measure since it captures the number of firms in the country i. More firms imply more competition, which drives these firms to lower their price to attract demand. Therefore, the price index will be lower where there is tougher competition and thus, the lower the demand for each variety. Finally, σ is the elasticity of substitution between varieties. It is higher than one. If the elasticity is high (varieties are more substitutable), the quantity will be more sensitive to changes in prices.

From the production side, the agricultural good uses constant returns to scale and the manufactured firms use increasing returns to scale. The technology is the same for all firms in the manufactured sector and involves a fixed input and a marginal input. We follow Fallah, Partridge, and Olfert (2011) in assuming that the two sectors use skilled and unskilled workers to produce their goods. However, we adopt Redding and Schott (2003) description of the dynamics of factor mobility within the unit of observation. They assume that everyone is born with an unskilled endowment and has the choice to transform this endowment into a skilled characteristic. They indicate that the transformation occurs through education and induces thus a cost, which is reduced if the individual has more "abilities"[3]. Moreover, unskilled workers do not only endure costs when they become skilled. They also benefit from higher wages. At the end of the day, individuals choose to become skilled (to stay unskilled) if the benefits are higher (lower) than costs.

This trade-off is an essential assumption. It gives social dynamics within the unit of observation that influences the income distribution. If unskilled workers become skilled, the inequalities of income increase or decrease depending on the initial share of the group in total population. Moreover, individuals consider *inter alia* wage differentials when deciding who to become. Therefore, prices affect the supply of the factor of production. On the contrary, the more of less important supply of skilled workers determines the wages. Consequently, there is a mutual impact between the two. Last but not least, mobility of workers across countries is not modeled for simplicity reasons.

In the Fujita, Krugman and Venables model, all firms applied the same price, which is a constant mark-up over the marginal cost. Free entry and exit imply that firms break-even when producing the optimal quantity $q*$, which is constant. The price needed to sell $q*$ is determined by the demand equation for firms (home and foreign demand). Combining the price equation (not transcribed here) and normalization assumptions[4], we obtain the wage equation:

[2]Fujita, Krugman, and Venables (1999), p.47.
[3]Redding and Schott (2003), p.8.
[4]See Fujita, Krugman, and Venables (1999), p.54-55.

$$(w_i^s)^\alpha \cdot (w_i^u)^\beta = \Big[\sum_{j=1} Y_j \cdot P_j^{\sigma-1} \cdot T_{ij}^{1-\sigma}\Big]^{1/\sigma} \tag{2.2}$$

This equation 2.2 emanates from Fallah, Partridge, and Olfert (2011)[5]. w_i^s is the wage in country *i* of skilled workers and w_i^u is the wage of the unskilled workers in country *i*. α and β are the share of each inputs used in the production function of the firm. T, which is equal or bigger than 1, stands for the transportation impediments. It captures the costs when a good is shipped. These costs may take the form of taxes, communication costs, etc. T_{ii} represents the cost of shipping a good in the country where it is produced. We will assume that this specific cost is null. Thus, in this case, *T=1*. Moreover, it implies that the delivery price of the variety in a foreign country will be higher than the one applied in the country where it is produced. This price is represented as $p_{ij} = p_i \cdot T_{ij}$.

Equation 2.2 postulates that wages from skilled and unskilled workers are linked to income, competition and transportation costs of all delivered countries (the country where goods are produced is also included). An increase in income, a decrease in the price index or in transportation costs will increase the skilled and unskilled wages a firm can afford to pay.

With this model, inferences of spatial economic activity on wages can be derived and tested. As the subject of this paper is wage disparities, we look for wage inequalities instead of two wages on the left-hand side of equation 2.2. Fallah, Partridge, and Olfert (2011) and Redding and Schott (2003) formalize it using this equation in two separate ways but agree on the same implications[6]. We will not rewrite what they find, only the economic intuition result behind it. The intuition is as follows: an increase in income, a decrease in the price index or in transport costs will increase the relative wage of skilled workers if the manufacturing sector employs relatively more skilled workers than unskilled workers compared to the agricultural sector. The reason is that the manufacturing sector expands with an increase in income. This raises the price at which a firm in the manufacturing sector can break-even without making any losses. This increases the demand for higher skilled workers than what the agricultural sector can "release"[7] (due to the fact that manufacturing sector is intensively skilled). In consequence, wages of skilled workers increase and wages of unskilled workers decrease.

Thanks to this hypothesis, we are able to predict the impact of a change in economic activity on wage differentials and thus on income distribution. We know that changes in wage differentials have an impact on mobility of workers, since we assume a specific link between the two through education above. Higher wages for skilled workers will increase the benefits to acquire education. The supply of skilled workers increases and the supply of unskilled workers decreases. The effect will not stop, since an increase in supply of skilled workers and a decrease in supply of unskilled workers will *decrease* the wage gap between them. This will have again an impact on skilled-unskilled mobility, etc... until another equilibrium is reached. The second effect does not cumulate with the first one, as they go in opposite directions. As long as the sectors produce in both regions, a *small* impulse from spatial economy change will affect the equilibrium, but will not produce catastrophic results, as full agglomeration. Prices of inputs serve as dispersion forces.

[5] Equation (12), p.869.
[6] P.870 for the former and p.15 for the latter.
[7] Redding and Schott (2003), p.15.

However, the interest in this paper lies in the connection between Market Potential and income distribution. As we will see in the next section 2.2, the right-hand side of equation 2.2 corresponds in fact to the MP. This means that every link or intuition that we establish between economic activity and wage differentials is related to the MP. In the next section, we will explain this appearance.

2.2 Market Potential

A Market Potential captures the *potential* of a unit (country, region, firm) in doing business with other units. When it produces a certain good or product, its obvious goal is to sell them to costumers. Does this unit have more "likelihood" to meet consumers when it stands in some places around the world? What is its "potential" to make business with other people around when it locates in certain places? Does it really matter where a productive unit locates? The MP tries precisely to measure if there are differences among locations in their potential to make business. It is precisely an example that location matters[8]. Indeed, MP catches presences (in the form of masses) in the space dimension and some presences count more than others because they are bigger in terms of number of consumers (potential buyers of the product) or they are closer and thus easier to reach. Being close to Paris or to Avignon is not really the same in terms of potential for a firm x.

Identifying a way to capture this measure is not an easy task. Bosker and Garretsen (2012) write that there are two ways to deal with the construction of the MP. The first one uses the intensity of commercial relations between countries. If a firm exports more to a country than another one, this means that it finds more profitable to exchange with the first one than the second. Or, that some elements facilitate the trade in the first case and impede trade in the second.

Commercial relations should reflect the decision process of agents (in the case where we consider the country as the unit of observation, the agents are all the firms that are in the considered country and that sell products). We concede that these agents look at diverse conditions when they decide to whom selling their products and where to locate (Head and Mayer (2004a)). If the MP captures the potential of a unit in doing business, it should pop up and appear as it is an important determinant of trade. Fortunately, this is feasible with the "gravity equation".

This equation links different elements. It reckons with the presence of the units of observation and the trade among them. "Gravity" stands for its link with the Law of gravity, where the intensity of attraction is explained by the mass of the two physical bodies subjected to their bilateral distance. The gravity equation "structures" commercial relations, where the intensity of trade flows is explained by the mass of the units of observation. It gives an idea of what the trade flows should be.

It was first presented by Tinbergen in 1962 and is a central equation in two points. First, it has multiple utilities and can also explain other flows, as the migration flow. Second, the MP and the gravity equation are closely linked. If the decision to export depends on the MP, the gravity equation, that states and structures what the trade flows should be, must include an element characterized as a MP. This approach has a *top-down* feature since the gravity equation embedded the MP in a framework represented by an equation.

The second way to deal with the construction of this variable attempts directly to measure it with economic variables. This methodology has a *bottom-up* approach,

[8] See Head and Mayer (2014).

because it constructs the MP with different components. These variables include proxies that estimate the mass of consumers and where they are located.

This latter approach consists of two stages. It first involves evaluating to whom the product is sold and what kind of environment there is. Does the country of arrival have a large market? Who sells also there? What are the rules to sell the good in that country, etc.? This first stage is often represented by two elements: a term capturing the mass of consumers and a second, called "multilateral resistance terms", catching the overall environment there.

The second step requires the measurement of the transportation costs of the good until it reaches the consumer. Where does the country sell its products? Is the good sold relatively close to its production location or is it far away? Does it cross a border and taxes have been applied? This part measures the "accessibility" of the market from the location where the product is sold to the store. More remote countries will be difficult to serve since the firm will have to pay a higher delivery price than firms that are close. Numbers of different variables are included to catch all the elements that impede trade to occur between two parters. Distance is the first variable that we think about, but borders also can impede trade among countries, trade agreements, also languages, etc.

We mention in the previous section that equation 2.2 already contained the MP. Indeed, Y_j measures the income in country *j*, $P_j^{\sigma-1}$ measures the multilateral resistance term, as it contains a piece of information on the number of firms in country *j*. This gives an idea on the toughness of this market. Finally, $T_{ij}^{1-\sigma}$ captures the accessibility of each market *j* since it represents the transport costs. This way, we just linked the MP with wages and thus with wage differentials.

The construction of the MP for each country is not an easy task when estimating it directly. Indeed, how can we estimate the multilateral resistance term? Which variables represent better the overall environment in a country? This is the reason why the gravity equation is a more appealing technique.

Instead of trying different elements to represent the MP, the gravity equation brings out another procedure. First, it states what the trade flows should be. Second, with the pieces of information that we collect with the second approach, there is only one element that we need to succeed in explaining trade flows. Once we include this piece, we shift from the Market Potential to the gravity equation. What we need to know concerns the exporter. Are many firms exporting in this country? Are they efficient? If they use intermediate inputs in their production, are they well supplied? At this point, we have two elements that capture the presence of two countries trading together (information on the importer and now on the exporter) and one element catching the trade impediments. This way, we can quantitatively measure for which reason countries trade and determine why specific countries count more than others when they trade. Following Head and Mayer (2014), the structural gravity equation[9] is:

$$X_{ij} = \underbrace{\frac{Y_i}{\Omega_i}}_{\mathrm{S}_i} \cdot \underbrace{\frac{X_j}{\Phi_j}}_{\mathrm{M}_j} \cdot \phi_{ij} \tag{2.3}$$

Where X_{ij} is the bilateral exports of country *i* to country *j*. S_i captures pieces of information from the exporter *i* and M_j from the importer. They are the supplier capacities and the market capacities for each bilateral commercial relation. Y_i is the

[9]Head and Mayer (2014), equation (2) p.138.

production of the exporter and Ω_i the multilateral resistance term for the exporter[10]. X_j is the importer's expenditure to the exporter i and Ω_j the multilateral resistance term for the importer[11]. ϕ_{ij} is the "phi-ness" of trade and is equal to $T_{ij}^{1-\sigma}$. It measures the transportation costs.

This gravity equation has the power to gather in a single equation, spatial interactions emerging from international relations. It will enable us to construct the MP. Taking the logs of 2.3 we obtain:

$$ln(X_{ij}) = ln(S_i) + ln(M_j) + ln(\phi_{ij}) \tag{2.4}$$

This equation will be our benchmark to estimate the MP. $ln(S_i)$ is estimated using exporter fixed effect and measures what kind of exporter the country i is. $ln(M_j)$ is evaluated by importer fixed effects. Finally, ϕ_{ij} is estimated by a vector of trade impediments and it weighs the demand of each importers to obtain true experience of trade. The MP for country i is the sum of all market capacities weighed by the trade impediments. It is defined as:

$$MP_i = \sum_j M_j \cdot \phi_{ij} = \sum_j X_j/\Phi_j \cdot \phi_{ij} \approx \sum_j Y_j/P_j^{1-\sigma} \cdot T_{ij}^{1-\sigma} \tag{2.5}$$

The far-right part occurs when we use the model described in section 2.1. We obtain exactly the left-hand side of equation 2.2.

Links between the MP on one side and mobility of workers and wage differentials on the other side have been discussed thanks to the model. The construction of the MP was presented through the gravity equation. The next section introduce the Gini index and how the income distribution dynamics are linked to the mobility of workers and the wage disparities. Implications of considering only two countries to remain in line with the theoretical framework will also be discussed.

2.3 Gini index

To capture the inequalities of income, we use the Gini index. As a consequence, we deviate from most of the papers that chose the wage ratio between two groups. One related to skilled workers and the second to unskilled workers[12]. Employing this index deviates us from the strict version of the model, since there is no left-hand side element representing *inequalities* in it. However, we mention before that the MP has an effect on the mobility of workers and the wage differential between groups. Our measure has the advantage of gathering in a single variable these two effects.

The Gini index evaluates the income dispersion in a given country. Its value is ranged between 0 and 1, where 0 stands for perfect equality and 1 for perfect inequality. The graphical representation of this index lies in figure 2.1[13].

Its dynamics are complex, since it sorts each individual from the poorest to the richest depending on its income. We cannot follow each individual and know how

[10]They express different variables depending on the model we use. In this paper, Y_i and Ω_i are represented by the number of firms and the price they applied.

[11]As for the supplier capacity, M_j might take different forms if we use separate models. They are composed by the manufacture expenditure and the price index in our model.

[12]Redding and Schott (2003), Ho, Wei, and Wong (2005), Hanson (2005) and Fallah, Partridge, and Olfert (2011) did it. De Oliveira Cruz and Naticchioni (2012) perform multiple regressions on different percentile of the Brazilian population. However, they analyze the wage premium with an urban economic approach. Fally, Paillacar, and Terra (2010) use state-industry fixed effect instead.

[13]This figure is taken from Li, Shi, Zhao, and Fred (2014), p.495.

much he earns every moment. We thus consider only two groups of people composing the population and stay in line with the theoretical framework by keeping the distinction between skilled and unskilled workers. Even with this extreme simplification, we will transcribe, as best as possible, the true dynamics of this index.

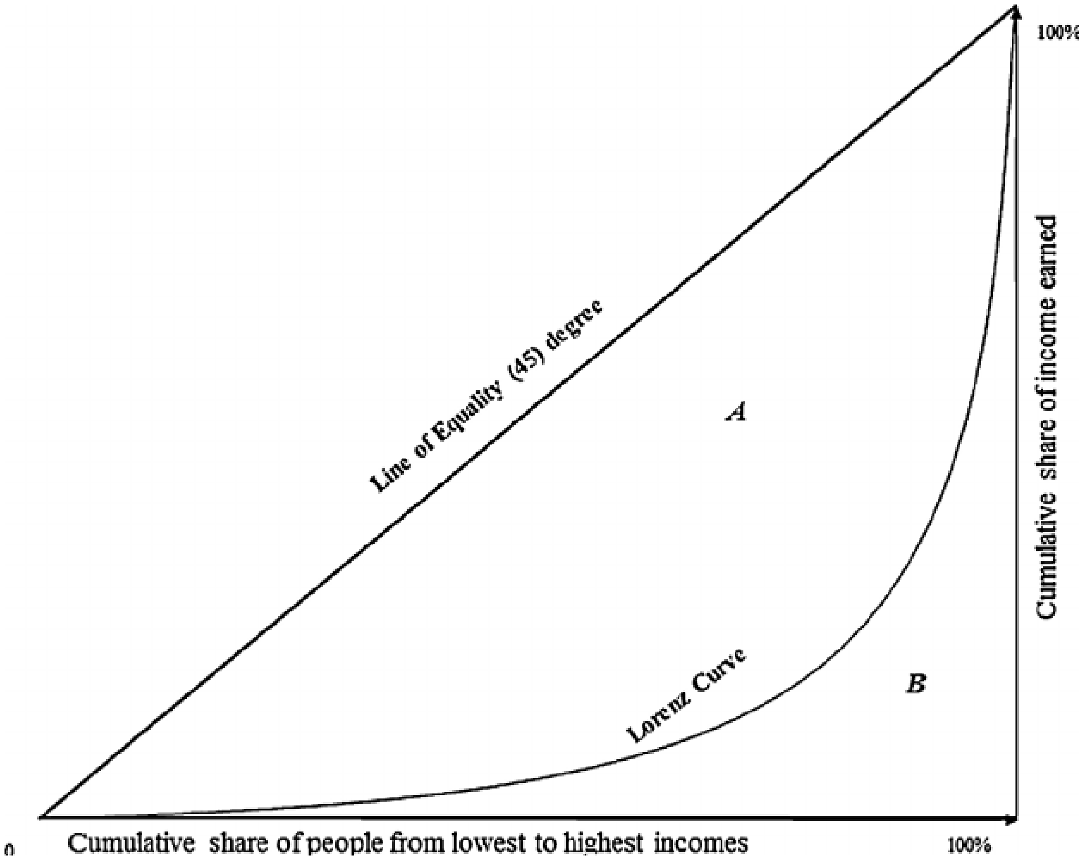

FIGURE 2.1: Gini coefficient (https://www.researchgate.net)

Let's say that all skilled individuals have the same wage. λ is their share of the overall population and ν is the cumulative share earned by them. Unskilled individuals have also the same wage, but lower than skilled workers one. With these assumptions, we can derive an equation for the Gini index. Graphically, we are calculating the area above the Lorenz curve and under the 45° curve[14]:

$$\begin{aligned} A_i &\approx 0.5 - \lambda \cdot (1-\nu) - (\nu \cdot \lambda)/2 - ((1-\nu)\cdot(1-\lambda))/2 \\ &= 0.5 - \lambda + \lambda \cdot \nu - (\lambda \cdot \nu)/2 - 0.5 + \nu/2 + \lambda/2 - (\lambda \cdot \nu)/2 \\ &= \lambda \cdot \nu - (\lambda \cdot \nu)/2 - (\lambda \cdot \nu)/2 - \lambda + \nu/2 + \lambda/2 \\ &= \nu/2 + \lambda/2 - \lambda \end{aligned} \tag{2.6}$$

We thus have:

$$G_i = 2A_i \approx \nu + \lambda - 2\lambda = \nu - \lambda \tag{2.7}$$

Two important points before engaging the estimation of the variables in our investigation. First of all, we obtain that the dynamics behind the income distribution

[14]For a survey of the diverse ways to write the Gini index, see Dorfman (1979) and Yitzhaki (1998).

are summarized only in two pieces of information: the cumulative share of income earned by the skilled group, ν, and their share of the overall population, λ. Lambert and Aronson (1993) find that the dynamics of the Gini index can be decomposed in three parts: a *within* variation, a *between* variation and a *residual* component that arises when two groups overlap. The last term means that there are individuals belonging to a poor group, that are richer than the poorest individuals belonging to a rich group. In our case, if an unskilled worker earns more than a skilled worker, the latter will have no incentive to become educated. This explains why we have only two terms in our equation 2.7 and why the residual component does not appear. It does not mean that this element will not affect the Gini index. Lambert and Aronson (1993) point out that unexpected results occur due to this component.

We assume that these two pieces of information can be characterized as vertical and horizontal. The vertical force is a change in price and corresponds to the *within* inequality variation. Equation 2.7 states that when the share of income of the richest group increases, the inequality will increase if skills mobility is given. Secondly, imagine that for some reason, we can move one individual from the first group, change its characteristics, and insert him in the second group. He will also earn the same income of this group. Inequalities will decrease if the income share is given. This movement will be called an horizontal change, which corresponds to the *between* variation. These two hypothesis allow to link the dynamics of the Gini index to the mobility of workers and to the wage differentials that are connected to the MP.

Finally, we obtain the perfect equality when $\lambda \rightarrow 0$ and $\lambda \rightarrow 1$. Consequently, for values of λ]0;1[, there exists at least one value of λ, where the Gini index reaches a maximum. An increase of λ have thus, at least, two effects on the evolution of the Gini index. The question is at which share the index will start decreasing. As shown in our appendix A, the maximum will be closer to λ=0 if the wage ratio between the skilled and unskilled workers is high (to see this, plot the combination of G_i and λ for which ν is constant, y-axis represented by G_i and x-axis by λ). In other words, with high wage disparities and starting at λ=0, the Gini index increases quickly to high values as λ increases and then decreases slowly to zero when $\lambda \rightarrow 1$.

The logical corollary is that countries with high (low) wage disparities between groups, an increase in the Market Potential will more *probably* decrease (increase) the inequalities. Indeed, there is a higher likelihood to be on the right (left)-hand side of the maximum, where the Gini index is decreasing (increasing) in λ.

This chapter presented the structure of the paper. We assume that Gini index dynamics are characterized by two forces that impact each other. They are linked through the assumption of the cost-benefit choice of each unskilled individual. Finally, the MP is able to influence both forces. This framework permits to make three hypothesis that we attempt to test in the following pages.

First, the MP has a non-linear relationship with the income distribution. The impact will be positive or negative depending on the interactions between the wage ratio and the share of each group in total population.

Second, an increase in MP increases the share of the richest group due to the vertical change.

Third, an increase in MP will decrease the Gini index through the horizontal dimension.

The next section introduces the multiple database that we use and the strategy that we adopt to estimate our variables and to test these hypothesis.

Chapter 3

Data and estimation strategy

In this section, we bring forward how we estimate the gravity equation and the wage equation. Estimating the first one helps us finding an approximation of the Market Potential and the second one determine if we construct it correctly. Indeed, if we find a positive relationship between MP and wages and that the coefficient obtained is not too far from what has been found in the literature, we can be confident on our approximation. Finally, we expose how we estimate the impact of the Market Potential on inequalities and explain which strategy we adopt to test the three previous hypothesis.

We use the UN Comtrade database on bilateral trade to construct our dependent variable in the gravity equation 2.4. This database gathers pieces of information on traded goods at the 4-digit level, but we only used the *each year total export value* for each country exporter-importer commercial relation as the dependent variable.

Then, capturing the trade impediments ϕ_{ij} is done using multiple and different covariates. For this purpose, we use the CEPII database available here[1]. Bosker and Garretsen (2012) affirm that there is no specific way to model these impediments in NEG model. They mention that the common way proceeds with the multiplicative form where covariates are in absolute term. As we use the log, they will be linearly related. Following this common estimation procedure, we include variables as the simple bilateral distance, D_{ij}, between the exporter and the importer; the border effect, B_{ij}; common language effect, CL_{ij}; two variables for colonial links: the first one captures if two countries where ever in a colonial relationship, CR_{ij} and the second one informs if two countries had a common colonizer after 1945, CO_{ij}. Finally, we include also data on RTA, RTA_{ij}, which captures if two countries are in a Regional Trade Agreement as established by the WTO in 2015[2]. We consider these covariates because they were often present in other papers. These variables are available for 224 countries. Log linearized, the trade impediments are estimated as:

$$\begin{aligned} ln(\phi_{ij}) = (1-\sigma)\cdot ln(T_{ij}) \approx \delta_1 \cdot ln(D_{ij}) + \delta_2 \cdot ln(B_{ij}) + \delta_3 \cdot ln(CL_{ij}) \\ + \delta_4 \cdot ln(CR_{ij}) + \delta_5 \cdot ln(CO_{ij}) + \delta_6 \cdot ln(RTA_{ij}) \end{aligned} \tag{3.1}$$

Exporter capabilities and importer's real demands are estimated using fixed effect for the exporter, FE_i and the importer, FE_j respectively. The coefficients attached to these fixed effects capture the effect of the production in country i, FE_i and the demand of the country j, FE_j. These country fixed effects are independent

[1] http://www.cepii.fr/CEPII/fr/bdd_modele/bdd_modele.asp. It is a French research center in international economics establish in 1978.

[2] https://www.wto.org/english/tratop_e/region_e/region_e.htm

from the country with which they trade and they control for unobserved characteristics that might "shift" the bilateral commercial relation[3]. The gravity equation 2.4 becomes:

$$\begin{aligned} ln(\widehat{X_{ij}}) = \widehat{FE_i} + \widehat{FE_j} + \hat{\delta}_1 ln(D_{ij}) + \hat{\delta}_2 ln(B_{ij}) + \hat{\delta}_3 ln(CL_{ij}) + \hat{\delta}_4 ln(CR_{ij}) \\ + \hat{\delta}_5 ln(CO_{ij}) + \hat{\delta}_6 ln(RTA_{ij}) + \epsilon_{ij} \end{aligned} \tag{3.2}$$

This equation is estimated for each year from 2000 to 2014. Two years (2004 and 2012) are then chosen to construct two MP, as it is enough to give a first insight of the MP implications on inequalities. In 2004, we obtain estimates of importer fixed effect for 213 countries and in 2012, for 212 countries. Thanks to the gravity equation, we obtain estimates for market capacities and trade impediments that we need to construct the MP as showed in equation 2.5:

$$\begin{aligned} \widehat{MP_i} &= \sum_j exp(\widehat{FE_j}) \cdot \hat{\phi}_{ij} = \sum_j exp(\widehat{FE_j}) \cdot exp[\hat{\delta}_1 ln(D_{ij}) + \hat{\delta}_2 ln(B_{ij}) \\ &\quad + \hat{\delta}_3 ln(CL_{ij}) + \hat{\delta}_4 ln(CO_{ij}) + \hat{\delta}_5 ln(CR_{ij}) + \hat{\delta}_6 ln(RTA_{ij})] \\ &= \sum_j exp(\widehat{FE_j}) \cdot D_{ij}^{\hat{\delta}_1} \cdot exp[\hat{\delta}_2 ln(B_{ij}) + \hat{\delta}_3 ln(CL_{ij}) \\ &\quad + \hat{\delta}_4 ln(CO_{ij}) + \hat{\delta}_5 ln(CR_{ij}) + \hat{\delta}_6 ln(RTA_{ij})] \end{aligned} \tag{3.3}$$

We obtained estimates for 128 countries in 2004 and 122 countries in 2012[4].

For the wage equation 2.2, Head and Mayer (2011) warned that the MP might be endogenous since it also includes the MP of the exporting country. Indeed, wages of the exporter are on the left-hand side of the equation and might cause some issues[5]. Moreover, we aim at modeling the influence of *external* elements on income distribution. As the MP of the exporter is an internal component, we must remove it. Therefore, we split the MP between a Foreign and a Domestic Market Potential and use only the Foreign component. However, we have bilateral trade data and our MP is already a Foreign Market Potential (hereafter FMP). Indeed, we do not possess data on a country's trade with itself. To obtain the overall MP, one should estimate the Domestic Market Potential. Since we do not consider the country itself in explaining its income distribution, we will not bother computing it. One should keep in mind that using only the FMP (now we use FMP and MP as interchangeable) is not in line with the theory, as the domestic size of the market also counts for the determination of local factors. The estimated FMP is:

$$\widehat{FMP_i} = \sum_{j \neq i} \hat{\phi}_{ij} \cdot exp(\widehat{FE_j}) \tag{3.4}$$

We do not report estimates of our FMP for 2004 and 2012, but we first compare them with those that are suggested by the CEPII here[6]. MP and FMP (from Redding and Venables (2004) and Head and Mayer (2004a)) are available from 1980 to 2003

[3] Head and Mayer (2014), p.152.

[4] More details on which countries were included and for each country with compute the MP can be transmitted upon request.

[5] These wages are often approximated by GDP per capita (Head and Mayer (2004b)). We would regress GDP per capita on income this way.

[6] http://www.cepii.fr/CEPII/fr/bdd_modele/presentation.asp?id=9

for 196 countries. We also compute the FMP between 2000 and 2003 (2000 and 2003 are included), but we obtained very inflated results for 2002 and slightly lower for 2001. This is why, computing the correlation between Head and Mayer or Redding and Venables coefficients and ours estimation, we get disappointing results: 0.36. However, removing first 2002 estimation and then 2001 increased the correlation to 0.66 and then 0.81. The following figures and the wage equation estimation are here also to provide evidence that our estimation of 2004 and 2012 of the FMP are correct.

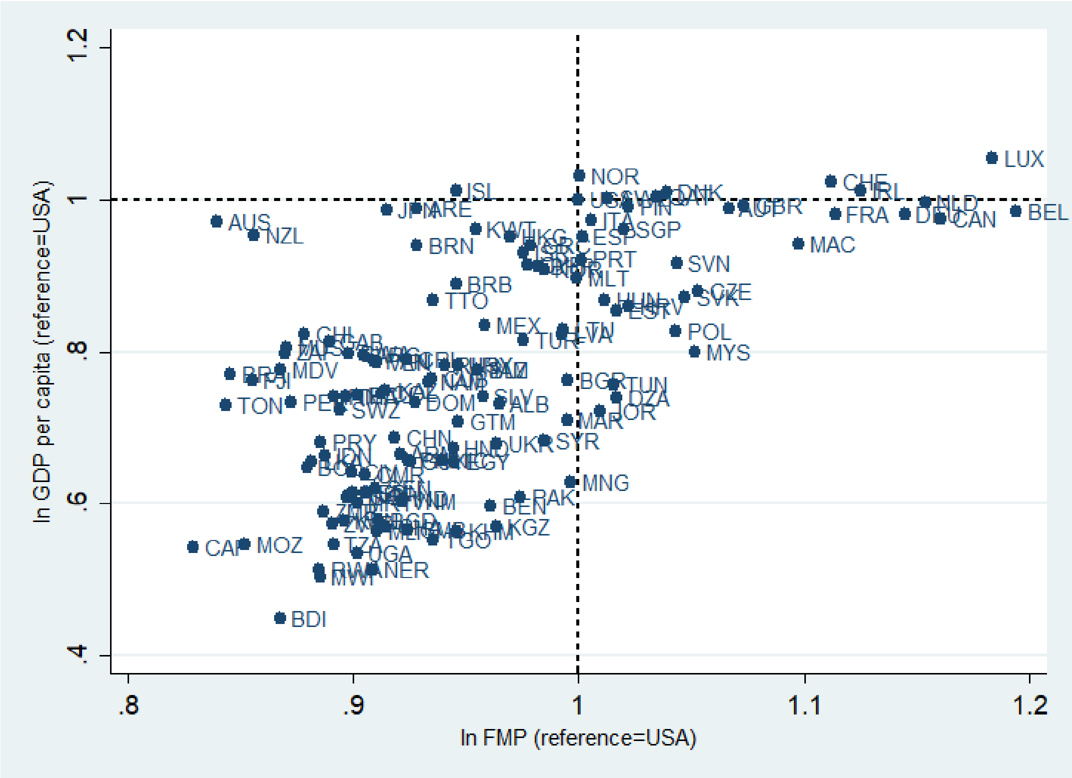

FIGURE 3.1: Foreign Market Potential and GDP per capita 2004

Secondly, we plot the natural log of GDP per capita determined by the natural log of the FMP for 2004 in figure 3.1 and 2012 in 3.2. As mentioned in the graphs, we evaluate the relationship taking the USA as a reference, meaning that all countries that are on the right of the vertical dotted line and above the horizontal dotted line (i.e.: LUX), have a FMP and a GDP per capita higher than USA. We clearly obtain a positive relation between the two variables for 2004 and 2012, with a correlation that should be smaller than 1, as all the observations do not lie in the bottom left and/or upper right rectangles. We notice that the countries seemed to concentrate around the same FMP through time producing a less clear relation in 2012. On both graphs, Belgium is the country with the highest FMP in 2004 and 2012 and the Central African Republic and Brazil with the smallest FMP in 2004 and 2012 respectively.

The World Bank database is used to obtain information on GDP and Population to proxy wages on equation 2.2. It gathers information on 248 countries. This will give an idea of the impact of MP on overall wages. Head and Mayer (2011) propose that productivity might influence the results of this equation. Moreover, Fally, Paillacar, and Terra (2010) control for individual characteristics to estimate the effect of MP on wages for Brazil, as they stated that individual characteristics are important determinants in explaining wage differentials. In our case, controlling for individual characteristics, not for one country has they do but for dozens of countries, might create issues of compatibility. However, differences in productivity and human capital can alter our coefficients both in the wage equation and when we explain the income distribution later. Therefore, we include information on these potential omitted variables and proxy them with average years of education, provided by Barro and Lee[7].

Their database gathers observations from 1950 to 2010 for 146 countries each five year. As we use estimates of FMP for the year 2004 and 2012, we thus select the years 2000, 2005 and 2010 and construct an approximation of 2004 and 2012 average years of education. This approximation is computed using a mean between 2000 and 2005 to proxy 2004 average years of education. We use 2010 to proxy 2012. The values of average years of education range from 1.1 in Mozambique in 2004 to 13.42 years in USA and Switzerland in 2012. We also did not include the supplier capacities in the regression, since there is a risk that Market Potential and the latter are correlated. To sum up, the wage equation is estimated using GDP per capita for the dependent variable and FMP and average years of education for the independent variables.

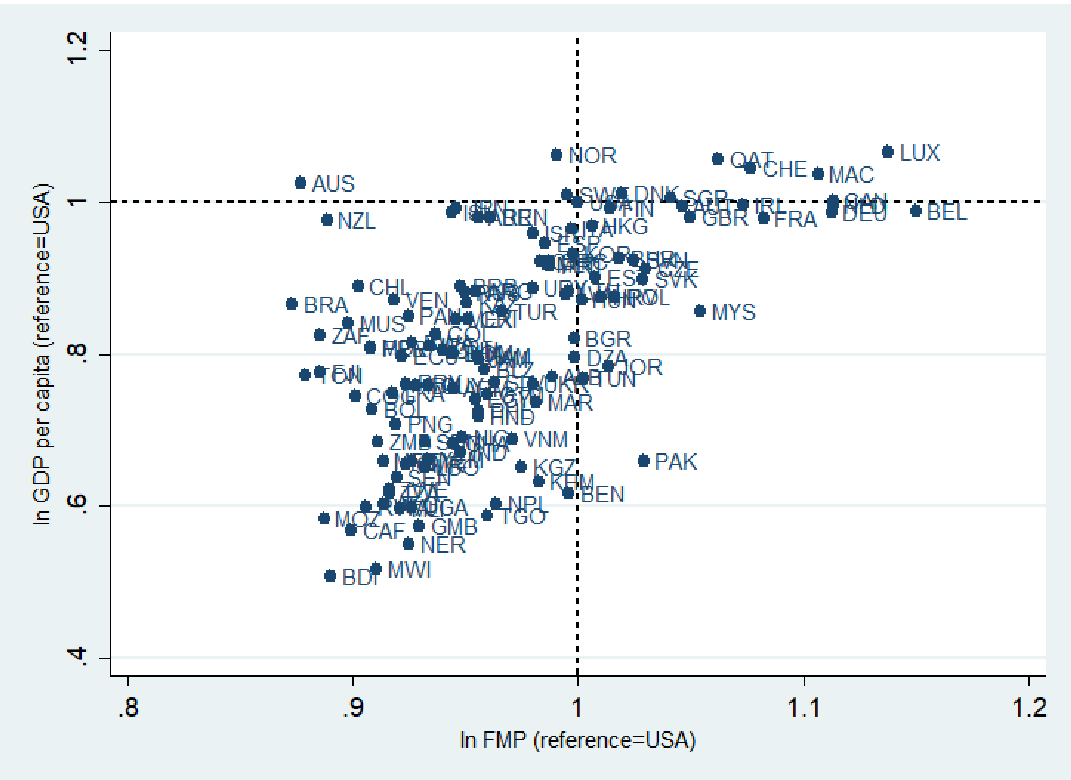

FIGURE 3.2: Foreign Market Potential and GDP per capita 2012

To obtain observations of the inequality in income for each country, we used in the first place the World Bank (thereafter WB) database providing the Gini index on their website[8]. This index contains information on 264 countries from 1960 to 2014, but most of the dataset carries missing values (we have observations from 1981 to 2014 for 155 countries). In addition, the number of observations for each year remained low and we were concerned about it. Indeed, most of them where concentrated between the years 2004 to 2012, with a mean observation for each year

[7] http://barrolee.com/
[8] http://data.worldbank.org/indicator/SI.POV.GINI

of around 70. This explains why we use 2004 and 2012 years to construct our MP. To increase this number, we used the World Income Inequality Database (WIID)[9]. It collects and pools information on income inequality and contains variables, including the Gini index, for 182 countries.

Estimating the impact of MP on income distribution is not an easy task. In addition, linking only the Gini index and the MP will give only a linear direction of the effect. We decided to adopt three strategies to decompose this impact.

First, we replace the Gini index by another measure of inequality, the share of income of the richest 40%. This share corresponds to ν in equation 2.7. This will consolidate the results obtain with the Gini index and helps us to disentangle the two effects that we supposed to be behind the dynamics of the Gini index. It will be difficult to separate them, but this paper will attempt to resolve this issue.

The division in two groups, 40% skilled and 60% unskilled, comes from a choice between three candidates, where we compare them in their ability to produce the real value of the World Bank Gini index. The purpose remains at finding a value for ν that produce the best approximation of the Gini index. We used 10%, 20% and 40% share as the candidates. Each of them gave an estimation of the index using the formula in equation 2.7[10]. The candidate that produced the lowest RMSE was 40%.

Second, the relationship between the MP and the Gini index might not be linear. Pyatt (1976) finds that the index had a non-linear behavior when different years of education are considered. He points out also the need to associate Gini index' dynamics with migration. Hence, we include in the first place the square of the MP to catch the non-linear pattern of this relationship and restrict our sample to advanced countries to identify if more egalitarian countries observe an increase in inequality when the MP increase (as we stated at the end of section 2.3).

Finally, we attempt in separating the vertical and the horizontal effect induced by a change in the MP. To do so, we attempt in controlling for the mobility of the workers using again the average years of education and also the employment share for the tertiary in 2005, provided by the World Development Report 2013 on Jobs and available here[11]. This way, we hope to remain with the vertical effect, but we stay cautious.

[9] https://www.wider.unu.edu/project/wiid-world-income-inequality-database

[10] We took the difference between the Gini index estimation and the correct one for all countries and all three candidates. The square of the difference is taken, then the mean and finally the root square.

[11] https://data.worldbank.org/data-catalog/wdr-2013-jobs. This piece of information contains data on 103 countries, from 9.2% in Burkina Faso to 90.3% in Kiribati.

Chapter 4

Methodology

In this paper, the standard OLS estimation procedure will be the benchmark to estimate the coefficients we are interested in. For this purpose, we decide to follow what most of the previous authors opted to include as control variable in their work and to reduce the risk of omitted variables. Then, the investigations are portrayed by our three hypothesis.

Firstly, fifteen OLS regressions (one regression per year) are performed to compute the coefficients from the variables constituting the vector of trade impediments. Amongst these regressions, we selected two of them and constructed the MP. We could have also transformed our sample into a panel one to control for all the individual characteristics fixed over time, but this way will have produce only one coefficient for each variable and for all years of our sample. Hence, we would have missed relevant variability in our analysis. As a matter of fact, having two MP constructed for different years allows not only to give a perspective to our results, but it permits to make direct comparisons. It also serves to control for import-year and export-year fixed effect and to generate statistical analysis of the dynamics behind each variable.

Choosing only one year has the risk of considering a year where an event could have occurred. Considering two years diminishes this risk. Indeed, we suspect for example that the years 2009 and 2010 are periods where our variables change considerably. Taking those two years to construct the MP would probably produce links between variables that would have been difficult to generalize. We do not say that studying those periods do not make sense, far from it. We remind that these periods of decrease in trade and growth appear less often in the business cycle. Taking them as a keystone to perform an analysis on income distribution through MP (which is not a theoretical question highly developed in the literature in addition) might be inappropriate. A general picture is more suitable for a paper analyzing primary effects. Therefore, we used OLS estimation and applied it on each year.

As said before, we had concerns about the lack of observations from the World Bank Gini index and decided to increase our sample with WIID database. The sample was first augmented with another method before adding this latter database. We computed missing values for 2004 and 2012 years with observations of the Gini index from other years. Table B.1 provides an example of two countries for which we did not have observations of the Gini index for 2004 and 2012.

This is an example using only WB database, but we also applied this method on WIID database. Column (3) depicts the Gini index as it appears in the WB database. The ISO3 country *ALB* has an observation for 2002 and 2005. To construct the 2004 observation, we put a weight of 2/3 for the observation that is the closest to 2004 and a weight of 1/3 for the second one. In this example, we have thus: 30.6*(2/3)+31.74*(1/3)=30.98. It can be applied when two observations, one for 2003 and the second for 2006 are reported. In this case, the 2/3 weight will be attached to the 2003 observation and 1/3 for 2006. For the second country *PAK*, where on observation follows and the other one precedes the year 2012, the same weight is applied to both observations. For this country, we have thus: 30.93*(1/2)+30.69(1/2)= 30.81.

Applying this method on observations that allow to do it (as in the table B.1) and including the WIID database permits to increase the sample from 66 to 78 observations for 2004 and from 61 to 78 observations for 2012, which is not negligible. In the results, the regressions are separated between the WB raw database and the augmented database with also WIID observations.

Table B.2 describes the way we select our WIID database with the criterion use in column (1). The arguments are in the Notes below the Table.

Chapter 5

Results

We introduce the results into three sections. The section 5.1 presents the estimation of the gravity equation, where the coefficients of the trade impediments are analyzed in detail. In section 5.2, results of the relationship between the Market Potential and wages are interpreted. Finally, section 5.3 submits our three hypothesis to econometric analysis.

5.1 Gravity equation

Gravity results are presented in table 5.1 and refer to the equation 2.4 estimated by OLS with equation 3.2. Column (1) presents the results for the 15 regressions that we perform each year between 2000 and 2014 using the log of the total export value as the dependent variable and fixed effects for importer and exporter plus trade impediments as independent variables. Column (1) reports only the mean coefficient of those 15 regressions. The results are slightly higher than those reported by Head and Mayer (2011). We obtain a coefficient of -1.6 and they find a mean distance coefficient of around -1. This increase in the effect of distance on trade is not new. Bosker and Garretsen (2012) get a coefficient of -1.96 for the 1993-2009 period, but they consider only exports of manufactured goods. Our coefficients regarding contiguity, common language, FTA and colony are also lower than what they detected.

We obtain a mean effect of FTA close to 0.5, which is higher than the one presented in Head and Mayer (2011). This statement also prevails for the common language coefficient of 0.7, that is higher than 0.4 found by the co-authors. Finally, an R-squared that is extremely high (close to unity) is reported. The columns (2) and (3) write down the regression for the years 2004 and 2012 that we did in the first column using the same variables. We employ these regressions to construct the MP for the two respective years. The coefficient of common colonizer for these two years were particularly high compared to the other years. To sum up, we obtain coefficients that are more or less confirmed by other papers and there is no unexpected result regarding each trade impediment coefficient.

In figure 5.1, we plot the coefficient of the variable's distance for each regression that we perform between 2000 and 2014. The y-axis stands for the natural log of the distance's coefficient and is represented with negative numbers. On the x-axis, we have the time. Two valuable pieces of information can be derived from this figure.

First, the change over time of the coefficient is significant when considering a confidence interval of 5%, meaning that the distance's evolution is a crucial element in trade between countries. Moreover, its variation across time matters, suggesting that the determination of the year to construct the MP is important. Different coefficients for the distance will appear through years. 2004 and 2012 seemed to be the average of the distance coefficient before 2009 and after 2009 respectively.

TABLE 5.1: Gravity equations

dep: ln export value	**(1)**	**(2)**	**(3)**
ln distance	-1.63***	-1.583*** (0.026)	-1.674*** (0.027)
contiguity	0.63***	0.704*** (0.110)	0.647*** (0.124)
com. lang.	0.71***	0.684*** (0.052)	0.745*** (0.056)
colony	1.15***	1.194*** (0.106)	1.052*** (0.111)
common colo	0.77***	0.882*** (0.068)	0.832*** (0.075)
FTA	0.49***	0.539*** (0.053)	0.438*** (0.050)
Time	Mean 2000-2014	2004	2012
Exporter Dummies	YES	YES	YES
Importer Dummies	YES	YES	YES
N (each year)	19474	19433	19579
R-Squared	0.983	0.983	0.982

Notes: Robust standard errors are in parenthesis. The coefficients in column (2) and (3) are used to compute the Market Potential for 2004 and 2012. Coefficients for each importer country that were not omitted were employed to construct the MP. *, **, *** dote the significance at 10, 5 and 1% respectively. The same method in Bosker and Garretsen (2012) was used to reported significance of the coefficients. *,** and *** stand for the significance at 1% level for the observations that are significant for at least 50%, 80% and 100% of the time respectively.

Another interesting result is the reversal of the path of the coefficient observed since 2010. We suspect the trade collapse of 2009 to be the underlying instigator of this change. Even if the reversal occurred one year later, the collapse produced a change in commercial relations between countries. Since then, they might bring forward other elements that where less important in the past (we observed a temporary increase in coefficients related to colonial links for example[1]).

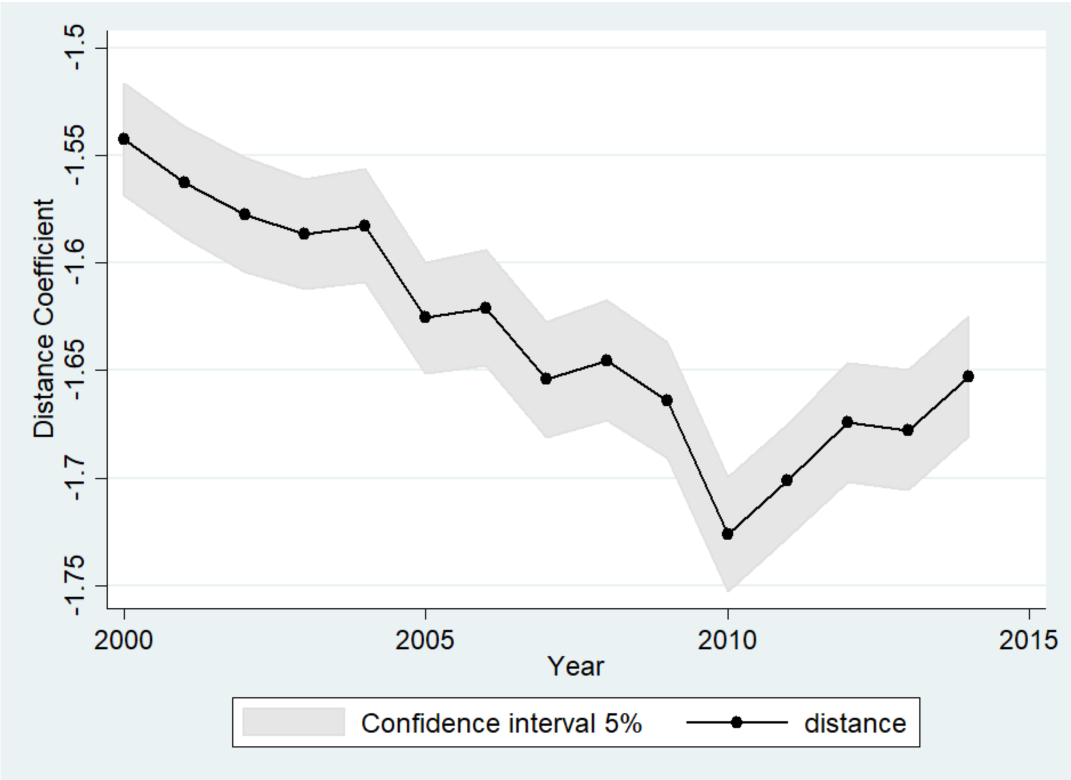

FIGURE 5.1: Distance effect from 2000 to 2014

Effects of Regional Trade Agreements and border effect through time are demonstrated in figure 5.2. Again, we observe that 2010 is a year of effect reversal. Border effects are increasing since then and show that exporting to closer partners is gaining importance. Combining this interpretation with the distance coefficient analysis, we obtain that the decrease in the distance effect might be related to commercial relations other than those with closed partners. Indeed, if the impact of the distance decreased in importance after 2010 and the border effect increased at the same time, the decrease for the former comes from countries that do not have a common border with the exporter. This result might derive from the fact that countries trade different goods depending on the distance with its partners. Moreover, the effect of RTA's coefficient shows that the trade collapse occurred also between members of the same Trade Agreement. The coefficient decreases meaning that the effect of being or not in a Trade Agreement do not have such powerful effect on exports than before.

[1]The coefficient of colonial relationship increased in 2013 and the common colony coefficient increased also substantially in 2012. These results are not reported here.

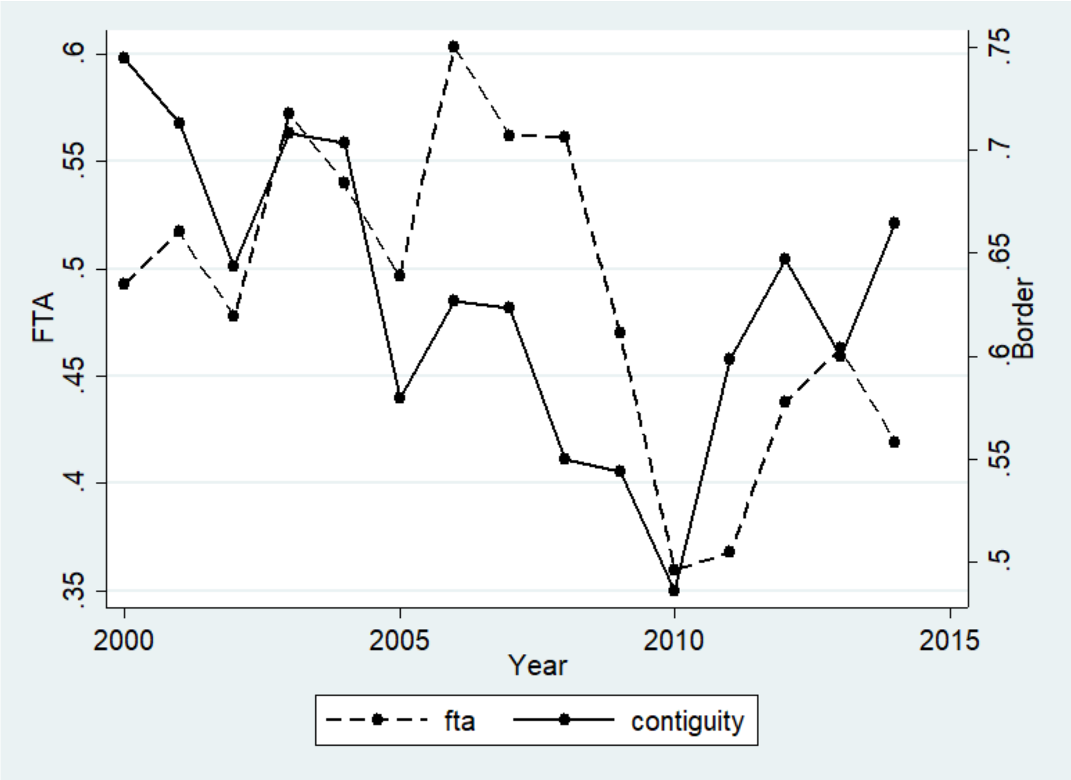

FIGURE 5.2: RTA and border effect from 2000 to 2014

5.2 Wage equation

To be confident on the construction of the MP, we estimate the log of the wage equation 2.2. This equation links the returns to the inputs factors (approximated thanks to GDP per capita) with its MP (see equation 2.5). We included only the Foreign Market Potential since our research studies the effect of other countries on the income distribution of the exporter country. As we mentioned, including its own market is also interesting, but raises a problem of endogeneity. The results of the four OLS regression are presented in Table 5.2. Here, not the same years are used in the regressions as Head and Mayer (they chose 1995). They also pooled the observations and included country fixed effect.

Columns (1) and (2) regress the log of GDP per capita on the log of the FMP in 2004 and the proxy of productivity (average years of schooling). The reason behind the inclusion of such variable comes from the fear that our estimate is biased due to omitted variables. The same regressions are presented for 2012 in columns (3) and (4). The coefficients obtained are smaller than those found by Head and Mayer (0.88, in their column 4). Their higher sample (180 observations) might be the explanation. The R-squared is however higher in this paper. It highlights that the impact of the FMP is smaller through time but explains more variance of the GDP per capita. This assertion is confirmed when we include the proxy of productivity in column (2) and (4).

In column (2), we obtain a coefficient of 0.44, which is confirmed by Head and Mayer (0.42). This coefficient corresponds to the inverse of the elasticity of substitution. A 1% increase in the FMP correspond to an increase in the GDP per capita of 0.44%. We have lower standard errors, since we use robust ones. Our R-squared is

TABLE 5.2: GDP per capita and Foreign Market potential

dep: ln GDP per capita	**(1)**	**(2)**	**(3)**	**(4)**
ln FMP	0.814*** (0.078)	0.444*** (0.073)	0.739*** (0.086)	0.314*** (0.079)
av. years schooling		0.310*** (0.033)		0.331*** (0.029)
constant	-4.651*** (1.283)	-1.113 (1.040)	-6.135*** (1.785)	-0.261 (1.446)
Time	2004	2004	2012	2012
N	128	128	122	122
R-Squared	0.388	0.638	0.349	0.699

Notes: Robust standard errors are in parenthesis. *, **, *** dote the significance at 10, 5 and 1% respectively.

smaller. Column (4) shows a better fit, even if the number of observations decreased. This increase in explaining the dependent variable does not come from the RMP, but the average years of schooling. Indeed, the coefficient of FMP remains significant for 2012, but it decreased more than a quarter and show more variability in its estimation. On the contrary, the positive effect linked to the proxy increases between the two periods. There is thus a sign that FMP decreases in importance in explaining wage disparities *between* countries.

5.3 Gini index and share of 40% richest individuals' income

Figure 5.3 and 5.4 plot the log of FMP on the WB Gini index. This first insight shows a negative relationship between the two with a slight weakening when higher values of FMP are considered. Moreover, the relation seems to last through time. The second insight refers to Fally, Paillacar, and Terra (2010) paper. They surprisingly find that a higher Market Potential in Brazil regions causes higher positive effect for unskilled workers than skilled workers. In figure 5.3 and 5.4, Brazil is described by the acronym *BRA* (top-left of the figures). Since Brazil is a country with substantial inequalities, we might expect its κ (ratio between skilled and unskilled wages[2]) to be probably high. This translates into a higher likelihood to observe a decrease in the Gini index after an increase in the share of skilled workers (which can be the effect of an increase in MP)[3]. In our framework, this can be explained by a mobility effect that is higher than the price effect. We do not know if it is the case, but the Gini index dynamics provide these explanations that are not totally pointless.

The following regressions in table 5.3 show the log of the WB Gini index as the dependent variable explained by the log of FMP and/or FMP squared. This table is divided in two parts to distinguish the case where the WB Gini index raw data is used and the second where we construct manually more observations. The robust standard errors are not significantly different between the two samples. The method we use is thus convincing. Moreover, the coefficients on the log of FMP and FMP

[2]See appendix A.

[3]The decrease in inequalities between 2004 and 2012 observed is supported by De Oliveira Cruz and Naticchioni (2012) findings.

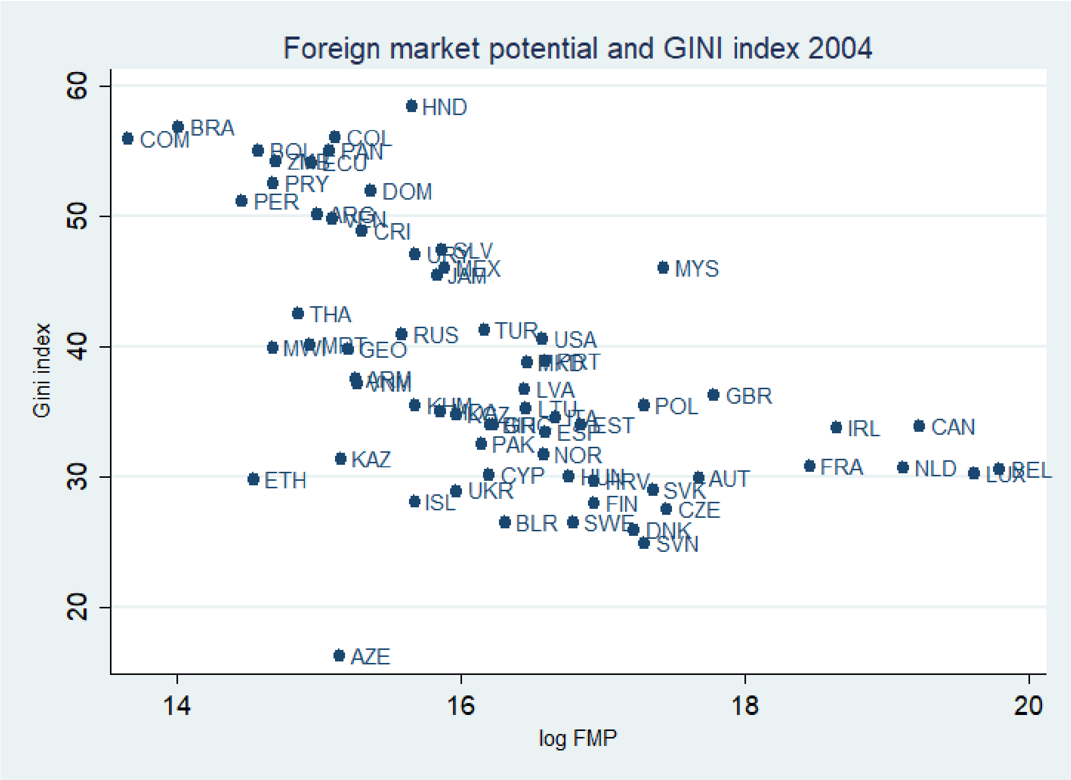

FIGURE 5.3: Foreign Market Potential and Gini index 2004

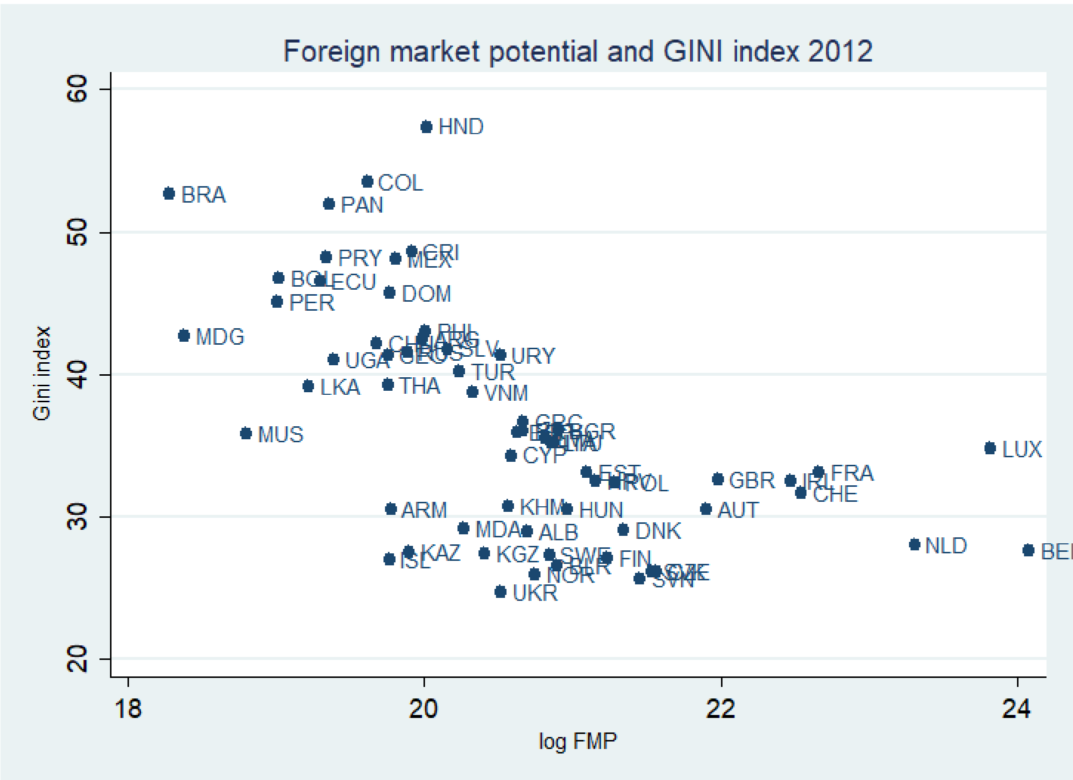

FIGURE 5.4: Foreign Market Potential and Gini index 2012

TABLE 5.3: WB Gini index and FMP

	Initial sample				Augmented sample			
dep: ln Gini ind.	**(1)**	**(2)**	**(3)**	**(4)**	**(5)**	**(6)**	**(7)**	**(8)**
ln FMP	-0.102*** (0.019)	-1.054*** (0.228)	-0.108*** (0.019)	-1.339*** (0.363)	-0.106*** (0.019)	-1.120*** (0.221)	-0.109*** (0.018)	-1.245*** (0.352)
ln2 FMP		0.029*** (0.007)		0.029*** (0.009)		0.030*** (0.006)		0.027*** (0.008)
constant	5.270*** (0.321)	13.153*** (1.969)	5.784*** (0.394)	18.685*** (3.842)	5.329*** (0.309)	13.719*** (1.899)	5.820*** (0.380)	17.697*** (3.708)
Time	2004	2004	2012	2012	2004	2004	2012	2012
N	66	66	61	61	71	71	66	66
R-Squared	0.275	0.343	0.343	0.423	0.291	0.367	0.339	0.404

Notes: Robust standard errors are in parenthesis. *, **, *** dote the significance at 10, 5 and 1% respectively.

squared for the two parts are close to each other. The R-squared increases through time and we observe that the constant is positive for the entire table.

Coefficient results are interesting. In all regressions, we obtain that an increase in FMP decreases the Gini index. This result is significant at 1%. With equation 2.7, we would be tempted to state again that the horizontal effect exceeds the vertical effect. Moreover, the coefficients of the log of FMP are 10 times larger when including log of FMP squared. The latter is also significant at 1% in all our regressions and the coefficients are positive. Therefore, the relationship between the two is non-linear and confirms the graphical representation in figures 5.3 and 5.4. This means that countries with high Foreign Market Potential might observe at least a slowdown of their decrease in inequalities or even an increase in inequalities if the FMP is high enough. Finally, the coefficients of FMP and FMP squared between 2004 and 2012 are not very different and suggests that the relationship is stable.

Finally, we took column (2) and (4) of table 5.3 and column (1) and (3) in table 5.2 to compute at which level of GDP per capita the relationship between FMP and Gini index change from negative to positive. For 2004, we obtain 25'370 $ per capita which correspond to New Zealand at that time. 26 out of 128 countries have a higher GDP per capita and most of them are advanced economies. For 2012, we obtain 55'587 $ per capita (Singapore) and only 8 out of 122 countries are above this level. The decrease in inequalities through the FMP does not seemed to apply to advanced economies.

The following table 5.4 shows interesting results. It is again divided in two parts to distinguish between the raw and the augmented sample. The share of income of the richest 40% is regressed on the log of the FMP and we obtain a coefficient that is negative and significant in all regressions. Previous results are thus confirmed. We also regressed the share of income on the log of FMP and the FMP squared[4] and obtained also a U-shape curve between FMP and inequalities. Moreover, we perform the same regressions with the 10% share[5]. The coefficients obtained had the

[4]Not reported.

[5]We only consider log of FMP this time. Results are not reported.

TABLE 5.4: Share of highest 40% and RMP

	Initial sample		Augmented sample	
dep: share highest 40%	**(1)**	**(2)**	**(3)**	**(4)**
ln FMP	-2.951*** (0.510)	-2.880*** (0.506)	-3.040*** (0.497)	-2.914*** (0.486)
constant	114.808*** (8.521)	124.835*** (10.485)	116.204*** (8.234)	125.678*** (10.085)
Time	2004	2012	2004	2012
N	66	61	71	66
R-Squared	0.321	0.358	0.335	0.351

Notes: Robust standard errors are in parenthesis. *, **, *** dote the significance at 10, 5 and 1% respectively.

same sign and were significant at 1%. Finally, we observe that the relationship is robust to the inclusion of more observations.

We might expect the coefficient of the log of FMP to be also positive. Indeed, the share of the 40% is not only a measure of inequalities, it captures the ν in our framework. We stated before that, as FMP increases, ν might increase because the wage of the skilled group increase. The regressions above do not show a positive effect unfortunately. However, they do not include any control variable. Therefore, we suppose that other effects might drive our dependent variable. The influence on wages of the increase in supply of skilled worker or the residual component that we assume to be null might be the instigators. These two effects need to surpass the positive effect due to the increase in wages of skilled workers and the mobility toward skilled characteristics. A second reason might come from a measurement error. The 40% share was the best candidate between the 10% and the 20%. However, its estimation of the Gini index was disappointing. The 40% share was in fact the "least worst" candidate. A third reason might be that we do not consider migration *across* countries. As our sample contains in majority developed countries, brain drain effects might be at play. As skilled workers observe their wage increasing, they might migrate in other countries to earn more.

Table 5.5 shows the same regression as in table 5.3 adding observations from WIID database. The coefficients in all columns are slightly lower than those in the latter table. The non-linear relationship is also less significant as we observe that the coefficient's significance of the log of FMP is now at 5% instead of 1% in column (2) and (4). The coefficient for the log of FMP squared is also less significant. It is now at 10% and 5% for 2004 and 2012 respectively. The R-squared is also lower. This is not due to a specific characteristic of the countries added, as they come from all the continents. We conclude that previous results are weakened by these findings, but we can maintain that the relationship between FMP and income distribution is negative, but non-linear.

Table 5.6 attempts to control for the horizontal dimension effect of the Gini index, i.e. the mobility of workers between groups. Moreover, we add an important temporal friction here to model the imperfect mobility of workers[6]. Indeed, unskilled

[6] Fally, Paillacar, and Terra (2010) recommend to consider labor market frictions.

TABLE 5.5: Gini index and FMP (2)

dep: ln Gini index	**(1)**	**(2)**	**(3)**	**(4)**
ln FMP	-0.096*** (0.017)	-0.680** (0.311)	-0.108*** (0.017)	-0.943** (0.397)
ln2 FMP		0.018* (0.009)		0.020** (0.009)
constant	5.156*** (0.286)	9.983*** (2.661)	5.794*** (0.356)	14.487*** (4.212)
Time	2004	2004	2012	2012
N	78	78	78	78
R-Squared	0.263	0.290	0.339	0.376

Notes: Robust standard errors are in parenthesis. *, **, *** dote the significance at 10, 5 and 1% respectively. We include observations from the WIID dataset in these regressions.

TABLE 5.6: Gini index and FMP (3)

dep: ln Gini index 2012	**(1)**	**(2)**	**(3)**	**(4)**	**(5)**
ln FMP (2004)	-0.116*** (0.015)	-0.104** (0.134)	-0.116*** (0.020)	-0.047* (0.025)	-0.027 (0.029)
av. years schooling (2005)		-0.016 (0.010)	-0.018 (0.011)		-0.002 (0.006)
employment tertiary (2005)			0.002 (0.002)		-0.010** (0.003)
constant	5.477*** (0.249)	5.429*** (0.248)	5.496*** (0.277)	4.267*** (0.454)	4.624*** (0.395)
Countries	All	All	All	Adv. Economies	Adv. Economies
N	47	47	47	16	16
R-Squared	0.480	0.518	0.532	0.159	0.524

Notes: Robust standard errors are in parenthesis. *, **, *** dote the significance at 10, 5 and 1% respectively.

workers cannot become skilled instantly. They must invest in education and it will take time to show the results. Therefore, we propose that the moment when the individual takes the decision to become skilled or not and the moment he actually becomes skilled or unskilled are two different times. The first comes before the second. Since we have pieces of information on 2004 and 2012, 8 years is the time we use to formalize the temporal friction. This might be too long, so we decided to keep the variable average years of schooling and employment tertiary unchanged. Namely, we do not try to approximate these variables for 2004 using a mean between 2000 and 2005 years observations and we keep the raw data that comes from year 2005.

In column (4) and (5), we restrict the sample to advanced economies[7] to capture the decrease or the reversal of the MP effect on the Gini index due to the higher probability for these countries to be on the left of the local maximum.

Column (1) replicates the third column in table 5.3 with FMP in 2004. We obtain quite the same coefficient for log of FMP, but a much better fit. Either it comes from the decrease in the number of observation or from the method we use to model the temporal friction. Market Potential and inequalities might be a process that produces higher results in the mid-term. In column (2) and (3), we perform the mobility control. Unfortunately, the coefficient of the log of FMP is not lowered and the two variables for skills mobility are not significant. This implies that the decrease in inequalities go through the vertical dimension or there are other elements that influence our regressions.

The two last columns show interesting results. The coefficient of the log of FMP in column (4) demonstrates a slowdown of the decrease in inequalities of the FMP between the two samples. Indeed, the coefficient is significant at 10% in column (4) and no more significant in column (5). This corroborates the non-linear relationship between the FMP and the Gini index. The negative effect of FMP on the Gini index does not operate when we analyse advanced economies. Moreover, the decrease in inequalities through skills mobility is at play in column (5). The horizontal effect seems to produce its effects only for particular countries and might catch a higher mobility of workers, in advanced economies than anywhere else, within the country. This is confirmed by Fallah, Partridge, and Olfert (2011) that find a positive link between the Market Potential and the share of educated workers in USA. Figure 5.5 also supports this analysis, since the relationship between employment in the tertiary sector and the GDP per capita depends on the level of the former. Indeed, advanced economies are mainly located in between 60% and 80% share, where we observe variation in the GDP per capita measure. A change in the share of employment in tertiary sector does not seem to influence the GDP per capita when we consider only *low* levels of the employment share.

[7] The 16 selected countries are: Austria, Belgium, Switzerland, Denmark, Spain, Finland, France, United Kingdom, Greece, Ireland, Italy, Netherlands, Norway, Portugal, Sweden, Turkey.

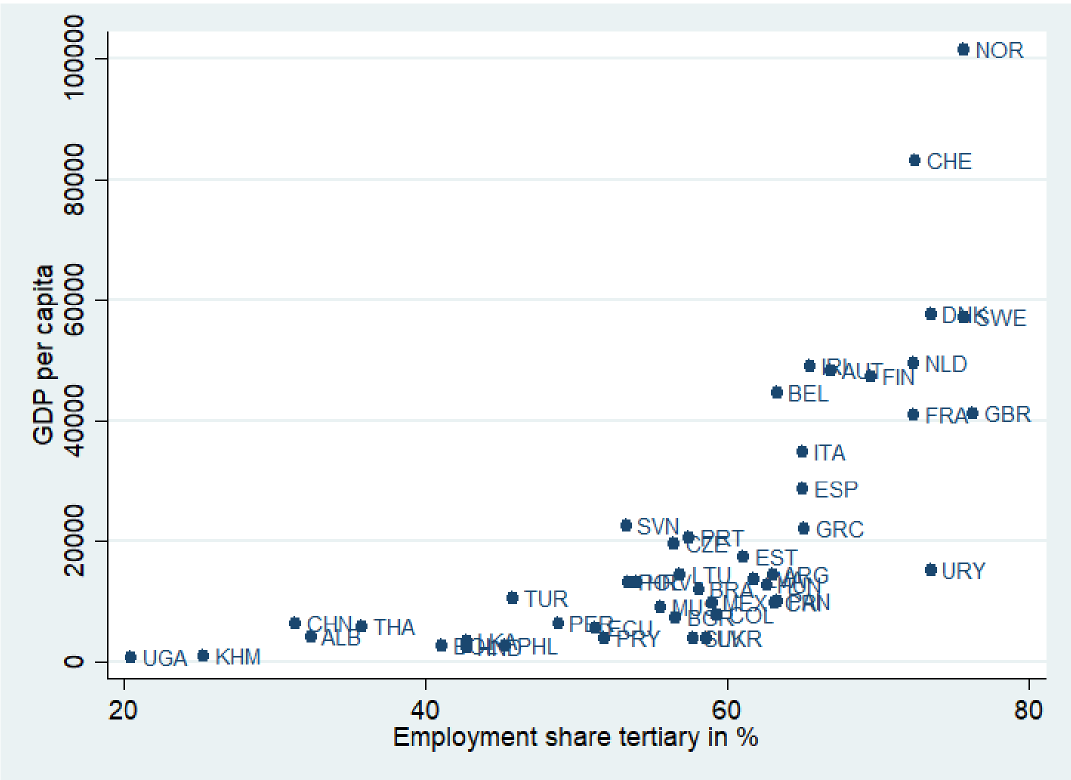

FIGURE 5.5: Employment in tertiary sector in % and GDP per capita in 2005

Chapter 6

Discussion and conclusion

The results show that an increase of the Market Potential has a decreasing effect on inequalities of income even after the sample increase with another database. The relationship between the two is also non-linear thanks to the positive effect of the squared of the FMP and its coefficient's sign on the regression including only advanced economies.

The consequences are that there is a more complex relation between the MP and income distribution. This remains in line with the NEG theory and is compatible with results of other authors. However, we stress the need to include the mobility of workers in inequalities analysis, which leads us to different conclusions in terms of potential impact of the Market Potential on inequalities.

Indeed, we provide another explanation to income distribution dynamics. The impact of the Market Potential on the Gini index is the result of the *combination* of two parameters. The wage ratio between skilled and unskilled workers and the share of the skilled group in total population. Countries with higher between group inequalities have more probability to obtain a decrease of their inequalities after an increase in MP. More equal countries, the opposite.

An increase in wage gap is not necessary a sign of increase in inequalities. Policies aiming at decreasing the inequalities through spatial reallocation of units make sense in more remote, less developed and less competitive countries in this framework. The development of better access to consumers, for example, would generate higher results in countries with high inequalities than more centrally located countries. Here, we only evaluate the MP on income distribution and not on welfare or on the income of specific groups.

The second hypothesis of this paper is questioned. Indeed, we find with two different approaches that the share of the richest group or the vertical change in inequalities induced a decrease of inequalities. Two explanations were provided to this interesting result: measurement error and the complexity behind the Gini index dynamics. It is important to remind that it does not alter the findings of previous authors since they employ *wage gap* instead of the share of overall income as we did. The framework that we provide is not enough to disentangle the dynamics behind income distribution. More investigation on the group composition of the population and the implications of overlapping groups need to be performed.

The third hypothesis is confirmed but only for specific countries : advanced economies. For them, the effect of FMP on their Gini index was lower than for other countries. Moreover, mobility of workers was the only effect that kept the Gini index from advanced in economies not becoming insignificant. Indeed, when we control for it, the relationship between MP and inequalities disappeared. This result implies that mobility of workers within countries should not necessary be a factor exacerbating the disparities. In our framework, it participates to a factor decreasing income inequalities.

Finally, the positive link between the Market Potential and the wages of a country was found at the beginning. Interestingly, we obtain that the link was shrinking between 2004 and 2012. To estimate the Market Potential, we had to estimate the gravity equation. Performing fifteen regressions allow us to obtain interesting dynamics concerning the variables estimating the trade impediments. The distance coefficient was found increasing after 2010. In consequence, we stress the importance in choosing carefully the specific years to compute the MP.

This paper is limited in three general points.

First, we do not integrate migration flows *across* countries. This might impede the results as migration flows are a sizable phenomenon in some countries. This has implications in terms of income distribution if this phenomena concerns only skilled or unskilled workers. An increase in MP might increase the outflow of skilled workers in other countries, which might increase or decrease the inequalities depending on the wage ratio between the skilled and unskilled workers and the initial share of skilled workers in total population. We could also control for population growth that definitely has an effect on inequalities. However, including more and more variables impedes identifying correctly what the FMP is catching with the remaining variation left. If we include population growth, migration magnitude and other variables a negative or positive coefficient of the FMP will be interpreted how? This problem of identification is linked to the following limitation.

Secondly, decomposing the country in only two groups is a weakness. Moreover, we also had to determine between multiple candidates which share of the population give for each skilled and unskilled group. Of course, assuming that all countries have the same share of skilled and unskilled workers is not appropriate. More variability should somehow be implemented at this point. We partially succeed in resolving this weakness by providing an analysis on a specific group of countries (advanced economies), but the results confirmed the need to include country specificities. Nonetheless, this assumption of no difference in endowment across countries allowed us to obtain interesting links between the wage differentials of the two groups, the share of skilled and the Gini index (appendix A) that would have been complicated to represent with more than two groups.

Finally, there is no theoretical NEG foundation in writing the Gini index on the left-hand side of the equations. Theoretically, wage disparities are model thanks to the wage equation, but not income distribution. We took advantage that the Gini index dynamics and the Market Potential implications were linked.

The key message of this paper remains that an increase of the Market Potential does not necessary increase inequalities within the unit of observation. Likewise, each country has its own dynamics in terms of labor market mobility and wage differentials that should be included in inequality analysis. The initial level at which the country stands should also be considered as an important aspect as it determines the direction of the correlation.

Bibliography

Behrens, K., D. Pokrovsky, and E. Zhelobodko (forthcoming). "Market size, occupational self-selection, sorting, and income inequality". In: *Journal of Regional Science.*

Behrens, K. and F. Robert-Nicoud (2014). "Survival of the fittest in cities: Urbanisation and inequality". In: *The Economic Journal,* 124.581, pp. 1371–1400.

Bosker, M. and H. Garretsen (2009). "Economic development and the geography of institutions". In: *Journal of Economic Geography,* 9.3, pp. 295–328.

— (2012). "Economic Geography and Economic Development in Sub-Saharan Africa". In: *The World Bank Economic Review,* 26.3, pp. 443–485.

Charnoz, P., E. Coudin, and M. Gaini (2011). "Decreasing Wage Inequality in France 1976-2004: Another French Exception". In: *Discussion paper, London School of Economics (LSE)-Centre for Economic Performance (CEP).*

De Oliveira Cruz, B. and P. Naticchioni (2012). "Falling urban wage premium and inequality trends: evidence for Brazil". In: *Regional Research-Investigaciones Regionales,* 24, pp. 91–113.

Dorfman, R. (1979). "A Formula for the Gini Coefficient". In: *The Review of Economics ans Statistics,* 61.1, pp. 146–149.

Fallah, B. N., M. D. Partridge, and M. R. Olfert (2011). "New economic geography and US metropolitan wage inequality". In: *Journal of Economic Geography,* 11.5, pp. 865–895.

Fally, T., R. Paillacar, and C. Terra (2010). "Economic geography and wages in Brazil: Evidence from micro-data". In: *Journal of Development Economics,* 91.1, pp. 155–168.

Fujita, M., P. R. Krugman, and A. J. Venables (1999). *The spatial economy: cities, regions and international trade*. Vol. 213. Wiley Online Library.

Hanson, G. H. (2005). "Market potential, increasing returns and geographic concentration". In: *Journal of International Economics,* 67.1, pp. 1–24.

Head, K. and T. Mayer (2004a). "Market potential and the location of Japanese investment in the European Union". In: *The Review of Economics and Statistics,* 86.4, pp. 959–972.

— (2004b). "The Empirics of Agglomeration and Trade". In: *Handbook of regional and urban economics,* 4, pp. 2609–2669.

— (2011). "Gravity, market potential and economic development". In: *Journal of Economic Geography,* 11.2, pp. 281–294.

— (2014). "Gravity Equations: Workhorse, Toolkit, and Cookbook". In: *Handbook of International Economics* 4, pp. 131–195.

Helpman, E., O. Itskhoki, M.-A. Muendler, and S. J. Redding (2017). "Trade and inequality: From theory to estimation". In: *The Review of Economic Studies* 84.1, pp. 357–405.

Hering, L. and S. Poncet (2010). "Market Access and Individual Wages: Evidence from China". In: *The Review of Economics and Statistics,* 92.1, pp. 145–159.

Ho, L. S., X Wei, and W. C. Wong (2005). "The effect of outward processing trade on wage inequality: the Hong Kong case". In: *Journal of International Economics,* 67.1, pp. 241–257.

Krugman, P. R. (1991). "Increasing Returns and Economic Geography". In: *The Journal of Political Economy,* 99.3, pp. 483–499.

Krugman, P. R. and A. J. Venables (1995). "Globalization and The Inequality of Nations". In: *The Quarterly Journal of Economics,* 110.4, pp. 857–880.

Lambert, P. J. and J. R. Aronson (1993). "Inequality decomposition analysis and the Gini coefficient revisited". In: *The Economic Journal,* 103.420, pp. 1221–1227.

Li, J., D. Shi, S. X. Zhao, and Y. Y. Fred (2014). "A study of the "heartbeat spectra" for "sleeping beauties"". In: *Journal of Informetrics* 8.3, pp. 493–502.

Piketty, T. (2014). *Capital in the twenty-first century / Thomas Piketty ; translated by Arthur Goldhammer*. Cambridge The Belknap Press of Harvard University Press.

Pyatt, G. (1976). "On the interpretation and disaggregation of Gini coefficients". In: *The Economic Journal,* 86.342, pp. 243–255.

Redding, S. J. (2013). "Economic Geography: a review of the theoretical and empirical literature". In: *Palgrave Handbook of International Trade,* pp. 497–531.

Redding, S. J. and P. K. Schott (2003). "Distance, skill deepening and development: will peripheral countries ever get rich?" In: *Journal of Development Economics,* 72.2, pp. 515–541.

Redding, S. J. and A. J. Venables (2004). "Economic geography and international inequality". In: *Journal of International Economics,* 62.1, pp. 52–82.

Strauss-Kahn, V. (2001). "Globalization, Agglomeration and Wage Premia". In: *CEPR Erwit Conference, London, mimeograph.*

Venables, A. J. (1996). "Equilibrium Locations of Vertically Linked Industries". In: *International Economic Review,* 37.2, pp. 341–359.

Verhoogen, E. A. (2008). "Trade, quality upgrading, and wage inequality in the Mexican manufacturing sector". In: *The Quarterly Journal of Economics* 123.2, pp. 489–530.

Yitzhaki, S. (1998). "More than a dozen alternative ways of spelling Gini". In: *Research in economic inequality.*

Appendix A

Gini index dynamics

Rewriting the Gini index function with only λ and one parameter.

$$
\begin{aligned}
G_i = \nu - \lambda = \frac{\kappa \cdot \omega \cdot \lambda}{\kappa \cdot \omega \cdot \lambda + (1-\lambda) \cdot \omega} - \lambda = \frac{\kappa \cdot \lambda}{\kappa \cdot \lambda + 1 - \lambda} - \lambda \\
= \frac{\kappa \cdot \lambda - \lambda(\kappa \cdot \lambda + 1 - \lambda)}{\kappa \cdot \lambda + 1 - \lambda} = \frac{\kappa \cdot \lambda - \kappa \cdot \lambda^2 - \lambda + \lambda^2}{\kappa \cdot \lambda + 1 - \lambda} \\
= \frac{\lambda \cdot (\kappa - \kappa \cdot \lambda - 1 + \lambda)}{(\kappa - 1) \cdot \lambda + 1} = \frac{\lambda \cdot ((1-\kappa) \cdot \lambda + \kappa - 1)}{(\kappa - 1) \cdot \lambda + 1}, \kappa \geq 1
\end{aligned}
\tag{A.1}
$$

Where κ is the wage ratio between skilled and unskilled workers and ω the real wage. We obtain that the Gini index equals zero when:

$$
G_i = 0, if \begin{cases} \lambda = 0, or \\ (1-\kappa) \cdot \lambda + \kappa - 1 = 0 \Leftrightarrow \lambda = 1 \end{cases}
$$

We are looking for a value of λ that corresponds to the critical point of the Gini index function.

$$
\begin{aligned}
\frac{\partial G_i}{\partial \lambda} = \frac{\kappa - 2\kappa \cdot \lambda - 1 + 2\lambda}{(\kappa - 1) \cdot \lambda + 1} + \frac{(\kappa \cdot \lambda - \kappa \cdot \lambda^2 - \lambda + \lambda^2)(-1)(\kappa - 1)}{[(\kappa - 1) \cdot \lambda + 1]^2} \\
= \frac{(\kappa - 2\kappa \cdot \lambda - 1 + 2\lambda)[(\kappa - 1) \cdot \lambda + 1] - 1(\kappa \cdot \lambda - \kappa \cdot \lambda^2 - \lambda + \lambda^2)(\kappa - 1)}{[(\kappa - 1) \cdot \lambda + 1]^2} \\
= \frac{\kappa - 2\alpha \cdot \lambda - 1 + 2\lambda - \kappa(\kappa - 1) \cdot \lambda^2 - \kappa \cdot \lambda + \lambda + (\kappa - 1) \cdot \lambda^2 + \kappa \cdot \lambda - \lambda}{[(\kappa - 1) \cdot \lambda + 1]^2} \\
= \frac{\kappa - 1 + \lambda(2 - 2\kappa) + \lambda^2[\kappa - 1 - \kappa(\kappa - 1)]}{[(\kappa - 1) \cdot \lambda + 1]^2} \\
= \frac{[(\kappa - 1) \cdot (1 - \kappa)] \cdot \lambda^2 + 2(1 - \kappa) \cdot \lambda + \kappa - 1}{[(\kappa - 1) \cdot \lambda + 1]^2} \\
= \frac{(\kappa - 1)[(1 - \kappa) \cdot \lambda^2 - 2\lambda + 1]}{[(\kappa - 1) \cdot \lambda + 1]^2}
\end{aligned}
\tag{A.2}
$$

The solutions to the equation above are λ_1 and λ_2:

$$\frac{\partial G_i}{\partial \lambda} = 0, if \begin{cases} \lambda_1 = \frac{1+\sqrt{\kappa}}{1-\kappa} < 0 \\ \lambda_2 = \frac{1-\sqrt{\kappa}}{1-\kappa} > 0 \end{cases}$$

If κ increase, λ_2 will decrease.

To show that λ_2 is a maximum, we derive the second derivative at the stationary point:

$$\begin{aligned} \frac{\partial^2 G_i}{\partial^2 \lambda}(\lambda_2) = \frac{2(\kappa-1)[(1-\kappa)\cdot\lambda_2 - 1]}{[(\kappa-1)\cdot\lambda_2+1]^2} - \frac{2(\kappa-1)^2[(1-\kappa)\cdot\lambda_2^2 - 2\lambda_2 + 1]}{[(\kappa-1)\cdot\lambda_2+1]^3} \\ = \frac{2(\kappa-1)}{[(\kappa-1)\cdot\lambda_2+1]^2} \\ \cdot\frac{[(1-\kappa)\cdot\lambda_2-1][(\kappa-1)\cdot\lambda_2+1] - (\kappa-1)[(1-\kappa)\cdot\lambda_2^2-2\lambda_2+1]}{(\kappa-1)\cdot\lambda_2+1} \\ = \frac{2(\kappa-1)}{\kappa\cdot\sqrt{\kappa}}\cdot\Big[-\sqrt{\kappa}\cdot\sqrt{\kappa} - \frac{(\kappa-1)}{1-\kappa}(1-\sqrt{\kappa})^2 - 2(1-\sqrt{\kappa}) + (1-\kappa)\Big] \\ = \frac{2(\kappa-1)}{\kappa\cdot\sqrt{\kappa}}\cdot[-\kappa + (1-\sqrt{\kappa})^2 - 2 + 2\sqrt{\kappa} + 1 - \kappa] = \frac{2(\kappa-1)}{\kappa\cdot\sqrt{\kappa}}\cdot(-\kappa) \\ = \frac{-2(\kappa-1)}{\sqrt{\kappa}} < 0 \end{aligned} \tag{A.3}$$

λ_2 is a local maximum.

Appendix B

Tables

TABLE B.1: World Bank Database example

ISO3 Exp.	Year	Gini index	New observation
ALB	2002	31.74	
ALB	2003	.	
ALB	2004	.	30.98
ALB	2005	30.6	
PAK	2011	30.93	
PAK	2012	.	30.81
PAK	2013	30.69	

Notes: We use the dot "." to show that there is no observation.

TABLE B.2: WIID3.4 Gini index sample

Gini data WIID3.4	Observations	Mean	Std. Dev.	Min	Max
Total	8782	39.30	10.90	12.1	78.6
Quality (high)	3970	39.56	10.85	18.5	77.1
Age (all)	3942	39.53	10.88	18.5	77.1
Area (all)	3233	38.14	10.89	18.9	77.1
Population (all)	3128	38.40	10.90	18.9	77.1
Unit of Analysis (person)	2775	38.47	11.34	18.9	77.1
Income Statistical Unit (Household)	2773	38.47	11.34	18.9	77.1
Income (income disposable)	2299	37.57	11.83	18.9	69.4
Scale Equivalence	2299	37.57	11.83	18.9	69.4
Years + balanced data	935	39.07	11.04	21.9	69.4
Countries	166	39.26	10.15	26.1	69.4

Notes: *QUALITY*: Only high quality data is preserved. The evaluation of the data quality was based on three criteria: if the concepts behind each observations is known, the coverage of income and/or consumption and the quality of the survey. *AGE*: Only the data that includes all categories of people was kept. *AREA*: We keep the data that was categorized as covering the entire Area. *POPULATION*: Only the data covering the entire population. Our assumption to divide the population in two categories is not affected by this filter. *UNIT OF ANALYSIS*: We decided to use the person as the unit of analysis, since the other option, household, do not take into account the different sizes among household. Moreover, our theoretical environment consider individuals as the unit of analysis. *INCOME STAT. UNIT*: We keep only the household as the statistical unit since most of the data use it and the authors recommended it. The World Bank based their survey on households. *INCOME*: "income disposable" was chosen as the measure of welfare. Indeed, this notion is used in the theory to catch the welfare. Most countries use also income instead of consumption (another welfare category in WIID) and the latter is difficult to measure, since it must not be confused with expenditure. We also discriminate between income "disposable" and income "gross", since the latter reported Gini results that were inflated comparing to other Gini results from WIID and World Bank database. The Gini mean result dropped compared to the previous line. Following the advices from WIID user guide, we choose then "income disposable" instead of "consumption". *SCALE EQUIVALENCE*: We follow again their advice in keeping the data that uses equivalent scale to weighted different statistical units. However, we did not discriminate among the different methods because there were only one category left. *YEARS*: We consider only the years that we were helpful for the construction of the observations: from 2002 to 2006 and 2010 to 2014. *COUNTRIES*: We remove all the countries for which we already had the two years (2004 and 2012) as observations from the WB database.