UNIVERSITÀ DEGLI STUDI DI BOLOGNA

DIPARTIMENTO DI SCIENZE STATISTICHE

Dottorato in Metodologia Statistica per la Ricerca Scientifica

XXIII ciclo

Export Productivity Premium:
A Statistical Approach based on Quantile Regression to compare International and Domestic Firms

Relatore:
Prof. MARZIA FREO

Co-relatore:
Prof. MARIA ROSARIA FERRANTE

Coordinatrice:
Prof. DANIELA COCCHI

Tesi di:
MATTEO ROCCHI

Settore scientifico-disciplinare:
SEC-S/03

Novembre 2011

ACKNOWLEDGMENTS

Many people have provided advice and support throughout this work, without whom I doubt it would have been possible.

Firstly, many thanks to my supervisor, Marzia Freo, for her support and advice, and to Maria Rosaria Ferrante, for her suggestions, kindness and for always having time. Their comments and suggestions have been indispensable for the good success of this work.

I am sincerely grateful to Emeritus Professor Leopold Simar, for his fundamental contribution and guidance, which I have been fortunate to appreciate from the first day I arrived at UCL (Université Catolique de Louvain). I am indebted to all the members of the CORE (Center for Operations Research and Econometrics), and all the other UCL students and staff with whom I shared my research interests, my doubts, or simply my free-time.

I am grateful, in general, to my teachers, and also to my colleagues in Bologna: Enrico Foscolo, Matteo Borrotti, Fabio Bertozzi, and Francesca Castelli, who made this period more fun and pleasant.

But most of all I would like to express my gratitude to my friends and in particular to my family, that helped me in many different ways, through their continuous support and encouragement in the hard and in the joyful moments of these years.

CONTENTS

LIST OF TABLES

LIST OF FIGURES

INTRODUCTION

The objective of this thesis is to study the relationship between internationalization of firms and productivity, focusing on Italian manufacturing firms. The topic is not easy to handle, and has already been discussed by several studies and researches. In the last years in particular a large number of works have studied the connection between internationalization and productivity of firms.

One of the first works which investigated this relationship was that by Bernard and Jensen (1995), focusing on US manufacturing firms and pointing out that firms involved in export markets are on average better than domestic firms across a large number of performance characteristics: in comparison to non-exporters, exporters are larger, more capital intensive, more likely to undertake R&D, and pay higher wages to workers. But Bernard and Jensen (1995) also showed that exporters are more productive than non-exporters, and that, "within the same industry, exporters do grow faster than non-exporters in terms of both shipments and employment. Exporting is associated with the reallocation of resources from less efficient to more efficient plants. In the aggregate, these reallocation effects are quite large, making up over 40 per cent of total factor productivity growth in the manufacturing sector. Half of this reallocation to more productive plants occurs within industries and the direction of the reallocation is towards exporting plants". The main concept which emerges from the work by Bernard and Jensen (1995) is anyway that productivity of firms is highly positively correlated to their degree of involvement into foreign markets. These results will be later confirmed by a large number of studies (e.g. Bernard and Wagner 1997, Girma et al. 2004). The approach proposed by Bernard and Jensen (1995) anyway considers a simple indicator which tests the difference in productivity between exporters and non-exporters. It roughly compares averages of productivity between two groups (international vs domestic firms), considering in this way an "unconditional" comparison based on a synthesis indicator of productivity. The procedure of Bernard and Jensen (1995) observes only the first moment of the productivity distribution, without considering all the productivity distribution and thus neglecting the heterogeneity of firms' productivity.
After Bernard and Jensen (1995), in order to evaluate the direction and the extension of the relationship between internationalization and productivity, the literature has proposed some alternative statistical-econometric tools.

For example, unlike Bernard and Jensen (1995), Delgado et al (2002) did not just focus their attention on the simple average of productivity between groups, but they considered the whole distribution, evaluating the hypothesis of the first order stochastic dominance between group distributions. But similarly to Bernard and Jensen (1995), even Delgado et al (2002) proposed an "unconditional" approach, neglecting the connection of productivity with other characteristics of firms.

Another strand of literature, including the works by Kimura and Kiyota (2006), Castellani and Zanfei (2007), Castellani et al (2009), also refers to "conditional" approaches, taking into consideration the connection of productivity with other characteristics of firms (e.g. size, industry, location, etc). These characteristics may be controlled through the estimation of a regression model which adopts productivity as outcome variable and both variables describing characteristics of firms and a dummy which differentiates international/domestic firms as covariates. The main drawback of this kind of approach based on the comparison of "conditional" averages is that it neglects the form of the productivity distribution.

In the very last years new approaches are proposed, referring to both "conditional" comparisons and considerations about the whole distribution. Ferrante and Freo (2010) in particular consider the two groups (international and domestic firms) as an "a priori" partition of the firms population, instead of considering them as covariates of a regression model. They identify productivity as the outcome variable, whilst groups are compared through the dummy which differentiates international/domestic firms. Hence Ferrante and Freo (2010) both consider the importance of characteristics, allowing in this way a "conditional" comparison between groups, and refer the comparison to the whole productivity distribution instead of limiting the comparison to synthetic indicators. Our work will follow this last strand of literature, aiming at increasing the body of literature about the relationship between internationalization and productivity.

We consider internationalization by defining two different groups of firms: "international firms" (INT) and "domestic firms" (DOM). The INT group includes multinationals, firms that export, firms involved in Foreign Direct Investments, and firms that have any other kind of internationalization status[1]. The DOM group includes only firms that just work domestically, without having any kind of interaction with foreign markets. We then consider productivity as the outcome variable, and as in Ferrante and Freo (2010) we let the international/domestic variable play the role of defining the groups to compare. In this way we consider the two groups as an "a priori" partition of the firm population, without including this variable in the vector of covariates of a regression model. Because of this our procedure will differentiate and stand out from the approaches proposed by Kimura and Kiyota (2006), Castellani and Zanfei (2007), and Castellani et al (2009), but it will anyway consider (as the mentioned authors do) that compared groups may

[1] Note that a firm which is involved in Foreign Direct Investments will be included into this group of "international firms", even in case the firm does not export

differ through many characteristics (e.g. as size, location, industry, etc.); we will afford in this way a "conditional" comparison between groups. In fact we believe that these characteristics may have some kind of association with the level of firm productivity. On the other hand if we did not consider these characteristics, our approach would be "unconditional" and the evaluation of the productivity gap between international and domestic firms would consist in a simple comparison which neglects the "ceteris paribus" condition. Moreover, we do not limit to synthetic indicators the comparison INT/DOM firms, but we take into consideration the whole productivity distribution.

The main approach used in the present work to deal with the objective of evaluating in direction and extension the relationship between internationalization of firms and productivity, is a quantile decomposition method; it allows us to compare the difference between group distributions through a quantile regression model after controlling for differences in individual characteristics. Through this procedure we will be able to distinguish what proportion of the overall productivity gap is due to differences in firms characteristics, and what proportion is due to the international status over the whole distribution.

Another approach that we use to decompose the productivity gap between international and domestic firms into components is a propensity score matching method. This approach refers to the stream of literature about impact evaluation of policy interventions. If the total (gross) productivity gap is considered an Average Treatment Effect of the openness to international market on productivity, this can be decomposed into a component due to characteristics of firms, and into a coefficient component identified by the Average Treatment Effect on the Treated. Since also this approach takes into consideration the influence of characteristics on the level of firm productivity, also this approach may be defined "conditional", but in this case we will not make considerations about the whole distributions.

The approaches outlined allow us to answer to a series of interesting research questions:

Considering the two groups of INT and DOM firms, are they different in terms of productivity? is their productivity gap significantly different from zero?

If the two groups (INT vs DOM) differ in terms of productivity, does the gap amount depend on the productivity level? Is it constant along the whole distribution, or does it vary quantile by quantile? is it lower or higher for firms with a lower or higher productivity?

Is the relationship between productivity and variables describing firms' heterogeneity different, in sign and amount, at different productivity distribution quantiles?

Does the productivity gap survive after controlling for a number of characteristics of firms? And, if the answer is affirmative, does the extent of this surviving gap depend on the amount of productivity?

Which portion of the productivity gap is due to characteristics of firms and which portion is due to a net productivity premium?

Will we obtain different results by applying different methods to decompose the total productivity gap? For example would the application of an "impact evaluation" approach offer the same results of a quantile decomposition technique?

The next chapters attempt to answer to all these questions aiming at increasing the body of literature about internationalization of firms and productivity.

The thesis is organized as follows.

Chapter 1 is entirely dedicated to internationalization of firms. Through this first chapter we attempt to offer to the reader the basis to understand the research field of the thesis. The objective is to illustrate a general point of view and a historical description of all the existing literature about micro-economic theories and models, as well as a general description of export markets with advantages and drawbacks that firms have to afford entering in international markets.

Chapter 2 presents the statistical methodologies useful to study the connection between internationalization and productivity of firms. In this chapter we will focus on the instruments that we need to deal with our economic problem. First of all we deal with productivity and the possible ways to measure it. Then we concentrate on methods that allow us to decompose differences between groups, focusing in particular on quantile decomposition techniques and impact evaluation methods.

Chapter 3 is dedicated to the full description of the method that we adopt to handle our economic problem, and to the application of this method to a dataset of 5,073 Italian manufacturing firms. Results of the survey, comments, and conclusions, complete this last chapter.

CHAPTER 1

INTERNATIONALIZATION OF FIRMS

0. **Introduction**
1. **A review of the micro-economic literature about internationalization of trade and firms**
2. **Self selection vs learning by exporting**
3. **Entry and exit from export markets**
4. **How to enter in foreign markets**
5. **Export and investments**

INTRODUCTION

Chapter 1 is entirely dedicated to internationalization of firms. The objective of the chapter is to offer to the reader a general point of view and a historical description of all the existing literature about micro-economic theories and models, as well as a general description of export markets with advantages and drawbacks that firms have to afford entering in international markets. Through this first chapter we attempt to offer to the reader the basis to understand the research field of the thesis.

The chapter is made up by five paragraphs.

Paragraph 1 offers a review of the micro-economic literature: starting from the point of view universally accepted until the mid-18th century (classical economics in opposition to mercantilism), we show how the theories modify and progress year by year until arriving to the contemporary internationalization trade theories.

After classical economics, we refer to Smith (1776), one of the first to oppose dirigisme and to see advantages in trade and specialization, Ricardo (1817) who explicitly articulated the theory of comparative advantage, Mill (1848), Marshall (1879), the early-20th century theories about the consequences of imposing impediments to international trade, Heckscher (1918), Ohlin (1924), and Haberler (1930) that broke with the labour theory of value, as well as the so-called “new trade theory” by Krugman (1979) and others. Then we cite the “proximity–concentration trade-off” by Brainard (1997), Clerides et al. (1998), and we arrive to the 2000’s, when an influent work by Melitz (2003) created a new strand of literature about causal links between exporting and industry productivity. Finally, we focus on the last developments of internationalization of firms (often born by the seminal work of Melitz), citing amongst others Jean (2002), Bernard Eaton et al. (2003), Ederington and McCalman (2004), Falvey et al (2004), Syverson (2004), Helpman Melitz and Yeaple (2005), Campbell and Hopenhayn (2005), Bernard Redding and Schott (2003, 2006), Melitz and Ottaviano (2003, 2008), Combes Duranton et al. (2009), Helpman Itskhoki and Redding (2009).

Paragraph 2 is dedicated to the two main hypothesis that can explain the productivity differences between international and domestic firms: self-selection and learning by exporting. Starting from the fact that international (exporting, or engaged in FDI[2]) firms are more productive than domestic (non-exporting and non-engaged in FDI) firms, as demonstrated by a large number of researchers e.g. Bernard and Jensen (1995), we focus on the literature that investigates the direction of causation between productivity and internationalisation, trying to understand if ex-ante differences in productivity lead to differences in internationalization choices (self selection), or if firms become more productive only after their entry in the foreign market, thanks to their international involvement through an exposure to best practice technology and learning (learning by exporting).

We will also show the results of some works which test the hypothesis of self selection and learning by exporting in several countries: the results that the researchers obtain are somewhat mixed: in some countries they find evidence of self selection but not of learning by exporting, in some others they find evidence of learning by exporting but not of self selection, whilst in most cases both effects are present.

Paragraph 3 deals with determinants and consequences of entry and of exit from export markets.

If it is demonstrated that entering in an export market in most cases gives to a firm the possibility to improve its productivity, it will be interesting to ask ourselves a question: how can a firm enter in export markets, and what determinants matter more? Thus one of the aim of this paragraph is to identify the main determinants of entry, that we identify in sunk costs, productivity, exchange rates, policy innovations, and agglomeration. A sub-section will be dedicated to each determinant.

A second part of the paragraph is dedicated to the positive consequences of entry, after which we will observe how it is possible to find similarities between determinants of entry and self-selection, as between consequences of entry and learning by exporting.

The last two sections of the paragraph are dedicated to the determinants of exit (if export improves productivity, why firms stop exporting?), and to the consequences of exit (what happens, in particular to productivity, after firms stop exporting?).

Paragraph 4 focuses on the different ways that firms have to internationalise themselves. Firms can enter in foreign markets in different ways: in particular the main two alternative ways to enter in foreign markets are through export and through investments abroad to sell on the foreign market (FDI). The first part of this fourth paragraph will be dedicated to these two alternatives.

But also other ways for firms to be international exist. One in particular may be considered interesting to study, because it involves a great degree of internationalisation. In fact some firms are

[2] Foreign Direct Investments

internationalized more than others, producing more products, exporting and being involved in FDI in more countries, using multiple stages of production. These particular kinds of firms may be defined "multinationals", and differ from all the others in that they are more likely to export and export more intensively. Because it is interesting to study this particular kind of internationalised firms, we will dedicate a large section of this paragraph to multinationals, focusing on the export decisions, on the positive impact of the choice of working as multinationals towards productivity, and on the policies introduced by governments in order to attract multinationals.

Finally, in the last section of the paragraph, we attempt to explain how firms take export decisions, describing the literature that deals with this topic. We cite amongst others the works by Motta and Norman (1996), Ekholm et al (2003), Grossman et al. (2003), Head and Ries (2004), and Girma et al. (2005).

Paragraph 5 is dedicated to the link between export and investments. As demonstrated by several authors (e.g. Lopez 2004) the correlation between export and investments is largely positive, so in this paragraph we attempt to understand what investments are correlated with export more than others, and why. We will focus in particular on the investments in research and development, technology, innovation, and in domestic market, citing works as Hirsch and Bijauoi (1985), Schlegelmilch and Crook (1988), Entorf Krader and Pohlmeier (1988), Bernard and Jensen (1995), Wagner (1996), Wakelin (1997, 1998), Ebling and Janz (1999), Baldwin and Gu (2004), Lopez (2004), Cassiman and Martinez-Ros (2007), Atkeson and Burstein (2007), Constantini and Melitz (2008), Bustos (2008), Verhoogen (2008), Lileeva and Treer (2008), Konings and Vandenbussche (2008), and Mayneris (2010) amongst others.

We conclude presenting a number of frameworks that in very recent years focus on investment, and in particular on the link between trade and credit constraints, referring to Rajan and Zingales (1998), Demirguc-Kunt and Maksimovic (1998), Beck Demirguc-Kunt and Maksimovic (2005), Manova (2008), Muuls (2008), Beck Demirguc-Kunt Laeven and Levine (2008), Mayneris (2010), Berman and Hericourt (2010), Manova (2010), and Bellone Musso Nesta and Schiavo (2010).

1.

A REVIEW OF THE MICRO-ECONOMIC LITERATURE ABOUT INTERNATIONALIZATION OF TRADE AND FIRMS

All the contemporary internationalization trade theories have been inspired by the classical economics developed in opposition to mercantilism, which was until the mid-18th century the most common and accepted view. Mercantilism referred to the interference in the economic activities by the governments through the imposition of prohibitions, quotas and tariffs in order to control international economics. The government had to intervene playing a protectionist role in the economy, by encouraging exports and discouraging imports. In fact in mercantilism economic assets or capital were represented by bullion (gold, silver, and trade value) held by the state, which could be increased only through a positive balance of trade with other nations (exports - imports).

Smith (1776) was one of the first to oppose dirigisme and to see several advantages in trade and specialization, providing the starting point for classical theories of trade in the work "Wealth of Nations" (1776). His main argument in favour of trade is the so-called "vent for surplus": Smith (1776) argued that, because of international trade "the narrowness of the home market does not hinder the division of labour in any particular branch of art or manufacture from being carried to the highest perfection. […] By opening a more extensive market for whatever part of the produce of their labour may exceed the home consumption, it encourages them to improve its productive powers and to augment its annual produce to the utmost, and thereby to increase the real revenue and wealth of the society".

Anyway Smith (1776) did not develop a theory of comparative advantage, even if he arrived very close to it noting that Britain was more productive in manufactures relative to Poland than it was in agriculture[3].

Ricardo (1817) was the first to explicitly articulate the theory of comparative advantage in the work “Principles of Political Economy and Taxation” (1817). Ricardo, as Smith, developed his views on trade within the framework of his model of how an economy should operate. Ricardo argued that lower costs in the production of goods was not a sufficient reason to produce them all, and that countries would take advantage to specialize in the production of those goods which they could produce better. Ricardo (1817) based his theory on a model of supply: price is the long-run supply price. Ricardo also considered the possibility of non-traded goods, and the idea that a country could produce and export more than one good if country size or demand patterns were highly uneven. In Ricardo’s trade theory specialization is fundamental, but there is no exact specification about how the gains from trade are divided between countries.

Mill (1848) contributed to the literature and explained how terms of trade were established between the limits set by comparative advantage, by inserting into the model the demand and its elasticity. In his analysis he even considered transport costs, more commodities, and more countries, finding that the gains from trade are greater for those countries where the opportunity cost of the exportable good is lower.

But whilst Mill (1848) developed his analysis just verbally, Marshall (1879) derived offer curves to depict graphically the general equilibrium in production and consumption. Marshall (1879) used the device of “bales” of goods, making the theory more general through a two-country multi-commodity analysis. Marshall developed detailed analysis about equilibrium and stability conditions, but he still did not consider elasticity.

During the 20th century the literature developed testing the theory of comparative advantage by Ricardo, by justifying economists’ generally free-trade position, and in the 1930s and 1940s by discussing about the gains and losses derived by imposing quotas, tariffs, and other impediments to international trade.

After the introduction of the “neoclassical” assumptions about technology and preferences in the analysis of comparative costs, Haberler (1930) broke with the labour theory of value, the production possibility frontier becoming standard in all economic theorizing, transforming in this way the theory of international trade.

[3] Even other authors arrived very close to the theory of comparative advantage some decade later, e.g. Torrens (1815)

Heckscher (1918) and Ohlin (1924) developed the same studies parallel to Haberler[4], addressing the issue of factor prices under free trade.

Heckscher (1918) developed the general equilibrium analysis, examining reasons and consequences of the Swedish migration to USA. He demonstrated that under certain circumstances trade in goods could substitute for movement of factors in equalizing factor prices, and if factor endowments were not too different factor prices would be equal throughout the world.

Ohlin (1924) refused the conclusion of the equalization of factor prices. He assumed that each country was completely specialized, and developed Heckscher's studies by using a more formal approach which referred to the Cassell's version of Walrasian general equilibrium analysis.

The Heckscher-Ohlin model of international trade referred to profit maximising firms, operating under constant returns to scale. However in their work the boundaries were not still well defined, and firms had no deterministic role in determining the pattern or commodity composition of trade.

Krugman (1979) and others[5] introduced the so-called "new trade theory", which gave us new insights into the determinants of trade. The "new trade theory" was born in contraposition to the "old trade theory" which referred to the "old" theories based on comparative advantage.

Whilst "old" trade theories emphasized reallocations of resources across industries, predicting the expansion of comparative advantage industries and the contraction of comparative disadvantage industries as trade costs fall, "new" trade theories are based on consumer love of variety and increasing returns to scale, predicting either that all firms export or that none do depending upon the level of trade costs.

The "new trade theory" is built on the Dixit-Stiglitz monopolistic competition[6]. Because each firm produces a unique variety that consumers need, all firms export. In this setting any trade costs just absorb a proportion of a firm's foreign revenue but do not stop it from exporting. All the fixed costs that occur to exporting firms are not considered by this theory; so the world represented by the

---

[4] Heckscher and his student Ohlin preceded Haberler in their original ideas, but they initially published their work only in Sweden, reducing drastically their visibility. Their work will then be published in English only in 1933 (Ohlin) and 1949 (Heckscher).

[5] After Paul Krugman, important contributes to the "new trade theory" are offered by Elhanan Helpman and William Ethier.
Helpman and Krugman (1985) will also integrate old and new trade theory by embedding horizontal product differentiation and increasing returns to scale in a model featuring endowment-based comparative advantage.

[6] The Dixit-Stiglitz monopolistic competition model (1977) has been used to study optimum product diversity. It is a simple general equilibrium model with *n* monopolistic goods and a numeraire good, which can be interpreted as labour (or leisure) time or as the aggregation of all the other goods in the economy.

model where all firms export, is still far from the real world where some firms export and others in the same industry do not.

Sunk costs did not enter in the theoretical microeconomic models until the 1990s. Clerides et al. (1998) were one of first to introduce the fixed costs into their model, claiming that firms had to face several sunk costs when they were entering in the export markets: in fact firms had to set up new distribution networks, complete option appraisals, do market research, modify existing products (e.g. repackaging of products to appeal to new consumers), and so on. In the Clerides et al. (1998) model not all firms export: firms have to face sunk costs, and only firms with sufficiently high profits to cover those costs have the possibility to start exporting.

This theory would apparently lead to the conclusion that firms self-select. Firms have to face several sunk costs, to do it they must be productive, and only those that have raised a sufficient productivity domestically can self–select to enter into export markets.

But Clerides et al. (1998) also consider the possibility of learning-by-exporting. After the firm has entered export markets, productivity increases just thanks to the fact of exporting.

This can be rationalised in various ways: actual involvement in export markets can sharpen incentives to innovate by raising returns to innovation (Holmes and Schmitz 2001), can lead to a re-engineering process, or again can force firms to reduce X-inefficiency because export markets are more competitive than domestic markets.

A similar point of view is given by Brainard (1997), not still rejecting a model with only country and industry effects, but finding that overseas production by multinationals increases relative to exports the higher are transport costs and trade barriers, and the lower are investment barriers and scale economies at the plant level relative to the corporate level.

The framework explains the internationalization of firms in terms of the so-called proximity–concentration trade-off, examining the extent to which multinational location decisions reflect a trade-off between achieving proximity to customers and concentrating production to achieve scale economies. Firms export if there are cost advantages to concentration, and establish foreign production facilities when proximity to local markets is more important.

A finding is that, contrary to conventional wisdom, multinational activity is more likely the more similar are the home and foreign markets, and firms usually produce abroad the same products they produce at home.

As Brainard (1997), traditional international trade theory asserted until the end of the 1990s that firms in the same industry or country adopt very similar behaviour in terms of international involvement. Firms produce the same products at home and in the foreign market, and they concentrate the production at home serving foreign markets via exports based on advantages to concentration.

After this strand of literature models have been extended to include internationalization forms that arise if the firm locates each stage of production in the country where it can reduce overall production costs. Hence, in this case firms produce abroad products different from those that they produce at home. One of the first papers to consider the interaction between productivity differentials across firms in the same industry and the fixed costs of exporting is that by Melitz (2003).

Melitz (2003) builds a dynamic industry model where heterogeneous firms operate in the Dixit-Stiglitz monopolistic competition, helping us to understand, despite its microeconomic structure, the correlation between export and growth observed at the macro level. In the model firms incur fixed costs to export, so a natural productivity threshold determines if a firm can or cannot export. First of all a rationalisation effect appears: exporting increases expected profit, which induces entry, pushes up the productivity threshold for survival and drives out the least efficient firms. Clearly this raises average industry productivity. Later, exporting allows the most productive firms to expand and causes less productive firms to contract. This reallocation effect acts to raise average industry productivity for the second time.

Anyway through Melitz (2003) a causal link between firm productivity and export is for the first time created, with only the most productive firms self-selecting into export markets. This idea will be confirmed by the future literature (e.g. Melitz and Ottaviano 2003, Bernard et al. 2003), leaving behind the assumption of a representative firm for each sector and providing theoretical foundations for the relationship between within-sector heterogeneity of firms and international trade in general equilibrium. After Melitz (2003) the literature which pointed to causal links between exporting and industry productivity became wider.

Ederington and McCalman (2004) developed a model of firm heterogeneity with an outcome which is the opposite than that of Melitz (2003). They consider heterogeneity as a consequence of the decision of some firms to start to export.

Bernard, Eaton, et al. (2003) presented an industrial organisation structure which is partially different than that by Melitz. As Melitz (2003) they derived rationalisation and reallocation effects, but while the framework by Melitz (2003) was driven by import competition, that by Bernard Eaton et al. (2003) is driven from exporters penetrating more markets.

Jean (2002) also identifies import driven and export driven contributors to industry productivity growth, in a two-country setting with differences in relative efficiencies across countries.

Helpman, Melitz and Yeaple (2005) extended Melitz (2003) to consider the decision to set up an overseas affiliate. As in Melitz (2003) increased globalisation is likely to lead to firm exit, where the probability is decreasing in whether the firm is an exporter or a multinational firm. The model introduced by Helpman, Melitz and Yeaple (2005) explains heterogeneity with respect to foreign trade, and, unlike Melitz (2003), includes the possibility that firms engage in FDI (Foreign Direct Investments). The productivity affects international choices, and the results in terms of export involvement depend on how productive the firms are: the most productive firms can enter in the foreign market affording the higher costs associated with FDI, the medium-productive firms find it profitable to reach foreign markets through export, and the least productive firms have no possibility to afford FDI or enter in the export market but can just serve the domestic market. The possibility to enter in the international market (through export or FDI) mainly depends on affording sunk costs. The sunk costs generally include fixed costs of market research into product compliance, new distribution networks, advertising, and specifically for FDI the acquisition of an existing firm or the built of new production facilities.

Many other papers extended Melitz (2003) considering asymmetries between countries.

Bernard, Redding and Schott (2003, 2006) develop a model which explains an alternative form of exit to death-industry switching. In their results they confirm the importance of productivity levels, and they find that more productive firms choose to produce products with higher sunk costs. Firms change their output mix towards industries in which they have a comparative advantage and therefore avoid competition from countries in industries where they do not. For OECD countries this is more likely towards the use of technologies with higher costs. Bernard, Redding and Schott (2003, 2006) explain through their framework why some countries export more in some industries (endowment-driven comparative advantage), why nonetheless two-way trade is observed within industries (firm-level horizontal product differentiation combined with increasing returns to scale), and why some firms export whilst others do not (self-selection).

Falvey et al (2004) focus on frontier technology and they find interesting results. First, if the degree of substitution across products is higher in a firm, there self selection is stronger. Second, the success of a firm in the export market is strongly correlated to the efficiency of the country where the firm is. The more the country is efficient, the more likely firms are to survive in the export market, and the less likely they are to survive in the more efficient country, which leads us to expect that trade structure is important.

Another strand of literature extends Melitz (2003) including “market size” as a variable: while in Melitz (2003) the market size does not affect the average firm size or the productivity distribution (at least in the closed economy and in the open economy with symmetric countries), some frameworks concentrate on this lack. In particular Syverson (2004) demonstrates that in sectors where physical output can be measured as well as sales (and hence prices recovered), larger markets imply higher average productivity, higher average plant size, less dispersed productivity distribution, lower prices, and less dispersed prices. Campbell and Hopenhayn (2005) focus on the market size variable finding that retail establishments in larger markets increase both sales and employment; the authors also give some evidence of greater dispersion when the market size is larger.

The introduction of the variable “market size”, together with new considerations on trade liberalization and new hypothesis, are introduced in a model created by Melitz and Ottaviano (2003, 2008) to develop studies about competition in different countries. Market size and trade affect the toughness of competition in a market, which feeds back to influence the productivity distribution. Both average productivity and average mark-ups respond to the opening of trade and influence one another. In other words Melitz and Ottaviano (2003, 2008) find that competition is “tougher” in the large countries, so that product choice is greater, average productivity higher, but firm survival lower because new entrants have a higher probability of failure. They also consider trade liberalisation, finding that it increases competition in particular in the big countries which attract a great number of firms. Hence firms selling in large markets are larger and more productive, since higher competition forces the mark-ups in a large market downwards. Like the Melitz (2003) model, the Melitz-Ottaviano (2003, 2008) model has been widely used in subsequent applications. In particular Combes, Duranton, et al. (2009) use the Melitz-Ottaviano framework to examine the empirical importance of agglomeration and selection in driving observed productivity differences across locations with different market size; another application is offered by Mayer, Thierry, Melitz and Ottaviano (2009), using the Melitz-Ottaviano framework to model multi-product firms.

Alternative new approaches are also offered by Behrens and Murata 2006 (monopolistic competition with CARA[7] preferences), Feenstra 2003, and Feenstra and Weinstein 2009 (monopolistic competition with homothetic translog preferences)

Helpman, Itskhoki, and S. Redding (2009), in their frameworks “Inequality and Unemployment in a Global Economy” and “Unequal Effects of Trade on Workers with Different Abilities” extend Melitz (2003) by concentrating on workers and wages. They introduce the “exporter wage premium” demonstrating that exporters pay higher wages than non exporters for given productivity. Helpman, Itskhoki, and S. Redding (2009) also find that the most productive firms screen workers more intensively in order to obtain information about match-specific abilities and exclude low-ability workers, hence most productive firms have the possibility to hire workforces of higher ability.

[7] Constant Absolute Risk Aversion

2.

SELF SELECTION vs LEARNING BY EXPORTING

2.1 Why international firms are more productive than domestic firms: the two hypothesis of self selection and learning by exporting

Even if it is not always possible to generalize, empirical evidence has shown how, as underlined in the previous paragraph, exporting firms are usually more productive than non-exporting firms (e.g. Bernard and Jensen 1995).

It is possible to find the reasons about this production difference through two alternative, but not mutually exclusive, hypothesis: self-selection and learning by exporting. These hypothesis increased the literature that investigates the direction of causation between productivity and internationalisation, which has always been controversial.

In the self-selection hypothesis the causality runs from productivity to export. The most productive firms "self-select" into export market, in the sense that they raise productivity before their entry in the foreign markets, not after. Only firms with a sufficiently high productivity have the chance to afford the initial fixed costs and hence to start export. Self selection is driven by all the costs that firms have to encounter because of the decision to sell their products in the foreign markets. The costs include distribution costs, transportation costs, personnel with skill to manage foreign networks, marketing costs, production costs in modifying current domestic products for foreign consumption. All these entry costs constitute an initial barrier to export, that can be overcome only by the most productive firms. Moreover the behaviour of firms might be forwardlooking in the sense that the desire to export tomorrow leads a firm to improve performance today to be competitive on the foreign market, too. Cross-section differences between exporters and non-exporters therefore may in part be explained by ex ante differences between firms (Wagner, 2007).

Learning by exporting is the alternative hypothesis to self-selection, and the causality runs now from export to productivity. Firms become more productive after their entry in the foreign market, thanks to their international involvement through an exposure to best practice technology and learning. Export starters, after entering in the foreign market, find the possibility to gain knowledge flows and productivity advantages through the contact with international competitors and customers. Exposition of new entrants to competition, in particular, stimulates them to work better improving the productivity. Export starters are exposed to an intense competition, so that they must learn how to improve themselves fast, while domestic firms do not need. In addition to that, interaction with foreign competitors and buyers provides information about process and product reducing costs and raising quality. Furthermore new exporters can also find advantages in the innovation acquired through the foreign market but domestically absent. In fact increased competition in foreign markets forces firms to be more efficient and stimulates innovation. Another consequence of learning by exporting is that the entry in the international market may lead firms to re-engineer their processes, enhancing again the productivity. Finally exporting allows firms to increase scale, and this may be an important source of efficiency change (Tybout and Westbrook 1995).

2.2 Testing self selection and learning by exporting in some countries: what literature shows

The literature about self-selection and learning by exporting, which tries to find the direction of causality between productivity and internationalisation has been very controversial, but in the last ten years the literature has rapidly increased finding applications and giving evidence on a large number of countries.

Helpman et al. (2004) point out the self-selection hypothesis, assuming that ex-ante differences in productivity lead to differences in the internationalization choices. A large number of studies confirm the assumption that the productivity of firms engaged in FDI or exporting dominates that of domestic ones, and show a presence of the self selection effect. Evidences on this relationship are now available for US (Bernard and Jensen, 1995, 1999, 2004), Germany (Bernard and Wagner 1997, Arnold and Hussinger 2005, Wagner 2005), Colombia Mexico and Morocco (Clerides et al. 1998, Isgut 2001), Taiwan and Korea (Liu et al. 1999, Aw et al. 2000, Hahn 2004), Italy (Castellani 2002, Castellani and Zanfei, 2007), Spain (Delgado et al. 2002, Farinas and Martin-Marcos 2007), Chile (Alvarez 2002, Pavcnik 2002, Alvarez and Lopez 2005), Thailand, Indonesia, Philippines and Korea (Hallward-Driemeier et al. 2002), Canada (Baldwin and Gu 2003), Republic of Ireland (Girma et al. 2004), UK (Greenaway an Yu 2004, Girma et al. 2005), Sweden (Greenway and Kneller 2005), and Japan (Kimura and Kiyota, 2006). A strong evidence for self-selection is shown in all the mentioned studies, but a singular exception is offered by Head and Ries (2003), which found opposite results studying Japanese firms. They find that if the foreign country offers some cost advantages, the least productive firms locate their production abroad whereas the more productive ones concentrate the production in the home country. Other exceptions are registered in Sweden (Hanssohn and Lundin 2004, Greenaway Gullstrand and Kneller 2005, Greenaway and Kneller 2007), and in Slovenia (Damijan et al. 2007), where self-selection is demonstrated to have insignificant effects.

The literature about learning by exporting is less wide, but there are studies about Colombia, Mexico and Morocco (Clerides et al. 1998), US (Bernard and Jensen 1999), China (Kraay 1999), Taiwan and Korea (Aw et al. 2000), Italy (Castellani 2001 2002), sub-saharian Africa (Bigsten et al. 2002), Spain (Delgado et al. 2002), Germany (Wagner 2002), UK (Girma et al. 2003, Crespi et al. 2008), Sweden (Hanssohn and Lundin 2004, Greenaway Gullstrand and Kneller 2005), and Slovenia (Damijan et al. 2007). Anyway evidence on learning by exporting is somewhat mixed. Some studies do not find any evidence for learning effects from exporting: these are the cases of Clerides et al. (1998), Bernard and Jensen (1999), and Wagner (2002). Other authors find some evidence for learning by exporting effects: Kraay (1999) finds evidence for China, and the same happens to Bigsten et al. (2002) studying firms in Sub-Saharan African countries, Hanssohn and Lundin (2004) and Greenaway Gullstrand and Kneller (2005) in Sweden, and Damijan et al. (2007) in Slovenia. Aw et al. (2000) perform studies about Taiwan and Korea. For Taiwan they find that new exporters outperform non-exporters before entry, and in some industries they also experience productivity improvements after entry. But after entry continuous exporters do not increase their productivity advantage towards non-exporters over time. These results are consistent with the self-selection hypothesis, but lend only limited support to the learning hypothesis. About Korea Aw et al. (2000) find that the correlation between export status and firm productivity is less crisp, but they find no support for the learning hypothesis. Castellani (2001, 2002) finds learning effects in Italian firms over a very high threshold of exposure to foreign markets, while below this threshold learning effects are absent. Finally Delgado et al. (2002) study a panel of Spanish firms applying non-parametric methods. In their results the evidence for learning effects is not significant. Anyway Delgado et al. (2000) find some evidence for learning effects through limiting the sample to the subset of young firms.

3.

ENTRY AND EXIT FROM EXPORT MARKETS

A large strand of literature is dedicated to investigate the reasons why firms decide to enter into the foreign markets, the reasons why they decide to exit, and the consequences of these choices.

3.1 Determinants of entry

The determinants of entry into an export market by a firm may be identified in sunk costs and productivity, exchange rates, policy innovation, and agglomeration.

3.1.1 Sunk costs and productivity

Melitz (2003) and others claim that participation decisions are mainly determined by a combination of firm productivity and sunk costs. Several firm characteristics are investigated to understand if they influence the probability of entering in the foreign markets. The set of firm characteristics include size, age, human capital, capital-intensity, ownership, etc. Studies revealed through robust results that most variables are correlated with export market entry. Hence it is possible to predict episodes of entry and exit by firms that show changes in these characteristics.

Studies find that one of the most important variable is time. The international choices of the firm tomorrow strongly depend on the international involvement today, even when other determinants of persistence have been controlled for. This time correlation is particularly strong in the Italian case: Bugamelli and Infante (2002) demonstrated that past participation increases the probability that a firm will continue to export by 90% in Italy. In other countries the percentage decreases (Bernard and Jensen 2004 estimated a 36% about the US case) but the correlation always remains strongly positive and significant. These result can be interpreted in terms of an important evidence of the sunk costs, as argued by Roberts and Tybout (1997), Bernard and Jensen (1999) and Bernard and Wagner (2001). According to this approach the best firms are the ones which are able to afford sunk costs associated to the entry into foreign markets, and this also leads to the possibility to create new positive profits abroad. In addition to that competition could be fiercer outside the home market, a feature that would again allow only the most productive firms to do well abroad. This explanation is in line with the assumption made in the theoretical literature of international trade with heterogeneous firms that high-performing firms self-select themselves into foreign markets (Arnold and Hussinger 2004). Hence entry is mainly determined by changes in sunk costs. As Das et al. (2001) show these are most relevant for those firms who export little, the "fringe players in export markets" (Tybout, 2003).

After sunk costs and productivity, Greenaway and Kneller (2007) identify other three changes that produce flows of entry and exit by export markets: exchange rates, policy innovation and agglomeration effects.

3.1.2 Exchange rates

Even if at macroeconomic level most papers (Pozo 1992, Chowdhury 1993, Parley and Wei 1993) claim that the effects of exchange rates are insignificant or small in magnitude so exchange rates play little or no role as a sunk cost, the microeconomic evidences consider exchange rates important. In the presence of sunk-costs the export responsiveness of exchange rate changes is likely to be higher amongst current exporters compared to non-exporters (Greenaway Kneller 2007). Changes in exchange rates contribute to changes in the intensive rather than extensive margin. This is confirmed by some frameworks, e.g. Bugamelli and Infante (2002), Bernard and Jensen (2004). The latter, studying the export response of US manufacturing plants to dollar depreciation in the 1980s, show that 87% of the expansion was from increased export intensity whilst 13% from entry of new firms. But a strand of literature claims that this approach does not provide a complete explanation of microeconomic responses. Das et al (2004) simulate a devaluation and find that the magnitude of industry response depends on previous export exposure, and homogeneity of expected profit flows between firms and their proximity to the export market entry threshold. Devaluation can also lead to substantial exit: Blalock and Roy (2007) show that although there was an expansion of export activity by established exporters and new entry by non-exporters, new activity was offset by cessation of exporting by previous exporters. An explanation can be found: firms that ceased exporting, compared to firms that continued to export, were not less productive and had no liquidity constraints or infrastructure problems. On the other hand they were less likely to be foreign and less likely to have made R&D or training investments. These same variables predicted which firms would become new exporters (Blalock Roy 2007).

3.1.3 Policy innovation

Policy changes may impact the environment where the firms work and therefore may influence export decisions. A pragmatic role is played by two kinds of policy innovation: trade liberalisation and export promotion.

Baldwin and Gu (2003) studied trade liberalisation and observed that the 4.5% reduction in Canadian-US tariffs increased the probability of exporting by 63%. Both the number of exporters (the share of plants that export increased from 37 to 53% between 1984 and 1990) and the export intensity (48% of exporters) increased.

Blalock and Gertler (2004) observed how trade liberalisation doubled the number of exporters in Indonesia in only 6 years (from 1990 to 1996).

Export promotion may imply the intervention of the government, providing infrastructure support and offering direct export subsidies. Bernard and Jensen (2004) find an insignificant effect from export promotion schemes, comparing exporters and non-exporters. Alvarez (2004) finds insignificant effects in the comparison between different typologies of exporters (permanent versus sporadic exporters), but finds significant effects in detail: the probability that a firm will become a permanent exporter does not depend on trade missions and trade shows, but it does positively depend on market studies and arranged meetings with clients, authorities and experts.

Moreover Alvarez (2004) finds that public instruments for export promotion are used by established exporters more than by sporadic exporters, and finally finds evidence of self-selection when evaluating export promotion schemes.

Görg et al. (2007) study the government intervention in the payment of grants to Irish firms. They refer to capital grants, loan guarantees, technology acquisition grants, employment grants, rent subsidies, training grants, research and development grants, and feasibility study grants. In their framework they find a success from intervention, but limited. Large grants lead to additional export inducing existing exporters to expand overseas sales further, but there is no evidence of additional entry in the sense that grants fail to encourage firms that did not export previously, to enter in the foreign market now.

3.1.4 Agglomeration

The literature about agglomeration does show mixed results. Even if many studies find strong positive spillover effects (Aitken et al. 1997, Kokko et al. 1997, Greenaway and Kneller 2003, Greenaway et al. 2004) some studies have found none or in some cases even negative effects (Sjoholm 2003, Barrios et al. 2003, Bernard and Jensen 2004, Ruane and Sutherland 2005), and again others have found mixed evidence, depending on the channel considered (Swenson 2005, Kneller and Pisu 2007). Whilst positive or insignificant effects may be considered easy to explain, negative effects are more puzzling. Some contribution to the explanation is offered by Ruane and Sutherland (2005, 2007) which study Irish firms in the period 1991-1998: considering Ireland as an export platform, Ruane and Sutherland claim that multinationals have less contact with indigenous firms, and this makes Irish firms less likely to export. Another contribution is offered by Swenson (2005): the competition with multinationals raises prices in product markets, forcing up as a consequence the average cost curves of domestic firms. Another possible explanation is that congestion of local infrastructure implies perhaps higher costs results.

3.2 Consequences of entry

The entry in foreign markets have several consequences and impacts on firms, most of them positive, such as the possibility to interact and to learn from foreign competitors and customers, to increase scale, to raise quality, to reduce costs, to gain innovation, to obtain additional information through the foreign markets, to raise efficiency, and so on.[8]

Because of the similarity between this strand of literature and the learning-by-exporting literature, in the previous paragraphs we have already showed most impacts and consequences of the entry of firms in export markets. Anyway a large strand of literature (e.g. Bernard and Jensen 2004, Hansson and Lundin 2004, Falvey et al. 2004) pointed out that the knowledge acquired by exporting firms is mainly due to an effect of "resource reallocation", which offers to firms the chance to raise productivity.

3.3 Similarity between determinants / consequences of entry and tests about self selection / learning by exporting

It is possible to find many similarities between the literature about determinants of entry and self-selection, as between consequences of entry and learning by exporting. Hence some research areas which have received a great attention in literature are in this case those which investigate the direction of causality between exporting and productivity, which test self selection vs learning by exporting, which compare new exporters and non-exporters, and which investigate the survival probability of exporters.

The main contribute to the literature about self-selection vs learning by exporting is offered by Bernard and Jensen (1999, 2004): they study US plants in the years 1983-1992 and, using both TFP (Total Factor Productivity) and LP (Labour Productivity) as a measure of productivity, they find

[8] Greenaway and Kneller (2007) find evidence that the probability of export market entry depends on spillovers associated with agglomeration, developing a study about manufacturing firms in the UK. Once entry has occurred the firms can reach additional productivity benefits; productivity improves for new export market entrants, but not for non-exporters with similar characteristics. Survival in the foreign market is then driven by size, TFP, and partly by industry characteristics.

evidence that self selection is stronger than learning by exporting. In fact they find that the difference in productivity growth is insignificant between exporters and non-exporters, while the pre-entry difference between new exporters and non-exporters is significant in the sense that new exporters were already better than non-exporters before entering in the export market.

After the framework by Bernard and Jensen (1999) the literature increased with tests about self selection vs learning by exporting applied to many other countries, so that the evidence obtained by Bernard and Jensen (1999) is now confirmed also in most countries out of the US. We have already mentioned some countries where we can find evidence of pre-entry differences e.g. China (Kraay 1999), Spain (Delgado et al. 2002), Italy (Castellani and Zanfei 2007), etc. There are also some exceptions, where pre-entry differences are demonstrated to be insignificant, while post-entry differences stronger: this happens in Sweden (Hanssohn and Lundin 2004, Greenaway Gullstrand and Kneller 2005), and in Slovenia (Damijan et al. 2007).

In most situations the tests about self selection vs learning by exporting are reduced to the comparison among established exporters, new exporters and non-exporters. Compared to non-exporters, future entrants have a higher productivity and better characteristics which give them the opportunity to start exporting. New entrants are different than established exporters, but after a short period in the export market they raise productivity becoming more similar to established exporters. One can test whether the surge in productivity is explained by the decision to become an exporter, or whether the productivity surge led to the export decision. This can be tested through instrumental variable ad matching techniques. Baldwin and Gu (2003) and Van Biesebroeck (2005) adopted instrumental variable approach using GMM, confirming that self selection is more important than learning by exporting.

Methods controlling for selection and methods without controlling for selection can be applied to compare new exporters and non-exporters. Studies about Germany (Bernard and Wagner, 1997) and UK (Girma, Greenaway and Kneller, 2004) are conducted: methods without controlling for selection show significant pre-entry differences in performance, while methods controlling for selection fail to report significant differences. Moreover Girma et al. (2005) show evidence of post-entry productivity changes in the UK, while Wagner (2002) do not report significant changes in Germany.

Other studies are conducted by Clerides et al. (1998) for Mexico Colombia and Morocco, Bigsten et al (2000) for some African countries, Blalock and Gertler (2004) for Indonesia, De Loecker (2004) for Slovenia, Greenaway Gullstrand and Knellar (2005) for Sweden. More studies report evidence for learning than fail to find such effects, although it is perhaps worth noting these tend to be studies that use matching.

This divergence may be explained by heterogeneity and timing.

Heterogeneity has been largely explored, even if the results obtained in literature are various. Delgado et al. (2002) and Fernandes and Isgut (2005) claim that learning is specific to young firms more than to old firms. Kraay (1999), Castellani (2002), Girma Görg and Strobl (2004), and Damijan et al. (2007) argue that learning strongly depends on the exposure to export markets: firms which are highly exposed to export markets learn more, and viceversa. Greenaway and Kneller (2003) claim that learning depends on past industry characteristics. Firms which have a high current exposure to foreign firms, through arms length trade and FDI, have lower productivity changes.

The timing issue has been explored by Lopez (2004) and Alvarez and Lopez (2005): they claim that a big part of the change in productivity in firms is due to the endogenous decision to start exporting. Because of the decision to start exporting, firms invest in new technologies, and "*learn to export*" rather than "*learn by exporting*". This point of view is confirmed first by Keller (2004), arguing that learning effects are not automatic but require important investments in domestic technologies, then by Blalock and Gertler (2004), Lopez (2004), Alvarez and Lopez (2005), and Van Biesebroeck (2005), testing empirically the same evidence. Moreover a recent paper by Bellone, Musso, Nesta, and Quere (2008) points at the fact that firms anticipate and prepare their entry on export markets by making specific investments[9], up to three years before the entry. Bellone, Musso, Nesta, and Quere (2008) also questions the possibility to disentangle properly selection from learning mechanisms on the basis of the timing of entry.

A different approach is adopted by Baldwin and Gu (2004) in their studies about Canadian firms: they combine micro data information with questionnaires about export behaviour, finding evidence about changes in scale and increased efficiency obtained through competition and learning.

[9] A large strand of literature in very recent years is dedicated at studying the strong connection existing between export and investment; we will focus on it in the last section of this chapter.

3.4 Determinants of Exit

Even if the literature about the determinants of exit is not so wide, as we could expect the determinants of exit from the export market are related to those of entry. It is less likely the larger, more productive and more human capital intensive the firm, and the lower the ratio of exports to domestic sales (Greenaway and Kneller 2003).

One of the common determinants of exit from the foreign market is related to the effect of exchange rate changes. Blalock and Roy (2005) study Indonesian firms and find evidence of export market exit in the presence of a depreciation of the Indonesia rupiah. Similar studies are conducted by Bernard and Jensen (2004), and Das et al. (2004): evidence shows how the exit from the foreign market is mainly due to the period of domestic currency depreciation while exports were likely to expand.

Other recurrent determinants of exit are import penetration, intra-industry trade, and finally industry sunk costs.

Greenaway and Kneller (2003) argue that, conditional on firm level variables, exit is more likely in industries with low sunk-costs, because re-entry is easier, and those with high levels of intra-industry trade.

3.5 Consequences of exit

The literature about consequences of exit is somewhat larger than that about determinants. Even in this case, as we could expect, the consequences of exit from the export market are related to entry. As with entry, self-selection is demonstrated to be important; firms seem to self-select out of the export market just as they do into them. Again as with entry, firms that exit from the export market are the weakest, but unlike entry there is little impact on productivity of this choice.

Aw et al. (2000), Baldwin and Gu (2003), and Girma et al. (2003) show that export quitters have lower productivity than firms that continue exports, whilst Bernard and Jensen (1999), Hansson and Lundin (2004) and Hahn (2004) argue that export quitters have no significant difference from non-exporters; in some cases productivity is demonstrated to be even lower in export quitters than in non-exporters.

The effect of exit on productivity produces mixed results. Bernard and Jensen (1999, 2004) study the US firms finding post-exit changes, not controlling for self selection. By contrast, Hansson and Lundin (2004) and Hahn (2004) find no obvious post-exit productivity changes, not conditioning for self-selection; their results are somewhat confirmed by Girma et al. (2003) and Blalock and Gertler (2004) who report similar results conditioning on self-selection.

4.

HOW TO ENTER IN FOREIGN MARKETS

Firms can work in foreign markets through export, through FDI, or as multinationals.

4.1 Export and FDI

Exporting and investing abroad to sell on the foreign market (FDI) are two alternative ways to enter in foreign markets. The decision of a firm whether to export or to do FDI reflects the models, already discussed in the previous paragraphs, by Brainard (1997) with the proximity-concentration trade-off theory, Melitz (2003), Helpman et al (2004), and all the subsequent literature. What emerges from this strand of literature is that the decision whether to do FDI or to export mainly depends on the involvement of the firm in the foreign markets.

Brainard (1997) argues that firms export to increase economies of scale, thus firms export when they may have cost advantages to concentration; on the other hand, when proximity to local markets is more important, firms prefer to establish foreign production facilities (theory of the proximity-concentration trade-off). This model (Brainard, 1997) finds empirical support: the share of exports is increasing in scale economies and decreasing in trade costs and foreign market size.

Helpman et al (2004) argue that the choice whether to do export or FDI is firstly based on considerations of market access. In general the involvement in foreign markets may be found through two different variables: size and costs. As the size of the foreign market increases, the involvement of the firm into foreign markets becomes more favourable. Similarly, as costs of exporting increase, the involvement of the firm in foreign markets is more favourable, but as costs of setting up foreign production grow it becomes less favourable.

A great relevance is given to sunk costs: as we previously explained, sunk costs are determined by costs in research, advertising, distribution networks, etc., but we did not focus our attention on which of these costs are more relevant when a firm is involved in FDI, and which are more relevant when a firm exports.

The relevance of sunk costs in firms involved in export and FDI are well defined in the model created by Helpman Melitz and Yeaple (2005). Whilst firms which export incur both fixed and variable costs, those involved in FDI incur only additional fixed costs. But the fixed costs of FDI are assumed to be higher than those of just exporting. Firms which export have to deal with high transportation costs, which are instead not required to firms involved in FDI. Firms which do FDI eliminate variable transportation costs, but on the other hand they are subject to high fixed costs which are the duplication of costs in establishing domestic production facilities as well as costs for building new production facilities or for the acquisition of existing firms; these FDI costs are supposed to be higher than those of just exporting. As a result, even if firms involved in FDI do not have to deal with high transportation costs, they are subject to higher fixed costs than firms which export.

All the recent theories about export and FDI (Helpman et al. 2004, Bernard et al. 2003, Melitz 2003, Helpman Melitz and Yeaple 2005) are based on considerations about heterogeneity: firms are heterogeneous in characteristics. Even if firms afford the same industry costs (within the same industry), they develop different choices about market entry. Heterogeneity of firms leads to self-selection: the most productive firms find it profitable to meet the higher costs associated with FDI, the not-so-productive firms prefer to serve foreign markets through exporting, while the least productive firms can only serve domestic markets.

Helpman Melitz and Yeaple (2005) confirm self-selection as the most important driver of results, but they observe that there are some firms that make different choices even if they have a similar productivity. According to the model, these firms lie in the “uncertainty” regions, where they may make different choices.

All the presented models can only be applied to single-product firms.

As we will understand in the next paragraphs about multinationals, multi-product firms face a more complex array of choices, considering that the decision whether to do export or FDI may be product-specific. If the conditions differ for each good produced, firms may find it profitable to produce some products domestically whilst some others may find it profitable to produce in the foreign countries, serving foreign markets through exports or FDI.

Considering also differences amongst foreign countries (e.g. market size, factor costs) it is straightforward that single-product firms may make different export/FDI choices to serve different countries or group of countries. This leads to the possibility that firms may be both multinationals and exporters (Girma Kneller and Pisu 2005).

4.2 Multinationals

4.2.1 Firms with a higher degree of internationalisation: multinationals

Another category of firms, after those involved in export and in FDI, should be considered: multinationals.

In microeconomics the first models of internationalization (e.g. Brainard 1997) refer to basic models of trade with representative firms; they help us to understand the differences between firms involved in export and in FDI, but they do not totally help us to detect firms which work as multinationals. To recognize them it is straightforward to consider both productivity and heterogeneity. Heterogeneity allows the choice to differ across firms within the same industry, and thus it allows us to understand which firms become multinationals and which just export. Similarly as explained before, heterogeneity and productivity imply self selection: the firms which are the least productive and heterogeneous do not work in foreign markets, but just in the domestic ones. Instead, among the self selected firms, those which are the most productive and heterogeneous become multinationals, whilst those whose productivity and heterogeneity fall in an upper-intermediate range are involved in FDI, and those whose productivity and heterogeneity fall in a lower-intermediate range are just involved in export.

The literature shows consistently that the most productive firms, contrary to the least productive, work as multinationals, relocating their production in foreign markets. But an exception is reported by Head and Ries (2003); they demonstrate that in particular situations, if the foreign country is small and if it offers cost advantages, the firms which locate abroad are the least productive, whilst the most productive firms only produce domestically. Thus the ordering of the productivity distribution between multinationals and non-multinationals can in particular cases be reversed.

Even if there is a large literature about multinationals vs non-multinationals, and also about exporters vs non-exporters, research which studies productivity levels in the comparison exporters vs multinationals is not so wide. The comparisons can be conducted through two different approaches. The first uses OLS to compare mean values (in some cases conditional on other firm and industry characteristics), and is introduced in a study about Japanese firms by Head and Ries (2003). The same method will be later applied also by Kimura and Kiyota (2004), and by Castellani and Zanfei (2007) in a research about Italian firms.

The second approach does not compare only the mean values, but it gives the possibility to compare the whole cumulative distribution of productivity for different types of firms. The method makes use of Kolmogorov-Smirnov tests of stochastic dominance, and the relative literature show evidences about UK (Girma et al. 2005), Ireland (Girma, Görg and Strobl 2004), and Germany (Arnold and Hussinger 2005, and Wagner 2005).

In the comparison multinationals vs exporters studied by these authors, the results show strong differences in productivity, with the only exception reported by Head and Rise (2003) where information about publicly listed firms is used.

The same authors (Head and Ries (2003), Kimura and Kiyota (2004), Castellani and Zanfei (2007), Girma et al. (2005), Girma, Görg and Strobl (2004), Arnold and Hussinger (2005), and Wagner (2005)) also compare exporters vs non-exporters, but in this case the results are somewhat mixed. In fact several of these studies report bias against finding significant productivity differences.

Additional studies report the importance of another variable: dispersion. Helpman et al (2004) study US firms adding the variable "industry dispersion" to the set of the traditional proximity-concentration variables. They consistently demonstrate that industries in which firm size is highly dispersed are associated with relatively more FDI than exports. Through their model Helpman et al (2004) predict FDI will be more common relative to exports, the greater the dispersion of productivity levels within industry.

4.2.2 Export decisions of multinationals

It is straightforward that the export decisions of multinationals should not be modelled as those of domestic firms, or as those of firms which are only partially involved in the foreign markets. The situation presented by multinationals is more complicated, because they can produce more products, export and be involved in FDI in more countries, and use multiple stages of production. Firms which produce a single product can decide to be involved in the foreign market just partially, but multinationals are more likely to export, and they export more intensively (Kneller and Pisu, 2004). The higher the international involvement of a firm is, the more it is difficult to understand and to model export decisions. But, even if it is not easy to model empirically export decisions of multinationals, some literature exists. The literature is divided into two strands, distinguished by the number of product lines that the firm is assumed to produce, thus by the degree of complication: export platform FDI and complementarity.

Export platform FDI (Motta and Norman 1996, Ekholm et al. 2003, Grossman et al. 2003) has the lower degree of complication and refers to exports of a single product line, where these are not to the home country. It involves the establishment of foreign production facilities and the allocation of the output to serve a third country. The degree of complication in export platform FDI can anyway become higher adding more countries and more stages of production to traditional theories.

The second and more complicated strand of literature is the so-called complementarity (Head and Rise, 2004). Complementarity refers to multi-product firms, to multiple stages of production, and to export and FDI flows from the home to foreign countries. Exports and FDI become positively correlated if there are horizontal or vertical complementarities across product lines.

4.2.3 Government policies attract multinationals

Governments often tend to attract multinationals through interventions and policies based on the conviction that FDI brings several benefits to host economies. Some examples of government interventions were registered in the US, Alabama, where the government paid $150,000 per employee to Mercedes for locating its new plant in the state in 1994, and in UK, where the British

Government provided $30,000 and $50,000 per employee to attract Samsung and Siemens respectively to the North East of England in the late 1990s.
Government interventions are due to the fact that there are direct positive effects arising from the presence of multinationals in the country: injection of additional capital in the host economy, increments in the demand for labour, R&D investments, and many other factors which have positive implications at both local and national level. Over all these direct implications the presence of multinationals even imply indirect effects arising from firm specific assets: improvements in management strategy, production techniques, know-how, and other benefits which multinationals are supposed to posses (Caves, 1996).

Spillovers deriving from FDI of multinationals can be classified into two categories: horizontal (intra-industry spillovers), and vertical (inter-industry spillovers). Blyde, Kugler and Stein (2005) claim that the main difference between the two categories is that the former are more likely to involve sector specific technical knowledge that would benefit competitors, hence there is greater incentive to prevent this type of spillovers. Vertical spillovers concern general rather than sector specific technological knowledge, and several authors underline how they can bring particular benefits to suppliers and buyers which foreign affiliates deal with. Dunning (1993) claim that foreign affiliates raise the quality of inputs produced and the productivity of suppliers. Whilst literature about vertical spillovers do find significant results, studies about horizontal spillovers are somewhat mixed. Some studies even report negative horizontal productivity spillovers (e.g. Aitken and Harrison, 1999). The lack of evidence of positive productivity spillovers may be due to the fact that foreign affiliates are successful in avoiding leakage of sector-specific technical knowledge on which their success is based (Kneller and Pisu 2005).

4.3 Export decisions

Motta and Norman (1996) show that export decisions can be largely dependent on trading blocks. They consider a single stage of production and start from an equilibrium between three identical countries which have exactly the same costs of production. The only difference between the three countries is that their trading costs are different, because two of them are allowed to raise external barriers against the third, and to enter a free trade agreement. The result is that after the initial equilibrium the third firm is encouraged to set up production facilities inside the free trade area, and to export to the other country in the block, demonstrating the importance of working inside trading blocks. As an additional result, neither of the inside countries choose export platform FDI as a strategy, because of identical costs.

This idea is not confirmed by Ekholm et al (2003), which argue that, when there are no vertical motives for FDI, the country inside the free trade area always has a motive to undertake export platform FDI. Ekholm et al (2003) consider multiple stages of production and start from two identical countries in the north and one in the south. One northern country and the southern are members of a free trade area; this makes optimal for that northern country to locate production in the south and export home (owing to the cost advantage from doing so). The other northern country has now three possible strategies, which depend on the size of the cost advantage to southern firms and on trade costs: no FDI, export platform FDI, and finally vertical FDI.

Grossman et al. (2003) develop the complex FDI model of Yeaple (2003) introducing an idea which has never been considered before: the importance of firm characteristics. In the same industry firms which are heterogeneous in productivity, even if they have the same costs of exporting and FDI, can make different choices. The results show the possibility for firms to undertake a great array of profitable FDI strategies; whilst the most productive firms choose complex strategies that involve a mix of FDI and exports, the least productive firms decide not to undertake FDI.

The literature also offers examples of complementarity, where firms have not only a single product line, but they are multi-product; hence new models are created to investigate export decisions of firms. An example is offered by Head and Ries (2004) who find that export and FDI are positively correlated if there are horizontal or vertical complementarities across product lines.

Some studies about export decisions of multinationals are offered by Lipsey and Weiss (1984), Swedenborg (1985), Lipsey et al. (2000), Head and Ries (2003), Kiyota and Urata (2005), Girma et al. (2005). The latter in particular study the UK case testing for export platform FDI; Girma et al. (2005) report that domestic firms which export tend to be selected and acquired by foreign multinationals. The authors also find consistent evidence of differences about future export intensity of the acquired firms: export intensity will decrease if the firm is inside EU, whilst it will increase if the firm is outside EU.

5.

EXPORT AND INVESTMENTS

5.1 The connection between export and investments in domestic market, R&D and innovation

Some studies are developed about the connection between export and investments.

Lopez (2004) investigates the correlation between export and domestic investments. He observes that in the period before the entry in the export market, domestic sales do not increase significantly, even if both investments and productivity rise. Lopez (2004) concludes that starting exports require important investments in technology for sales to foreign markets, but not for sales to domestic market.

More recently Mayneris (2010)[10], attempting to find the reasons why some firms are more/less productive, finds a key role in investment, a variable that as well as export consistently influences the productivity gains of firms. In fact Mayneris (2010) observes how, after a trade liberalization episode, initially more productive new exporters experience larger productivity gains than less productive ones in some countries, while it is the opposite in some others. Starting by this apparent contradiction, Mayneris (2010) builds a model that can reconcile both results: in fact Mayneris (2010) shows that the link between initial productivity and investment (and consequently growth) depends on the relative shape of the returns to investment and of the investment cost with respect to firm size. Increasing marginal returns to investment with respect to size generate higher incentives

[10] At the moment of writing this work, the framework of Mayneris (2010) is still a draft, so all the information given about his model and results cannot be considered definitive because they can be subsequently revised by the author

for initially big firms to invest. In the same way, if marginal cost of investment decreases with firm size, initially bigger firms will have more incentives to invest. Mayneris (2010) focuses then on the impact of credit constraints on firm-level investment and growth. Firms that decide to start exporting invest more due to higher expected sales. Conditioning on this decision, if profits are concave with respect to firm size/productivity, smaller firms will tend to invest more. However, if small firms face more severe credit constraints than big firms, this negative relationship between size and investment (and thus growth) is attenuated and can even be reversed. In a nutshell, Mayneris (2010) argues that when returns to investment decrease with productivity/size, less productive firms invest more and grow more than the others, conditioning on export decision, but if small firms face more severe credit constraints than big ones, this negative relationship could be attenuated or even reversed.

Other studies explore the connection between export and investments in R&D, whose correlation is demonstrated by a large number of papers; all the studies find that firms involved in export markets have higher levels of R&D. There are evidences about USA (Bernard and Jensen, 1995), UK (Bleaney and Wakelin 2002, and Roper and Love 2002), and Canada (Baldwin and Gu 2004).

All these mentioned papers study the correlation between export and R&D without establishing the direction of causality, which is instead explored by Aw et al. (2006), showing the evidence about Taiwan. Aw et al. (2006) use an econometric approach to study the evolution of productivity and R&D for exporters in Taiwanese electronics. They claim that investments in R&D are necessary for firms to benefit from their exposure to international markets. As a conclusion firms that both export and invest in R&D are the bests in terms of productivity growth (in comparison to firms that just export without investing in R&D, and to firms that just invest in R&D without exporting). They also find that firms that do not invest in R&D have lower productivity growth than those that just export.

Another contribute to this literature is offered by a very recent draft by Roberts and Xu (2010), that emphasize the simultaneity between R&D and export decisions, but we do not give a detailed description of their work because it is still a work in progress.

A parallel strand of literature investigates the connection between export and innovation. Most authors claim that R&D and innovation are strictly related, and consider R&D expenditures as an indirect measure of innovations (e.g. Hirsch and Bijaoui 1985, Schlegelmilch and Crook 1988, Kumar and Siddharthan 1994, Braunerhjelm 1996, Basile 2001). Only more recently a number of studies make use of explicit information about the actual innovations (e.g. Wakelin 1998, Bernard and Jensen 1999, Roper and Love 2002, Lachenmaier and Wossmann 2006, Cassiman and Martinez-Ros 2007). What emerges from all these studies is that a strong positive impact of innovations on exports exists. Only a little number of studies making use of the less preferable R&D expenditures as an indirect measure of innovations fail to find such a positive impact (see Cassiman and Martinez-Ros 2007 for a survey).

Even if it is difficult to generalize, most papers arrive to common conclusions: firstly larger and exporting firms are more inclined towards innovating, secondly innovations display a persistent pattern, and finally specific obstacles to innovations matter.

Hirsch and Bijauoi (1985) and Schlegelmilch and Crook (1988), in their first studies about effects of innovations on exports, arrived to mixed evidences using measures of innovation input.

Entorf, Krader, and Pohlmeier (1988)[11] estimated a simultaneous equation system of exports, innovation, and labor demand. They identified a positive impact of innovations, captured by an indicator variable, on exports, and even one of exports on innovations.

Wagner (1996) and Wakelin (1997, 1998) developed firm-level studies using more direct measures of innovation output (i.e. actual innovations). They both arrived to the conclusion of positive impact of innovation on exports, the former studying firms in Lower Saxony (Germany), the latter studying British firms.

Ebling and Janz (1999)[10] studied the impact of innovations, captured by a binary variable, on the extensive margin of exports in the service sector. Their results showed a positive impact of innovations on exports, but not vice versa.

Atkeson and Burstein (2007) and Constantini and Melitz (2008) conducted recent studies analyzing dynamic industry models to formalize linkages between firm-level productivity and the choices of both to export and to invest in R&D or adopt new technology. In their models productivity distinguishes heterogeneous firms and its evolution is endogenous and affected by innovation decisions at the firm level (apart from a stochastic component). As in the these last papers, in the very recent years research has slightly reoriented, focusing more on the heterogeneous response of firms to the internationalization of their commercial activities.

Bustos (2008) introduces a dichotomic technology choice in a monopolistic competition model with CES preferences and selection on export markets. In this framework, initially more productive new

[11] Entorf, Krader, and Pohlmeier (1988), and Ebling and Janz (1999) referred to endogenous innovations, in opposition to all the other authors that referred instead to exogenous innovations

exporters are the only ones to adopt the high technology after trade liberalization. Her analysis of technology adoption by Argentinean firms following the MERCOSUR regional trade agreement corroborates these theoretical findings.

Verhoogen (2008) finds similar theoretical and empirical results, studying the quality upgrading behaviour of Mexican plants after the 1994 peso crisis.

Quite surprisingly, Lileeva and Treer (2008) find completely opposite results: In Canada, following the Canada-United States free trade agreement, less productive new exporters grow more in terms of labour productivity, engage more in product innovation and have higher adoption rates of advanced manufacturing technologies. Lileeva and Treer rationalize these findings by assuming that firms are initially heterogeneous along two dimensions: initial productivity, and conditioning on initial productivity, returns to investment. Their model incorporates fixed investment cost and firm selection on export markets: initially less productive new exporters have thus necessarily higher returns to investment, since they were initially further from the export threshold.

Konings and Vandenbussche (2008) obtain similar results on French domestic firms that benefit from antidumping protection. Commercial protection is for domestic firms equivalent to trade liberalization for exporters in terms of business opportunities and incentives to invest. They show that initially less productive domestic firms have higher productivity growth during the protection period than more productive ones; these productivity gains are shown to be, at least partly, linked to higher investments over the same period.

These results suggest that the correlation between initial productivity and firm-level investment is different between countries. As far as inferences can be made out of four case studies, it seems that entry on export markets induces higher performance growth for initially better firms in developing countries, while the reverse is true in more developed ones.

After these papers, a number of frameworks in very recent years focus on investment, and in particular on the link between trade and credit constraints.

5.2 Literature about credit constraints and export

The growing body of literature focusing on the connection between credit constraints and export includes amongst others Manova (2008, 2010), Muuls (2008), Berman and Hericourt (2010) Bellone Musso Nesta and Schiavo (2010), Rajan and Zingales (1998), Demirguc-Kunt and Maksimovic (1998), Beck Demirguc-Kunt and Maksimovic (2005), Beck Demirguc-Kunt Laeven and Levine (2008), and Mayneris (2010).

Manova (2008) shows that equity market liberalizations increase exports disproportionately more in financially vulnerable sectors that require more outside finance or employ fewer collateralizable assets, suggesting that the presence of credit constraints affect negatively the intensive margin of trade at the country-industry level.

In Manova (2010), credit constraints are also shown to impact on the extensive margin of trade. By introducing firm-level credit constraints in a Melitz-type model of trade, she shows that financial frictions distort both the extensive and the intensive margins of trade, and that they can account for the prevalence of zero bilateral exports, the selection of firms into exporting, the number of products traded, and the extent of product turnover. Theoretical predictions are corroborated by the analysis of bilateral trade ows at the country-industry level.

While these two papers rely on semi-aggregated data, Muuls (2008) (on Belgian firm-level data), Berman and Hericourt (2010) (on firm-level data from 9 developing countries) and Bellone, Musso, Nesta, and Schiavo (2010) (on French firm-level data) show that financial constraints negatively impact on firm-level probability to become an exporter, but not on the quantity shipped by firms.

On the other hand, the negative impact of credit constraints on growth has been widely emphasized. Rajan and Zingales (1998) show that industrial sectors that are more dependent on external finance grow disproportionately faster in countries that are better financially developed. The underlying idea is that financial services, when efficient, allow an allocation of capital to the highest value use.

At a finer level, Demirguc-Kunt and Maksimovic (1998) show that in countries with better legal and financial systems, a greater proportion of firms use long-term external financing.

Moreover, Beck, Demirguc-Kunt, and Maksimovic (2005) and Beck, Demirguc-Kunt, Laeven, and Levine (2008) show that credit constraints are heterogeneous across firms, affecting disproportionately smaller firms. Small firms consequently benefit more, in terms of growth, from better financial institutions.

This innovative strand of literature surely includes also the framework by Mayneris (2010) that we have already described previously. In fact Mayneris (2010) builds a bridge between the literature on export participation and firm-level growth and the literature on credit constraints and firm-level growth.

CHAPTER 2

THE STATISTICAL APPROACH

INTRODUCTION

The objective of studying the relationship between internationalization and productivity of firms is handled by an innovative approach, which refers to a "conditional" comparison between groups (international INT vs domestic DOM) and to considerations about the whole distribution, as previously outlined.

The two groups (INT and DOM firms) are considered as an "a priori" partition of the firms population, and not as covariates of a regression model. Productivity is identified as the outcome variable, whilst groups are compared through the dummy which differentiates INT/DOM firms.

In the approach that we adopt, we consider the importance of characteristics, allowing in this way a "conditional" comparison between groups, but we also refer the comparison to the whole productivity distribution, instead of limiting the comparison to synthetic indicators.

In applying such an approach we anyway encounter a number of problems: how can we measure productivity and what is the best method of estimation for this purpose? What statistical-econometric tool is suitable to deal with our multi-object task? The next paragraphs attempt to answer to these questions, explaining how we can estimate productivity, and what statistical tool may be adopted to handle our problems. To answer to the first question we will refer to the Levinsohn and Petrin (2003) method of TFP estimation, while to answer to the second question we will refer to models that decompose differences, in particular to the "quantile decomposition" approaches described by Machado and Mata (2005) and Melly (2006), and to "impact evaluation" approaches[12] (Rosenbaum and Rubin, 1983) : even though they do not directly deal with the comparison of INT/DOM firms nor with firms productivity, but with the comparison of wage distributions referred to different racial groups (quantile decomposition) or with the comparison of treated and non-treated groups (impact evaluation), their procedure implies a decomposition which may allow us to decompose the productivity gap between INT and DOM groups of firms. The quantile decomposition approach, in particular, allows us to compare the difference between group distributions in a quantile regression model framework after controlling for differences in individual characteristics. Through this procedure we will be able to distinguish what proportion of the overall productivity gap is due to differences in firms characteristics, and what proportion is due to the international status over the whole distribution.

[12] The impact evaluation approaches that we adopt deal with the decomposition problem, but not with the consideration of the whole distribution

All these methods of decomposition of differences refer to the estimation of a quantity which is not directly observable, the so-called “counterfactual”.

Chapter 2, which is entirely dedicated to the statistical methodologies useful to study the connection between internationalization and productivity of firms, includes three paragraphs.

Paragraph 1 deals with productivity and the ways to measure it. In the first section we show how the current research about productivity estimation was born, thanks to old contributions as those of Von Thuenen (1820’s), and Flux (1913). The second section is centered on the most common measure of productivity: Total Factor Productivity (TFP). We will offer a review about the main methods of estimation of TFP, paying attention to the main problems that need to be addressed in estimation, i.e. simultaneity and selection bias. We will focus on OLS, fixed effect estimators, instrumental variables approaches, after which we will describe the seminal methods introduced by Olley and Pakes (1996) formerly, and Levinsohn and Petrin (2003) later. Because this last approach may be considered the best for our purpose, in the third section we will give an overview with an emphasis on the mechanics of the Levinsohn and Petrin (2003) estimator, with a subsection about estimation and another one about standard errors. To conclude the paragraph, we will refer to novel developments, such as Criscuolo and Martin (2005), Martin (2008), and the “factor share approach” discussed by Baily et. al (1992), Klette (1999), and Martin (2005).

Paragraph 2 is dedicated to the techniques used to decompose differences. Here we show the statistical approaches that we could use for our purpose (we will refer in particular to Melly, 2006). Most techniques originally refer to the comparison of wage distributions about different racial groups, but they can be easily applied to decompose the productivity gap between international and domestic groups of firms. In particular we start mentioning the pioneering work by Blinder (1973) and Oaxaca (1973): their decomposition technique provides a way of decomposing changes or differences in mean wages into a wage structure effect and a composition effect, and further dividing these two components into the contribution of each covariate. After Blinder’s (1973) and Oaxaca’s (1973) work, more flexible analytical methods, based on a quantile regression technique, have been used to decompose differences in log wages between two groups. To our knowledge Koenker and Bassett (1978) are the firsts to introduce “regression quantiles”, but important contributions are also offered by Juhn, Murphy and Pierce (1993), DiNardo Fortin and Lemieux (1996), Donald Green and Paarsch (2000), Gosling Machin and Meghir (2000), Barsky Bound Charles and Lupton (2002), Firpo Fortin and Lemieux (2007). Moreover Machado and Mata (2005) propose an innovative method to decompose the changes in the wage distribution over a period of time in several factors contributing to those changes. Melly (2006) presents an alternative estimator of counterfactual unconditional distributions that is similar to that of Machado and Mata (2005), but more efficient. We will pay a particular attention to this estimator since it will be used in our work; this is the reason why we will dedicate a section to the formal derivation of Melly's (2006) estimator. The final part of the paragraph is about the last developments: we will cite in particular the work of Chernozhukov, Fernandez-Val and Melly (2009) who do show a wide class of estimators (note that the Melly 2006's estimator may be interpreted as a special case of this class of estimators), and the work of Rothe (2010). All these approaches will allow us to compare the difference between group distributions in a quantile regression model framework after controlling for differences in individual characteristics. Through this procedure we will be able to distinguish what proportion of the overall productivity gap is due to differences in firms characteristics, and what proportion is due to the international status.

Paragraph 3 is dedicated to impact evaluation of reforms and interventions. Even though this topic is not directly connected to the micro-economic world of firms, it is anyway strictly connected to the world of the "counterfactual", the unknown quantity that must be estimated to allow a decomposition. Hence, we may say that it offers a valid alternative to the techniques of Machado and Mata (2005) or Melly (2006). Methods for impact evaluation may be divided into experimental and non-experimental. We will describe them focusing on the instrumental variables estimator, the Heckman selection estimator, the difference-in-differences estimator, and in particular on the matching estimator with use of propensity score. We identify in this last method the best instrument that we may use to handle our econometric problem. We will offer a general overview of this approach, also dedicating a section to the conditions needed, in particular assumptions (Conditional Independence Assumption, and Common Support Condition) and data requirements. In a last section we explain why we consider matching with propensity score the best method for our dataset.

1.

ESTIMATION OF PRODUCTIVITY

The first problem encountered in our approach deals with productivity, which relates output to inputs: what are the methods used in econometric literature to estimate productivity, and what is the best method for our purpose? This paragraph attempts to answer.

1.1 Birth of the literature about production function estimation

Economists have been relating output to inputs since at least the early 1800's. Much of the early applied work exploring this relationship was pioneered by agricultural economists like Von Thuenen[13], who collected data at his farm in the 1820's to measure the marginal product of inputs and the substitutability between inputs. Flux (1913), using one of the first available manufacturing censuses, painstakingly details relationships between inputs and output for manufacturing firms in England. Since then, economists have developed a large literature on production function estimation, in part because much of economic theory yields testable implications that are directly related to the production technology and optimizing behavior. In addition to estimating the marginal productivity of an input (such as fertilizer or skilled labor) and the elasticity of substitution between inputs, this literature has focused on obtaining measures of returns to scale and in particular on obtaining measures of productivity, calculated as the residuals to production functions (Levinsohn and Petrin, 2000).

[13] A colleague of Cournot

1.2 TFP and main methods of estimation

Total Factor Productivity (TFP) is the most common measure of productivity used in literature and may be defined as a variable which accounts for effects in total output not caused by inputs[14]. It is not directly observable but it may be deduced as a residual from a production function (for example from the Cobb-Douglas production function). Since Total Factor Productivity (TFP) is not directly observable, a variety of frameworks have been developed to infer a firm's underlying productivity level from observable input and output data. Obtaining consistent estimates of TFP (through the consistent estimates of the parameters of the production function) has frequently proven to be challenging, usually because important inputs are not observed, and thus are omitted. Indeed, since at least as early as Marschak and Andrews (1944), applied researchers have worried about the potential correlation between input levels and the unobserved firm-specific shocks in the estimation of production function parameters. The economics underlying this concern are intuitive. Firms that have a large positive productivity shock may respond by using more inputs.

There are various ways of estimating total factor productivity as the residual in the production function, ways which differ vastly in approach and, potentially, outcome[15]. In its simplest form, the production function may be estimated using ordinary least squares (OLS) or panel fixed effects techniques. Such estimates, however, may suffer from simultaneity and selection biases.

Simultaneity arises because productivity is known to the profit maximizing firms (but not to the econometrician) when they choose their input levels (Marschak and Andrews, 1944). Firms will increase their use of inputs as a result of positive productivity shocks. OLS estimation of production functions will yield biased parameter estimates (and, by implication, biased estimates of productivity) because it does not account for the unobserved productivity shocks. A fixed effect estimator would solve the simultaneity problem only if we are willing to assume that the unobserved firm-specific productivity is time-invariant. In fact the fixed effects solution has the unappealing feature of requiring a component of the productivity shock to be fixed over time.

Other methods, including instrumental variables approaches, have also been proposed to control for this bias when estimating the parameters of production functions. But valid instruments need to be

---

[14] The definition is deduced through that by Hornstein and Krusell (1996): "Growth in total-factor productivty (TFP) represents output growth not accounted for by the growth in inputs."

[15] For an overview of the history of this discussion, see Holtz-Eakin et al. (1988), Arellano and Bond (1991), Arellano and Bover (1995), Griliches and Mareisse (1998), Blundell and Bond (1998, 2000), Pavcnik (2002), Levinsohn and Petrin (2003), and Wooldridge (2005).

correlated with firm-level input choices and need to be orthogonal to the productivity shock. In many cases, there simply are no valid instruments that can overcome these requirements.

Another issue that one needs to address when estimating production function parameters is selection bias. Selection bias results from the relationship between productivity shocks and the probability of exit from the market. If a firm's profitability is positively related to its capital stock, then a firm with a larger capital stock is more likely to stay in the market despite a low productivity shock than a firm with a smaller capital stock because the firm with more capital can be expected to produce greater future profits. The negative correlation between capital stock and probability of exit for a given productivity shock will cause the coefficient on the capital variable to be biased downward unless we control for this effect.

Olley and Pakes (1996) proposed an innovative approach to address the simultaneity and selection problems while estimating the production function parameters and firm-level productivity. The former issue is addressed by using investment to proxy for an unobserved time-varying productivity shock, and the latter is addressed using survival probabilities. Olley and Pakes (1996) show how to use investment to control for correlation between input levels and the unobserved firm specific productivity process.

More recently, Levinsohn and Petrin (2003) prove that, like investment, intermediate inputs (those inputs which are typically subtracted out in a value-added production function) can also solve the simultaneity problem, and point to the evidence from firm-level datasets that suggest investment is very lumpy (that is, that there are substantial adjustment costs). If this is true, the investment proxy may not smoothly respond to the productivity shock, violating the consistency condition. Levinsohn and Petrin (2003) show the conditions under which intermediate inputs can also solve this simultaneity problem. Remarkably, in most applications (as in our), these inputs are not used beyond subtracting them from the gross-output number to get value added, so the approach comes at no additional cost in data or computation. Levinsohn and Petrin (2003) discuss the theoretical benefits of extending the proxy choice set in this direction and provide substantial empirical evidence that these benefits are important.

One benefit is strictly data driven. It turns out that the investment proxy is only valid for plants reporting nonzero investment [16]. Using intermediate input proxies instead of investment avoids truncating all the zero investment firms. In the dataset that we will use firms almost always report positive use of intermediate inputs. In our work Intermediate inputs which will also be subtracted from the gross-output number to obtain value added, will be calculated as the sum of Raw materials and consumables, Services, Expenses for leased assets to third parties, and Variation of Materials.

[16] This is due to an invertibility condition

Levinsohn and Petrin (2003) also argue that, to the extent that adjustment costs are an important issue, intermediate inputs may confer another benefit. If it is less costly to adjust the intermediate input, it may respond more fully to the entire productivity term than investment. For example, if adjustment costs lead to kink points in the investment demand function, plants may not respond fully to productivity shocks, and some correlation between the regressors and the error term can remain. Another nice feature of the intermediate input is that it provides a simple link between the estimation strategy and the economic theory, primarily because intermediate inputs are not typically state variables. Levinsohn and Petrin (2003) develop this link, showing the (mild) conditions that must hold if intermediate inputs are to be a valid proxy for the productivity shock. They suggest three specification tests for evaluating any proxy's performance. In addition, they derive the expected directions of bias on the OLS estimates relative to Levinsohn and Petrin's intermediate input approach when simultaneity exists.

1.3 Levinsohn and Petrin (2003)'s Productivity Estimation

In this section, we give an overview with an emphasis on the mechanics of the Levinsohn and Petrin (2003) estimator[17], which may be implemented in Stata through the routine "*levpet*".

Let us assume the production technology to be Cobb–Douglas

$$y_t = \beta_0 + \beta_l l_t + \beta_k k_t + \beta_m m_t + \omega_t + \eta_t$$

where y_t is the logarithm of the firm's output, measured as value added; l_t and m_t are respectively the logarithm of the freely variable inputs labor (or cost of labor) and the intermediate input, and k_t is the logarithm of the state variable capital (or cost of capital).

The error has two components: the transmitted productivity component given as ω_t and η_t, an error term that is uncorrelated with input choices. The key difference between ω_t and η_t is that the former is a state variable and, hence, impacts the firm's decision rules. It is not observed by the econometrician, and it can impact the choices of inputs, leading to the well-known simultaneity problem in production function estimation. Estimators ignoring this correlation between inputs and this unobservable factor (like OLS) will yield inconsistent results.

Demand for the intermediate input m_t is assumed to depend on the firm's state variables k_t and ω_t:

$$m_t = m_t(k_t, \omega_t)$$

[17] A more detailed exposition can be found in Levinsohn and Petrin (2003a)

Making mild assumptions about the firm's production technology, Levinsohn and Petrin (2003, appendix A) show that the demand function is monotonically increasing in ω_t.

This allows inversion of the intermediate demand function, so ω_t can be written as a function of k_t and m_t:

$$\omega_t = \omega_t(k_t, m_t)$$

The unobservable productivity term is now expressed solely as a function of two observed inputs.

A final identification restriction follows Olley and Pakes (1996). Levinsohn and Petrin (2003) assume that the productivity is governed by a first-order Markov process

$$\omega_t = E[\omega_t|\omega_{t-1}] + \xi_t$$

where ξ_t is an innovation to productivity that is uncorrelated with k_t, but not necessarily with l_t; this is part of the source of the simultaneity problem.

1.3.1 Estimation

Letting v_t represent value added we can write the production function as

$$\begin{aligned} v_t &= \beta_0 + \beta_l l_t + \beta_k k_t + \omega_t + \eta_t \\ &= \beta_l l_t + \phi_t(k_t, m_t) + \eta_t \end{aligned}$$

where

$$\phi_t(k_t, m_t) = \beta_0 + \beta_k k_t + \omega_t(k_t, m_t)$$

Substituting a third-order polynomial approximation in k_t and m_t in place of $\varphi_t(k_t,m_t)$, makes it possible to consistently estimate parameters of the value-added equation using OLS as

$$v_t = \delta_0 + \beta_l l_t + \sum_{i=0}^{3}\sum_{j=0}^{3-i} \delta_{ij} k_t^i m_t^j + \eta_t$$

where β_0 is not separately identified from the intercept of $\phi_t(k_t, m_t)$

Another restriction is necessary to separately identify β_0 from the intercept of $\phi_t(k_t, m_t)$.

This completes the first stage of the estimation routine from Levinsohn and Petrin (2003), from which an estimate of β_l and an estimate of φ_t (up to the intercept) are available.

The second stage of the routine identifies the coefficient β_k. It begins by computing the estimated value for φ_t using

$$\begin{aligned} \widehat{\phi}_t &= \widehat{v_t} - \widehat{\beta}_l l_t \\ &= \widehat{\delta}_0 + \sum_{i=0}^{3} \sum_{j=0}^{3-i} \widehat{\delta_{ij}} k_t^i m_t^j - \widehat{\beta}_l l_t \end{aligned}$$

For any candidate value $\widehat{\beta_k}$, we can compute (up to a scalar constant) a prediction for ω_t for all periods t using

$$\widehat{\omega}_t = \widehat{\phi}_t - \beta_k^* k_t$$

Using these values, a consistent (nonparametric) approximation to $E[\omega_t|\omega_t-1]$ is given by the predicted values from the regression

$$\widehat{\omega}_t = \gamma_0 + \gamma_1 \omega_{t-1} + \gamma_2 \omega_{t-1}^2 + \gamma_3 \omega_{t-1}^3 + \epsilon_t$$

which Levinsohn and Petrin (2003) call $E[\widehat{\omega_t|\omega_{t-1}}]$.

Given $\widehat{\beta}_l$, β_k^*, and $E[\widehat{\omega_t|\omega_{t-1}}]$, Levinsohn and Petrin (2003) write the sample residual of the production function as

$$\widehat{\eta_t + \xi_t} = v_t - \widehat{\beta}_l l_t - \beta_k^* k_t - E[\widehat{\omega_t|\omega_{t-1}}].$$

The estimate $\widehat{\beta_k}$ of β_k is defined as the solution to

$$\min_{\beta_k^*} \sum_t \left(v_t - \widehat{\beta}_l l_t - \beta_k^* k_t - E[\widehat{\omega_t|\omega_{t-1}}]\right)^2$$

The relative Stata command, *levpet*, that will be later used for the elaborations, uses a golden section search algorithm to minimize that function. A bootstrap approach is used to construct standard errors for $\widehat{\beta}_l$ and $\widehat{\beta_k}$.

1.3.2 Standard errors

The estimators described above involve two main stages of estimation. In each of these stages, a number of preliminary estimators are used. The covariance matrix of the final parameters must account for the sampling variation introduced by all of the estimators used in the two stages. Although deriving an analytic covariance matrix may be feasible, this calculation is not trivial. Instead, Levinsohn and Petrin (2003) substitute computational power for analytic difficulties, employing the bootstrap to estimate standard errors. Because Levinsohn and Petrin (2003) use panel data, they sample with replacement from firms, using the entire time series of observations for that firm in the bootstrapped sample when the firm's id number is randomly drawn. The variation in the point estimates across the bootstrapped samples provides an estimate for the standard errors of the original point estimates. Bootstrapping, when overidentifying restrictions are imposed, is slightly different. The sample moments computed using the original dataset will, in general, not equal zero, even though the population moments do (by assumption)[18].

As Horowitz (2001) and others have noted, this means that, for each of the bootstrapped samples, one must "recenter" the moment conditions by subtracting the values of the sample moments calculated using the original dataset (at the minimum).

The Levinsohn and Petrin (2003) Stata implementation accomplishes this by first performing the estimation on the original dataset, then storing the values of the sample moments using a series of global macros. Their value is then subtracted from the bootstrapped sample's moments when minimizing the objective function for that bootstrapped sample. This restores the consistency of the bootstrap approach in the construction of standard errors.

1.3 Last developments

Recently novel developments, such as Criscuolo and Martin (2005) and Martin (2008) among others, have proposed alternative methods to Olley and Pakes (1996) investment-based approach for estimating TFP or to Levinsohn and Petrin (2003) intermediate inputs-based approach.

Compared to Olley and Pakes (1996) the main innovation introduced by Criscuolo and Martin (2005) is to use profits instead of investment. This has a number of advantages. Firstly, one criticism of the Olley and Pakes framework may be that investment is a poor predictor of the fixed

[18] See Horowitz (2001) for an overview of the bootstrap and a discussion of the necessity of recentering.

component of the TFP[19]. Criscuolo and Martin (2005) approach, similarly to Levinsohn and Petrin (2003) who use intermediate inputs instead of investment, does not suffer from this problem. Secondly, differently from Levinsohn and Petrin (2003), Criscuolo and Martin (2005) identify all relevant parameters from a moment condition on capital without having to assume separability in intermediate inputs or relying on instrumental variable techniques. Also, Criscuolo and Martin (2005) do not require any assumptions on the substitutability between variable production factors.

Martin (2008) proposes a more refined framework for computing TFP by combining an improved version of the methodology of Olley and Pakes (1996) with the revenue production framework introduced by Klette and Griliches (1996). The Martin (2008) framework allows for imperfect competition, a flexible production technology, non constant returns to scale, it addresses the endogeneity of inputs problem in production function estimation and controls for measurement error in labour inputs. Martin (2008) provides Monte Carlo evidence showing that the suggested framework is more precise and robust to misspecification than competing approaches.

Summarizing: the approach proposed by Olley and Pakes (1996) could address the simultaneity problem by using investment to proxy for an unobserved time-varying productivity shock, but the use of investment itself created some problems: both the Markov assumption and the rather strange capital accumulation are essential to the Olley and Pakes (1996) model, and an additional problem arises with convex adjustment costs e.g. some fixed cost to install new machines. This might induce firms not to invest all in some periods. Levinsohn and Petrin (2003) answer to these issues using an indicator that is more likely to be correlated with short run variations such as intermediate inputs. This solves several problems of Olley and Pakes (1996) but leads however to new complications. Levinsohn and Petrin (2003) have to resort to instrumental variables using lagged levels of intermediate inputs. The procedure may be compared to the GMM approach where instead of using a fixed effects model to account for the more persistent aspects of TFP Levinsohn and Petrin (2003) use a particular function. Note that in this case we do not instrument differences with levels, so that no “poor instrument problem” will arise. The approach of Levinsohn and Petrin (2003) also imposes some conditions on the production function for monotonicity. Namely, labour must not react more elastically to TFP shocks than materials. Finally Martin (2008) proposes to use net revenue instead of intermediate inputs. This requires some assumptions on the firms output demand function, namely that there is constant markup pricing across firms. It might seem restrictive at first to complicate the problem with assumptions about output demand, rather than analyzing the problem entirely from the supply side. Of course even this last approach is not exempt from problems: indeed to make progress Martin (2008) must make assumptions on output demand.

[19] If firms are essentially in the steady state, and the capital stock in period t reflects the firm's knowledge about ω_{it} at t – 1, then the variation in investment reflects primarily adjustments to news about ω from period t.

As a conclusion both the approaches introduced by Levinsohn and Petrin (2003) and Martin (2008) still require assumption on the dynamic evolution of TFP and capital for identification, but they both are however more flexible than the Olley and Pakes (1996) method regarding these assumption.

An alternative and completely different approach to production function estimation is the "factor share approach" discussed for example by Baily et. al (1992), Klette (1999), and Martin (2005)[20]. This approach mainly differentiates from the others by the fact that it does not run any regression. The factor share approach offers a simple way to allow for a more flexible production technology than Cobb Douglas, and it is easy to implement, therefore it may be an important reference case. However the factor share approach requires strong assumptions, namely linear homogeneity (constant returns to scale), differentiability of the production function, and perfect competition[21].

[20] M. N. Baily, C. Hulten, and D. Campbell (1992), "Productivity dynamics in manufacturing plants", Brookings Papers: Microeconomics, pages 187–267
T. J. Klette (1999), "Market power, scale economies and productivity: estimates from a panel of establishment Data", Journal of Industrial Economics, XLVII(4):451–476, December 1999
R. Martin (2005), "Computing the true spread", CEP Discussion Paper 0692, year 2005

[21] Even if in particular cases the approach may be extended to relax these assumptions

2.

DECOMPOSITION OF DIFFERENCES

The statistical approaches that we may use in our work are presented in this paragraph. The instruments that we are going to present mainly refer to the comparison of wage distributions in different racial groups. The purpose of these approaches seems only apparently far from our interests, since all these methods imply a decomposition technique which may allow us to decompose the productivity gap between international and domestic groups of firms, even though the methods were not originally born with this specific purpose.

2.1 Blinder (1973) and Oaxaca (1973) Decomposition

To quantify the components of a wage gap between two groups Blinder (1973) and Oaxaca (1973) first developed a decomposition technique that detects the sources of the difference in the means. This approach is demonstrated to be particularly useful in explaining the differences in average log wages $\bar{Y}_t$ between two groups t = (0;1), e.g. between whites and blacks, between men and women[22], or more generally between a favoured group indexed with t=1 and a discriminated group indexed with t=0. But as talking about treated and non-treated groups is more appropriate instead of favoured and discriminated groups in our context, the international firms will form the treated group whereas the domestic firms will represent the non-treated group. By assuming that the expected value of Y conditionally on X is a linear function of X, E[Yt|T=t] can be estimated consistently via OLS by $\bar{X}_t\hat{\beta}_t$, where the groups' average characteristics $\bar{X}_t$ can be obtained by

[22] The same procedure can be applied to the productivity gap, instead of the wage gap, as well as groups such as domestic/international firms can be compared instead of whites/blacks (or men/women)

$\frac{1}{n_t}\sum_{i:T=t} X_i$ and the corresponding coefficients $\hat{\beta}_t$ are resulting from the regressions of Yt on Xt . Then, since $\bar{Y}_t = \bar{X}_t\hat{\beta}_t$, the difference between $\bar{Y}_1$ and $\bar{Y}_0$ can be written as

$$\bar{Y}_1 - \bar{Y}_0 = \bar{X}_1\hat{\beta}_1 - \bar{X}_0\hat{\beta}_0 \qquad (1)$$

Addition and simultaneous subtraction of the counterfactual $\bar{X}_0\hat{\beta}_1$ gives

$$\bar{Y}_1 - \bar{Y}_0 = \bar{X}_1\hat{\beta}_1(-\bar{X}_0\hat{\beta}_1 + \bar{X}_0\hat{\beta}_1) - \bar{X}_0\hat{\beta}_0 \qquad (2)$$

Then, the Oaxaca-Blinder Decomposition (1973) is given by

$$\bar{Y}_1 - \bar{Y}_0 = \underbrace{(\bar{X}_1\hat{\beta}_1 - \bar{X}_0\hat{\beta}_1)}_{\text{characteristics}} + \underbrace{(\bar{X}_0\hat{\beta}_1 - \bar{X}_0\hat{\beta}_0)}_{\text{coefficients}} \qquad (3)$$

Alternatively, the difference between $\bar{Y}_0$ and $\bar{Y}_1$ can be decomposed in an analogous way as

$$\bar{Y}_0 - \bar{Y}_1 = \underbrace{(\bar{X}_0\hat{\beta}_0 - \bar{X}_1\hat{\beta}_0)}_{\text{characteristics}} + \underbrace{(\bar{X}_1\hat{\beta}_0 - \bar{X}_1\hat{\beta}_1)}_{\text{coefficients}} \qquad (4)$$

By introducing the counterfactuals $\bar{X}_0\hat{\beta}_1$ as well as $\bar{X}_1\hat{\beta}_0$, Blinder (1973) and Oaxaca (1973) showed that not only the characteristics of individuals but also the simple belonging to a group determines the magnitude of the resulting wages. In the literature these two effects are commonly known as “characteristics effect”, given by the term in the first bracket of equations (3) and (4), and “coefficients effect” given by the term in the second bracket of (3) and (4). The “characteristics effect” reflects the justified wage differential between both groups due to different productivities depending on the groups’ characteristics whereas the rest of the observable wage gap is contributable to the “coefficients effect” which honors the simple belonging to the treated group or punishes the simple belonging to the non-treated group. This becomes clearer if equation (3) is rewritten as

$$\bar{Y_1} - \bar{Y_0} = \underbrace{(\bar{X}_1 - \bar{X}_0)\hat{\beta}_1}_{\text{characteristics}} + \underbrace{(\hat{\beta}_1 - \hat{\beta}_0)\bar{X}_0}_{\text{coefficients}}$$

2.2 After Blinder (1973) and Oaxaca (1973)

After Blinder's (1973) and Oaxaca's (1973) work, more flexible analytical methods, based on a quantile regression technique, have been used to decompose differences in log wages between two groups. These innovative methods overcome the large waste of information because they do not only consider means of variables, but they also analyze differences at various quantiles of the distributions. Another important feature of quantile regression is its robustness against outliers.

This strand of literature has rapidly increased in the last decades and distributional issues have attracted a lot of attention (in particular in labor economics). One important factor behind the resurgence of interest for distributional issues is the large increase in wage inequality in the United States and in several other countries. But this rapid increment in literature is mainly carried on by labor economics with works that look at wages differentials between subgroups that goes beyond simple mean comparisons[23]. More generally, there is an increasing interest for distributional impacts of various programs or interventions (we will discuss about this strand of literature - impact evaluation of interventions - in the following paragraph). In all these cases, the key question of economic interest is which factors account for changes (or differences) in distributions. For example, did wage inequality increase because education or other wage setting factors became more unequally distributed, or because the return to these factors changed over time? In response to these important questions, a number of decomposition procedures have been suggested to untangle the sources of changes or differences in wage distributions.

Popular methods used in the wage inequality literature include the seminal work of Koenker and Bassett (1978), the "plug-in" procedure of Juhn, Murphy and Pierce (1993), the reweighting procedure of DiNardo, Fortin and Lemieux (1996, 1998), the approach of Gosling, Machin, and Meghir (2000), the method of Donald, Green and Paarsch (2000), the procedure of Barsky, Bound, Charles, and Lupton (2002), and, more recently, the quantile-based decomposition method of

[23] For example several papers such as Albrecht, Björklund and Vroman (2003) look at whether the gender gap is larger in the upper tail than in the lower tail of the wage distribution. Albrecht, Bjorklund, and Vroman (2003) analyze the effect of observable differences between male and female workers on the gender gap at various quantiles of the Swedish wage distribution, finding evidence for a "glass ceiling" effect.

Machado and Mata (2005), that of Melly (2005, 2006), and the works of Chernozhukov, Fernandez-Val and Melly (2009) and Rothe (2010).

Most papers have proposed procedures to decompose intra-group differences in general distributional features into a composition and a structural effect, thus generalizing the Oaxaca-Blinder technique. A further advantage of these methods is that they also allow for more complex nonlinear relationships between the covariates and the outcome variable, whereas the Oaxaca-Blinder procedure critically relies on the linear model.

2.2.1 Koenker and Bassett (1978) and quantile regression

Koenker and Bassett (1978) introduce “regression quantiles” to estimate conditional quantiles of a response variable Y given regressors X. They complement Laplace's (1818) median regression (least absolute deviation estimator) and generalize the ordinary sample quantiles to the regression setting.

Assuming linearity between the quantiles of the dependent variable Y and the covariates X, then the τ^{th} conditional quantile of Y is given by

$$Q_Y(\tau|X) = X\beta(\tau), \qquad \forall\tau \in (0,1) \tag{5}$$

Koenker and Bassett (1978) solve by minimizing in $\beta(\tau)$

$$\hat{\beta}(\tau) = \min_{\beta\in\mathbb{R}^K} n^{-1}\left[\sum_{i}^{n} \rho_\tau(Y_i - X_i\beta)\right], \qquad (i = 1, \ldots, n) \tag{6}$$

where the check function ρ_τ weights asymmetrically the residuals u_i so that

$$\rho_\tau(u_i) = \begin{cases} \tau u_i & \texttt{for } u_i \geq 0 \\ (\tau - 1)u_i & \texttt{for } u_i < 0 \end{cases} \tag{7}$$

The method proposed by Koenker and Bassett (1978) is extremely simple if compared to the modern instruments we know nowadays (e.g. Machado Mata 2005, Melly 2006, etc.), but it introduced great innovations if compared to the main model used in that period: the linear regression model. While the linear regression model specifies the change in the conditional mean of the dependent variable associated with a change in the covariates, the quantile regression model specifies changes in the conditional quantile. Since any quantile can be used, it is for the first time possible to model any predetermined position of the distribution. Thus, researchers can choose positions that are tailored to their specific inquiries. Poverty studies concern the low-income population; for example, the bottom 11.3% of the population lived in poverty in 2000 (U.S. Census Bureau, 2001). Tax-policy studies concern the rich, for example, the top 4% of the population (Shapiro & Friedman, 2001).

Conditional-quantile models offer the flexibility to focus on these population segments whereas conditional-mean models do not. Since multiple quantiles can be modeled, it is possible to achieve a more complete understanding of how the response distribution is affected by predictors, including information about shape change. A set of equally spaced conditional quantiles (e.g., every 5% or 1% of the population) can characterize the shape of the conditional distribution in addition to its central location. The ability to model shape change provides a significant methodological leap forward in social research on inequality (Hao and Naiman, 2007).

A decade and a half after Koenker and Bassett first introduced quantile regression, empirical applications of quantile regression started to grow rapidly. Empirical researchers took advantage of quantile regression's ability to examine the impact of predictor variables on the response distribution.

Two of the earliest empirical papers by economists (Buchinsky, 1994; Chamberlain, 1994) provided practical examples of how to apply quantile regression to the study of wages. Quantile regression allowed them to examine the entire conditional distribution of wages and to determine if the returns to schooling, experience and the effects of union membership differed across wage quantiles. The use of quantile regression to analyze wages increased and expanded to address additional topics such as changes in wage distribution (Machado & Mata, 2005; Melly, 2006), wage distributions within specific industries (Budd & McCall, 2001), wage gaps between whites and minorities (Chay & Honore, 1998) and between men and women (Fortin & Lemieux, 1998), educational attainment and wage inequality (Lemieux, 2006), and the intergenerational transfer of earnings (Eide & Showalter, 1999). The use of quantile regression also expanded to address the quality of schooling (Bedi & Edwards, 2002; Eide, Showalter, & Sims, 2002) and demographics' impact on infant birth weight (Abreveya, 2001). Quantile regression also spread to other fields, notably sociology (Hao, 2005, 2006a, 2006b), ecology and environmental sciences (Cade, Terrell, & Schroeder, 1999; Scharf, Juanes, & Sutherland, 1989), and medicine and public health (Austin et al., 2005; Wei et al., 2006).

2.2.2 Juhn, Murphy and Pierce (1993)

The Juhn, Murphy and Pierce (1993) decomposition method[24] simply extends the Oaxaca decomposition by taking into account the residual distribution. The innovation in the Juhn, Murphy, and Pierce (1993) extension is to decompose the “unexplained” or “residual” portion of the gap from the Oaxaca decomposition into two distinct components. In other words, this technique maintains the component of Oaxaca’s decomposition which captures the differential due to the differences in characteristics (endowments), but the second component (due to differences in coefficients) is divided into two effects: one reflects the differential attributable to the respective mean percentile ranks, and the second part captures the differential due to dispersion. As other future frameworks, this by Juhn, Murphy and Pierce (1993) has the advantage to allow us to look at how differences in characteristics affect the entire distribution, and not just the variance. We can identify how differences in the distribution affect other inequality measures, or how the effects on inequality are different below and above the mean. This method, however, requires assumptions about the functional form of the conditional expected function, which may be avoided making use of a decomposition technique introduced by DiNardo, Fortin, Lemieux (1996)[25] that we present in the next section. In addition to that one of the main problems of the decomposition proposed by Juhn, Murphy and Pierce (1993) is that it does not account for heteroscedasticity. In the original paper, Juhn, Murphy and Pierce (1993) formally allow for the distribution of residuals to depend on the covariates, but they do not explain how to do it empirically and give no details. Most other applications of this decomposition do not condition on the covariates. If the error term is really independent and normally distributed, this procedure is efficient. However, if the location model is inappropriate, this decomposition can produce misleading results. The whole conditional distribution of wages and not only the first two moments can depend on the covariates. This lack of flexibility have motivated new estimators which are less restrictive.

[24] see Appendix A2 of Sierminska, Frick and Grabka (2008) for more details and for a formal presentation of the Juhn, Murphy and Pierce (1993) procedure

[25] Used for example by Cobb-Clark and Hildebrand (2006)

2.2.3 DiNardo, Fortin and Lemieux (1996)

The reweighting procedure proposed by DiNardo, Fortin and Lemieux (1996) and extended by Lemieux (2002) provides consistent estimates of the wage structure and composition effects for any distributional statistic of interest under a set of assumptions. The counterfactual weights are chosen in a way that makes the distribution of skills constant across time. The main advantage of this approach is the lack of restrictions on covariates effects and density shapes. Analyzing the effect of various factors on changes in the distribution of wages, the procedure yields a visually clear representation of precisely where in the distribution various factors have their greatest impact. It could be applied to any other problem where it is useful to know what part of the distribution is affected. What this type of procedure does not provide, however, is a general way of further dividing up the composition effect into the contribution of the respective covariates. One exception is the special case of a dummy covariate: in this particular case the authors propose a conditional reweighting procedure based on a sequential conditioning argument. This approach is not easily extended to covariates other than dummy variables. Furthermore, with many dummy covariates, one would have to compute a large number of conditional reweighting factors to account for the contribution of each covariate. Contrary as other procedures, this approach does not encompass the Oaxaca-Blinder procedure as a special case.

2.2.4 Donald, Green and Paarsch (2000)

Donald, Green and Paarsch (2000) construct a flexible-functional-form estimator of cumulative distribution functions for non-negative random variables which admits large numbers of covariates. The estimator adopts and extends techniques from the spell-duration literature for estimating hazard functions to distribution functions for wages, earnings, and income. Donald, Green and Paarsch (2000) show an application of these methods to investigate sources of wage inequality for full-time male workers between Canada and the United States.

2.2.5 Gosling, Machin, and Meghir (2000)

Gosling, Machin, and Meghir (2000) provide a simple characterization of the way that the distribution of wages has evolved, describing the distribution of wages using a set of quantiles. They show that each of the chosen quantiles of the distribution of wages can be modeled as an additive function of cohort, life-cycle and time effects (constructed to average to zero within the sample period).

For the estimation of the quantiles Gosling, Machin, and Meghir (2000) use a two step estimation procedure. The first step regressions include general functions of cohort and age and interactions thereof. These quantiles are estimated using the SLAD (Smoothed Least Absolute Deviations) estimator suggested by Horowitz (1998); since the criterion function is differentiable, optimization can take place using standard gradient type methods. In the second step the predictions from the first stage regression are regressed, using weighted least squares where the weights are based on the covariance matrix of the predicted quantiles from the first stage. The estimated conditional quantiles are then used to construct the entire conditional distribution of wages, allowing to construct counterfactual distributions of wages and consider the within and between group contributions to changes in dispersion.

The approach later proposed by Machado and Mata (2005) will be similar to this of Gosling, Machin, and Meghir (2000), using quantile regressions to model the conditional distribution. However, the two approaches differ in the method used to construct the unconditional distributions implied by the conditional model. Having estimated the conditional quantiles, Gosling Machin and Meghir (2000) derive the unconditional quantiles through the following procedure: first, the conditional quantile function is inverted to produce the conditional cumulative distribution function; then, the conditional cumulative distribution function is averaged with respect to the empirical distribution of the covariates to yield the unconditional distribution function[26]; finally, the unconditional cumulative distribution function is inverted again to produce the respective quantiles.

The alternative proposed by Machado and Mata (2005) estimates unconditional wage distributions from a given conditional quantiles function, and may be considered a simpler approach. We will anyway discuss and give a deeper presentation of the approach of Machado and Mata (2005) in the next sections.

[26] If, in this second step of the Gosling Machin and Meghir (2000)'s procedure, one takes a distribution different from the empirical, the resulting marginal could be used for the counterfactual comparisons constructed in the framework by Machado and Mata (2005). This makes these two different frameworks so similar for the approach they propose.

2.2.6 Barsky, Bound, Charles, and Lupton (2002)

Noting and exposing some reasons why the Blinder–Oaxaca Decomposition (1973) may yield misleading results, Barsky, Bound, Charles, and Lupton (2002) suggest a solution that both provides a more reliable answer to the original problem and affords a richer examination of the sources of intergroup differences in the variable of interest. The conventional application of the Blinder–Oaxaca Decomposition requires a parametric assumption about the form of the conditional expectation function. Furthermore, it often uses estimates based on that functional form to extrapolate outside the range of the observed explanatory variables. Barsky, Bound, Charles, and Lupton (2002) show that misspecification of the conditional expectation function is likely to result in nontrivial errors in inference regarding the portion attributable to differences in the distribution of explanatory variables, a problem compounded by the computation of conditional expectations outside the observed range of the conditioning variables. They propose a nonparametric alternative to the Blinder–Oaxaca Decomposition that reweights the empirical distribution of the outcome variable using weights that equalize the empirical distributions of the explanatory variable. Barsky, Bound, Charles, and Lupton (2002) apply their approach to the role of earnings in explaining the black-white wealth difference.

2.2.7 Firpo, Fortin, and Lemieux (2007)

Firpo, Fortin, and Lemieux (2007) consider a procedure based on their Recentered Influence Function regression approach (Firpo, Fortin, and Lemieux, 2009b) to determine the contribution of a covariate to the composition effect. The approach allows us to decompose intra-group differences in general distributional features into a composition and a structure effect, and to further divide the composition effect into the contribution of each covariate. The method is made up by two stages. In the first stage, distributional changes are divided into a wage structure effect and a composition effect using a reweighting method. The reweighting allows them to estimate directly these two components of the decomposition without having to estimate a structural wage setting model. In the second stage, the two components are further divided into the contribution of each explanatory variable using Recentered Influence Function regressions. These regressions estimate directly the impact of the explanatory variables on the distributional statistic of interest. The method is mainly based on the Recentered Influence Function regression technique in Firpo, Fortin, and Lemieux

(2009b), which provides a clever way of computing partial effects of changes in the distribution covariates on some functional of the unconditional distribution of the outcome variable of interest. The approach can be seen as a generalization of the Oaxaca-Blinder procedure to general distributional statistics. The Firpo, Fortin, and Lemieux (2007) paper indeed generalizes the popular Oaxaca-Blinder decomposition method by extending the decomposition to any distributional measure besides the mean and by allowing for a much more flexible wage setting model. Firpo, Fortin, and Lemieux (2007) apply their work to analyze how polarization of U.S. male wages that took place between the late 1980s and the mid 2000s was affected by factors such as de-unionization, education, occupations and industry changes. As a drawback due to a particular parametric restriction inherent in the approach, the method will always suggest that a covariate does not contribute to the full composition effect if its mean does not change. This property can lead to unexpected biases even for simple data generating processes like linear models.

2.3 Machado and Mata (2005)

Machado and Mata (2005) propose a method that extends the traditional Oaxaca-Blinder decomposition of effects on mean wages (Oaxaca 1973, Blinder 1973) to the entire wage distribution, through an estimator of counterfactual unconditional wage distributions based on quantile regressions. Basing their procedure on the estimation of the marginal density function of wages in a given year implied by counterfactual distributions of some or all the observed attributes, Machado and Mata (2005) propose an innovative method to decompose the changes in the wage distribution over a period of time in several factors contributing to those changes. The method is based on the estimation of marginal wage distributions consistent with a conditional distribution estimated by quantile regression as well as with any hypothesized distribution for the covariates. Comparing the marginal distributions implied by different distributions for the covariates, one is then able to perform counterfactual exercises. The approach enables the identification of the sources of the increased wage inequality observed in most countries. Specifically, it decomposes the changes in the wage distribution over a period of time into several factors contributing to those changes, namely by discriminating between changes in the characteristics of the working population and changes in the returns to these characteristics.

In principle, considering the regression model for the τ^{th} quantile of Y conditional on the covariates X (τ goes from 0 to 1), that we have already presented while introducing the model of Koenker and Bassett (1978):

$$Q_\tau(Y|X) = X\beta(\tau) \tag{5}$$

running such quantile regressions for all possible quantiles should describe the whole conditional distribution of wages. One can then use the $\beta(\tau)$ estimated for one group to construct a counterfactual distribution for the other group, and then use this counterfactual distribution to compute the overall composition and wage structure effect[27].

Furthermore, if one plugs in the $\beta(\tau)$ pertaining to a single covariate only, it is then possible to estimate the contribution of this covariate to the wage structure effect, as in the Oaxaca-Blinder decomposition.

Indeed the difference of the θ^{th} unconditional quantile between two groups' distributions can be decomposed according to Blinder and Oaxaca (1973) as

$$\hat{F}_{Y_1}^{-1}(\theta|T=1) - \hat{F}_{Y_0}^{-1}(\theta|T=0) = \underbrace{\hat{F}_{Y_1}^{-1}(\theta|T=1) - \hat{F}_{Y_1}^{-1}(\theta|T=0)}_{\text{characteristics}} + \underbrace{\hat{F}_{Y_1}^{-1}(\theta|T=0) - \hat{F}_{Y_0}^{-1}(\theta|T=0)}_{\text{coefficients}} \tag{8}$$

or inversely as

$$\hat{F}_{Y_0}^{-1}(\theta|T=0) - \hat{F}_{Y_1}^{-1}(\theta|T=1) = \underbrace{\hat{F}_{Y_0}^{-1}(\theta|T=0) - \hat{F}_{Y_0}^{-1}(\theta|T=1)}_{\text{characteristics}} + \underbrace{\hat{F}_{Y_0}^{-1}(\theta|T=1) - \hat{F}_{Y_1}^{-1}(\theta|T=1)}_{\text{coefficients}} \tag{9}$$

where $\hat{F}_{Y_t}^{-1}(\theta|T=t)$ denotes the θ^{th} unconditional quantile of group t's wage. Again, the unconditional counterfactual quantiles $\hat{F}_{Y_1}^{-1}(\theta|T=0)$ as well as $\hat{F}_{Y_0}^{-1}(\theta|T=1)$ in the terms on the right hand side of (8) and (9) are needed to detect the mentioned effects at any unconditional

[27] As Machado and Mata (2005), also Albrecht et al. (2003) use conditional quantile regressions to construct counterfactual unconditional wage distribution. Machado and Mata (2005) draw *n* numbers at random to choose the quantiles, estimate the conditional quantile coefficients from the first group, then for each quantile draw a random sample from the covariates of the alternate group and generate the counterfactual wages. Albrecht et al. (2003) modify this procedure by choosing quantiles 1 through 99, and by taking 100 draws for each quantile.

quantile. Even though an appropriate method of consistently estimating the variance is not presented in Machado's and Mata's (2005) pioneer work, several authors make use of this decomposition technique in their applications (Albrecht et al. 2001, Albrecht et al. 2004, Kohn 2006). But more importantly, Melly (2006) shows that their estimator only yields good MSE-properties if the number of quantile regression coefficients *m* is large or goes at best to infinity since its variance vanishes. Indeed if $m \to \infty,$ the MSE of Machado's and Mata's estimator (MSE_{MM}) reduces to the bias that does not depend on *m*. So if a dataset is relatively small, one can increase *m* without losing too much computation time. However, many applications are based on large or even huge datasets for which choosing the right *m* is a sensitive question since estimation time depends crucially on *m* and *n.* The situation even worsens if the standard errors need to be bootstrapped in order to obtain reliable inference statistics.

Machado and Mata (2005) apply their methodology to the analysis of the changes in the distribution of wages in Portugal from 1986 to 1995. For instance, they estimate the wage density that would have prevailed in 1995 if education had been distributed as in 1986 and the other covariates as in 1995. By comparing it with the actual marginal distribution in 1995, they can isolate the contribution of changes in education to the observed changes in the distribution of wages. The counterfactual nature of the exercise requires the estimation of the wage distribution conditional on the variables of interest. They accomplish this first step by means of quantile regressions, that is, by estimating models for the quantiles of the conditional wage distribution. Quantile regressions capture the impact of changes in covariates upon a conditional wage distribution, in very much the same way that mean regression measures the impact of changes in covariates upon the mean of the conditional wage distribution. However, Machado and Mata (2005) go beyond a mere conditional model. Indeed, a conditional distribution does not reflect the variability of the covariates in the population. In other words, it is the distribution that would prevail if all workers had the same observed characteristics. The second step of their approach, and its major methodological contribution, is thus to marginalize the conditional distribution estimated in the previous step using different scenarios for the distribution of workers' attributes. The basic gist of their approach resembles DiNardo Fortin and Lemieux (1996) in that their methodology also estimates counterfactual densities and yields a decomposition of the factors that explain the changes in the marginal distribution of wages. However, whilst the approach proposed by DiNardo, Fortin and Lemieux (1996) is essentially based on nonparametric weighted-kernel methods, Machado and Mata (2005) depart from the DiNardo et al. (2002) methodology on two counts: the cornerstone of the method of Machado and Mata (2005) is a parametric model for the quantiles of the conditional distribution, and Machado and Mata (2005) resort to resampling procedures to obtain a marginal distribution consistent with both the conditional model and the covariate densities. It is difficult to say which approach is better. Certainly, resorting to a parametric model is necessarily restrictive. Yet, this weakness buys some additional information. Indeed, it enables the identification, in the changes in the wage density that are not explained by the changes in the distribution of the covariates (coined 'unexplained' in DiNardo Fortin and Lemieux, 1996), of the part that is due to the changes in the quantile regression coefficients. Consequently, the Machado and Mata (2005) approach decomposes the change in distribution of wages that is explained by the statistical model in part due to changes in the distribution of workers' attributes and, also, due to changes in returns to those attributes.

There are, however, a number of drawbacks to the procedure of Machado and Mata (2005). First and foremost, the approach does not provide a way of dividing up the composition effect into the contribution of each single covariate. To solve this problem however Machado and Mata (2005) suggest to use an unconditional reweighting procedure to compute the contribution of a covariate to the composition effect. Second, the procedure is computationally difficult to implement as it involves estimating a large number of quantile regressions, and conducting large scale simulations. Third, as in the case of the mean, the decomposition is only consistent if the right functional form is used for quantiles. Since the right functional form has to be chosen for each and every quantiles, making sure that the specification is correct is a very difficult empirical exercise. Furthermore, if the correct functional form is not linear, it is then difficult to compute the contribution of each covariate to the wage structure effect, since there is no longer a single $\beta(\theta)$ coefficient associated to a given covariate, as in the linear case.

2.4 Melly (2006)

2.4.1 The method

As previously explained Melly (2006) shows that the estimator proposed by Machado and Mata (2005) has good MSE-properties only if *m*, the number of quantile regression coefficients, is sufficiently large; indeed the MSE reduces to the bias that does not depend on *m* if *m* goes to infinity[28]. So if a dataset is relatively small, one can increase *m* without losing too much computation time. Most applications are anyway based on large datasets for which choosing the right *m* is a sensitive question since estimation time depends crucially on *m* and *n.* The situation even worsens if the standard errors need to be bootstrapped in order to obtain reliable inference statistics. For this reason Melly (2006) presents an alternative estimator of counterfactual unconditional distributions that copes with this challenge. On the one hand he shows that Machado's and Mata's estimator is numerically equivalent to his own estimator if *m* goes to infinity. On the other hand, and most importantly for applications using large datasets, he proves that the MSE of his estimator (MSE_{Melly}) does, in contrast to MSE_{MM}, not depend on *m*, and thus $MSE_{Melly} \leq MSE_{MM}$. Indeed a comparison of the Mean Squared Errors (MSE) of both estimators displayed as Relative MSE $\frac{MSE_{MM}}{MSE_{Melly}}$ shows that for $m = n = 400$ the MSE_{MM} is more than twice as large as the MSE_{Melly} and respectively for $m = 1000$ still 1.5-times as large. In a nutshell, decomposition analysis based on quantile regression technique using large datasets become feasible (Heinbach and Spindler, 2007). Since Melly's estimator of counterfactual unconditional distributions is the basis of our application, the formal proceeding is briefly presented in the following section.

[28] The Machado and Mata (2005) estimator and the estimator proposed by Melly (2006) will be numerically identical if the number of simulations used in the Machado and Mata procedure goes to infinity. Hence, Melly's (2006) asymptotic results apply also to Machado's and Mata's (2005) estimator and, since it is never possible to compute an infinite number of simulations, Melly's (2006) estimator actually uses more information.

2.4.2 Formal derivation of Melly's (2006) estimator

After having estimated all conditional quantiles of Y given X by linear quantile regression, Melly (2006) executes several calculation steps in order to obtain the unconditional quantiles of interest. For this purpose, he first estimates the conditional distribution of Y_t given X_i at q[29] by

$$\hat{F}_{Y_t}(q|X_i) = \int_0^1 1(X_i\hat{\beta}_t(\tau) \leq q)\, d\tau = \sum_{j=1}^{J}(\tau_j - \tau_{j-1})1(X_i\hat{\beta}_t(\tau_j) \leq q) \tag{10}$$

since is not possible to simply integrate the conditional quantile function for lack of monotonicity. The magnitude of the expression $(\tau_j - \tau_{j-1})$ [30] in equation (10) diminishes by nature with growing *m*.

Having once estimated the conditional distribution of Y_t , the unconditional distribution functions can easily be computed in a second step by

$$\hat{F}_{Y_t}(q|T = t) = \frac{1}{n_t}\sum_{i:T_i=t}\hat{F}_{Y_t}(q|X_i) \tag{11}$$

Then, the unconditional quantiles $\hat{q}_t(\theta)$ as well as the unconditional counterfactual quantiles $\hat{q}_{c_1}(\theta)$ based on $X_0\hat{\beta}_1(\tau)$, and $\hat{q}_{c_0}(\theta)$ based on $X_1\hat{\beta}_0(\tau)$, are given by equations (12), (13) and (14) respectively:

$$\hat{q}_t(\theta) = \inf\{q : \frac{1}{n_t}\sum_{i:T_i=t}\hat{F}_{Y_t}(q|X_i) \geq \theta\} \tag{12}$$

$$\hat{q}_{c_1}(\theta) = \inf\{q : \frac{1}{n_0}\sum_{i:T_i=0}\hat{F}_{Y_1}(q|X_i) \geq \theta\} \tag{13}$$

$$\hat{q}_{c_0}(\theta) = \inf\{q : \frac{1}{n_1}\sum_{i:T_i=1}\hat{F}_{Y_0}(q|X_i) \geq \theta\} \tag{14}$$

[29] *q* is used as auxiliary tool for the estimation of the conditional distribution function

[30] As we can consider that m equals 100 in our application, we can assume $(\tau_j - \tau_{j-1})$ to take a constant value of 0.01

Finally, the difference between the θ^{th} unconditional quantiles of both groups can be decomposed in analogy to Blinder (1973) and Oaxaca (1973) as

$$\hat{q}_1(\theta) - \hat{q}_0(\theta) = \underbrace{(\hat{q}_1(\theta) - \hat{q}_{c_1}(\theta))}_{\text{characteristics}} + \underbrace{(\hat{q}_{c_1}(\theta) - \hat{q}_0(\theta))}_{\text{coefficients}} \tag{15}$$

or alternatively as

$$\hat{q}_0(\theta) - \hat{q}_1(\theta) = \underbrace{(\hat{q}_0(\theta) - \hat{q}_{c_0}(\theta))}_{\text{characteristics}} + \underbrace{(\hat{q}_{c_0}(\theta) - \hat{q}_1(\theta))}_{\text{coefficients}} \tag{16}$$

Following the approaches of Juhn, Murphy and Pierce (1993) and Lemieux (2006) amongst others, Melly (2006) also considers an extension to his approach which underlines the effect of residuals. The obtained result is a decomposition of the differences in distribution into three factors: coefficients, characteristics and residuals. In fact, since there is a theoretical interest in several applications as in our to identify these three sources of differences in distribution, Melly (2006) shows how to extend his decomposition in order to separate the effects of coefficients into the effects of median coefficients and residuals. This additional step was developed and applied independently by Melly (2005) and Autor, Katz and Kearney (2005).

Melly (2006) takes the median as a measure of central tendency of a distribution. While the effect of characteristics can be estimated similarly to as Melly (2006) did in the previous two-factors decomposition, the τ^{th} quantile of the residuals distribution conditionally on X_i can consistently be estimated by $X_i\left(\hat{\beta}_t(\tau) - \hat{\beta}_t(0.5)\right)$, so that it is possible to separate the effect of coefficients from the effect of residuals.

Defining

$$\hat{\beta}_{m1,r0}(\tau_j) = \left(\hat{\beta}_1(0.5) + \hat{\beta}_0(\tau_j) - \hat{\beta}_0(0.5)\right),$$

Melly (2006) estimates the distribution that would prevail if the median return to characteristics were the median return in the treated group but the residuals were distributed as in the control group by $\hat{q}_{m1,r0}(\theta)$

The difference between $\hat{q}_{m1,r0}(\theta)$ and $\hat{q}_c(\theta)$ is due to differences in coefficients since characteristics and residuals are kept at the same level, while the difference between $\hat{q}_1(\theta)$ and $\hat{q}_{m1,r0}(\theta)$ is due to residuals.

Finally, the difference between the θ^{th} unconditional quantiles of the groups can be decomposed into three factors as

$$\hat{q}_1(\theta) - \hat{q}_0(\theta) = \underbrace{(\hat{q}_1(\theta) - \hat{q}_{m1,r0}(\theta))}_{\textbf{residuals}} + \underbrace{(\hat{q}_{m1,r0}(\theta) - \hat{q}_{c_1}(\theta))}_{\text{coefficients}} + \underbrace{(\hat{q}_{c_1}(\theta) - \hat{q}_0(\theta))}_{\text{characteristics}}$$

or alternatively as

$$\hat{q}_1(\theta) - \hat{q}_0(\theta) = \underbrace{(\hat{q}_0(\theta) - \hat{q}_{m1,r0}(\theta))}_{\textbf{residuals}} + \underbrace{(\hat{q}_{m1,r0}(\theta) - \hat{q}_{c_0}(\theta))}_{\text{coefficients}} + \underbrace{(\hat{q}_{c_0}(\theta) - \hat{q}_0(\theta))}_{\text{characteristics}}$$

If compared to the previous literature (e.g. Juhn, Murphy and Pierce 1993, DiNardo, Fortin and Lemieux (1996), Machado and Mata 2005, amongst others), since the basic idea of estimating the conditional distribution function by parametric quantile regression and integrating it to obtain the unconditional distribution is not new, Melly (2006) proposes an estimator that is faster than others to compute. Melly (2006) also derives the asymptotic distribution of his parametric estimator and uses the asymptotic results to propose an analytical estimator of its variance. Bootstrapping the results is time consuming and sometimes simply impossible if the number of observations is very large. The Monte-Carlo simulations show that the asymptotic results are useful approximations in finite sample. The analytical standard errors perform better than the bootstrap standard errors in the simulations. In addition to that, Melly (2006) proposes a new estimator based on nonparametric quantile regression that does not require any parametric restriction. *n* consistency, asymptotic normality and achievement of the semiparametric efficiency bounds are proven. This procedure can be seen as the quantile equivalent of the estimator proposed by Heckman, Ichimura and Todd (1998) for the mean. A consistent procedure for the estimation of the variance is also presented always in Melly (2006). The estimators perform well in Monte Carlo simulations. Melly (2006) also shows an application of his estimators to issues concerning racial discrimination in the USA.

Finally Melly provides a routine, "*rqdeco3*", which allows us to implement his 3-factors decomposition in Stata[31].

2.5 Last developments

2.5.1 Chernozhukov, Fernandez-Val and Melly (2009)

Chernozhukov, Fernandez-Val and Melly (2009) develop procedures for performing inference in regression models about how potential policy interventions affect the entire marginal distribution of an outcome of interest. These policy interventions consist of either changes in the distribution of covariates related to the outcome holding the conditional distribution of the outcome given covariates fixed, or changes in the conditional distribution of the outcome given covariates holding the marginal distribution of the covariates fixed. Under either of these assumptions, Chernozhukov Fernandez-Val and Melly (2009) obtain uniformly consistent estimates and functional central limit theorems for the counterfactual and status quo marginal distributions of the outcome as well as other function-valued effects of the policy, including for example the effects of the policy on the marginal distribution function, quantile function, and other related functionals. Chernozhukov, Fernandez-Val and Melly (2009) construct simultaneous confidence sets for these functions; these sets take into account the sampling variation in the estimation of the relationship between the outcome and covariates. The procedures rely on the main regression approaches for modeling and estimating conditional distributions, focusing especially on classical, quantile, duration, and distribution regressions. The procedures are general and accommodate both simple unitary changes in the values of a given covariate as well as changes in the distribution of the covariates or the conditional distribution of the outcome given covariates of general form. Chernozhukov, Fernandez-Val and Melly (2009) show an application by applying these procedures to examine the effects of labor market institutions on the U.S. wage distribution.

2.5.2 Rothe (2010)

Rothe (2010) proposes a method to evaluate the effect of a counterfactual change in the marginal distribution of a single covariate on the unconditional distribution of an outcome variable of

[31] Even a routine for implementing the same model in R is provided by Melly: in this case the routine is "*rqdeco*"

interest. Rothe (2010) shows that such effects are point identified under general conditions if the covariate affected by the counterfactual change is continuously distributed, but typically only partially identified if its distribution is discrete. For the latter case, Rothe (2010) derives informative bounds making use of the available information. Rothe (2010) allows us to study the question how the marginal distribution of an outcome variable of interest is affected by a counterfactual change in marginal distribution of a single covariate, holding “everything else” constant. Rothe (2010) argues that the natural way to render this ambiguous notion of a ceteris paribus change in one of the marginals of a multivariate distribution more precise is by imposing a rank invariance condition: that is, constructing the counterfactual change in the marginal distribution of one of the covariates in such a way that the joint distribution of ranks remains unaffected. This notion is equivalent to requiring the copula function of the joint distribution of the covariates to remain constant. Having defined the objects of interest, Rothe (2010) studies identification of the effect induced by the counterfactual experiment. Under certain restrictions on the support of the covariates and a conditional exogeneity condition, the distribution of the outcome variable after the counterfactual experiment can be obtained through a two-step procedure.

In the first step, a new counterfactual covariate distribution is constructed by connecting the joint distribution of ranks to the new set of marginal cumulative distribution functions. In the second step, the conditional cumulative distribution function of the outcome variable given the covariates is integrated with respect to the counterfactual covariate distribution, yielding the cumulative distribution function of what we call the counterfactual outcome distribution. This new cumulative distribution function can then be used to directly calculate counterfactual population parameters of interest, such as moments, quantiles or inequality measures. Rothe (2010) calls the difference between these quantities and their counterparts in the distribution of observed outcomes a partial composition effect. This can be interpreted as the contribution of the respective covariate to the full composition effect. A particular complication arises when analyzing the contribution of a discrete covariate, such as a binary indicator of union membership. In this case, the rank of an individual in the respective marginal distribution is not uniquely determined by the data. Rothe (2010) shows that as a result the effect of a change in the marginal distribution of a discrete covariate is typically only partially identified; the data generating process reveals some nontrivial information about this effect, but does not allow for an exact quantification. Rothe (2010) also shows how to implement the identification results in practice, proposing flexible procedures to either directly estimate the object of interest under point identification, or the respective identified set in the partially identified case. These turn out to be easy to implement, requiring only the estimation of a univariate cumulative distribution function, a univariate quantile, and a conditional cumulative distribution function. Depending on the requirements of the application, estimates can be fully nonparametric or make use of certain parametric restrictions. Adapting arguments in Chernozhukov, Fernandez-Val and Melly (2009), Rothe (2010) derives a complete asymptotic theory for his estimators in both cases, and shows how to conduct valid inference based on resampling methods like to bootstrap.

Rothe (2010) applies his results to study the role of changes in the distribution of labor force characteristics for the polarization of the US labor market. In particular Rothe (2010) investigates the role of composition effects in the polarization of the US labor market. The results suggest that changes in the distribution of education and labor market experience had only a minor impact, whilst de-unionization is shown to have contributed to the increase in both overall wage inequality and inequality at the top-end of the wage distribution.

3.

IMPACT EVALUATION

One of the main problems to afford in our statistical approach is the construction of the counterfactual. In the previous paragraph we have shown one way to construct it (e.g. Melly, 2006). Here we show an alternative: in fact a large strand of literature about the construction of the counterfactual is that related to the impact evaluation of a policy program, reform or intervention.

"Impact evaluations compare the outcomes of a program against a counterfactual that shows what would have happened to beneficiaries without the program. Unlike other forms of evaluation, they permit the attribution of observed changes in outcomes to the program being evaluated by following experimental [non-experimental] and quasi-experimental designs" (definition by the World Bank's DIME Initiative).

The counterfactual analysis developed by the impact evaluation literature refers to "a comparison between what actually happened and what would have happened in the absence of the intervention" (White, 2006). In the strand of literature about impact evaluation of policies, it is straightforward to identify the causal effect of a treatment on a certain outcome of interest relative to an unobserved counterfactual for the population of interest. It would be for example interesting to measure the outcome of an individual that would be treated whilst it still had not been treated, or the outcome of an individual that would not have been treated whilst it had been treated. In these cases we encounter a missing data problem, due to the fact that we want to measure something which is unobservable. Thus the only solution is to construct an appropriate counterfactual that can be used to solve the missing data problem. Individuals are usually identified by some observable variables (e.g. age, gender, education), thus the problem is to evaluate the impact of the policy on each type of individual. The construction of the counterfactual is the central issue to address, because we are interested in observing the outcome variable for those individuals who were not involved in the policy intervention, if they were involved, or vice versa[32].

3.1 Experimental and non-experimental methods

Evaluation methods proposed in literature, which implicitly enable us to construct the counterfactual, may be divided in experimental (or non-observational) and non-experimental (or observational) methods.

3.1.1 Experimental methods

Experimental methods randomly exclude from the treatment (intervention) a group of individuals, which will be later used as a benchmark to evaluate the policy intervention, providing in this way the best estimate of the counterfactual outcome. Experimental methods provide a consistent impact estimator when there are no randomization bias (the experiment does not distort the environment in its absence), no substitution bias, no drop-out bias, and no displacement effects[33] (Diaz and Handa, 2004).

In social more than economic research, experimental methods are largely used; even if they may require high costs to be implemented and may raise ethical concerns regarding the denial of treatment for randomized-out units, when they are correctly applied they can produce the best estimates of program impacts. Experimental methods eliminate the evaluation problem, by providing the correct missing counterfactual. They rule out self-selection as a source of bias. Let us consider a randomly chosen group of individuals that undergo an intervention (treatment). The group is randomly chosen, it is therefore independent of the treatment effect because within the group assignment to treatment is independent of any outcome variable. Thus the chosen treated group and the remaining non-treated group must be statistically identical in all but the treatment status. All this is valid only if there is a total absence of side-effects, drop-out bias, or substitution bias, which would alter the characteristics of the sample and would make the process non-random, invalidating the final results. One way to understand how similar or different are the two created groups is to conduct a comparison among the observable characteristics of the two distinct groups: the more the treated group differs from the non-treated group, the more the assignment may be

[32] See Ashenfelter (1978), Ashenfelter and Card (1985), and Heckman and Robb (1985, 1986) for a review about the development of the analysis about the evaluation problem in the labour market area.

[33] See Heckman and Smith (1995) and Heckman, Lalonde and Smith (1999) for a discussion

considered non-random, and the more the experimental method is unable to identify the treatment effect. In addition to that the researcher should pay attention to other factors which may influence the behaviour of experiment participants, e.g. the experiment itself when selecting treated and non-treated. As a contrary, if the assignment can be considered random, if no other factors influence the individuals, and if we can hypothesise that the treatment impact on the individual *i*, α_i , is the same for all the *i* individuals (homogeneous treatment effects), the treatment effects may be identified through a simple comparison of mean outcomes $\hat{\alpha}$:

$$\hat{\alpha} = \overline{Y}_t^{(1)} - \overline{Y}_t^{(0)}, \quad t > k$$

where $\overline{Y}_t^{(1)}$ is the treated mean outcome at time *t* after the intervention (occurred at time *k*), and $\overline{Y}_t^{(0)}$ is the non-treated mean outcome at time *t* after *k*.

This "pure randomised social experiment" method of evaluation can be considered "the most convincing method of evaluation since there is a control (or comparison) group which is a randomised subset of the eligible population" (Blundell and Dias, 2000). Even if many frameworks underline the advantages of experimental methods, mainly regarding the fact that such methods can easily overcome the missing data problem (see for a review Cochrane and Rubin 1973, Fisher 1951, Bassi 1983 1984, Hausman and Wise 1985, Card and Robins 1998), experimental methods are quite rare in economics, and report several drawbacks: the most important drawback is that their implementation usually requires very high costs. In addition to that they require the control group to be totally unaffected by the intervention, and it is difficult to use such methods in ex-ante analysis of policy reform proposals, because experimental methods are not amenable to extrapolation. Finally, they have much to offer in enhancing our knowledge of the possible impact of policy reforms. A comparison of results between experimental and non-experimental data may help to assess appropriate methods where experimental data are not available. For example the studies developed by LaLonde (1986), Heckman Smith and Clements (1997) and Heckman Ichimura and Todd (1997) use experimental data to assess the reliability of comparison groups used in the evaluation of training programmes (Blundell and Dias, 2000).

3.1.2 Non-experimental methods

Non–experimental methods have the advantage to estimate program impacts overcoming selection bias problems, but they may anyway produce biases because of self-selection and environment differences such as differences in data sources, quality and in local labour markets. Moreover non-experimental methods require non-testable hypothesis, although most of them may be tenable in actual data. Anyway non-experimental data require special care and are difficult to deal with. The main problem which may be encountered when handling non-experimental data is the so-called “econometric selection problem”: let us consider a dataset which is divided into a treatment group from a given programme and a comparison group from the whole population. Because of the econometric selection problem we cannot be sure about the absence of differences in unobservables that are related to the programme participation. In addition to that even the choice of the comparison group is not easy, because it should respect the strict comparability rules based on observable information which are sometimes impossible to guarantee.

The estimator $\hat{\alpha} = \overline{Y}_t^{(1)} - \overline{Y}_t^{(0)}, \quad t > k$ proposed for experimental data is not suitable anymore in this context, because it leads to a “non-identification problem”. In fact the estimator would identify for large samples

$$E(\hat{\alpha}) = \alpha + [E(U_{it} \mid d_i = 1) - E(U_{it} \mid d_i = 0)]$$

where *d* is a dummy variable such that $d_i = 1$ if the individual *i* has participated in the programme while $d_i = 0$ if the individual *i* has not participated in the programme, and U_{it} is the error term of mean zero and uncorrelated with the exogenous variables[34].

Unless $E(U_{it} d_i) \neq 0$, the expectation $E(\hat{\alpha})$ will differ from α, thus we will need alternative estimators.

The methods exploitable for this purpose depend on the underlying model, on the parameters of interest, and on the type of information available to the researcher. For example models which may

[34] For clearness consider the outcome

$$Y_{it} = X_{it}\beta + d_i\alpha + U_{it} \quad \text{if} \quad t > k$$

$$Y_{it} = X_{it}\beta + U_{it} \quad \text{if} \quad t \leq k$$

where *Y* is the outcome variable which depend on a set of exogenous variables *X* and on the already described dummy variable *d*, whereas β identify the set of parameters which define the relationship between the exogenous variables *X* and the dependent variable *Y*.

make use of longitudinal or repeated cross-section data use datasets which are richer of information than models which just adopt single cross-section data, thus they can support less restrictive estimators. The nature of the data hides a number of issues: the dataset must contain information on individuals both before and after their participation in the programme, and the two types of information must be homogeneous, e.g. whether the researcher uses questionnaires to have data about the potential comparison groups, or uses other already existing survey datasets, the data must be severely homogeneous.

The appropriate method also depends on the nature of the questions to be answered, which can refer to the overall impact, the effect of the treatment on the treated, or the extrapolation to a new policy reform, and on the nature of the programme, which can be small scale or global, local national or international.

Following Blundell and Dias (2000) we will focus on four non-experimental methods: the instrument variables (IV) estimator, the two-step Heckman selection estimator, the difference-in-differences estimator, and the matching estimator. Whilst the first two estimators can be used when only single cross-section data is available, the third requires data in longitudinal or repeated cross-section format, and the last may be used with either cross-section or longitudinal data

3.2 The Instrumental Variables (IV) Estimator

The instrumental variables method use one or more variables which are unrelated to the outcome, but related to the programme participation decision, so that the impact of the treatment is identified avoiding selection problems. Indeed the method satisfies the required randomness since the instrumental variables are assumed to be unrelated to the outcome, except through the participation to the programme.

The IV method is quite simple to implement but has some drawbacks, mainly due to the strong initial hypothesis on which the method is based. Firstly the IV method refers only to homogeneous frameworks, whilst it is unable to evaluate the impact of training in a heterogeneous framework because this method would violate in the heterogeneous framework case the predefined fundamental assumptions. Secondly the choice of the instrumental variable(s) is not easy: in fact it is difficult to find a variable which satisfies all the assumptions required to identify α, in particular the simultaneous assumptions of “participation determination” and “non-influence on the outcome participation”. One possible solution to this drawback is, in presence of longitudinal data, to consider lagged values of some determinant variables, although these lagged values will be strongly correlated to their future values included in the outcome regression. Thus, even if this solution is the most common used in this situation, it does not totally solve the problem.

To shortly illustrate the method consider the "homogeneous treatment effect" case, and consider one or more regressors Z^* exclusive to the decision rule.
Z^* is/are the instrumental variable(s), used by the method as a source of exogenous variation to approximate randomised trials.

Z^* must respect three assumptions:

1. Z^* determines the programme participation, thus it has a significant coefficient in the decision rule equation;

2. a transformation of Z^*, $g(Z^*)$, which is uncorrelated with the error U given the exogenous variables X, exists;

3. Z^* is not completely determined by the exogenous variables X.

The instrumental variable(s) Z^* provide variation correlated with the participation decision, but do not directly affect the potential outcomes from the treatment. The method is simply applied by substituting the treatment indicator with $g(Z^*)$ and running a regression. An alternative procedure makes use of both Z^* and X to predict d, and builds a new variable $\hat{d}$ which substitutes d in the regression.

3.3 The Heckman Selection Estimator

The Heckman selection estimator was initially introduced by Heckman in 1979, and later integrated into the "impact evaluation" literature by Heckman and Robb in 1985. It is a two-step procedure which makes use of an explicit model of the selection process to control for the part of the participation decision correlated with the error term in the outcome equation. Unlike many other methods, the Heckman selection estimator accounts for selection on the "unobservables", not on the "observables". The procedure relies on an exclusion restriction, which requires a variable that determines participation in the programme but not the outcome of the programme itself (Blundell and Dias 2000). The Heckman selection method requires initial assumptions which are even stronger than those required by the IV method, but it leads to more consistent results. Considering the "homogeneous treatment effect" case, the main hypothesis requires the existence of one or more additional regressors in the decision rule. The regressor must have a significant coefficient in the decision rule equation, and it must be uncorrelated with the error term V_i[35]. Finally, the joint density of the distribution of the error terms U_{it} and V_i, say $h_i(U_{it}, V_i)$, must be known or consistently estimable. The estimator will focus on the error term of the outcome equation, directly finding which part of it is correlated with the participation dummy variable. The procedure is made up by two steps:

1. we estimate the part of the error term U_{it} correlated with the participation dummy variable d_i;

2. we include the estimate found in point 1. into the outcome equation, and we estimate the effect of the programme.

The remaining part of the error term included in the outcome equation is by construction uncorrelated with the participation decision.

Considering the "heterogeneous treatment effect" case, the treatment effect would vary across individuals; the Heckman selection method would be in this case unable to identify the effect of training if individuals were randomly assigned to the treatment, and only the treatment-on-the-treated impact would be identifiable.

3.4 The Difference-In-Differences Estimator

The difference-in-differences method considers the policy programme as an experiment and seek a comparison group which shows properties similar to the control group in the properly designed experimental context. The name of the method "difference-in-differences" is due to the fact that this

[35] The error term V_i pertains to a parameterization of the participation decision which refers to IN_i, an index calculated for each individual i, which depends on a set of variables Z and of parameters γ:

$$IN_i = Z_i\gamma + V_i$$

When the index IN_i rises above zero, enrolment occurs, so the dummy variable d_i is equal to 1. Otherwise, if $IN_i < 0$, the dummy variable d_i becomes null.

approach compares the differences in average behaviour before and after the programme for the eligible group with the before and after contrast for the comparison group (Blundell and Dias, 2000). As for other methods, even this suffers the drawback of imposing some initial assumptions, which in this case create difficulties in the choice of a proper comparison group: in particular the difference-in-differences method relies on the two critical hypothesis of “common time effects across groups” and of “no composition changes within each group” (we will not focus on these assumptions here, but they are widely described in the framework by Blundell, Duncan and Meghir, 1998). It is anyway important to note that the requirement of longitudinal or repeated cross-section data instead of simple cross-section data enables the method to have less restrictive initial hypothesis in comparison to other approaches such as the IV or the Heckman selection methods. For example, unlike both the IV and the Heckman selection methods, the difference-in-differences method does not require any exclusion restriction, because there is no need for any regressor in the decision rule. In addition to that even the outcome equation does not have to be specified as long as the treatment impact enters additively. When the assumptions are respected, the difference-in-differences approach enables the removal of unobservable individual effects and common macro effects. In this way the procedure measures the average effect of the treatment on the treated, by recovering the average effect of the programme on the “treated” individuals who entered into the programme. As a beginning of the procedure the researcher must identify one pre-programme set of observations and one post-programme set of observations. The difference-in-differences estimator measures the excess outcome growth for the treated compared with the non-treated.

Abstracting from other regressors besides the treatment indicator, the difference-in-differences estimator may be expressed as

$$\hat{\alpha}_{DID} = (\overline{Y}_{t_1}^{T} - \overline{Y}_{t_0}^{T}) - (\overline{Y}_{t_1}^{C} - \overline{Y}_{t_0}^{C})$$

where $t0$ and $t1$ identify the pre-programme and the post-programme periods for which data are available, $\overline{Y}^{T}$ is the mean outcome for the treatment group, and $\overline{Y}^{C}$ is the mean outcome for the comparison (non-treatment) group.

An alternative difference-in-differences estimator is the differential-trend-adjusted estimator proposed by Bell, Blundell and Van Reenen (1999):

$$\hat{\alpha}_{TADID} = [(\overline{Y}_{t_1}^{T} - \overline{Y}_{t_0}^{T}) - (\overline{Y}_{t_1}^{C} - \overline{Y}_{t_0}^{C})] - [(\overline{Y}_{t_{**}}^{T} - \overline{Y}_{t_*}^{T}) - (\overline{Y}_{t_{**}}^{C} - \overline{Y}_{t_*}^{C})]$$

where t_* and t_{**} are the extremes of a time interval over which a similar macro trend has occurred.

The variations introduced by Bell, Blundell and Van Reenen (1999) are due to the fact that the difference-in-differences estimator does not work good when the treatment group and the comparison group have some characteristics that distinguish them and make them react differently to common macro shocks, so that the macro effect has a differential impact across the two groups. The model by Bell, Blundell and Van Reenen (1999) requires as additional hypothesis that the treatment group and the comparison group are affected by macro shocks in exactly the same way, but the procedure produces consistent estimates ($E(\hat{\alpha}_{TADID}) = \alpha$), contrary to the simple difference-in-differences estimator.
Another weakness suffered by the difference-in-differences estimator is the lack of control for unobserved temporary individual-specific components that influence the participation decision, because in such situation the procedure would overestimate the impact of the treatment.

3.5 The matching estimator with propensity score

3.5.1 The method

The matching method is a procedure which does not require any particular specification to be assumed. The aim of the matching estimator is to handle the problem of identifying the treatment impact on the outcomes, re-establishing the conditions of an experiment although we do not possess any experimental data. The missing information about the treated individuals is inferred through the members of the created comparison group. Under the matching assumptions the treated and the non-treated group are identical for everything except the participation to the programme, thus the remaining difference between the two groups may be interpretable in terms of counterfactual. The approach selects some observable factors so that any individuals with the same values will display no systematic difference as reaction to the programme. If any treated individual can be matched with a non-treated individual which shows similar matching variables, it is possible to measure the impact of the programme on those individuals. The choice of the matching variables is fundamental to obtain good final estimates, otherwise the counterfactual effect can not be correctly measured.

The choice of matching variables may even be supported by experimental data, that can help to evaluate the choice of variables as underlined in the work by Heckman, Ichimura and Todd (1997). The matching method can be applied with either cross-section or longitudinal data, but it usually requires a large amount of data for both the treated and the non-treated group, before the programme and after the programme.

If the amount of data is sufficiently large, it will be possible to apply a "propensity score" method of matching, which can easily produce acceptable results. The "propensity score" method is introduced after modelling the probability of participation to the programme, by enabling us to estimate a value for each individual, allowing in this way to match similar individuals just by observing their propensity scores (similar individuals will show numerically similar propensity scores). In this way we will have constructed a comparison group of non-treated individuals with observable characteristics similar to those of treated individuals.

The main drawback of this approach may be considered the requirement of strong initial assumptions, but it is possible to allow for less restrictive assumptions by combining this matching method with other methods. In this case even more precise estimates would be created. In particular a combination between the matching method and the difference-in-differences method has the potential to significantly improve the quality of the final results.

Formally, let us consider the specification of different outcome functions according to the participation decision:

$$Y^T = g^T(X) + U^T$$
$$Y^C = g^C(X) + U^C$$

where Y^T is the outcome of the treated group, Y^C is the outcome of the comparison group, X is a set of observable variables, U^T is the unobservable term of the treated group equation, and U^C is the unobservable term of the comparison group equation.

Identifying the impact of the treatment may be considered the main goal of all evaluation methods, and it is reached by the matching method imposing some initial assumptions. The first of them is the initial hypothesis of conditional independence between non-treated outcomes and programme participation, which states that, after one controls for the observable variables X, the outcomes of the non-treated are independent of the participation status, *d*:

$$\alpha_T = E(Y^T - Y^C \mid X, d = 1)$$

For each treated observation Y^T, we can find in the comparison group a non-treated observation, Y^C, which shows similar realisations of the observable variables X: Y^C may be considered as a possible counterfactual.

A second important initial hypothesis that must be assumed by the matching approach, which in this case guarantees that all the treated individuals have a counterpart in the non-treated population, is that:

0 < Prob (d = 1 | X) < 1

This assumption does confirm that any individual constitutes a possible participant, but it does not ensure that the same happens within any sample.

Considering *S,* the so-called ”support of *X*”, as the set of all possible values that the vector *X* may assume, and *S**, the common support of *X*, as the space of *X* that is simultaneously observed among participants and non-participants, a subset of comparable observations may be composed from the original sample.

A consistent estimator for the treatment impact on the treated, α_T, is the empirical counterpart of

$$\frac{\int_{S*} E(Y^T - Y^C \mid X, d=1) \mathrm{d}F(X \mid d=1)}{\int_{S*} \mathrm{d}F(X \mid d=1)}$$

The matching method requires that, given *X*, the observations of non-participants are statistically identical to those that participants would have shown if they had not participated.

It has the advantage to avoid specifying any particular form for the outcome equation, for the decision process or for the unobservable term.

We may say that the matching method decomposes the treatment effect in the following way:

$$E(Y^T - Y^C \mid X, d=1) = [E(Y^T \mid X, d=1) - E(Y^C \mid X, d=0)]$$
$$-[E(Y^C \mid X, d=1) - E(Y^C \mid X, d=0)]$$

Where the latter right-hand-side term is the bias conditional on *X*, assumed to be zero.

The method replaces the unobserved outcomes of the participants with the outcomes of non-participants that show the same characteristics as observable variables *X*.

The choice of the similar individuals between the treated and the comparison group is often not easy, but it may be aided through the use of the “propensity score”, a technique introduced by Rosenbaum and Rubin (1983, 1984).

The introduction of the propensity score is motivated by the fact that in some situation, e.g. when a wide range of variables *X* is in use, the matching estimator can be very difficult to be implemented because of the high dimensionality of the problem.
The propensity score approach suggests to match on a function of *X,* carrying out on the propensity to participate given the set of characteristics *X*:

$$P(X_i) = \text{Prob}(d_i = 1 \mid X_i)$$ (propensity score)

It is possible to demonstrate that under the matching assumptions

$$(Y^T, Y^C) \perp d \mid X$$

and

$$0 < \text{Prob}(d = 1 \mid X) < 1,$$

the conditional independence remains valid when controlling for *P*(*X*) instead of *X*:

$$(Y^T, Y^C) \perp d \mid P(X)$$

It can be shown that the propensity score is ancillary for the estimation of the average effect of the treatment on the population (Hahn, 1998), and that the knowledge of the propensity score may improve the efficiency of the estimates of the average effect of the treatment on the treated. Its value for the estimation of this latter parameter lies in the “dimension reduction” feature (Blundell Dias, 2000).
The comparison group must be chosen for each treated individual through a pre-defined criterion of proximity (hence through a pre-defined kind of measure of proximity).
After defining the neighbourhood for each treated individual, we will have to choose the appropriate weights to connect the set of non-treated individuals and the non-participants.
It is possible to rely on several possibilities: equal weights to all observations, a weight 1 to the nearest observation and 0 to the others, or kernel weights which account for the relative proximity of the non-participants’ observations to the treated ones in terms of *P*(*X*).

The general form of the matching estimator may be expressed as

$$\hat{\alpha}_{MM} = \sum_{i \in T} \left(Y_i - \sum_{j \in C} W_{ij} Y_j \right) w_i$$

where W_{ij} is the weight that refers to the comparison observation j for the individual i, and w_i refers to the reweighting that reconstructs the outcome distribution for the treated sample.

For example, in the nearest neighbour matching case, the estimator becomes

$$\hat{\alpha}_{MM} = \sum_{i \in T} (Y_i - Y_j) \frac{1}{N_T}$$

where j is the nearest neighbour in terms of $P(X)$ in the comparison group to i in the treatment group.

In general, kernel weights are used for W_{ij} to account for the closeness of Y_j to Y_i.

3.5.2 Conditions needed by matching with propensity score: assumptions and data requirements

As previously described, the general idea of the matching estimator is straightforward: in absence of an experimental design, assignment to treatment is frequently non-random, and thus, units receiving treatment and those excluded from treatment may differ not only in their treatment status but also in other characteristics that affect both participation and the outcome of interest. To avoid the biases that this may generate, matching methods find a non-treated unit that is "similar" to a participating unit, allowing an estimate of the intervention's impact as the difference between a participant and the matched comparison case. Averaging across all participants, the method provides an estimate of the mean program impact for the participants.

The matching estimator will not necessarily work in all circumstances: specific conditions have to be met to produce valid impact estimates. First, if the condition requiring one to find untreated units that are similar in all relevant characteristics to treated units is to be satisfied, it is clear that these characteristics must be observable to the researcher. In other words, matching method with propensity score requires selection on observables, the inability of the researcher to measure one or more relevant characteristics that determine the selection process results in biased estimations of the impact of the program. Second, in order to assign a comparison unit to each treated unit, the probability of finding an untreated unit must be positive.

We will now focus on these issues, emphasizing in the next two paragraphs the theoretical assumptions underlying the matching estimator, and the data requirements for implementing it.

3.5.2.1 Assumptions

In experimental methods, randomization ensures that all the relevant characteristics, either observable or unobservable, of the studied units are balanced[36] between treatment and control group and, because of this, the difference in mean outcomes correctly estimates the impact of the intervention. In the absence of randomization, however, the groups may differ not only in their treatment status, but also in their values of X. In this case, it is necessary to account for these differences (in econometric jargon, “to control for X” or “to condition on X”) to avoid potential biases.

To implement the matching estimator with propensity score the conditions that must be satisfied are mainly two. First, the variables (X) on which the treated and untreated groups differ must be observable to the researcher. This assumption is known as the “conditional independence” or “unconfoundedness” assumption, and it becomes subtle when a large number of variables may be potentially affecting the selection into the program. Second, in order to calculate the difference in mean outcomes for each value of X, for each possible value of the vector of covariates X, there must be a positive probability of finding both a treated and an untreated unit to ensure that each treated unit can be matched with an untreated unit. If some units in the treatment group have combinations of characteristics that cannot be matched by those of units in the comparison group, it is not possible to construct a counterfactual, and therefore, the impact for this subgroup cannot be accurately estimated. This is commonly known as the common support or overlap condition (Heinrich, Maffioli and Vázquez, 2010).

We now present a summary of these two assumptions in more technical terminology.

Assumption 1 *(“Conditional Independence Assumption” or CIA)*

There is a set X of covariates, observable to the researcher, such that after controlling for these covariates, the potential outcomes are independent of the treatment status:

$(Y_1, Y_0) \perp D \mid X$

[36] This means they are equally distributed

This is simply the mathematical notation for the idea that the potential outcomes are independent of the treatment status, *given X*. Or, in other words: after controlling for *X*, the treatment assignment is "as good as random". This property is also known as "unconfoundedness" or "selection on observables". The CIA is crucial for correctly identifying the impact of the program, since it ensures that, although treated and untreated groups differ, these differences may be accounted for in order to reduce the selection bias. This allows the untreated units to be used to construct a counterfactual for the treatment group.

Assumption 2 *(Common Support Condition)*

For each value of X, there is a positive probability of being both treated and untreated:

$$0 < P(D = 1| X) < 1$$

This equation implies that the probability of receiving treatment for each value of *X* lies between 0 and 1. By the rules of probability, this means that the probability of not receiving treatment lies between the same values (this is because $P\ (D=0\ |\ X) = 1 - P\ (D=1|\ X)$). Then, a simple way of interpreting this formula is the following: the proportion of treated and untreated individuals must be greater than zero for every possible value of *X*. The second requirement is also known as "overlap condition", because it ensures that there is sufficient overlap in the characteristics of the treated and untreated units to find adequate matches (or a "common support"). When these two assumptions are satisfied, the treatment assignment is said to be "strongly ignorable" (Rosenbaum & Rubin, 1983).

3.5.2.2 Data Requirements

The data (variables) available for matching are critical to justify the assumption that, once all relevant observed characteristics are controlled, comparison units have, on average, the same outcomes that treated units would have had in the absence of the intervention. Since in many cases the researcher does not know precisely the criteria that determine participation, it is common to control for all the variables that are suspected to influence selection into treatment (although controlling for too many variables could generate problems with the common support). As a result, the researcher should have access to a large number of variables to be able to correctly characterize the propensity score.

Prior evaluation research has also shown that it is important for data for both the treatment and comparison units to be drawn from the same sources (i.e., the same data-collection instruments), so that the measures used (for control and outcome variables) are identical or similarly constructed. In cases where the data on treated units and comparison units derive from different sources, it is critical to attempt to ensure that the variables are constructed in the same way (e.g., under the same coding conventions). Any missing data should also be handled similarly for treated and untreated units. Although data errors are always a potential issue, the bias in impact estimates may be relatively small if data errors have the same structure for treated and comparison units. In contrast, if there are systematic differences in the way that errors are introduced, particularly for outcome measures, even small differences may induce substantial biases in impact estimates.
Finally, to obtain impact estimates that can be generalized to the population of interest, it is necessary for the pool of comparison units to have a sufficient number of observations with characteristics corresponding to those of the treated units. If the comparison pool is large enough, adequate matches may be possible even if the average unmatched characteristics are very different. If the variables in question are of substantial importance, however, it may be necessary to discard treated units whose characteristics cannot be matched in estimating impacts.
Because of its large data requirements regarding both the number of variables and the sample size, the matching method with propensity score is often described as a “data-hungry method”. When data are scarce, the appropriateness of this technique should be carefully analyzed (Heinrich, Maffioli and Vázquez, 2010).

3.5.3 Why matching with propensity score is the best method for our dataset

In the previous sections we have described various non-experimental methods used and discussed in the literature (i.e. Heckman selection, Difference in Differences, Instrumental Variables, etc.), but not all they are appropriate for our purpose, and many have been shown to include drawbacks and strong assumptions. Good properties are shown by some methods that refer to a single cross-section dataset, in case that is the only type of available data, but since the analysis carried out in this framework has access to panel data, it makes cross sectional methods inappropriate, as several authors (e.g. Blundell and Dias, 2006) argue that panel methods are always preferred, because panel datasets enable researchers to control for fixed unobservable characteristics. In addition to that most methods cannot account for the fact that the treatment and comparison groups may be composed of a very different range of characteristics. This is where matching techniques are preferred. As Bryson et al. (2002) discussed, matching estimators have the advantage of highlighting the above problem of “common support”. This is important because the common support assumption ensures that the characteristics of the treatment group are also seen in the comparison group. Moreover, when a large number of variables may be potentially affecting the selection into the program, introducing a “conditional independence” assumption becomes straightforward. The main disadvantage of the matching method with propensity score is that it is very “data hungry”. In other words, to make sure that close matches are found, as well as to estimate precise propensity scores, both the number of variables and the number of observations need to be large. But in our case the size of the dataset should not be an issue since 5.073 firms will be analyzed.

CHAPTER 3

THE COMPARISON

BETWEEN

INTERNATIONALIZED AND DOMESTIC FIRMS

INTRODUCTION

Chapter 3 is dedicated to the full description of the method that we adopt to study the relationship between internationalization of firms and productivity.

Paragraph 1 offers a description of the procedure used to handle our economic problem. We illustrate an innovative approach (Ferrante and Freo, 2010), that is conditional on firms characteristics, and that also takes into account the whole productivity distributions. We show how the proposed procedure enhances the work of Bernard and Jensen (1995), taking the best aspects of the Delgado et al. (2002) model, and mixing them with the best aspects of the Kimura and Kiyota (2006), Castellani and Zanfei (2007), and Castellani et al. (2009) models. This permits to disentangle the “gross” productivity premium from the component due to structural composition of the two groups of DOM/INT firms, providing a measure of the “net” productivity premium actually imputable to the internationalization status.

Paragraph 2 illustrates the dataset of 5,073 Italian manufacturing firms. Summary statistics about firms involved in internationalization are here presented, focusing on compositional characteristics of INT versus DOM groups. We find in what industry sectors firms open to international markets operate more frequently, and we discover if it is possible to find differences between INT and DOM firms about regional aspects, firm size, firm age, and other features.

Paragraph 3 presents the results of the application of the procedure (described in paragraph 1) to the dataset (described in paragraph 2). We present the results in sections regarding the main parts in which our economic problem may be decomposed: the estimation of productivity, the conditioning on structural variables, and the decomposition of differences between the INT and DOM groups. Section 1 deals with the estimation of productivity through the most common measure used in literature: Total Factor Productivity. The estimation is developed through the semi-parametric technique introduced by Levinsohn and Petrin (2003) which makes use of an instrumental variable (intermediate inputs, assumed as a proxy of capital). We then describe the results about the TFP estimates observing them over the two groups of INT and DOM firms, finding a significant productivity premium for INT firms. We then offer some highlights about the proportion of export on turnover, the distance of export, and a series of international activities that firms can develop.

In section 2 a quantile regression technique is applied, in order to adjust for the compositional effects observed, and to identify how much of the TFP gap is explained by characteristics of firms.

We choose the covariates that better identify characteristics of firms that may influence the internationalization status, and we observe the estimates of the coefficients in a median regression and in an inter-decile regression, differentiating between the DOM and the INT group.

In section 3 we decompose the TFP differences between INT and DOM groups through two distinct methods: the Quantile Decomposition Method (Melly, 2006), and the Propensity Score Matching Method. We observe the results obtained through these two distinct methods about the productivity premium decomposition, and we interpret the results in terms of gross productivity premium, effect of self selection, and net productivity premium.

Paragraph 4 concludes.

1.

THE PROCEDURE

The aim of the present work is to study the relationship between internationalization of firms and productivity, in both direction and extension.

The empirical literature about international trade already predicts the presence of a productivity premium for internationally involved firms with respect to the domestic ones.

In literature the evaluation of the direction and of the extension of the relationship between internationalization of firms and productivity is conducted through a large number of different statistical and econometric tools.

To understand them, let us first of all define the international status of firms by considering two groups of firms.

One group (INT group) includes firms operating in the international market: we include all firms that belong at least to one of the following groups: firms that export, firms that do commercial penetration abroad, firms that have commercial agreements with foreign firms, firms that receive assistance abroad (from Italian institutes), firms that realize their business activities abroad (at least partially), and firms that do Foreign Direct Investments.

The second group (DOM group) includes only firms that just operate in the domestic market, without having any kind of contact with the international markets.

The first contributions to the strand of literature about the relationship between internationalization of firms and productivity, include simple approaches, as that offered by Bernard and Jensen (1995): the authors test the difference in productivity between DOM and INT firms by roughly comparing group productivity averages, focusing on an unconditional comparison based on a synthesis indicator of productivity, the average.

Some years later some authors have begun to enhance the work of Bernard and Jensen (1995) by considering the whole productivity distributions, instead of the simple averages. In this sense, the main contribution is offered by Delgado et al. (2002), that evaluated the hypothesis of the first order stochastic dominance between group distributions, through the use of the whole distribution. As that of Bernard and Jensen (1995), also the Delgado et al (2002) approach may be considered unconditional, since it neglects the connection of productivity with any other firm characteristic but internationalization.

In the following, authors have begun to consider also the importance of a conditional approach, by including into the model a series of characteristics that may influence the firms' productivity (i.e. size, location, industry. etc). In particular Kimura and Kiyota (2006), Castellani and Zanfei (2007), Bernard et al. (2007), and Castellani et al. (2009), adopted a way to control for these characteristics: they evaluated the productivity difference between INT and DOM firms by estimating a regression model that adopts as outcome variable the productivity, and as covariates both the dummy INT/DOM and variables describing firms characteristics. Note that this approach is based on the comparison of conditional averages, neglecting in this way the form of the productivity distribution. In other words the authors offer a model based for the first time on a conditional approach, but with the disadvantage of considering the simple averages instead of the whole productivity distributions.

What we aim to do is to use an innovative approach, as Ferrante and Freo (2010), that takes the good aspects of the Delgado et al. (2002) model, and those of the Kimura and Kiyota (2006), Castellani and Zanfei (2007), and Castellani et al. (2009) model: it means we consider an approach that is conditional on firms characteristics, and that also takes into account the whole productivity distributions.

In other words we disentangle the "gross" productivity premium from the component due to structural composition of the two groups of firms, providing a measure of the "net" productivity premium actually imputable to the internationalization status, and we observe the results along the whole distribution of productivity.

Firstly, considering that our main aim is to compare the productivity with reference to different groups defined by the international status (in our example INT and DOM), we identify the productivity (in our case TFP) as the outcome variable and we leave to the INT/DOM variable the role of defining groups to compare. In other words, instead of including this variable into the vector of covariates of a regression model, we consider the two INT and DOM groups as an a priori partition of the firm population. In this way the INT/DOM dummy variable plays a leading role by identifying an a priori partition of the firm population into two groups, instead of being one of the covariates in a regression model as in Bernard and Jensen (1999), Kimura and Kiyota (2006), Bernard et al. (2007), Castellani and Zanfei (2007), Serti and Tomasi (2008), and Love and Mansury (2009).

The comparison of productivity between groups could anyway lead to biased conclusions if it is singled out without considering that the compared groups can differ with respect to many characteristics defining heterogeneity of firms, such as size, location, industry, and so on. In fact, each of these characteristics could have some association with the level of firm productivity, so the evaluation of the between groups productivity gap (the so called "unadjusted" comparison) would consist in a comparison neglecting the "ceteris paribus" condition. Consequently, we single out a "conditional" comparison between groups, that allows us to estimate the "net" productivity premium.

Moreover at the aim to not limit the comparison to synthetic indicators, we refer the comparison to the whole productivity distribution. In this way we do not limit our productivity comparison to the average as Bernard and Jensen (1995) and others do, but we have the possibility to make comparisons along all the different quantiles.

The statistical approach we adopt to deal with this multi-object task is that proposed by Melly (2006)[37] and outlined in Chapter 2. Based on the idea that a full understanding of the gross productivity premium requires a disentangling of the effects of differences in the structural characteristics of firms from the effects of differences in returns to the characteristics by internationalization choice, this decomposition approach permits us to decompose the productivity gap between INT and DOM groups, and to distinguish what proportion of the overall productivity gap is due to differences in firms' characteristics, and what proportion is due to internationalization status. Briefly, the method allows us to compare the difference between groups distributions in a quantile regression model work after controlling for differences in individual characteristics.

Note that this approach allows us to unify the two main streams of literature about the relationship between productivity and international status: that based on (conditional) linear regression models and that comparing distribution function based on the concept of the first order stochastic dominance.

We also aim at evaluating the extent of productivity differential between DOM and INT, but we do not aim at singling out causes or consequences of these differences. Then we do not deal with the evaluation of the causal relationship between internationalization and productivity but we focus on their connection.

Melly (2006) allows us to decompose the total differences into three components: the component due to characteristics may be interpreted as the effect of self-selection factors on the choice of openness to international markets; the component due to coefficients, measuring the between-group difference, may be interpreted as the net internationalization productivity premium, and may

[37] Melly (2006) introduces his approach in a framework about the comparison of wage distributions referred to different racial groups or genders. Similar approaches were before proposed by Autor et al. (2005), and Machado and Mata (2005).

include an effect of learning by exporting. The last component measures the residual within-group difference.

Finally we use a different approach to evaluate the direction and the extension of the relationship between internationalization of firms and productivity: referring to the stream of literature about impact evaluation of policy interventions, we apply a propensity score matching method to decompose total differences into components. If the total differences are considered an Average Treatment Effect of the openness to international market on productivity, the coefficient component is the Average Treatment Effect on the Treated. Obviously also in this case it is possible to interpret the results in terms of total effect (or “gross” internationalization productivity premium), effect of self-selection, and net internationalization productivity premium (that may include an effect of learning by exporting).

2.

THE DATASET

The analysis is conducted on firm level data from the tenth Capitalia survey, which contains accounting information from balance-sheets as well as information on geographical location, export, sub-contracting, and innovation. It includes all Italian firms with more than 500 employees and a stratified sample of those with less than 500 employees. In this analysis 5,073 manufacturing firms were selected.

Information on internationalization behavior is collected in a specific section of the questionnaire. Firms are requested to report whether they exported or not during the 2004-2006 period and, if they exported, in what proportion of turnover. Furthermore, other internationalization patterns are surveyed, such as a firm's involvement in commercial penetration and/or agreements, in realizing business activities abroad (at least partially), in buying materials from abroad, and so on.

Manufacturers involved in export during the three years period were 61.1% of the whole: 22.2% of firms export less than 25% of turnover, 18.9% of firms export between 25% and 50% of turnover, and 20.4% export more than half of their turnover. Focusing on different internationalization forms, 29.4% of firms buy their materials in foreign markets, 14.3% of firms are involved in commercial penetration, 9.6% of firms realize commercial agreements with foreign firms, while other internationalization forms are less frequent (Table 1).

As a whole 64.6% of the firms are involved in at least one internationalization choice while the other 35.4% are purely domestic firms, not involved in any form of internationalization.

Table 1 – Firms involved in internationalization

	N	%
International	3279	64.6%
Domestic	1794	35.4%
Exporting	3102	61.1%
Non-exporting	1848	36.4%
Exporting < 25%	1128	22.2%
Exporting > 25%, <50%	961	18.9%
Exporting > 50%	1035	20.4%
Commercial Penetration	724	14.3%
Buying licenses from abroad	40	0.8%
Selling licenses abroad	23	0.5%
Receiving assistance abroad	69	1.4%
Realizing (at least partially) business activity abroad	351	6.9%
FDI	38	0.7%
Realizing commercial agreements with foreign firms	487	9.6%
Buying materials from abroad	1493	29.4%
Total	5073	100.0%

Firm heterogeneity may be found in compositional characteristics of INT versus DOM groups (Table 2). Within the DOM group more than one third of firms (35.1%) operate within Mineral and Metal products industries, while the percentage arrives only to 20.8% in the INT group; also industries that deal with wood, paper, printer and publishing are typically domestic firms (15.4% in the DOM group vs 5.2% in the INT group). Firms open to international markets, instead, operate more frequently in non-electric machinery industries, vehicles and transportation industries, and textile clothing and leather industries.

With respect to regional aspects, INT firms are more often found in Northern Italy.
Other features associated with international openness are the skill intensity of employees (36.9% of white and blue collars over total employment in INT firms, vs 32.4% in DOM firms), the presence of employees with an university degree (13.7% in INT firms vs 12.2% in DOM), the age of the firm (31 years for INT firms, vs about 27 for DOM). Finally, international firms tend to be larger than domestic firms, in terms of employment.

Table 2 – Summary statistics for INT and DOM firms

	DOM	**INT**
Log of employment (mean)	3.20	3.78
Age[38] (mean)	26.90	31.04
Degree[39] (mean)	12.3%	13.7%
Intensity of labor[40] (mean)	0.32	0.37
Area NCS[41] (mean)	20.4%	34.2%
Food and beverages[42]	9.8%	6.8%
Textile, clothing and leather	10.5%	14.1%
Wood, paper, printer and publishing	15.4%	5.2%
Chemicals, rubber and plastic	9.6%	10.6%
Minerals and metals	35.1%	20.8%
Non-electric machinery	8.5%	17.5%
Electric electronic and medical instruments	9.1%	8.9%
Vehicles and transportation	1.8%	16.0%

[38] Calculated in years

[39] Employees with an university degree over total employment

[40] White collars and managers over total employment

[41] Firms that belong to the northern regions of Italy

[42] This and all the next variables identify the industry sectors

3.

THE RESULTS

Our economic problem may be divided into some parts: the estimation of productivity, the conditioning on structural variables, and the decomposition of differences between the INT and DOM groups.

3.1 Estimation of productivity

Productivity has been widely used as a performance benchmark to rank producers or to measure their rate of performance over time[43]. Several indices of productivity exist, but in a general multi-output and multi-input setting the most common measure of productivity is TFP (Total Factor Productivity), a variable which is not directly observable but which may be calculated using a wide variety of methods. In our work we will consider a parametric approach to calculate productivity from an estimated production function, and we will refer in particular to the Cobb-Douglas production function. TFP will be deduced from its formulation as a residual.

Let us assume a two factor Cobb-Douglas production function

$$y_{it} = A_{it} \cdot l_{it}^{\beta_l} \cdot k_{it}^{\beta_k}$$

[43] Since Solow's decompositions of output growth into the contribution of input growth and a residual productivity term

where y_{it} is a measure of output, l_{it} and k_{it} respectively represent the usage of labor and capital, and A_{it} is the TFP which increases the marginal product of all factors simultaneously. In this framework the main problem of estimating productivity at micro-level arises from the simultaneity problem. In fact, through the standard estimation of the log-transformed production function

$$\ln y_{it} = \beta_l \ln l_{it} + \beta_k \ln k_{it} + u_{it}$$

the residuals become estimates of $\ln A_{it}$. But at least a part of the TFP is known by the firm and influences a firm's input decisions, so that error terms and the regressors of the production function are correlated and estimates are biased.

To remedy this, the error term has to be split into two components, ω_{it}, and ε_{it}

$$\ln y_{it} = \beta_l \ln l_{it} + \beta_k \ln k_{it} + \omega_{it} + \varepsilon_{it}$$

where ω_{it} measures the ln TFP and is the part observed by the firm with influence on the decision, while ε_{it}, is a pure noise component.

If the TFP may be considered plant-specific invariant over time, the equation may be estimated by the semi-parametric technique introduced by Levinsohn and Petrin (2003). The technique uses an instrumental variable assumed as a proxy of capital: intermediate inputs[44], deflated by proper index.

The TFP is measured at firm level by estimating nine Cobb-Douglas production functions by industry, with value added as output, total costs of labor as labor input and the book value of fixed and intangible assets as capital input by aggregation of industry. Also in this case all nominal variables are deflated by proper index numbers.

Since the level of TFP cannot be measured in any meaningful units, movements relative to a representative firm need to be computed.

To this purpose within the Levinshon and Petrin (2003) estimation, firm specific TFPs are obtained as averages of the exponential transformations of the $\widehat{\omega}_{it}$ divided by industry means.

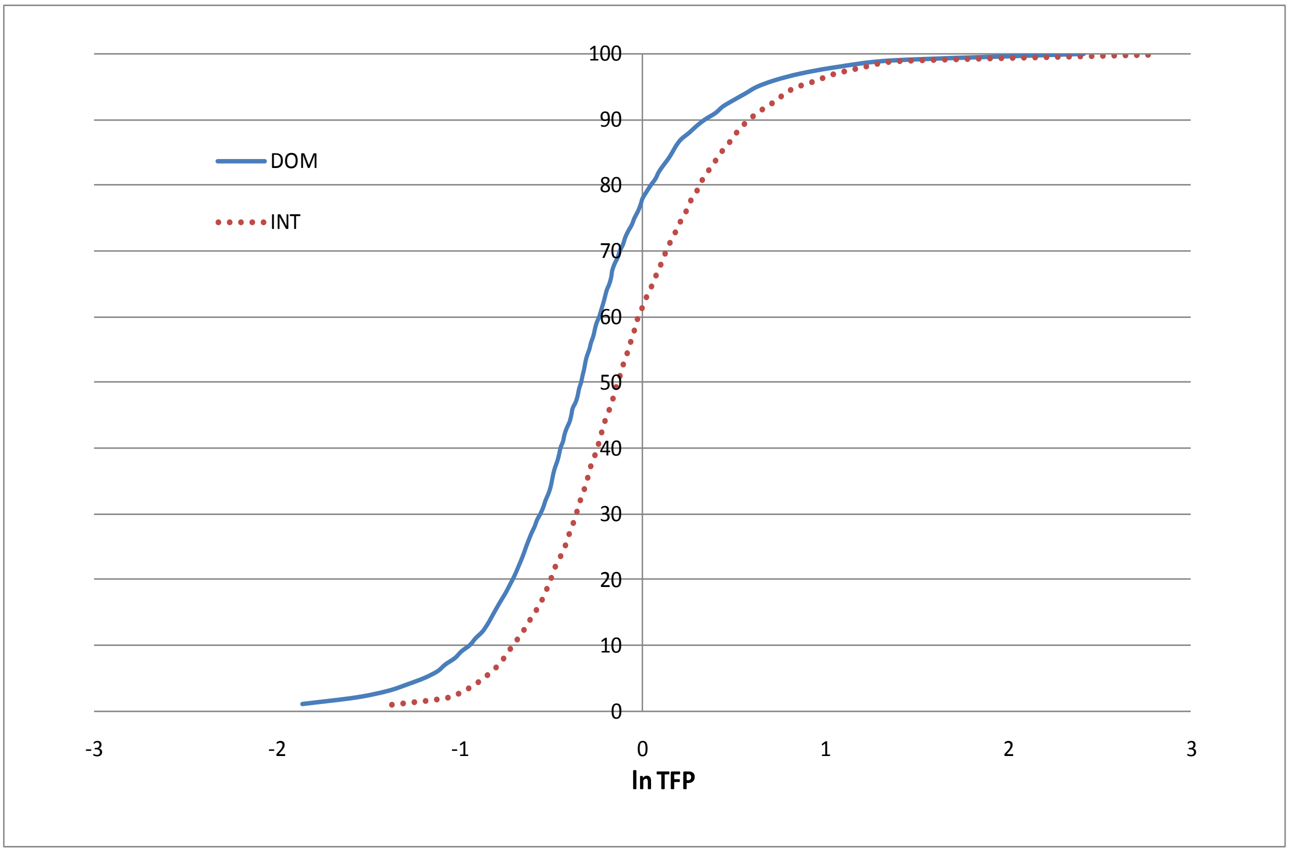
Finally, these scaled TFPs are log transformed and provide relative measures of how firm specific TFP diverges from the average during the three year period.

The results show a clear stochastic dominance of TFP distribution for INT firms. The group of INT firms have a significant gross productivity premium, if compared to the group of the DOM firms. INT firms are found more productive than DOM firms of 15.2% (average).

[44] Intermediate inputs are calculated as the sum of Raw materials and consumables, Services, Expenses for leased assets to third parties, and Variation of Materials.

Stochastic dominance of TFP distribution for international firms is evident also graphically (Figure 1): the TFP gap between INT and DOM groups is always positive in sign (the INT curve is always positioning right to the DOM curve along the TFP distribution) and varies along the distribution from 12.6% at the first quartile to 18.5% at the third quartile.

Figure 1 – Empirical cumulative TFP distributions by international status



It is also important to note (Table 3) that the higher is the quantity of export as proportion on turnover, the higher is the gross productivity premium: firms that export less than 25% of turnover are only 10.1% more productive than DOM firms, firms that export between 25% and 50% of turnover are 17.2% more productive than DOM firms, and firms that export more than half of their turnover arrive to be 19.4% more productive than DOM firms.

Table 3 – Summary statistics of TFP distribution by international status

	mean	1st quartile	median	3rd quartile
international vs domestic firms	15,2%	12,6%	13,8%	18,5%
exporters(<25%) vs domestic firms	10,1%	8,7%	9,5%	12,9%
exporters(>25%,<50%) vs domestic firms	17,2%	13,4%	14,4%	22,0%
exporters(>50%) vs domestic firms	19,4%	16,4%	17,8%	24,0%

As well as the proportion of export on turnover, also another variable seems to importantly influence the gross productivity premium: the distance of export (Table 4). Firms that export in far foreign markets seem to be more productive than firms that export near their original country. Firms that export from Italy to UE15 [45] show a productivity that is just 4.4% higher than that of the DOM group, but firms that export further increase their productivity premium: they arrive to be 21% more productive than DOM if they export to other European countries, 24.4% if they export to Africa, 23.9% if they export to Asia (25.4% to China), 20.4% if they export to Center or South America, 23.1% if they export to North America Mexico or Canada, and even 30.5% if they export to Oceania.

Table 4 – Firms exporting to foreign countries

	mean	1st quartile	median	3rd quartile
exporting to UE 15 vs DOM	4,4%	-2,4%	4,7%	11,6%
exporting to countries entered in UE in 2004 vs DOM	21,2%	8,3%	18,9%	29,4%
exporting to Russia vs DOM	25,3%	19,3%	21,0%	32,8%
exporting to other european countries (including Turkey) vs DOM	20,6%	12,2%	15,7%	24,4%
exporting to Africa vs DOM	24,4%	17,8%	22,3%	27,3%
exporting to Asia (excluding China) vs DOM	23,9%	17,7%	21,4%	31,7%
exporting to China vs DOM	25,4%	15,1%	22,6%	32,6%
exporting to USA Canada and Mexico vs DOM	23,1%	17,7%	21,4%	30,5%
exporting to Center-South America vs DOM	20,4%	17,8%	17,9%	23,4%
exporting to Australia and Oceania vs DOM	30,5%	21,4%	30,4%	40,4%

45 UE15 = Austria, Belgium, Denmark, Finland, France, Germany, Greece, Ireland, Luxembourg, Netherlands, Portugal, Spain, Sweden, UK.
Countries entered in UE in 2004 = Cyprus, Estonia, Latvia, Lithuania, Malta, Poland, Czech Republic, Slovakia, Slovenia, Hungary

Not only export enhances the gross productivity premium of firms, but also any other kind of internationalization (Table 5).

Firms that do commercial penetration are 24.6% more productive than DOM firms, firms that realize their business activities abroad are 29.3% more productive, firms that realize commercial agreements with foreign firms are 22.1% more productive, and so on.

Table 5 – Productivity premium by International status

	mean	1st quartile	median	3rd quartile
commercial penetration vs DOM	24,6%	20,6%	24,0%	-22,7%
realizing commercial agreements with foreign firms vs DOM	22,1%	19,8%	21,8%	-23,5%
buying licenses from abroad vs DOM	31,6%	13,7%	38,4%	-29,6%
selling licenses abroad vs DOM	30,6%	29,0%	24,4%	-14,3%
receiving assistance abroad vs DOM	25,8%	19,3%	27,3%	-24,0%
realizing business activities abroad vs DOM	29,3%	23,7%	26,5%	-19,6%

3.2 Conditioning on structural variables

After having obtained an estimation of productivity and thus an estimation of the gross productivity premium of INT firms towards DOM firms, we apply a quantile decomposition technique, in order to adjust for the compositional effects observed, and to identify how much of the TFP gap is explained by characteristics of firms.

For each international status the log of TFP is regressed on a set of influent attributes. The chosen attributes are the most widely used in the empirical literature on firm heterogeneity: structural covariates (geographical area and industry dummies, as defined in Table 2), age of the firm, and firm size calculated as logarithm of employment. We do not insert into the model the covariates “degree” and “intensity of labor”, even if significant, because they contain a great amount of missing data and their insertion into the model would affect the significance of results.

For the covariates definition see Table 6.

Table 6 – Variables definition

Covariates	***Description***
Area NCS	1 if the firm is geographically located in the North of Italy, 0 otherwise
Age	age of the firm, calculated in years
Log of employment	firm size as logarithm of employment
Dummy 1,2,3….	1 if the firm belongs to a specific industry sector, 0 otherwise

We run two distinct model specifications for the two groups of DOM and INT firms, to the aim of singling out the role of the covariates in the two distinct groups.

In Table 7 the estimates of median and inter-decile regressions are reported. A significant coefficient in the median regression indicates that a significant effect of the specific variable on the median level of response is detected for that group. A significant coefficient in the inter-decile regression means that a significant different impact between the ninth and the first decile, that is between high and low values of response variable to the specific covariate, is detected in that group. The variables inserted in the specification are not all significant in each regression, but each variable is significant in at least one regression. In other words each variable inserted in the model is found to have an impact on the log-transformations of the TFP in one of the two groups, or a different impact on the TFP within one of the two groups.

All variables have expected signs. Location in the north of Italy has a positive association with TFP. As the age of the firm increases, median levels of TFP increase, as well as the heterogeneity in TFP distribution within the groups. At the end, TFP has a significant and positive association with the firm size.

Table 7 – Median regression and interdecile ranges coefficients by international status [46]

	beta(0,5)		*beta(0,9)-beta(0,1)*	
	INT	*DOM*	*INT*	*DOM*
areaNCS	**0.026186**	**0.038424**	-0.04303	-0.02512
lnempl	**0.209585**	**0.224433**	0.018908	-0.02673
age	**0.000185**	0.000217	0.00072	**0.003134**
dummy1	**-0.15842**	0.014307	**0.573595**	0.286666
dummy2	**-0.23576**	0.015982	**0.459011**	**0.190857**
dummy3	**-0.17325**	**0.06615**	**0.338083**	-0.08998
dummy4	**-0.28728**	-0.04093	**0.350776**	0.077632
dummy5	**-0.22516**	-0.028	**0.355667**	0.002469
dummy7	**-0.32618**	-0.05551	**0.211977**	-0.07261
dummy8	**-0.24494**	-0.01231	**0.408915**	0.342034
_cons	**-0.53603**	**-0.87011**	**0.341624**	**0.789389**

3.3 Decomposition of differences between INT and DOM groups

3.3.1 Quantile decomposition by Melly (2006)

Considering the results of quantile regressions, we decompose differences in TFP distributions following the methodology of Melly (2006) previously presented. Results are presented in Table 8 and Figure 2.

What we observe is that the gross productivity premium can be decomposed into a factor due to characteristics (average +12.3%) that play a key role, into a factor due to coefficients (+5.2%) that has a lower impact on the total premium, and into residuals that are obviously negligible (-1.0%).

The most interesting results are those regarding the observation of the overall distribution. The gross productivity premium is positive and significant along the whole distribution of TFP, but its composition is not uniform: in the lower part of the TFP distribution the total difference is almost

[46] In bold significant coefficients at 95%

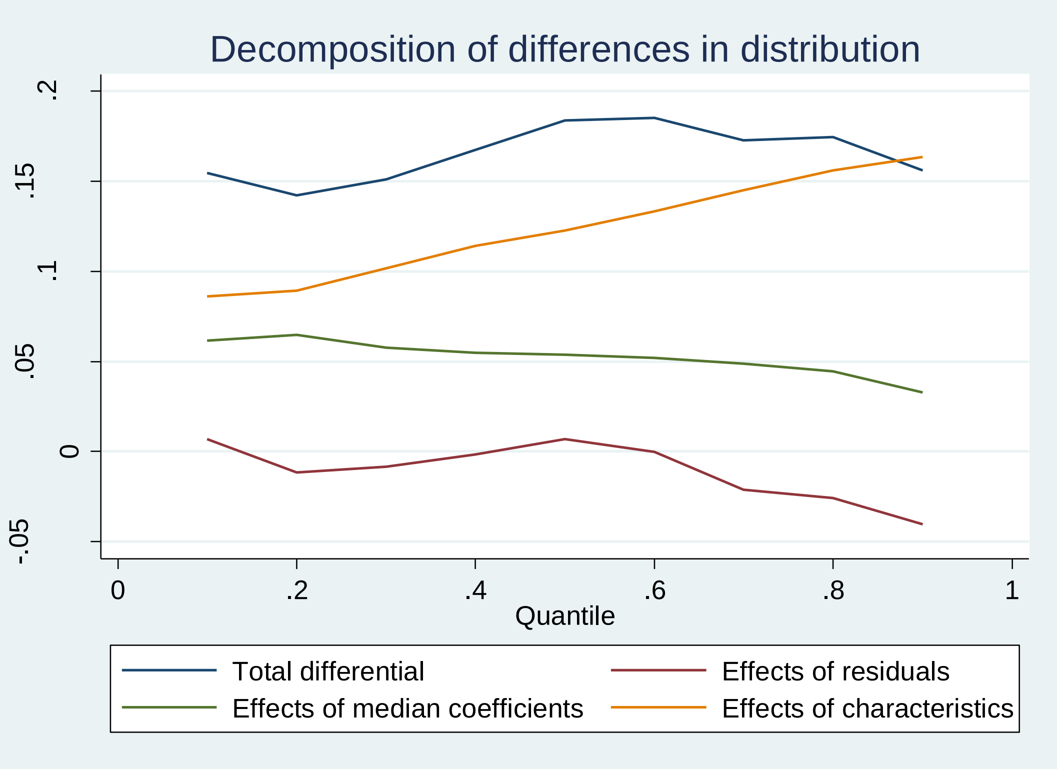
equally due to both characteristics and coefficients, but ascending the distribution the effect of characteristics becomes predominant whilst the effect of coefficients significantly reduces. This effect is particularly evident in the upper tail of the distribution, but the reduction of net productivity premium is already clear from the first quantiles of the distribution.

In conclusion it is possible to claim that a gross productivity premium surely exists, but if we clean it from the effect of self selection (effect due to characteristics), the remaining part interpretable as net productivity premium is very low and even lower (almost absent) in the upper tail of the distribution. In other words even if the effect of self selection is present for any kind of firm (less or more productive firms), it is not possible to assert the same about the effect of learning by exporting. We just can affirm that its effect, if it exists, can be found for firms with a lower level of TFP (less productive firms), but it is harder to find it for firms with a higher level of TFP (more productive firms).

Table 8 – Productivity premium decomposition

Quantile	Total difference	difference by residuals	difference by coefficients	difference by characteristics
0.1	0.154642	0.0070466	0.0615452	0.0860502
0.2	0.1422125	-0.0116047	0.0647002	0.0891171
0.3	0.1511833	-0.0082235	0.0575647	0.1018420
0.4	0.1672337	-0.0016692	0.0547223	0.1141805
0.5	0.1835184	0.0068981	0.0539037	0.1227165
0.6	0.1849856	-0.0002405	0.0520559	0.1331702
0.7	0.1725749	-0.0211135	0.0486776	0.1450108
0.8	0.1745469	-0.0258622	0.0444701	0.1559391
0.9	0.1559441	-0.0403682	0.0328856	0.1634266

Figure 2 – Decomposition of differences in distribution



3.3.2 Propensity Score Matching Decomposition

We finally apply another approach to estimate the gross productivity premium and to decompose it in order to find an effect of self-selection and a net productivity premium that may include an effect of learning by exporting.

We run a matching method, that permits us to adjust for pre-treatment[47] observable differences between a group of treated (INT) and a group of untreated (DOM).

We calculate approximate standard errors on the treatment effects assuming independent observations, fixed weights, homoskedasticity of the outcome variable within the treated and within the control groups, and that the variance of the outcome does not depend on the propensity score.

[47] The term "treatment" derives from the fact that the matching method is often applied to epidemiology, so the two groups are "treated" or "untreated" through a specific medicine. In our case the treatment is simply identified by "internationalization", so the treated group can be identified by the group of the INT firms, whilst the untreated group can be identified by the group of the DOM firms

A probit regression is firstly used. We consider exactly the same variables that we previously described in Table 2: structural covariates such as geographical area and industry dummies, and firm size calculated as logarithm of employment (see Table 2 for a review of these characteristics).

A significant coefficient in the probit regression indicates that a significant effect of the specific variable is detected.

The estimates of the coefficients offer us the results that we expected: each variable has a significant impact on the internationalization status, and all variables[48] have expected signs. The age of the firm, together with firm size calculated as logarithm of employment, and with location in the north of Italy, have a positive association with internationalization status. Even all industry dummies are strongly significant.

We present the results in Table 9.

Table 9 – Probit regression[49]

	Coef.	Std.Err.	z	P>|z|	[95%Conf.Interval]	
areaNCS	**0.317146**	0.045511	6.97	0	0.227947	0.406346
lnempl	**0.281998**	0.02063	13.67	0	0.241563	0.322433
age	**0.00448**	0.000962	4.66	0	0.002594	0.006366
dummy1	**-0.95728**	0.126298	-7.58	0	-1.20482	-0.70974
dummy2	**-0.56082**	0.121182	-4.63	0	-0.79833	-0.32331
dummy3	**-1.39754**	0.124218	-11.25	0	-1.641	-1.15407
dummy4	**-0.77408**	0.123561	-6.26	0	-1.01625	-0.5319
dummy5	**-1.14797**	0.113486	-10.12	0	-1.3704	-0.92554
dummy6	**-0.40781**	0.121358	-3.36	0.001	-0.64567	-0.16995
dummy7	**-0.80061**	0.125264	-6.39	0	-1.04613	-0.5551
_cons	-0.12261	0.132846	-0.92	0.356	-0.38298	0.137764

48 We consider significant variables with $\alpha > 95\%$

49 In bold significant coefficients at 95%

For each firm we obtain a propensity score, interpretable in terms of conditional treatment probability. The results are presented in Figure 3. As we could expect, INT (treated) firms mainly obtain high values of propensity score, and DOM (untreated) firms mainly obtain low values. Since it is also possible to find some DOM firms with high propensity score (DOM firms similar to INT firms in terms of structural characteristics), and INT firms with low propensity score (INT firms similar to DOM firms in terms of structural characteristics), we will have the possibility to compare these firms in order to estimate the counterfactual (what would happen to the TFP of an INT firm if it did not start any internationalization process? And what about a DOM firm if it started some internationalization process?).

The algorithm we use to associate DOM and INT firms is the nearest-neighbor algorithm, so that each firm in the DOM group is matched to a firm in INT group basing on the closest propensity score.

Figure 3 - Propensity score histogram by treatment status

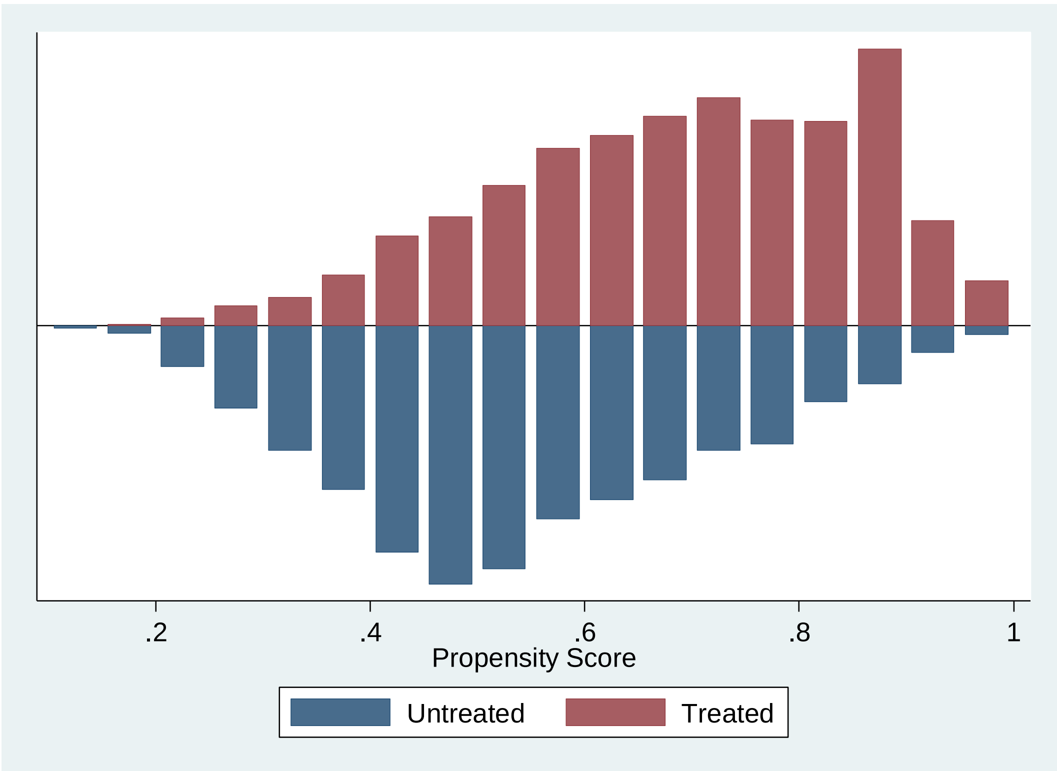


Table 10 – Reduction of the dataset[50] (n)

	Propensity Score Matching	Melly (2006)
DOM	1,639	1,639
INT	2,884	2,944
Total	**4,523**	**4,583**

The decomposition through propensity score matching creates results that are mainly similar to those obtained by the Melly's quantile decomposition technique: the first evidence in results is the presence of a significant gross productivity premium (15.84%) for the group of INT firms.

If we decompose the total difference, we realize that the gross productivity premium is almost totally due to an effect of self selection (15.76%), whilst only a very little and non-significant part (0.08%) is due to a net productivity premium. Basing on these results we can claim that the effect of self selection is high and significant, but the effect of learning by exporting is absolutely null (or anyway non-significant). Results are presented in Table 11.

Table 11 – Productivity premium decomposition through Propensity Score Matching

	Difference	**S.E.**	**T-stat**
Effect of self selection	0.157615577	0.013730479	11.48
Net productivity premium	0.000748694	0.031375728	0.02

[50] Starting from a full dataset of 5.073, Propensity Score Matching uses only a part of it: 4.523 observations (2.884 INT firms and 1.639 DOM firms). The approach of Quantile Decomposition (Melly, 2006) used 4583 observations (2944 INT firms and 1639 DOM firms). Whilst the main reduction (for both Quantile Decomposition and Propensity Score Matching) is due to missing data for covariates, the reduction in Propensity Score Matching of 60 INT firms is due to the fact that it is impossible for those firms to find "similar" DOM firms.

4.

CONCLUSIONS

Through our work we have attempted to give a contribute to the strand of microeconomic literature which studies the heterogeneity in firms and to the microeconomic and macroeconomic literature that studies the relationship between internationalization and productivity.

In general most works in literature demonstrate that international involvement implies a productivity premium, but some questions on this topic are still open. Some doubts entail the amount of productivity premium, some others entail the direction of the causality link between productivity and international openness.

Does a self-selection mechanism induce the more productive firms to enter into the international market or do internationalized firms, under the pressure of the global competition, become more productive by means of a learning-by-exporting type process?

The present work does not provide a definitive answer to these questions but it surely provides some insights on them.

In our work TFP is estimated by a method (Levinson and Petrin, 2003) which overcomes the simultaneity bias affecting standard estimations of production functions. The results clearly show, in terms of productivity, a predominance of INT firms on DOM firms. TFP levels are strongly influenced by the internationalization status, but in addition to that, also the proportion of export on turnover, the distance of export, together with a series of internationalization activities, are fundamental to determine TFP levels.

TFP estimates are then decomposed by using two different approaches: a quantile decomposition technique and a propensity score matching method.

Owing to these techniques, the overall productivity gap between INT and DOM firms is disentangled in the part due to differences in firm characteristics (self selection), and the part actually due to internationalization status (net productivity premium).

The main findings highlight that, after compositional effects are accounted for, the net productivity premium have difficulties to survive. In fact, compositional effects are found to play a crucial role in determining the gross productivity premium for internationalized firms; in particular the decomposition technique's results show that, after controlling for these effects, the net productivity premium is substantially reduced to lower levels, and it even decreases in the higher part of the TFP distribution (more productive firms). The propensity score matching approach shows that net productivity premium is generally absent (positive in sign, but not significant).
Moreover, while the difference of the gross productivity premium among groups is quite uniform along the distribution, the spread is no more uniform but higher for the self selection effect, whilst it is lower for the net productivity premium. The net premium is estimated positive for the less productive firms and negligible for the most productive.
Since the net productivity premium could be partially due to a learning by exporting hypothesis, our results seem to confirm the previous evidence of Lileeva and Trafler (2007): the learning by exporting effect mainly entails less productive (and often smaller) firms, more than more productive (often bigger) firms.
Clearly our findings do not give a definitive answer to the causality direction, but they restate the open question about the level and direction of the net TFP gap, as we estimate that net premiums mostly regard the less productive firms. It means that in this work, the learning by exporting hypothesis could be only confirmed for the less productive firms, not for the more productive ones. Firms do not find it particularly convenient (in terms of productivity) to enter into the foreign markets if they belong to the group of the most productive firms. On the other hand the entry in the foreign market is not easy, in particular for high levels of the TFP distribution, since new exporters have to face entry barriers (e.g. sunk costs). These barriers seem to be higher for more productive firms, lower for less productive firms. After entering, less productive firms have good possibilities to increase their productivity, but more productive firms have no possibilities to increase productivity so easily, probably because they have already achieved a productivity threshold.
Less productive firms seem to engage in a catching-up type behavior and take more competitive advantage from openness to international markets.
From this point of view, in a macroeconomic context, since policies targeting internationalization could result in an increase of firm productivity, the internationalization policies should mainly target firms with a low level of TFP. In fact firms with a lower level of TFP find it easier to enter in the foreign market, and if they can enter they have the possibility to gain more productivity benefits (they perhaps gain know-how and other benefits that more productive firms already have). At the

end, internationalization policies should foster the firms' performances and their productivity instead of just promoting internationalization itself.

BIBLIOGRAPHY

- J. M. Arnold, K. Hussinger (2006), “Exports versus FDI in German manufacturing: firm performance and participation in international markets”, Deutsche Bundesbank, Discussion Paper, Series 1: Economic Studies, No 04/2006

- J. M. Arnold, K. Hussinger (2004), “Export Behavior and Firm Productivity in German Manufacturing: A Firm-level Analysis”, ZEW – Zentrum fur Europaische Wirtschaftsforschung GmbH, Discussion Paper No. 04-12

- D. H. Autor, L.F. Katz, M. S. Kearney (2005), “Rising wage inequality: the role of composition and prices”, NBER working paper series.

- B. Aw, S. Chung, M. J. Roberts (2000), “Productivity and Turnover in the Export Market: Micro-level Evidence from the Republic of Korea and Taiwan (China)”, World Bank Economic Review, 14, 65-90.

- R. Barsky, J. Bound, K. Ko” Charles, J. P. Lupton. (2002), “Accounting for the Black–White Wealth Gap”, Journal of the American Statistical Association, September 1, 2002, 97(459): 663-673, doi:10.1198/016214502388618401

- A.B. Bernard, J. Jensen (1995), “Exporters, jobs and wages in US manufacturing: 1976-87”, Brookings Papers on Economic Activity: Microeconomics, 67-112

- B. Bernard, J. B. Jensen, S. J. Redding, P. K. Schott (2006) “Firms in International Trade”

- B. Bernard, J. Jensen (1999), “Exceptional Exporter Performance: Cause, Effect, or Both?”, Journal of International Economics, 47, 1, 1–25.

- B. Bernard, J. Jensen (2004), “Exporting and Productivity in the USA”, Oxford Review of Economic Policy, 20, 3, 343–57.

- B. Bernard, S. J. Redding and P. Schott (2007), “Comparative Advantage and Heterogeneous Firms”, The Review of Economic Studies, 74, 1, 31-66

- S. Blinder (1973), “Wage Discrimination: Reduced Form and Structural Estimates”, The Journal of Human Resources, 8(4), p. 436–455.

- R. Blundell, M. C. Dias (2000), “Evaluation Methods for Non-Experimental Data”, Institute of Fiscal Studies (2000) vol. 21, no. 4, pp. 427–468

- R. Blundell, M. C. Dias (2002), “Alternative approaches to evaluation in empirical microeconomics”,
- Portuguese Economic Journal (2002) 1: 91–115

- S. Bougheas and H. Görg (2008), “Organizational Forms for Global Engagement of Firms”, Discussion Papers 08/33, University of Nottingham, GEP.

- L. Brainard (1997), An Empirical Assessment of the Proximity-Concentration Trade Off between Multinational Sales and Trade, American Economic Review, 87, 4, 520-44.

- J. Campbell, H. Hopenhayn (2005) “Market Size Matters", Journal of Industrial Economics, 53, 125

- D. Castellani and A. Zanfei (2007), “Internationalization, innovation and productivity: how do firms differ in Italy”, The World Economy, 30,1, 156-176.

- D. Castellani, F. Serti and C. Tomasi (2009), “Firms in International Trade. Importers and Exporters Heterogeneity in the Italian Manufacturing Industry”, The World Economy, forthcoming.

- V. Chernozhukov, I. Fernandez-Val, B. Melly (2009), “Inference on Counterfactual Distributions”, papers.ssrn.com

- V. Chernozhukov (2005), “Extremal Quantile Regression” Annals of Statistics - 2005

- P. Combes, G. Duranton, L. Gobillon, D. Puga, S. Roux (2009) “The Productivity Advantages of Large Cities: Distinguishing Agglomeration from Firm Selection"

- Criscuolo, R. Martin (2005), “Multinationals and US Productivity Leadership: Evidence from Great Britain”, CEP (Centre for Economic Performance) Discussion Paper No 672, January 2005

- M. A. Delgado, J. C. Farinas, S. Ruano (2002), Firm productivity and export markets: a non-parametric approach, Journal of International Economics, 57, 2, 397-422.

- J. DiNardo, N. M. Fortin, T. Lemieux (1996), “Labor Market Institutions” and the Distribution of Wages, 1973-1992: A Semiparametric Approach”, Econometrica, Vol.64, No.5 (September 1996), 1001-1044

- S. G. Donald, D. A. Green, H J. Paarsch (2000), “Differences in Wage Distributions Between Canada and the United States: An Application of a Flexible Estimator of Distribution Functions in the Presence of Covariates”, Review of Economic Studies, Volume 67, Issue 4, pages 609–633, October 2000

- R. C. Feenstra (2003) “A Homothetic Utility Function for Monopolistic Competition Models, Without Constant Price Elasticity", Economic Letters, 78, 2003, 79-86.

- R. C. Feenstra, D. E. Weinstein (2009) “Globalization, Competition, and the U.S. Price Level"

- M. R. Ferrante, M. Freo (2009), “International versus domestic firms: decomposing the TFP differentials using a regression quantile approach”, draft

- M. R. Ferrante, M. Freo (2010), “Total Factor Productivity gap between internationalized and domestic firms: net premium or heterogeneity effect?”, draft

- M. R. Ferrante, M. Freo (2009), "Adjusting the Internationalization Productivity Premium for Firms Heterogeneity by using the Quantile Regression Decomposition"

- S. Firpo, N. Fortin, T. Lemieux (2007), "Decomposing Wage Distributions using Recentered Influence Function Regressions"

- S. Girma, D. Greenaway, R. Kneller (2004), "Does Exporting Increase Productivity? A Microeconometric Analysis of Matched Firms", Review of International Economics, 12(5), 855–866, 2004

- S. Girma, H. Görg, E. Strobl (2004), "Exports, international investment and plant performance. Evidence from a non-parametric test", Economics Letters, 83, 317-324.

- S. Girma, R. Kneller, M. Pisu (2005), "Exports versus FDI: An Empirical Test", Review of World Economics/Weltwirtschaftliches Archiv, 141, 2, 193-218.

- Gosling, S. Machin, and C. Meghir, (2000), "The Changing Distribution of Male Wages in the U.K", Review of Economic Studies, Blackwell Publishing, 67, 4, 635-66.

- Greenaway, R. Kneller (2007), "Firm Heterogeneity, Exporting and Foreign Direct Investments", The Economic Journal, 117 (February), F134 – F161

- Greenaway, R. Kneller (2007), "Exporting, productivity and agglomeration", European Economic Review 52 (2008) 919–939

- Greenaway, J. Gullstrand and R. Kneller (2005), "Exporting May Not Always Boost Firm Level Productivity", Review of World Economics, 141, 4, 561–582.

- D. Greenaway, R. Kneller (2004), "Exporting and Productivity in the United Kingdom", Oxford Review of Economic Policy, 20, 3, 358-371.

- D. Greenaway, R. Kneller (2007), “Firm Heterogeneity, Exporting and Foreign Direct Investment: A Survey”, Economic Journal, Royal Economic Society, 117, 517, F134-F161.

- L. Hao, D. Q. Naiman (2007) “Quantile Regression”

- K. Head, & J. Ries, (2003) “Heterogeneity and the FDI versus export decision of Japanese manufacturers”, Journal of the Japanese and International Economies, Elsevier, 17, 4, 448-467.

- Heinrich, A. Maffioli, G. Vázquez (2010), “A Primer for Applying Propensity-Score Matching”, Impact-Evaluation Guidelines, Technical Notes No. IDB-TN-161, August 2010, Inter-American Development Bank

- Helpman, M. Meliz, S. Yeaple (2004), “Export versus FDI with Heterogenous Firms”, American Economic Review, 94, 1, 300-316

- Juhn, K. M. Murphy, B. Pierce (1993), “Wage inequality and the rise in return to skill”, The Journal of Political Economy 101, 410–442.

- Kimura, and K. Kiyota (2006), “Exports, FDI, and Productivity Dynamic: Evidence from Japanese Firms”, Review of World Economics (Weltwirtschaftliches Archiv), Springer, 142, 4, 695-719.

- R. Koenker, G. Bassett (1978), “Regression Quantiles”, Econometrica, 46(1), p. 33–50

- A.N. Kolmogorov, (1933), “Sulla Determinazione Empirica di una Legge di Distribuzione”, Giornale dell”Istituto degli Attuari, 4, 83-91.

- J. Levinsohn, A. Petrin (2000), “Estimating production functions using inputs to control for unobservables”, NBER Working Paper No. 7819, August 2000

- Lileeva, and D. Trefler (2007), “Improved access to foreign markets raises plant-level productivity”, Working Paper 13297, NBER.

- J. H. Love, M. A. Mansury (2009), “Exporting and productivity in business services: Evidence from the United States”, International Business Review 18, 630–642

- J. A. F. Machado, J. Mata (2005), “Counterfactual decomposition of changes in wage distributions using quantile regression”, Journal of Applied Econometrics, 20, 445–465.

- R. Martin (2008), “Productivity Dispersion, Competition and Productivity Measurement”, CEP Discussion Paper No 692, June 2008

- T. Mayer and G. Ottaviano (2008), “The Happy Few: The Internationalisation of European Firms”, Intereconomics: Review of European Economic Policy, 43, 3, 135-148.

- F. Mayneris (2010), “Entry on Export Markets, Heterogeneous Financial Constraints and Firm-Level Performance Growth: Intra-Industrial Convergence or Divergence?”, draft

- P. Medina, M. Meléndez, K. Seim (2003), “Productivity Dynamics of the Colombian Manufacturing Sector” , Documento CEDE (2003), economia.uniandes.edu.co

- J. M. Melitz (2003), “The Impact of Trade on Intra-Industry Reallocations and Aggregate Industry Productivity”, Econometrica, 71, 6, 1695-1725.

- J. M. Melitz, and G. Ottaviano (2008), “Market Size, Trade and Productivity”, Review of Economic Studies, 75, 295-316.

- Melly (2006), “Estimation of Counterfactual Distributions Using Quantile Regression”, Discussionpaper, SIAW, University of St. Gallen.

- Melly (2005), “Decomposition of differences in distribution using quantile regression”, Labour Economics 12 (2005) 577-590

- R. Oaxaca (1973) “Male-Female Wage Differentials in Urban Labor Markets”, International Economic Review, 14, 3, 693-709.

- Palgrave Macmillan (2008), “The New Palgrave Dictionary of Economics”, edited by Lawrence E. Blume and Steven N. Durlauf

- Petrin, B.P. Poi, J. Levinsohn (2004), “Production function estimation in Stata using inputs to control for unobservables”, Stata Journal

- P. R. Rosenbaum, D. B. Rubin (1983), “The Central Role of the Propensity Score in Observational Studies for Causal Effects”, Biometrika, Vol. 70, No. 1 (Apr., 1983)

- Rothe (2010), “Decomposing Counterfactual Distributions”

- P. Selby (2008), “Net impact evaluation of the Department for Work and Pensions Working Neighbourhoods Pilot”, Department for Work and Pensions, Working Paper No 51, ISBN 978 1 84712 398 5

- F. Serti, C. Tomasi, (2008), “Self-Selection and Post-Entry Effects of Exports: Evidence from Italian Manufacturing Firms”, Review of World Economics (Weltwirtschaftliches Archiv), 144, 4, 660-694

- M. Sierminska, J. R. Frick, M. M. Grabka (2008) “Examining the Gender Wealth Gap in Germany”, Discussion Paper Series – Forschungsinstitut zur Zukunft der Arbeit, Institute for the Study of Labor - June 2008

- N. Smirnov, (1939) “On the Estimation of the Discrepancy Between Empirical Curves of Distribution for Two Independent Samples”, Bulletin Mathematique de l”Université de Moscou, 2, 2.

- Syverson (2004) “Market Structure and Productivity: A Concrete Example" Journal of Political Economy, 112, 11811222

- J. Van Biesebroeck (2005), “Exporting Raises Productivity in Sub-Saharan African Manufacturing Firms”, Journal of International Economics, 67, 2, 373-391

- J. Wagner (2007) “Exports and Productivity: A survey of the evidence from firm level data”, The World Economy, Blackwell Publishing, vol. 30(1), pages 60-82, 01

- J. Wagner, (2006) “Exports, Foreign Direct Investment, and Productivity: Evidence from German Firm Level Data”, Applied Economics Letters, 13, 347–349.

- J. Wagner, (2006) “Export Intensity and Plant Characteristics: What Can We Learn from Quantile Regression?”, Review of World Economics, 142, 1, 195–203.

- White (2006), “Impact Evaluation: The Experience of the Independent Evaluation Group of the World Bank”, World Bank, Washington D.C.

- M. Yasar, C. H. Nelson and R. M. Rejesus (2006), “Productivity and Exporting Status of Manufacturing Firms: Evidence from Quantile Regression”, Review of World Economics, 142, 4, 675-694.

- M. Yasar, R. Raciborski, B.P. Poi (2008), “Production Function Estimation in Stata Using the Olley and Pakes Method”, Stata Journal