Essays in Growth, Development and International Trade

Author: Luigi Pascali

Boston College Electronic Thesis or Dissertation, 2010

Boston College

The Graduate School of Arts and Sciences

Department of Economics

ESSAYS IN GROWTH, DEVELOPMENT AND INTERNATIONAL TRADE

a dissertation

by

LUIGI PASCALI

submitted in partial fulfillment of the requirements

for the degree of

Doctor of Philosophy

August 2010

ESSAYS IN GROWTH, DEVELOPMENT AND INTERNATIONAL TRADE

- THESIS ABSTRACT -

by

LUIGI PASCALI

Thesis Committee:

PROFESSOR FABIO SCHIANTARELLI

PROFESSOR JAMES ANDERSON

PROFESSOR SUSANTO BASU

The thesis is composed of the following three distinct papers.

1. *Banks and Development: Jewish Communities in the Italian Renaissance and Current Economic Performance*

Do banks affect long-term economic performance? I answer this question by relying on an historical development that occurred in Italian cities during the 15th century. A sudden change in the Catholic doctrine had driven the Jews toward money lending. Cities that were hosting Jewish communities developed complex banking institutions for two reasons: first, the Jews were

the only people in Italy allowed to lend for a profit; second the Franciscan reaction to Jewish usury led to the creation of charity lending institutions that evolved into many of the current Italian banks. Using Jewish demography in 1450 as an instrument, I estimate large effects of current banking development on the income-per-capita of Italian cities. Additional firm-level analyses suggest that well-functioning local banks exert large effects on aggregate productivity by reallocating resources toward more efficient firms. Controlling for province effects, using additional historical data on Jewish demography and exploiting the expulsion of the Jews from the Spanish territories in Italy in 1541, I argue that my results are not driven by omitted institutional, cultural and geographical characteristics. In particular, I show that the difference in current income between cities that hosted Jewish communities and cities that did not exists only in those regions that were not Spanish territories in the 16th century. These difference-in-difference estimates suggest that the Jewish Diaspora can explain at least 10% of the current income gap between Northern and Southern Italy.

2. *Contract Incompleteness, Globalization and Vertical Structure: an Empirical Analysis*

This paper studies the effects of international openness and contracting institutions on vertical integration. It first derives a number of predictions regarding the interactions between trade barriers, contracting costs, technology intensity, and the extent of vertical integration from a simple model with incomplete contracts. Then it investigates these predictions using a new dataset of over 14000 firms from 45 developing countries. Consistent with theory, the effect of technology intensity of domestic producers on their likelihood to vertically integrate is decreasing in the quality of domestic contracting institutions and in international openness. Contract enforcing costs are particularly high in developing countries and their effects on the vertical structure of technological intensive firms may have significant welfare costs. If improving domestic contracting institutions is not feasible an equivalent solution is to increase openness to international trade. This would discipline domestic suppliers reducing the need for vertical integration.

3. *Productivity, Welfare and Reallocation: Theory and Firm-Level Evidence (joint with Susanto Basu, Fabio Schiantarelli and Luis Serven)*

We prove that in a closed economy without distortionary taxation, the welfare of a representative consumer is summarized to a first order by the current and expected future values of the Solow productivity residual in level and by the initial endowment of capital. The equivalence holds if the representative household maximizes utility while taking prices parametrically. This result justifies TFP as the right summary measure of welfare (even in situations where it does not properly measure technology) and makes it possible to calculate the contributions of disaggregated units (industries or firms) to aggregate welfare using readily available TFP data. We show how these results must be modified if the economy is open or if taxes are distortionary. We then compute firm and industry contributions to welfare for a set of European OECD countries (Belgium, France, Great Britain, Italy, Spain), using industry-level (EU-KLEMS) and firm-level (Amadeus) data. After adding further assumptions about technology and market structure (firms minimize costs and face common factor prices), we show that welfare change can be decomposed into three components that reflect respectively technical change, aggregate distortions and allocative efficiency. Then, using the appropriate firm-level data, we assess the importance of each of these components as sources of welfare improvement in the same set of European countries.

Contents

Acknowledgements

I am profoundly grateful to my advisors Fabio Schiantarelli, James Anderson and Susanto Basu for their guidance, patience and support during these years.

I also wish to thank my family for giving me unconditional love and support and my great friend Matteo Cacciatore with whom I have shared every moment of my PhD experience at Boston College.

Chapter 1

Banks and Development: Jewish Communities in the Italian Renaissance and Current Economic Performance

1.1 Introduction

Do banks affect long term economic performance? This question is central to our understanding of the role of financial institutions in explaining cross-country and cross-regional differences in per capita income. A causal relationship has implications for both economists and policymakers. In terms of policy, if better functioning among banks has large effects on economic performance, then this increases the importance of legal and regulatory reforms designed to stimulate banking development. However, the economic theory is divided. A large body of literature dating back to Schumpeter emphasizes the positive influence of the development of a country's financial sector on the level and the rate of growth of its per capita income. The main argument is that financial intermediaries reduce the cost of acquiring information and allow for the better assessment, selection and monitoring of investment projects. For example, in Greenwood and Jovanovic (1989), the ability of financial intermediaries to improve information collection results in an increase in the efficiency of resource allocation and hence in economic performance[1]. However, according to some theoretical contributions (for example Bencivenga and Smith (1991); King and Levine (1993b)), an improvement in the reallocation of resources that results in an increase in the return to savings may actually depress saving rates and harm future economic growth. Because the theoretical literature is divided, it remains the task of the empirical literature to shed light on the effect of finance on development. In the 1990s, starting with the studies by King and Levine (1993a, 1993b), a new body of empirical evidence began to indicate a positive relationship between the level of development achieved by the banking system and economic performance, both at the national [2] and at the regional level[3]. Italy has represented a good "laboratory" for these empirical studies[4] for two reasons. First of all, focusing on Italy allows researchers to isolate the role of banks in fostering economic performance. The Italian

[1] See also Townsend (1979), Diamond (1996) and Boyd and Prescott (1985).

[2] On a pure cross-country basis, the first work that documents a positive correlation between finance and growth dates back to Goldsmith (1969). King and Levine (1993a) and King and Levine (1993b) extend this work adding more countries and controls and examining in details two channels through which finance might affect growth: capital accumulation and productivity growth. Levine and Zervos (1998) show that both stock market liquidity and banking development positively predict growth. Using similar data, Levine (1998), Levine (1999) and Levine, Loayza, and Beck (2000) provide evidence of a causal relationship from finance to growth using the legal origin of the country as an instrument for financial development. A substantial literature has documented the same causality direction using panel data techniques instead (for example see Beck, Levine, and Loayza (2000); Levine, Loayza, and Beck (2000); Loayza and Ranciere (2006)). At industry level, Rajan and Zingales (1998) show that the same industries that rely on external financing in US grow faster in financially developed countries. Finally a large literature has used time-series techniques. For example, Rousseau and Wachtel (1998) use a series of tests to determine the Granger causality direction between finance and growth in 5 countries and document that the dominant direction of causality runs from financial development to economic growth. Xu (2000) uses a VAR approach in a broad study of 41 countries to identify the long term cumulative effects of finance on growth.

[3] For example, at the US state level, Jayaratne and Strahan (1996) find that economic growth increases in states that relax intrastate bank branching restrictions. At European regional level, Hasan, Koetter, and Wedow (2009) show that more profitable banks spur regional growth.

[4] Lucchetti, Papi, and Zazzaro (2001) examine how the efficiency of local banks affects regional economic development. Angelini and Cetorelli (2003) study the effects of regulatory reforms on bank mark-ups. Bonaccorsi di Patti and Dell'Ariccia (2004) focus on firm creation. Guiso, Sapienza, and Zingales (2004) present evidence of the effect of local financial development on a wide set of outcomes, such as business formation, firm entry and growth. Guiso, Sapienza, and Zingales (2006) study the effect of banking regulation on the cost and access to credit. Benfratello, Schiantarelli, and Sembenelli (2008) concentrate on the effects of branch density on the probability that firms engage in R&D.

financial system can be characterized as bank-based. The capitalization of the Italian stock market is low compared to that of most of the other developed countries, and Italian firms have traditionally used debt rather than equity to finance their activity. Therefore, banking development is likely to be particularly important for Italian firms. Second, there is considerable spatial diversity in the degree of banking development. Until the early nineties, the competition among Italian banks was dampened by restrictions on lending and branching across geographical areas. This led to the development of deep-seated differences in the local credit markets of Italian cities.

A possible objection could be that the local conditions of the credit market become irrelevant as long as individuals and firms can tap markets other than the local one. There is a growing body of literature, however, documenting that distance matters in the provision of funds, especially for small firms (Petersen and Rajan (2002); Bofondi and Gobbi (2004); Lerner (1995)). Moreover, the fact that distance is an important barrier to lending is consistent with the views of bankers. Guiso, Sapienza, and Zingales (2004) report that "the president of the Italian Association of Bankers (ABI) declared in a conference that the banker's rule of thumb is to never lend to a client located more than three miles from his office."

Although a large amount of empirical literature has documented a strong correlation between banks and development, assessing the direction of causality has proved to be a difficult task. There is little agreement on what determines banking institutions, making it difficult to isolate exogenous sources of variation and estimate their effects on performance. In this paper, I conduct an empirical analysis of Italian municipalities and argue that the presence and size of a Jewish community in the Renaissance could have been a source of exogenous difference in the local credit market. My argument rests on the following premises:

(1) Jews arrived in Rome at the time of the Roman Empire as a result of mass deportation following the defeat of rebels in Judea by the Roman Empire. For commercial reasons, because of temporary expulsions from Rome and especially because they were deported as slaves, they spread from Rome to the rest of Italy. For centuries, they lived mostly in the proletarian sectors, and their religion prevented them from acquiring economic and social prominence in Italy. However, at the end of the fourteenth century a sudden change in the Catholic doctrine prohibited the Catholics from lending for a profit while allowing the Jews to do so.

(2) Cities that hosted a Jewish community developed complex credit markets. This happened for two reasons. The first was that several Jewish bankers were competing on the local market. According to Shulvass (1973), the Italian Jewish communities used to derive their livelihood mainly from usury, pawnshops and lending. Second, Italian Renaissance society was still devoutly Christian and thus by definition hostile to Jews. The Franciscan propaganda against usury, particularly ferocious during the fifteenth century, led to the creation of charitable loan banks, called "Monti di Pietà" (mount of piety), in cities where the Jewish minority was more influential; they were intended to drive the Jews out of the financial market.

(3) Finally, I argue that banking institutions tend to be very persistent over time. As I will document later in the paper, a significant portion of current Italian banks traces its origins to the period between 1470 and 1570, directly from the experience of the Monti. An intuitive reason for this persistence is that a bank's major asset is its reputation, which

usually appreciates over time.

Consider for example the cities of Ivrea and Chivasso. Ivrea has 23,714 residents and is located thirty miles east of Turin; Chivasso has 23,649 residents and is located fifteen miles south of Ivrea. These two cities have a very similar demographic history[5] and shared the same rulers for at least eight centuries[6]. Today, they share the same legislators and the same courts because they both belong to the province of Turin. However, Ivrea hosted a Jewish community in 1450 (tourists can still visit an ancient Jewish cemetery and a synagogue), while Chivasso did not. In 1591, a Monte di Pietà was created in Ivrea; its constitution act was motivated by the need to protect the Catholic masses from Jewish usury. Although the Jewish community disappeared at least a century ago, the Monte operated until 1984 as the main lending institution in the city. In 2005, based on the two measures of local banking development that I will use in my analysis, Ivrea dominated Chivasso. The ratio of private credit to GDP is 98% in Ivrea and 42% in Chivasso; the ratio of bank branches to residents is 0.001 in Ivrea and 0.0006 in Chivasso.

To conduct the analysis in a more systematic way, I identify the largest towns in Italy in 1861, for which I reconstruct the size of the Jewish population in 1450 and collect several measures of current banking and economic development. Then, I study the effect of local banking development on current GDP per capita in Italian municipalities, using data on Jewish demography in 1450 A.D. as an instrument for banking development. The exclusion restriction implied by my instrumental variable regression is that, conditional on the controls

[5] According to Malanima, they had less than 5 thousands residents until the early 19th century. According to the first Italian census, in 1861 Ivrea had around 6 thousands residents and Chivasso few hundreds less.

[6] In the 13th century they are both under the domain of the emperor Frederick II who assigned them to the marques of Monferrato. In the 14th century, they passed under the House of Savoy where they remain until the unification of Italy.

included in the regression, Jewish demography five centuries ago has no effect on GDP per capita today other than through its effect on current banks.

A formal test of this exclusionary restriction is impossible. However, let me argue that the restriction is plausible. Starting from the early 17th century and for at least two centuries, Jews lived segregated from the rest of the Italian population. Often, Catholics were only allowed to interact with Jews for business-related reasons and Jews lived almost exclusively on money-lending. Not until 1848 did the Kingdom of Piedmont establish equal civil and political rights for all citizens independent of their religion but, by this time, most of the Jewish communities had disappeared. Therefore, Jewish demography in 1450 could hardly affect current economic development if not through its effects on banks.

Notice, however, that if there is a location advantage (not captured by eventual controls) that led the Jews to settle in a particular city and at the same time fostered local banks and economic development, then my instrument will still be inappropriate. To address this issue, I use a difference-in-difference approach based on an historical counterfactual. Between 1493 and 1541, Jews were fully expelled from the regions under the Aragon crown (Southern Italy, some provinces in Central Italy, some cities close to Rome, Sicily and Sardinia), to which they would not be allowed to return for three centuries. This event is exogenous with respect to the Italian social and economic situation of that period. It was in fact the result of the attitude of the Spanish crown toward the Jews: the edict was promulgated in Spain and then extended to the Spanish possessions in Italy. We would expect that in these regions, the presence of a Jewish community in 1450 should have no effect on current credit institutions and economic performance. I document that within these regions, there is no difference in current credit availability or economic development between cities that previously hosted Jewish communities and cities that did not. Instead, in those regions where Jews were not banned, cities that used to host Jewish communities nowadays have larger credit-to-GDP ratios and GDP per capita. I interpret this result as an indication that there are no geo-morphological variables that affect both the presence of Jewish communities in 1450 and the current banking and economic development.

Finally, to rule out the possibility that institutional features are driving my results, I will use province fixed effects. Cities within Italian provinces have shared the same rulers (with a small number of exceptions) for at least ten centuries and, moreover, they still share the same courts and legislators.

Having established the validity of my instrument, I can use it to estimate the impact of banks on economic development. I find that an increase in credit availability of 1% (as measured by credit over GDP ratio) increases GDP per capita by at least 0.20%[7]. The effect of branch density is even stronger: an increase in the ratio of bank branches to total residents of 1% increases GDP per capita by at least 0.7%. These estimates support the view that credit institutions have strong positive effects on economic development.

Interestingly, according to the estimates coming from the difference-in-difference estimation, at least one-third of the gap in current credit availability between the North and the South of Italy can be attributed to the expulsion of the Jews from the Aragon kingdom. Based on the IV results, this implies that at least 10% of the north-south gap in GDP per capita is attributable to the lower current credit availability for which this event was responsible.

In the last part of the paper, I illuminate a particular channel through which the im-

[7] These estimates refer to the years 2002-2004.

provement of credit institutions affects economic development. At the beginning of the twentieth century, Joseph Schumpeter argued that innovations drive economic development (e.g. "different employment of existing services and labor and land", Schumpeter 1934). The so-called "Shumpeterian view" in the literature in finance and growth is based on the idea that banking institutions affect economic performance through their ability to foster aggregate productivity (e.g., the total output produced by the economy for a given set of inputs) rather than capital accumulation. My results validate this view. An increase in credit availability of 1% increases aggregate productivity by at least 0.11% (0.52% for branch density). Most of the theoretical literature focuses on two channels through which banks could affect aggregate productivity. First, banks produce ex-ante information about possible investments, and this causes a reallocation of capital towards more productive firms. Second, banks monitor investments ex-post and exert corporate governance, and this implies an average increase in firm productivity. In order to distinguish between these two channels, it is helpful to break down productivity figures in Italian cities into two parts as suggested by Olley and Pakes (1996): the unweighted average productivity of the firms in the city and a reallocation term that captures whether higher shares of value added go to more productive firms in the city.

The effects of banks on aggregate productivity seem to come into being through the reallocation of resources towards more productive firms rather than through a boost in the average productivity of firms. This seems to validate theories that stress the importance of the role of the banks in exploiting ex-ante information on investment opportunities to select the most promising ones.

This paper is organized as follows. Section 2 goes briefly into the previous literature on the effects of financial development on economic performance. Section 3 presents some historical background on the Italian Jewish communities; in particular, it focuses on the origin of Jewish money-lending during the Renaissance. Section 3 tests the hypothesis that the presence of a Jewish community in 1450 caused an improvement in current credit institutions. Section 4 examines the relationship between current credit institutions and GDP per capita in Italian cities using Jewish demography in 1450 as an instrument. Finally, in section 5, I switch to firm-level data and examine how local aggregate productivity is affected by financial development. Some concluding remarks close the paper.

1.2 Previous literature

In the 1990s, starting with the studies by King and Levine (1993a, 1993b) a new body of empirical evidence began to consider the effects of the financial system on economic performance at both the national and the regional level. There are four main approaches used in this literature.

In an important contribution, King and Levine (1993a) show that on a cross-country basis, the predetermined component of financial development is a good predictor of growth over the next 10 to 30 years. However, skeptics offer two arguments against this methodology for analyzing causality.

First, there could be some omitted variable, like the propensity of households to save, driving both financial development and economic development. Second, there could be a reverse causality problem because the usual measures of financial development (capitalization of the stock market and availability of credit to the private sector) may respond to expectation of future growth.

The second approach aims to rule out omitted country-level factors by focusing on interaction effects rather than on the main effects of financial development. On a cross-country basis, using industry-level data, Rajan and Zingales (1998) test the idea that financial development should disproportionately help industries that are relatively more dependent on external finance for their growth. As a proxy for external need of finance in a certain industry, the authors use data on the difference between investments and cash flow in the analogous industry in the US. This variable is directly interacted with the usual proxies of a country's financial development and is then regressed on measures of growth at the industry-country level. The main problem with this approach is that the magnitude of the coefficient on the interaction term is hard to interpret without making some dubious assumptions (that all countries share the same technologies and perform the same tasks within each industry, and that capital markets in the US are perfect).

The third approach focuses on the time series dimension and studies the effect of one-time exogenous financial liberalization. For example, at the US state level, Jayaratne and Strahan (1996) find that economic growth increases in states that relax intrastate bank branching restrictions. The main problem with this approach is that normally, changes in financial institutions tend to be associated with changes in other institutions, and this makes it difficult to disentangle the effect of financial development alone. According to Fry (1995), the simultaneity of reforms appears binding for researchers: "Most clear-cut cases of financial liberalization were accompanied by other economic reforms (such as fiscal, international trade and foreign exchange reforms). In such cases it is virtually impossible to isolate the effects of financial components of the reform package".

The last approach that has been widely used in the literature is the instrumental variable one. Most authors have used GMM estimators developed for panel data where the instruments come from lagged values of a financial development proxy. Levine, Loayza, and Beck (2000) use data on a panel of 77 countries over the period 1960-1995. The main advantage of this methodology is that it controls for cross-sectional fixed effects. On the other hand, the procedure is data-intensive, and researchers cannot normally count on long time series. As Levine (2005) notes: "Levine, Loayza, and Beck (2000) employ data averaged over a five-year period, yet models we are using to interpret data are typically models of steady state growth. To the extent that five years do not adequately proxy for long-run relationship, the panel methods may imprecisely asses the finance growth link". In order to overcome this problem, the literature has searched for "external" instruments that could explain cross-sectional differences in financial development without requiring long time series of data.

For example, a large body of literature has exploited the fact that historical and geographical factors could be exogenous driving forces of local financial institutions. Because the latter tend to be very persistent over time, the effect that legal tradition, colonial history and cultural factors had on the initial development of local financial markets may have persisted until the present day.

Levine (1998), Levine (1999) and Levine, Loayza, and Beck (2000) use the Porta, de Silanes, Shleifer, and Vishny (1998) measures of the legal origin of the country as instruments for current financial development. Because the legal origin of a country could emerge through occupation and colonization, this variable is treated as exogenous. Stulz and Williamson (2003) argue that different religions may imply different attitudes toward finance. For example, historically, Catholics had deep misgivings about anything related to finance and this could have prevented the brightest individuals in a Catholic country from entering finance-related professions.

These kinds of analysis have three main drawbacks. 1) It is difficult to exclude the possibility that these instruments have affected not only financial institutions but also other institutions. For example, the legal origin of the country could have strong effects on the contractual institutions of the country (see La Porta et al, 2001). The religion of the country could have effects on educational institutions and human capital; for example, the positive effects of the Protestant Reformation on the literacy of the European masses are well known. 2) It is difficult to rule out missing geo-morphological variables that could drive both instrumental and instrumented variables. 3) Generally, these studies are based on a small number of observations (usually less than eighty).

In conclusion, although there is a very large body of empirical literature on the effects of financial development on growth, further empirical analysis on the direction of causality is necessary. This paper will use an instrumental variable approach on a pure cross-section dimension. I will argue that the usual drawbacks of these kinds of analysis will not apply in my case for the following reasons:

1) I will study the effects of credit availability on the economic development of Italian cities. Concentrating on a single country and using regional and province fixed effects, I will be able to rule out the presence of other institutional changes that could be correlated with both my instrument and local financial development. I will argue that my instrument, Jewish demography in the early Renaissance, had effects on financial development and nothing else: in this period, almost all Jews lived off of money-lending; they lived segregated

in ghettos and were allowed to speak with Christians only for business-related reasons. Moreover, they were not allowed to hold any public positions and did not participate in the government of the cities. Their presence was crucial in the development of local financial markets but did not persist until today as a consequence of the migrations in the nineteenth century and the Nazi persecution. 2) A difference-in-difference analysis, based on the fact that Jews were expelled in the sixteenth century from some but not all Italian regions, will allow me to exclude the possibility that some missing geo-morphological characteristic is driving my results. 3) Moving to cross-city analysis rather than cross-country analysis will allow me to increase the number of available observations.

1.3 The Jews in the Italian Renaissance

Jews were already present in Italy in the second century B.C.E.[8] The first large communities were the result of mass deportations following the Jewish struggle and defeat in Judea by the Roman Empire[9]. Bonfil (1991) describes their role in the Roman society in the following way:

> "The fact that the Jews in Italy were of petty bourgeois or even servile origin and that they were not infrequently suspected of opposing Roman policy abroad prevented individual Jews from attaining prominence in economic or social life.

[8] The first evidence of Jewish presence in Italy dates back to 168 B.C.E. A Jewish general, Maccabees was leading the struggle to free Palestine from the Syrian domination and sent an embassy to Rome asking for military support.

[9] The first large wave of Jewish prisoners arrived in Italy in 61 B.C.E. after Pompey and the Roman legions had submitted Judea under the Roman Empire and conquered Jerusalem. In 66 C.E., Judea rebelled against the invaders: the war lasted four years and ended up with the complete defeat of the Jews. Again, a large portion of the Jewish prisoners was brought to Italy. According to later sources, 1500 arrived in Rome alone and 5000 in Apulia. The last mass deportation of Jewish prisoners in Italy dates back to 134, when the Jewish struggle against the Romans ended up with the wholesale destruction of Jerusalem and more than one thousand of other Jewish towns.

[..] They engaged in humble occupations and lived in the proletarian sections. Cultural standards were not high, although there were painters, actors, and poets."

It has been estimated that around fifty thousand Jews were living in Italy during the first century. For commercial reasons, because of temporary expulsions from Rome and especially because they were deported as slaves, the Jews spread from Rome to the rest of Italy. Whenever possible, they established themselves in more cosmopolitan cities where the local population was more tolerant of their religious convictions and customs. For these reasons, we find them concentrated in cities with important ports or where commerce was a prominent activity (Milano (1963), p. 29).

Even after the fall of the Roman Empire, the strong opposition of the Christian Church confined the Jews to the margins of Italian society. According to Bonfil, until the end of the 13th century Jews remained a group of petty bourgeois, mainly artisans (especially dyers and silk weavers) and small merchants. Typically, they owned houses in towns, but occasionally, some Jews also engaged in farming[10].

This situation dramatically changed in the 14th century. During this period, Jews in Italy engaged in a new sphere of economic activity as money-lenders. There were three main motives that drove the Jews towards the loan business. First, during the Middle Ages, the Catholic Church, through several Ecclesiastical Councils, had banned the practice of lending to earn a profit[11]. This prohibition, which had previously been limited to the

[10] In the middle of the thirteenth century, Saint Thomas Aquinas wrote that, unlike in other countries, Jews in Italy earned their livelihood through their own work and not through money-lending.

[11] The Christian prohibition to lend for a profit tracks its origin in the ancient times and is inspired by two principles. First, the Aristotelian maxim "Pecunia pecuniam parere non potest (money cannot beget money)" excluded the possibility that investing for future profits could be beneficial for the society. Second,

Catholic clergy, was extended to the Catholic laics. On the other hand, the Lateran Council in 1215, having forbidden the Jews from lending for high and immoderate interest rates, silently allowed them to lend in exchange for normal interest. Second, between 1260 and 1340, the Italian peninsula experienced a strong expansion of merchant and craft guilds (Morelli (2008)). These organizations acquired full control of the main economic activities in the largest Italian cities. Because membership required adherence to Catholicism, a large number of Jews had to leave their traditional occupations. Moreover, Jews could not continue their farming activities because they were not allowed to own land in a majority of the Italian states. Third, some Jews in Central Italy who had engaged in trade during the Middle Ages had accumulated sizable wealth and had both the capital and the expertise to become money-lenders.

These three factors drove the Jews *en masse* toward money-lending. By the start of the 15th century, the geographic expansion of the loan business by the Jews was complete and had become a general economic phenomenon in all parts of Italy. According to Shulvass (1973), Italian Jews in this period primarily derived their livelihood from usury, pawnshops and lending[12]. This led to the accumulation of small fortunes in the hands of several Jewish bankers. A large number of Jews adopted the manners of the gentile upper class, with

lending for a profit was considered at odds with the principle "Mutuum date nihil inde sperantes (give without hoping to receive anything in return)", enunciated in the Gospel according to Saint Lucas.

[12]In 1320, Kalonymos ben Kalonymos (1286-1328 A.D.), a Spanish Jewish philosopher, wrote in his Maseket Purim: "no usurious loans are to take place on Purim that is in the land of Israel, but it is permitted in Babylonia and in Greek Italy [..]. Jews of Babylonia and Italy have nothing else but usury upon which to rely [for their support]". Two centuries later, Jehiel Nisim da Pisa (1507-1574 A.D.), a rich Italian Jewish banker, also attested that "in these lands [Italy] more than everywhere else in the entire Diaspora has the custom of lending to non-Jew become widespread". Famous rabies were also moneylenders and, according to Sonne (1948), most of the North Italian rabbis were bankers even at a time when they functioned as heads of rabbinical schools. Leon da Modena (1638) charged that "in our generation all interest lenders are regarded honorable and not only are they not ineligible to testify and to judge, it is quite the reverse, namely, their word is as reliable as a hundred of witnesses, they are our leaders and judges".

a taste for the letters and the arts: this period is remembered as one of unprecedented prosperity of the Italian Jewry. These achievements, however, were undermined by two factors.

The first was the attitude of the Spanish Aragon crown toward its Jewish subjects. In March 1492, the Aragon crown promulgated an edict of expulsion of the Jews from its territories. At that time, Sicily and Sardinia were under the Aragon rule, thus the edict applied there as well. Then, in 1503 the Kingdom of Naples (which included all of Southern Italy, the region of Abruzzi and some cities close to Rome) came under the Aragon crown, and in 1510 the expulsion of the Jews from these territories was ordered. The opposition to the edict by both the Christian masses and the local aristocracy led to some exceptions. In particular, about 200 wealthy families were formally permitted to remain. However, in 1541, these exceptions were abrogated, and the law excluding Jews from the Kingdom remained in force for over three centuries.

Meanwhile, other Italian states experienced increasing opposition towards Jewish loan-banking from among the Christian population. According to Shulvass,

> "The economic depression of the masses caused by endless wars waged throughout all of Italy, contrasted with a rise in the living standard of Jewish pawnbrokers, aroused strong anti-Jewish feelings. The movement was led by the Franciscans, who during this period had a number of outstanding itinerant preachers with tremendous influence upon the masses. [..] They believed that the abolition of the Jewish loan business would heal all social ills. The masses also believed that the loan business was ruining the country."

With the explicit intention of keeping Christians in need of loans away from Jewish moneylenders, Franciscan leaders such as Bernardino da Siena (1380-1444), Giacomo della Marca (1391-1476), Giovanni da Capistrano (1386-1456), and Bernardino da Feltre (1439-1494) laid the foundations for the "Monti di Pietà", lending institutions sponsored by wealthy Christians that would extend credit on a non-profit basis. Dependent upon the largesse of wealthy Christians and fueled by the anti-Jewish sermons of the Franciscan preachers, the "Monti" flourished in Umbria, Marches, Veneto, Lombardy, Emilia, Tuscany and beyond; an estimated twenty institutions were founded in northern Italy between the years 1462 and 1496. Consider Florence, for example: here, the propaganda of a Franciscan preacher, Girolamo Savonarola, urged the wealthy to contribute to the creation of a Monte. In 1495, his sermons led the city council to authorize the creation of a Monte di Pietà. The text of the law motivates the Monte's creation, citing the high interest rates imposed by Jewish bankers. A few years later, Jewish pawnshops closed, and the Jews were expelled from the city. In her study on the Florentine Monte, Carol Bresnahan Menning explains: "As brokers of small loans against pawns, Italian Monti di Pietà were expected not only to replace Jewish moneylenders but also to set up the conditions in which all Jews could be expelled."[13] In 1539, a Monte was established in Naples, and in the following two decades, the Monti expanded their activities to Southern and Central Italy as well.

[13] One of the main Franciscan preacher of that time, Bernardino da Siena, used to depict Jewish lenders as bloodsuckers. In his sermon 43 on usury he says: "It is usually the case that when wealth and money are concentrated into fewer and fewer hands and purses, it is a sign of the deteriorating state of the city and the land. This is similar to when the natural warmth of the body abandons the extremities and concentrates only in the heart and the internal organs; this is seen as the clearest indication that life is slipping way and that the person is soon to die. And if this concentration of wealth in the hands of the few is dangerous to the health of the city, it is even more dangerous when this wealth and money is concentrated and gathered into the hands of the Jews. For in that case, the natural warmth of the city—for this is what its wealth represents—is not flowing back to the heart to give it assistance but instead rushes to an abscess in a deadly hemorrhage, since all Jews, especially those who are moneylenders, are the chief enemies of all Christians."

By the end of the 16th century, local credit markets in Italian cities could be sorted into two groups.

The first group was composed of cities that did not host Jewish communities. Here, most of the credit extended to the private sector came from inter-household loans in which, at least formally, no interest could be charged. Sometimes, a Jewish lender was invited to move into the city through the mechanism of the "condotta". The condotta was a bilateral contract of limited duration, usually lasting from three to five years (in rare circumstances for fifteen or more) stipulated by the rulers of the city and a Jewish lender. These charters regulated the number of Jews who could move into the city (normally, the limit was one person or one family) and the interest rates that could be charged (ranging from thirty to sixty percent annually). Jewish lenders were guaranteed to operate in a monopoly, and in exchange, they had to pay an annual tax and agree to lend (sometimes under favorable terms) to the government.

The second group was composed of cities that were hosting Jewish communities. Here, the financial markets were far more complex. First of all, several Jewish lenders were competing in the local credit market. Moreover, in these cities, the Monti di Pietà were particularly successful in raising charity funding: it was here that the Franciscan preachers concentrated their efforts against usury and where anti-Jewish feelings had grown stronger[14]. The Monti di Pietà certainly succeeded in lowering the interest rates imposed by the Jewish pawn banks. However, the lack of a firm business base undermined their stability. They were continuously dependent on charity for financing, and the lack of any profit motive

[14] Daniele Montanari (1991), using historical data on the Monti di Pieta' and Jewish bankers in the 16th century documented the effects that Jewish presence had on the location and the initial endownment of the Monti di Pietà.

made them particularly inefficient. The size of the loan that each person was allowed to ask for from the Monte was limited, and people had to turn to the Jews for larger amounts. Moreover, in periods of general hardship, it was difficult for the Monti di Pietà to raise sufficient funds to satisfy the demand for loans because all their depositors lived in the same town and were subjected to correlated shocks. On the other hand, Jewish lenders were able to provide access to credit even in the presence of negative aggregate shocks. Through a network of family ties, social relationships and economic partnerships, Jewish lenders in different cities shared risk and thereby were able to provide the citizens with access to external sources of credit[15].

1.4 Data description

I combine three sets of data: one including historical data for Italian cities; one including geographic, demographic, educational, economic and financial characteristics of Italian cities; and one including detailed characteristics of Italian firms.

The first dataset contains Italian demography data from the early Renaissance. The historical Jewish demography data come from the work of an Italian scholar, Attilio Milano. His book, "Storia degli Ebrei in Italia" (e.g. History of the Jews in Italy, 1963), includes a map of the Jewish communities in Italy in 1130 A.D. and in 1450 A.D. (the map is reproduced in Figure 1-2). In particular, he distinguishes among three types of communities: small (a dozen families), medium (some dozen families) and large (several dozen families)[16]. Particularly interesting is the original source of most of the data about the communities

[15] An interesting discussion on the complementarities between Jewish money lenders and Monti di Pietà can be found in Botticini (2000). See also Montanari (1999) (p. 10).

[16] The same map is reported by Bonfil (1991).

in 1130. They come from the chronicles of a Spanish merchant, Benjamin de Tudela, who traveled around Italy (and many other countries) at the beginning of the 12th century and wrote detailed descriptions of the Jewish communities he visited, including their total populations and the names of notable community leaders. He also provided some information about the communities he did not visit but had heard of. Data on Jewish communities in 1450, in contrast, come from several historiographic studies that have examined Jewish communities in different Italian regions. The historical urban population data come from Malanima (1998). Malanima compiled a dataset with urban population estimates for over 500 Italian cities on a centennial basis over the period 1300-1861, relying heavily on the seminal work on Italian population history by Beloch (1963). The sample comprises all of the Italian cities with an estimated average population of at least five thousand people for a century or more in the historical period considered. Table 1a reports the summary statistics for this dataset. There are 544 Italian cities: 41 percent of them used to contain a Jewish community. Of the 223 Jewish communities in 1450, approximately half were small (115), while the other half was equally divided among medium-sized (55) and large (53) communities. An interesting observation from the urban population data is that the average city population increased nearly threefold in the period 1300-1861, but decreased markedly in the fourteenth century: most likely, this is explained by the epidemics of the plague.

The second set of data contains the current information on Italian cities. The geomorphological data come from the Italian Geographical Institute De Agostini. The information on population and average years of education in each city comes from the Italian National Statistical Institute (ISTAT). The same source provides me with value-added data. These are available not at the city level but rather at the level of the "local labor system" (LLS). This unit is defined on the basis of the Population Census data and is composed of a set of contiguous municipalities with a high degree of self-containment of daily commuter travel and similar economic and geographic characteristics. There are a total of 854 LLSs in Italy, and all cities in my sample are located in separate ones. In the rest of my analysis, I will assume that the GDP per capita of each city is the same as that of the LLS where the city is located. Financial data on branch density and private credit come from the Bank of Italy. Table 1b reports summary statistics for these city-level data. It is interesting to notice the large variation in the level of economic development across Italian cities. The richest city has a GDP per capita that is more than eight times that of the poorest city. Looking at financial data, the private credit to GDP ratio has a surprisingly large mean (0.69) and standard deviation (0.60). For example, using a sample of 75 countries and a similarly constructed measure of private credit to GDP ratio, Levine, Loayza and Beck (2000) report a mean of 0.4 and a standard deviation of 0.29. This reinforces the idea that Italy has a bank-based financial system and features a very large degree of variation in the level of financial development across cities.

The third dataset contains current information about Italian firms. The main source of information is Amadeus, a comprehensive firm-level pan-European database developed by Bureau Van Dijk. For every firm, it provides data on the industry in which the firm operates (at the 4-digit NACE level), the location, the year of incorporation, the ownership structure and the number of employees, in addition to the complete balance sheets and the profit and loss accounts. The data set includes both publicly traded and non-traded companies and accounts for nearly 90 per cent of the sales reported in the national accounting data. In order to deflate firms' sales, materials, intermediates and capital, I have merged this dataset with an industry-level dataset that comprises output and input prices for industries at roughly the 2-digit level of aggregation coming from the EU-KLEMS project. Table 1c reports summary statistics for firms' deflated quantities.

1.5 Jewish settlements in the Italian Renaissance and current credit institutions

This paragraph documents the effects of Jewish demography in 1450 on the level of current banking development in Italian cities. Table 1.3 reports an ordinary least squares (OLS) regression of the ratio of private credit to GDP against the presence of a Jewish community in 1450. The linear regressions are for the following equation:

$$F_i = \alpha_1 J_i + \beta X_i + \epsilon_i$$

where F_i is the current financial development of city i, J_i is a dummy that identifies those cities where there were Jewish communities in 1450 and X_i is a vector of covariates. In column 1, I report estimates of α_1 without adding any control variable other than the year dummies. Having had a Jewish community in the city in 1450 is related to an increase of 0.40 in my measure of credit availability, which corresponds to a 58% increase with respect to the average level. This effect is statistically significant at 1% and remains significant when controlling for province dummies (column 2), a series of geomorphological characteristics (column 3), a dummy for whether the city is a province capital (column 4) and the extension of the municipality (column 5).

This positive correlation, however, does not necessarily indicate a causal effect. It is still possible, in fact, that some unobserved or poorly measured characteristics might drive these results. In particular, it could be that the same local advantage that led Jews to settle in a particular city is also responsible for a higher level of current credit availability. To address this issue, I use a difference-in-difference approach based on an historical event: the expulsion of the Jews from the lands under the Aragon crown between 1493 and 1541. This event is exogenous with respect to the Italian social and economic situation during the period. It was in fact the result of the attitude of the Spanish crown toward the Jews; the edict was promulgated in Spain and then extended to the Spanish territories in Italy.

A useful framework for studying the effect of Jewish communities during the Italian Renaissance on actual financial development is provided by the following matrix, which divides Italian cities along two dimensions: the presence of a Jewish community in 1450 and the definite expulsion of Jews in the following century. F represents today's level of average financial development in each cell.

Table 1.1: Difference in difference

		Jewish community in 1450	
		No	Yes
Region where Jews were expelled in 1500	Yes	F_{00}	F_{01}
	No	F_{10}	F_{11}

A simple test for the magnitude of the effect of a Jewish community during the Renaissance can therefore be conducted by concentrating on those regions that were not under Aragonese rule (and where Jews were not expelled) and determining the degree to which cities that hosted Jewish communities in 1450 are more financially developed today, or $F_{11} - F_{10}$. This estimate is analogous with the regression above and may suffer from omitted variable bias. A more compelling test would be to see whether the difference in today's financial development of cities that hosted Jewish communities in 1450 versus cities that did not is higher in regions where the Jews were not expelled compared to regions where they were, or:

$$(F_{11} - F_{10}) - (F_{01} - F_{00})$$

This difference-in-difference estimate is consistent under the assumption that the factors that led to the creation of a Jewish community were the same in regions where the Jews were subsequently expelled as in regions where they were not.

As usual in the literature, I can express the difference-in-difference results in a regression format. My measure of financial development, F_i, is regressed on a dummy that identifies the cities where there were Jewish communities in 1450, J_i, a dummy that identifies cities where Jews were able to stay after the end of the fifteenth century, S_i, and the interaction between these two dummies, plus a vector of cities covariates, X_i.

$$F_i = \alpha_0 S_i + \alpha_1 J_i + \alpha_2 J_i * S_i + \beta X_i + \epsilon_i \tag{1.1}$$

Table 1.4 reports the results. In column 1, the only controls that I use are a set of year dummies and a dummy that identifies province capitals and is motivated by the fact that until the late nineties, national banks could open their branches only in these cities. There are two striking results. The first is that having a Jewish community in 1450 does not have any effect on current banking development per se. This seems to suggest that, after distinguishing the capital from the other cities within the province, there are no missing variables that systematically drive the Jewish population in the Renaissance and credit today. The second striking result is that the coefficient of the interaction term is statistically significant at 1 per cent and very large. Having had a Jewish community in 1450 in regions from which the Jews were not expelled increases private credit to GDP today by 0.26, which corresponds to an increase of almost 40% with respect to the average level. In contrast, Jewish demography has no effect in regions from which Jews were subsequently expelled. This result is robust to the inclusion of a set of province dummies and a set of geomorphological characteristics (columns 3 and 4).

A possible concern is that the factors that led to the creation of a Jewish community were different among the different Italian regions. For this reason, in column 4, I limit my analysis to Central Italy, focusing on a set of more comparable cities in terms of history and geography. Qualitatively, the results do not change, although the estimate for the coefficient of the interaction term increases by a third. This increase is probably related with the fact that Franciscan preachers began their crusade against Jewish usury specifically in Central Italy. Here, Jewish lenders started to compete with Catholic lending institutions earlier than elsewhere in Italy, and this probably accentuated their effect in fostering local banking development.

In column 5, I rerun the regression controlling for the size of the urban population in 1300, 1400 and 1500. Cities that were larger in the Renaissance could today be more economically developed as a result of having inherited a higher level of human and social capital (see Guiso, Sapienza, and Zingales (2008); Percoco (2009)). This could bias my results if the Jews in the Aragonese regions were living in smaller cities as compared to the Jews in other Italian regions. This is plausible because most of the large Italian cities were concentrated in the Central and Northern Italy. The ancient urban population is clearly an endogenous regressor in equation 1.4: there may be some omitted features that both drive credit today and affected the urban population five centuries ago. However, the fact that the coefficient of the interaction term is affected very marginally by its inclusion suggests that my results are not driven by the distribution of the urban population in the Renaissance.

Most of the Jews were able to read and write during the Renaissance while most of the other Italians were not. In order to exclude the possibility that the effects of Jewish communities on credit were driven by human capital factors (in particular if Jews in a region from which they were expelled were less educated than Jews in a region from which they were not), I add average years of schooling in 2000 to the regressors (columns 6 and 7). Again, this variable is clearly endogenous, but it is reassuring the fact that its inclusion does not significantly affect the coefficient of the interaction term.

1.5.1 What if Jews had not been expelled from the South?

The estimates of equation 1.1 suggests that in those regions that were not under the Aragon crown, the current credit over GDP ratio in the cities that used to host Jewish communities is at least 40% higher than in those cities that did not.

Imagine increasing the credit availability in those regions that were under the Aragon crown by 40% only in those cities that used to host Jewish communities (that were subsequently expelled). Since, most of the regions that were under the Aragon crown in 1500 are concentrated in the South of Italy, in this way we can infer how much in the North-South gap in current Italian banking institutions can be attributed to the expulsion of the Jews.

The answer is surprising. At least one third of the gap in current credit availability between the North and the South of Italy can be explained by this event. I don't have reasonable elements to explain the remaining gap. However, it is possible (in fact, likely) that this estimate on the effect of the expulsion of the Jews is conservative. Probably, through temporary migration (due to the mechanism of the condotta), the positive effect of Jewish communities spilled over into neighboring towns, further contributing to the development of financial institutions in the Center-North.

Therefore, the expulsion of the Jews from the regions under the Aragon crown can be read as an exogenous negative shock on current banking development. It would be interesting to know how much of the North-South gap in economic development is explained by this shock. In the next section, I will estimate the effects of a shock on credit availability on GDP per capita. For now, let me anticipate that the most reliable result in the next section implies that an exogenous increase in credit availability by 1% causes an increase in GDP per capita by at least 0.2%. This implies that the expulsion of the Jews is responsible of at least 10% of the north-south gap in GDP per capita[17].

1.6 Financial development and income

1.6.1 OLS regressions

Table 1.5 reports the ordinary least-squares (OLS) regressions of per capita income on the private credit-to-GDP ratio. I focus on all Italian cities that were not under the Aragon crown in the Renaissance and estimate the following equation:

[17] There is a large literature in history that attributes the decline of Spain and South of Italy at the beginning of the Renaissance to the edict of expulsion of the Jewish communities.

$$\log Y_i = \alpha \log F_i + X_i' \beta_Y + \upsilon_{1,i} \tag{1.2}$$

where Y_i is income per capita in city i and X_i is a set of covariates that affect economic performance. The coefficient of interest throughout the paper is α, which captures the effect of increasing the availability of credit on per capita income.

In column 1, there are no covariates. As expected, the effect of credit availability on GDP per capita is positive and significant. An increase in the private credit to GDP ratio of 1% is associated with an increase in GDP per capita of 0.1%. However, the addition of province fixed effects (column 2) induces a tenfold reduction in the coefficient of interest, which even becomes negative when I control for the political and economic importance of the city by including among the covariates: the extension of the municipality (column 4), a dummy for regional capitals (column 5) or a dummy for province capitals (column 6). This result is surprising and at odds with the previous findings of the literature, but it should be kept in mind that this correlation cannot be interpreted as a causal relationship. The coefficient is probably strongly downward-biased due to the measurement error in the GDP at the city level, which creates a spurious negative correlation between GDP per capita and the ratio of credit over GDP. However, two other potential biases lean in the opposite direction. First, there is a reverse causality problem because richer economies may be able to afford better banks and ask for more credit. Second, there are many omitted determinants of income differences that will be naturally correlated with differences in financial institutions. All these problems could be resolved using a valid instrument for credit availability. Such an instrument should be able to account for variations in the availability of credit that have no direct effect on economic performance.

1.6.2 IV regressions

Consider a system of equations that (in addition to equation 1.2, which describes the relationship between current credit availability and economic performance) includes the following:

$$\log F_i^{1450} = \alpha_1 J_i^{1450} + X_i' \beta_J + \upsilon_{3,i} \tag{1.3}$$

$$\log F_i = \gamma_C \log F_i^{1450} + X_i' \beta_C + \upsilon_{2,i} \tag{1.4}$$

where J_i^{1450} is a dummy variable that identifies cities that used to host a Jewish community in 1450, F_i^{1450} is a measure of credit availability in 1450 and X' is a vector of covariates that affect all variables.

Equation 1.3 captures the fact that the presence of a Jewish community during the Renaissance was able to foster credit availability in the city. As we have seen, this was mainly for two reasons. First, only Jews were allowed to lend for a profit and the presence of a Jewish community tended to be associated with greater competition among Jewish moneylenders. Second, with the explicit objective of counteracting the influence of Jewish money-lending, the Franciscan movement had promoted the creation of the Monti di Pietà. These institutions were particularly successful in cities that hosted Jewish communities because it was here that Franciscan preachers had concentrated their efforts.

Equation 1.4 is motivated by the hypothesis that financial institutions tend to be very persistent. After all, the main asset of a bank is reputation, and this is an asset that strongly appreciates over time. In a different paper (Pascali (2009)), I systematically document the long-term persistence of banking institutions in Italian cities. Branch density and credit availability are higher today in cities that had a Monte di Pietà or a Jewish banker in the 16th century. This relationship remains robust even when an instrumental variable approach is used to rule out the possibility of omitted variables that could both have driven financial institutions in the 16th century and be driving them today. In this case, the instrument for banking institutions in early Renaissance is the presence of a bishop in the year 1000 A.D. and is motivated by the fact that in the Renaissance, cities with a deep-rooted Catholic tradition were more likely to challenge Jewish bankers by founding alternative charity loaning institutions. In fact, some of the largest current Italian banks trace their origin to the Monte di Pietà that were created in the early Renaissance. In 1995, the largest Italian banks by number of branches in Southern, Central and Northern Italy were respectively Banco di Napoli, Banca di Roma and San Paolo. Banco di Napoli was funded through the merger of eight Catholic institutions that opened in Naples between 1539 and 1640[18], while Banca di Roma and San Paolo come, respectively, from the Monte

[18] In 1539, the Monte di Pietà of Napoli was founded with the philanthropic purpose of providing interest-free pawn loans. Later, the Monte di Pietà opened a depository bank that was recognized with a viceregal proclamation in 1584. In the next 50 years other seven Catholic institutions were founded in Naples: the Sacro Monte e Banco dei Poveri (1600); the Banco Ave Gratia Plena or Banco della Santissima Annunziata (1587); the Banco di Santa Maria del Popolo (1589); the Banco dello Spirito Santo (1590); the Banco di Sant' Eligio (1592); the Banco di San Giacomo e Vittoria (1597); and the Banco del Santissimo Salvatore (1640). These eight banks prospered for over two hundred years until they were merged to create the "Banco Nazionale di Napoli" in 1794 by Ferdinand IV of Bourbon.

di Pietà of Rome [19] and that of Turin[20]. In general, there are hundreds of Italian banks that can be traced back to a Monte di Pietà[21].

Based on the set of relationships identified by equations 1.2, 1.3 and 1.4, I will use the presence of a Jewish community in 1450 to instrument for current availability of credit. This identification strategy will be valid as long as the instruments are uncorrelated with the error term in equation 1.2; i.e., $Cov(J_i, \upsilon_{1,i}) = 0$. Let me decompose the residual $\upsilon_{1,i}$ into three parts: $\zeta_{[-\infty;1450],i}$, $\zeta_{[1450;2000],i}$ and ϵ_i so that

$$\upsilon_{1,i} = \epsilon_i + \zeta_{[1450;2000],i} + \zeta_{[-\infty;1450],i}$$

where ϵ_i represents exogenous shocks and measurement errors in the current economic development of city i; $\zeta_{[-\infty;1450],i}$ is the set of unobserved features of city i that affect current economic development and that were already in place before 1450A.D.; and $\zeta_{[1450;2000],i}$ is the set of unobserved features of city i that affect current economic development and that can be traced to after 1450 A.D. The three sufficient conditions for the exclusion restriction

[19] Banca di Roma regrouped the histories of several notable Rome-based financial houses. The oldest of these was the Monte di Pietà di Roma, founded by a papal bull in 1539 in the aftermath of the sack of Rome in 1527 and the famine of 1538. The rebuilding effort drained the city of credit capital and increases the interest rates placed by the Jewish moneylenders. In response, Pope Paul III issued a bull establishing the Monte di Pietà di Roma, which was placed under the protection of the Franciscan Order . Another Italian bank that participated in the development of what became Banco di Roma is the Banco di Santo Spirito created in 1605 in order to raise funding for the charitable operations of the Arch-hospital Santo Spirito.

[20] The "Campagnia della Fede Cattolica di San Paolo" was created in 1563 after Piedmont had countered the invasion of Phillip II of Spain. The long war had aggravated an already difficult economic situation, increasing famine and poverty in the city of Torino, and the initial aim of the Compagnia was to centralize the collection and distribution of alms. Also in this case, with the formal intent of fighting Jewish moneylenders, the Compagnia created a Monte di Pieta' in 1579 that has operated uninterrupted (with an exception of less than 10 years during the Napoleonic domination of Piedmont) to the present day.

[21] There is no sufficient space in this article to give a complete list of the current Italian banks that tracks their origin in the 16th century and before. As an example, let me cite Banca Monte dei Paschi di Siena (1473), Rolo Banca (descendent of "Banca del Monte di Bologna e Ravenna", 1473), Banca del Monte di Lucca (1516), Banca Monte Parma (1488), Cassa di Risparmio di Udine e Pordenone (descendent of "Monte di Pieta' di Udine, 1496), Banca Carige (descendent of "Monte di Pieta' di Genova, 1483), Banca del Monte di Lombardia (from merging "Banca del Monte di Milano", 1483 and "Banca del Monte di Pavia e Bergamo, 1493).

to be valid are: $Cov(J_i, \epsilon_i) = 0$, $Cov(J_i, \zeta_{[1450;2000],i}) = 0$ and $Cov(J_i, \zeta_{[-\infty;1450],i}) = 0$.

Although the first condition seems plausible, the other two require further discussion. As a first step, let me argue that $Cov(J_i, \zeta_{[1450;2000],i}) = 0$. This assumption would be invalidated if Jewish demography in the early Renaissance had an effect on current economic performance that was not a consequence of its effect on current credit availability. As a matter of fact, Jews have traditionally displayed a high level of literacy; moreover, historically, Jews used be employed in occupations that were particularly skill-intensive (Botticini and Eckstein (2005)). It could be that either Jewish communities have persisted until now and still affect the level of human capital in their cities or that they did eventually disappear but not before transmitting their knowledge to the rest of the population, thereby contributing to the present level of human capital. However, both cases seem implausible. First, it is well-documented that the distribution of the Jewish population in Italy in the last two centuries has completely changed and, with three notable exceptions (Rome, Florence, Venice), most of the ancient Jewish communities have disappeared or are insignificant in size[22]. Second, it is very unlikely that Jewish communities in the Renaissance could have contributed to the cultural, institutional and economic development of their cities (if not through their effects on financial development). In fact, from the beginning of the 17th century through the middle of the 18th century, Jews lived segregated from the rest of the population in most Italian states and derived their livelihood almost exclusively from money-lending. The Lateran Council forbade Catholics from interacting with the Jews, if

[22] Bonfil writes that "[Between 1815 and 1938], the structure of the Jewish community changed radically. In 1840 there existed about 70 organized communities, in 1938 only 23. [..] The distribution of the Jewish population also changed. Many small rural communities disappeared, while medium-sized urban ones suffered through migration to the large centers.". Some years later the Nazi persecutions in Italy during the Second Word War decimated the Italian Jewry. Through deportations, conversion to other religions and emigration, Italy lost in less than 5 years, almost half of its Jewish population.

not for business reasons. The Jews lived in a dedicated part of the city, the ghetto. They were not allowed to leave the ghetto at night, and during the day they were obliged to wear a distinguishing badge. Moreover, they were excluded from all professions (with some exceptions in medicine), from academia and from all public offices.

The final step in defending the exclusion restriction is arguing that $Cov(J_i, \zeta_{[-\infty;1450],i}) = 0$. This condition would be violated if there were omitted variables driving both Jewish demography in the Renaissance and current economic development. Let me divide these potential omitted variables into four categories: economic, institutional cultural and geo-morphological.

First, I will deal with omitted economic features of a city in 1450. If Jews moved to rich cities, where banks were needed, and if there were some persistence in the level of economic development of Italian cities, then the instrument would not be valid. However, during the early Renaissance, two factors largely prevented the creation of new Jewish communities in Italian cities: first, Jewish communities needed strong links with local aristocracy in order to be protected from the frequent waves of intolerance; and second, Jews could only marry amongst themselves, which kept small groups of families from moving into new cities permanently[23]. In fact, all of the large communities (with the exceptions of Florence and Reggio Emilia) and most of the medium-sized communities that existed in continental Italy in 1450 A.D. were already there at least three centuries earlier (according to the reports of Benjamin de Tudela), well before Jews had been allowed to become money-lenders.[24]

A second possible class of omitted variables is the institutional features of the city

[23] I thank Maristella Botticini for clarifying this point to me.

[24] However, single Jewish families were usually invited from temporary periods in some cities that needed financial services through the mechanism of the "condotta".

in 1450. The main argument is that the same "good rulers" that attracted the Jewish population to a city in 1450 are now responsible for better institutions or higher levels of social capital. I address this possible omitted variable by conducting a city-level analysis and using province fixed effects to rule out the possibility that a Jewish community in the Renaissance could be a proxy for better institutions in a given city. Cities within each Italian province have shared (with very few exceptions) the same rulers during the last 8 centuries. Moreover, today they share the same legislators and courts.

Note, however, that there could be some cultural differences even across cities within the same province. It could be that values or beliefs in some cities were particularly favorable for the establishment of Jewish communities in the Renaissance (for example, a higher level of tolerance towards diversity) and are responsible for better economic outcomes today. It is difficult to rule out this possibility in the absence of data on how tolerant Italian cities were towards the Jewish minority. It is plausible, however, that local cultures were more favorable to the Jewish presence in cities where Jews had been living for a longer period of time than they were in cities where the Jewish presence was more recent. Following this line of reasoning, I examine cities that hosted Jewish communities in 1450 and that were ruled by the Aragons in 1500. Jews were expelled from these cities because of the attitude of the Spanish king rather than that of the local population. Table 1.6 reports the results of the following regression:

$$Y_i = \alpha_1 J_i^{1130} + \beta X_i + \epsilon_i$$

where J_i^{1130} is a dummy variable that identifies cities that were already hosting a Jewish community in 1130. If local tolerance towards the Jewish minority in the Renaissance is correlated with current economic performance, we should expect that cities that already hosted a Jewish community in 1130 will have a larger income today. However, in all the specifications, independently of whether I control for province fixed effects, geomorphological characteristics, province and regional capitals, the coefficient of J_i^{1130} is very small and is not significantly different from zero. In the last column, I add a dummy that identifies cities with archeological evidence of a Jewish presence at the time of the Roman Empire. Again, the coefficient of this variable is negative and insignificant. This means that even cities under the Aragon crown with a Jewish presence twenty centuries ago (and presumably a culture very favorable to the Jews) are not richer today.

Finally, the last possibility that would invalidate the empirical strategy is that there could be some unobserved geographical and morphological characteristics of Italian cities that drove both the Jewish demography in 1450 and economic performance today. For example, the existence of some amenities could have influenced a Jewish community to settle in a city several centuries before and could also be responsible for the current economic performance of the city. In order to address this issue, I will use a difference-in-difference approach based on the fact that the Jews were expelled from the Aragon-controlled regions. As before, I make the assumption that the factors that led to the establishment of a Jewish community in a particular territory were the same both in regions from which Jews were subsequently expelled and in regions from which they were not. I run a regression similar to equation 1.1 but using GDP per capita as a dependent variable instead of credit availability:

$$Y_i = \alpha_0 S_i + \alpha_1 J_i + \alpha_2 J_i * S_i + \beta X_i + \epsilon_i$$

The estimates are reported in Table 1.7. In column 1, the only covariates are a set of year and province fixed effects. The coefficient of J_i is positive and both economically and statistically significant, suggesting that there could be some omitted geomorphological characteristics that were associated with the presence of a Jewish community in 1450 and that affect current income. However, after the inclusion of some observed geomorphological characteristics among the covariates, this result disappears: the coefficient of J_i becomes insignificant, while the coefficient of $J_i * S_i$ is positive and significant. This suggests that after the inclusion of some *observed* geomorphological characteristics as well as a set of province fixed effects, Jewish demography becomes a good instrument of credit availability because there are no other missing variables correlated both with current GDP per capita and with the presence of a Jewish community in the early Renaissance. This result is robust to the several specifications that are reported in the table.

Having discussed the validity of my instruments, I can move to the two-stage least-squares (2SLS) estimates of equation 1.2. I limit the analysis to cities in regions that were not under the Aragon crown. The results are presented in Table 1.8. In Panel A, the availability of credit is treated as endogenous and modeled as follows:

$$\log F_i = \xi_1 J_i^S + \xi_2 J_i^M + \xi_3 J_i^L + X_i' \delta + \epsilon_i$$

where J_i^S, J_i^M and J_i^L indicate, respectively, the presence of a small, medium-sized or large community in 1450. The estimate of the elasticity of income per capita with respect to the credit-to-GDP ratio is 0.25 when the only covariates are a set of year dummies; this estimate is highly significant (with a t-statistic of 5.71). As expected, in contrast with the OLS estimate, it has a positive sign. This suggests that the influence of measurement error on GDP, which creates a downward bias, is likely to be more important than the reverse causality and omitted variable bias. The first-stage regression in Panel B shows that the effect of the size of the Jewish community in 1450 on current credit availability is positive and strongly significant. Interestingly, it is not only the presence of the Jewish community that has effects on the current credit to GDP ratio but also its size: the larger the Jewish community, the larger the current credit availability. The F-test result for the excluded instruments in 0.6, suggesting that the estimates do not suffer of a weak instrument problem. In column 2, I add province fixed effects, and in column 3, I add the usual set of geomorphological characteristics. Again, the estimates of the impact of credit availability are virtually unchanged, and they remain so when I add the extension of the municipality (column 4), a dummy for regional capital (column 5) and a dummy for province capital (column 6) to ensure that my results are not driven by a spurious correlation between the size of the Jewish community in the Renaissance and the current size of the city.

A possible concern related to this result is that the size of the Jewish community could be driven by the need for banking services during the Renaissance. I have already mentioned that most of the medium-to-large-sized Jewish communities of 1450 were already there three centuries before, well before Jews became money-lenders. However, it could still be that their relative size changed depending on the local demand for credit. For this reason, I repeat the 2SLS regressions using the presence of a Jewish community (instead of its size) as an instrument for current credit availability. The results are very similar to the previous ones, with the estimate of the impact of credit availability on income per capita ranging from 0.2 to 0.25 (see Table 1.9).

In table 1.10, I use branch density instead of the credit to GDP ratio as a measure of local banking development. The results are surprising. The effect of banks on income appears much larger than in any other previous result: a 1% increase in branch density increases GDP per capita by 0.7-1.3% depending on the econometric specification. This seems to validate the view that banks affect economic performance not only through the provision of credit but also by supplying a larger variety of services.

Table 1.11 reports a set of robustness checks. In the first three columns, I limit the analysis to Central Italy: surprisingly, it seems that in these regions the effect of credit availability on GDP per capita is even stronger. The coefficient of interest almost doubles. Based on the first-stage estimates, notice that the effects of Jewish demography in 1450 on current credit availability are also much stronger in Central Italy, with coefficients increasing by a third on average. This is probably because the creation of the "Monti di Pietà" started in Central Italy and it was in these regions that the Franciscan preachers concentrated their efforts against Jewish usury. In columns 4 to 6, I add the populations of the city in 1300, 1400 and 1500 among the covariates. The main concern that motivates this robustness check is that the presence of a Jewish community in the Renaissance could work as a proxy for the size of the city in this period. The effect of credit availability on income drops slightly from 0.20 to 0.17-0.18 (depending on the other covariates). Finally, in the last three columns, I add the difference-in-difference analysis that I used to validate my instrument earlier in this paragraph directly to the IV regressions. For the sample of all Italian cities, I estimate the following regression:

$$\log Y_i = \alpha \log F_i + \psi_1 J_i^S + \psi_2 J_i^M + \psi_3 J_i^L + X_i' \beta_Y + \upsilon_{1,i} \tag{1.5}$$

where availability of credit, $\log F_i$, is considered endogenous and modeled as

$$\log F_i = \xi_1 J_i^S * S_i + \xi_2 J_i^M * S_i + \xi_3 J_i^L * S_i + X_i'\delta + \epsilon_i$$

Practically, the dummies J_i^S, J_i^M and J_i^L should control for those city-level characteristics that are common across the cities where Jewish communities were hosted in 1450. At the same time, the interaction terms between these dummies and a dummy that identifies territories that were not under the Aragon crown are used as instrumental variables in the 2SLS regression. The estimate of α is larger compared to my baseline model, increasing from 0.20 to 0.30. As expected, the direct effect of the size of Jewish communities on income is statistically insignificant (and negative most of the time). This confirms that the cities where there were Jewish communities in the Renaissance do not have any advantage today in terms of income per capita if Jews were expelled. In contrast, from the first stage estimates, the effect of the the size of Jewish communities in regions where they were not expelled has been positive and significant.

Overall, the 2SLS results show a large effect of local banking development on economic performance. This effect is robust to different measures of banking development, different samples and different econometric specifications.

1.7 Financial development and technology

Joseph Schumpeter argued that economic development is driven by innovations (e.g., "different employment of existing services and labor and land", Schumpeter 1934). The so-called "Shumpeterian view" in the literature on finance and growth is based on the idea that banking institutions affect economic performances via their ability to foster aggregate productivity (e.g., the total output produced by the economy for a given set of inputs) rather than capital accumulation. According to this literature, financial intermediaries are able to identify the more innovative entrepreneurs and the more productive production processes and provide them with the necessary purchasing power by diverting the means of production from their previous uses. By selecting the more promising investments within a firm and across different firms, banks are able to foster aggregate productivity.

To test the Shumpeterian hypothesis, I use a detailed dataset for Italian firms in the manufacturing sector to study the effect of local banking development on aggregate productivity in Italian cities. First, I will infer the productivity of each firm in the sample as the residual of an estimated production function. Then, I will compute a measure of the aggregate productivity in Italian cities from the productivity of the local firms. Finally, I will provide evidence that financial institutions matter for aggregate productivity, shedding light on two channels through which local banks could affect aggregate productivity.

Assume that the (gross) production function in industry j is a Cobb-Douglas:

$$\log Y_{fit} = \varepsilon_L^j \log L_{fit} + \varepsilon_K^j \log K_{fit} + \varepsilon_M^j \log M_{fit} + \delta_{it} + \eta_{jt} + \alpha_f + \omega_{fit} \tag{1.6}$$

where Y_{fit} denotes the total sales of firm f in city i, L_{fit}, K_{fit} and M_{fit} are the firm's production factors, δ_{it} is a city-specific component of productivity, η_{jt} an industry-specific common component of productivity, α_f a time-invariant firm level component and ω_{fit} an idiosyncratic component. I have estimated equation 1.6 (at the 3-digit industry level) using several methodologies: OLS, Difference OLS, Olley and Pakes, Difference GMM and System

GMM. The advantages and disadvantages of each choice are well known, although there is no agreement on which estimator should be used[25]. The results in this section are robust to these different methodologies.

Having obtained the estimates of the output elasticity to each production factor, I recover the total factor productivity of firm f, t_{fi}, as follows:

$$\log t_{fi} \equiv \log Y_{fi} - \widehat{\varepsilon}_L^j \log L_{fi} - \widehat{\varepsilon}_K^j \log K_{fi} - \widehat{\varepsilon}_M^j \log M_{fi}$$

Finally, I compute a measure of the aggregate productivity of city i as a weighted average of the productivity of the firms operating within the city:

$$\log T_i \equiv \sum_{f \in i} w_{fi} \log t_{fi}$$

where the weights are $w_{fi} \equiv VA_f / \sum_{f \in i} VA_f$ and VA_f is the value added produced by firm f. T_i is a valid measure of aggregate productivity in city i because it captures whether the economy is able to produce more output for a given set of inputs. Note that it does not generally coincide with the usual Solow residual, which is computed from aggregate data on value added and primary inputs. A large body of literature has shown that the Solow residual is a good measure of aggregate technology under very restricted hypotheses.

[25] One fundamental estimation problem is the endogeneity of the input variables, which are likely to be correlated both with α_f and ω_{fit}. Correlation with ω_{fit} may reflect both simultaneity of input choices or measurement errors. Given the shortness of the panel, elimination of α_f through a within transformation is not the appropriate strategy. Differencing of (1.6) and application of the difference GMM estimator (Arellano and Bond (1991)) is a possibility, but appropriately lagged values of the regressors may be poor instruments if inputs are very persistent. Application of the GMM System estimator (Blundell and Bond (1998) and Blundell and Bond *(2000)* is probably a better option. An alternative approach is the one proposed by Olley and Pakes (1996). This estimator addresses the simultaneity (and selection) problem by using firm investment as a proxy for unobserved productivity and requires the presence of only one unobserved state variable at the firm level and monotonicity of the investment function. A recent survey of different methodologies to estimate the production function is provided by Beveren (2007).

For example, Hall (1988, 1990) notes that with imperfect competition, the Solow residual rises when the use of primary inputs rises. Basu and Fernald (2002) note that if firms have different markups of price over marginal cost or face different wages, then reallocation of resources towards firms with higher mark ups or those that pay higher wages raises the Solow residual.[26]

The measure of aggregate productivity, T_i, is then regressed on local credit availability according to the following equation:

$$\log T_i = \alpha \log F_i + X_i' \beta_Y + \upsilon_{1,i} \tag{1.7}$$

OLS estimates are reported in Table 1.12: local credit availability seems to have no effect on aggregate technology. The results change dramatically when we move to 2SLS using the usual instruments for local banking development. The results are reported in Table 1.13. Column 1 reports the 2SLS estimates for the coefficient of interest α when the only covariates are the province dummies. The fact that this coefficient is positive and both economically and statistically significant seems to validate the Shumpeterian theory: banks have strong effects on local productivity. This result is consistent across different specifications (namely adding geographic characteristics, a region capital dummy, province dummies and population figures for 1300, 1400 and 1500). However, the effect of local banks on productivity does not fully account for their effect on the GDP per capita of Italian cities: while an increase in credit availability of 1 percent increases GDP per capita by at least 0.2 percentage points, the effect on aggregate technology in the manufacturing sector is much

[26] For a detailed discussion on the differences between the measure of aggregate productivity in this paper and the Solow residual see Basu and Fernald (2002) and Basu, Pascali, Schiantarelli, Serven (2009).

smaller, ranging between 0.12 and 0.13 percentage points. Table 1.14 repeats the same analysis using branch density as a measure of local banking development. The effect of branch density on productivity is statistically and economically significant, but even in this case, it does not fully account for the effect on GDP per capita. While an increase in branch density of 1 percent increases GDP per capita by at least 1 percent, the effects on aggregate technology range between 0.6 and 0.7 percentage points. One possible interpretation of this result is that while the Shumpeterian channel is responsible for the bulk of the effect of banks on GDP per capita, other channels could be operating as well. For example, it could be that better banking institutions increase the propensity of households to save and therefore boost the accumulation of capital in the economy. Note, however, that my results regarding the effect of banks on firm productivity could be downward-biased because of sample selection. First, it could be that banks have stronger effects on innovation in non-manufacturing sectors (which are not covered in this analysis because the estimation of a production function is problematic). A second possible interpretation is that although Amadeus covers 86 percent of the total revenues of Italian firms operating in manufacturing, it does not capture the smallest firms (with revenues smaller than one million), which are probably those that benefit more from local financial development.

Most of the theoretical literature focuses on two channels through which banks could affect firm productivity. First, banks produce ex ante information about possible investments; this implies a reallocation of capital towards more productive firms. Second, banks monitor investments ex post and exert corporate governance; this implies an average increase in firm productivity. In order to distinguish between these two channels, it is helpful to decompose productivity figures in Italian cities into two parts as suggested by Olley and Pakes (1996):

$$\log T_{i,t} = \sum_{f \in i} \Delta \log t_{fit} \Delta w_{fit} + \overline{\log t_{it}} \tag{1.8}$$

where:

$$\Delta \log t_{fit} \equiv \log t_{fit} - \overline{\log t_{fit}} \quad \text{and} \quad \Delta w_{fit} \equiv w_{fit} - \overline{w_{it}}$$

and:

$$\overline{\log t_{it}} \equiv \sum_{f \in i} \log t_{fit} \quad \text{and} \quad \overline{w_{it}} \equiv \sum_{f \in i} w_{fit}$$

The first term in equation 1.8 represents the sample covariance between productivity and value added. The larger this covariance, the higher the share of value added that goes to more productive firms and the higher city i's productivity. The second term is the unweighted average of firm-level productivity figures.

Table 1.16 presents 2SLS estimates for the following equation:

$$\operatorname{Re} all_i = \alpha \log F_i + X_i' \beta + \epsilon_i$$

where $\operatorname{Re} all_i \equiv \sum_{f \in i} \Delta \log t_{fit} \Delta w_{fit}$ and X_i is the usual set of covariates. Local financial development has a positive and statistically significant effect on the variations in aggregate productivity that are due to the reallocation of resources towards more efficient firms. This seems to validate the original Shumpeterian view that banks exert their effects on growth by identifying the best entrepreneurs and diverting resources to finance their innovations. Ceteris paribus an increase in credit availability by 1% increases the reallocation term by

0.5-0.6%. Again, the effect of branch density is much stronger because an increase in the latter of 1% increases the reallocation term by 1.3-1.8%. Thus, the effect of local banking development on the reallocation of resources toward more productive firms is very sizeable from an economic point of view. In contrast, I do not find the same effect when looking at the unweighted average of firm productivity figures.

Table 1.15 presents 2SLS estimates for the following equation:

$$\overline{\log t_i} = \alpha \log F_i + X_i' \beta + \epsilon_i$$

Credit availability has a small and statistically insignificant effect on unweighted average firm-level productivity. The effect of branch density is roughly fivefold larger but still statistically insignificant.

In conclusion, local banking development has strong positive effects on city-level aggregate productivity. Moreover, the effects of banks on aggregate productivity seem to operate by reallocating resources towards more productive firms rather than by boosting the average productivity of firms. This seems to validate theories that stress the importance of the role of the banks in exploiting ex ante information on investment opportunities to select the more promising ones.

1.8 Conclusion

Many economists believe that differences in the quality of banking institutions are the root of large differences in per capita income. Several empirical works have emphasized the presence of a positive correlation between the quality of banking institutions and economic development both at the country and at the regional level. However, assessing causality in this regard has also proved difficult because of the obvious 'chicken-and-egg' problem of circular causality that bedevils any study confined to a short time frame. I depart from the standard approach by looking for an exogenous source of variation in banks that dates back over five centuries.

My argument rests on the following premises:

1) Local financial markets were extremely heterogeneous across Italian cities during the Renaissance. At one extreme, there were cities in which several Catholic charity institutions and Jewish bankers were competing in the local lending market. At the other extreme, there were cities where lending activity was limited to intra-household transfers.

2) The presence and the size of a local Jewish community had strong effects on the complexity of the local financial markets. This happened for two reasons: first, because the Jews were the only ones allowed to lend for a profit; and second, because of the Franciscan reaction to Jewish usury, which led to the creation of charity lending institutions called Monti di Pietà. These institutions, which were dependent on charity for financing, were particularly successful in those cities where Jewish communities used to live: there, Franciscan preachers had concentrated their efforts against usury, and anti-Jewish feelings had fueled the donations to the Monti.

3) Finally, I argue that financial institutions tend to be very persistent over time. Most of the Italian banks trace their origin in the Monti of the fifteenth and sixteenth century.

First, I document that Jewish demography in the early Renaissance had strong effects on current financial development. Toward this end, I use a difference-in-difference approach based on the fact that in the sixteenth century, the Jews were expelled from some Italian regions but not from others. In practice, I use the regions from which they have been expelled as a control group. This allows me to rule out the possibility that my results are driven by some missing variables that are correlated with the presence of a Jewish community in the Renaissance and could affect current financial development.

Second, I exploit differences in Jewish demography in the Renaissance across Italian municipalities to estimate the impact of banking institutions on economic performance. I find surprisingly large effects, which validates the view that good banks are decisive determinants of development.

Third, I shed light on the channels through which banks affect economic performance. I find that higher availability of private credit implies higher aggregate productivity. In particular, the effect of banks on aggregate productivity seems to operate through a reallocation of resources towards more productive firms rather than by boosting the average productivity of firms. This seems to validate theories that stress the importance of the role of banks in exploiting ex-ante information regarding investment opportunities as they seek to select the more promising opportunities.

Let me conclude by pointing out that my findings do not imply that banking institutions today are predetermined by local historical events and cannot be changed. I emphasize Jewish demography as one of the many factors affecting Italian local financial institutions; because it is arguably exogenous, it is useful as an instrument for isolating the effects of banks on development. In fact, my reading of the results of this paper is that improvements in financial institutions may be substantially beneficial to the economic environment.

Bibliography

Angelini, P., and N. Cetorelli (2003): "The Effects of Regulatory Reform on Competition in the Banking Industry," *Journal of Money, Credit and Banking*, 35(5), 663–84.

Arellano, M., and S. Bond (1991): "Some Tests of Specification for Panel Data: Monte Carlo Evidence and an Application to Employment Equations," *Review of Economic Studies*, 58(2), 277–97.

Basu, S., and J. G. Fernald (2002): "Aggregate productivity and aggregate technology," *European Economic Review*, 46(6), 963–991.

Beck, T., R. Levine, and N. Loayza (2000): "Finance and the sources of growth," *Journal of Financial Economics*, 58(1-2), 261–300.

Beloch, K. J. (ed.) (1963): *Bevolkerungsgeschichte Italiens*. Walter de Gruyter and Co.

Bencivenga, V. R., and B. D. Smith (1991): "Financial Intermediation and Endogenous Growth," *Review of Economic Studies*, 58(2), 195–209.

Benfratello, L., F. Schiantarelli, and A. Sembenelli (2008): "Banks and innovation: Microeconometric evidence on Italian firms," *Journal of Financial Economics*, 90(2), 197–217.

Bofondi, M., and G. Gobbi (2004): "Bad Loans and Entry into Local Credit Markets," Temi di discussione (Economic working papers) 509, Bank of Italy, Economic Research Department.

Bonaccorsi di Patti, E., and G. Dell'Ariccia (2004): "Bank Competition and Firm Creation," *Journal of Money, Credit and Banking*, 36(2), 225–51.

Bonfil, R. (1991): *Jewish life in Renaissance Italy*. University of California Press.

Botticini, M. (2000): "A Tale of "Benevolent" Governments: Private Credit Markets, Public Finance, and the Role of Jewish Lenders in Medieval and Renaissance Italy," *The Journal of Economic History*, 60(1), 164–189.

Botticini, M., and Z. Eckstein (2005): "Jewish Occupational Selection: Education, Restrictions, or Minorities?," *The Journal of Economic History*, 65(4), 922–948.

Boyd, J. H., and E. C. Prescott (1985): "Financial intermediary-coalitions," Discussion paper.

Diamond, D. W. (1996): "Financial intermediation as delegated monitoring: a simple example," *Economic Quarterly*, (Sum), 51–66.

Fry, M. J. (1995): *Money, interest, and banking in economic development*. John Hopkins University Press.

Goldsmith, R. W. (1969): *Financial structure and development*. Yale University Press.

Greenwood, J., and B. Jovanovic (1989): "Financial Development, Growth, and the

Distribution of Income," NBER Working Papers 3189, National Bureau of Economic Research, Inc.

Guiso, L., P. Sapienza, and L. Zingales (2004): "Does Local Financial Development Matter?," *The Quarterly Journal of Economics*, 119(3), 929–969.

——— (2006): "The Cost of Banking Regulation," CEPR Discussion Papers 5864, C.E.P.R. Discussion Papers.

——— (2008): "Long Term Persistence," Economics Working Papers ECO2008/30, European University Institute.

Hasan, I., M. Koetter, and M. Wedow (2009): "Regional growth and finance in Europe: Is there a quality effect of bank efficiency?," *Journal of Banking & Finance*, 33(8), 1446–1453.

Jayaratne, J., and P. E. Strahan (1996): "The Finance-Growth Nexus: Evidence from Bank Branch Deregulation," *The Quarterly Journal of Economics*, 111(3), 639–70.

King, R. G., and R. Levine (1993a): "Finance and Growth: Schumpeter Might Be Right," *The Quarterly Journal of Economics*, 108(3), 717–37.

King, R. G., and R. Levine (1993b): "Finance, entrepreneurship and growth: Theory and evidence," *Journal of Monetary Economics*, 32(3), 513–542.

Lerner, J. (1995): "Venture Capitalists and the Oversight of Private Firms," *Journal of Finance*, 50(1), 301–18.

Levine, R. (1998): "The Legal Environment, Banks, and Long-Run Economic Growth," *Journal of Money, Credit and Banking*, 30(3), 596–613.

——— (1999): "Law, Finance, and Economic Growth," *Journal of Financial Intermediation*, 8(1-2), 8–35.

——— (2005): *Finance and Growth: Theory and Evidence*. Elsevier.

Levine, R., N. Loayza, and T. Beck (2000): "Financial intermediation and growth: Causality and causes," *Journal of Monetary Economics*, 46(1), 31–77.

Levine, R., and S. Zervos (1998): "Stock Markets, Banks, and Economic Growth," *American Economic Review*, 88(3), 537–58.

Loayza, N. V., and R. Ranciere (2006): "Financial Development, Financial Fragility, and Growth," *Journal of Money, Credit and Banking*, 38(4), 1051–1076.

Lucchetti, R., L. Papi, and A. Zazzaro (2001): "Banks' Inefficiency and Economic Growth: A Micro-Macro Approach," *Scottish Journal of Political Economy*, 48(4), 400–424.

Malanima, P. (1998): "Italian Cities 1300-1800. A quantitative approach," *Rivista di Storia Economica*, 2(19), 92–126.

Milano, A. (ed.) (1963): *Storia degli Ebrei in Italia*. Giulio Einaudi Editore.

Montanari, D. (ed.) (1999): *Monti di Pieta' e presenza ebraica in Italia*. Bulzoni Editore.

Morelli, L. (2008): "Guilds Reappraised: Italy in the Early Modern Period," *International Review of Social History*, 53(16), 159–178.

Olley, G. S., and A. Pakes (1996): "The Dynamics of Productivity in the Telecommunications Equipment Industry," *Econometrica*, 64(6), 1263–97.

Pascali, L. (2009): "Long-term persistance: the case for financial development," *mimeo*.

Percoco, M. (2009): "History, Geography and Economies of Agglomeration: Evidence from Italian Cities," *mimeo*.

Petersen, M. A., and R. G. Rajan (2002): "Does Distance Still Matter? The Information Revolution in Small Business Lending," *Journal of Finance*, 57(6), 2533–2570.

Porta, R. L., F. L. de Silanes, A. Shleifer, and R. W. Vishny (1998): "Law and Finance," *Journal of Political Economy*, 106(6), 1113–1155.

Rajan, R. G., and L. Zingales (1998): "Financial Dependence and Growth," *American Economic Review*, 88(3), 559–86.

Rousseau, P. L., and P. Wachtel (1998): "Financial Intermediation and Economic Performance: Historical Evidence from Five Industrialized Countries," *Journal of Money, Credit and Banking*, 30(4), 657–78.

Shulvass, M. A. (1973): *The Jews in the world of the Renaissance*. Spertus College of Judaica Press.

Stulz, R. M., and R. Williamson (2003): "Culture, openness, and finance," *Journal of Financial Economics*, 70(3), 313–349.

Townsend, R. (1979): "Optimal contracts and competitive markets with costly state verification," Discussion paper.

Xu, Z. (2000): "Financial Development, Investment, and Economic Growth," *Economic Inquiry*, 38(2), 331–44.

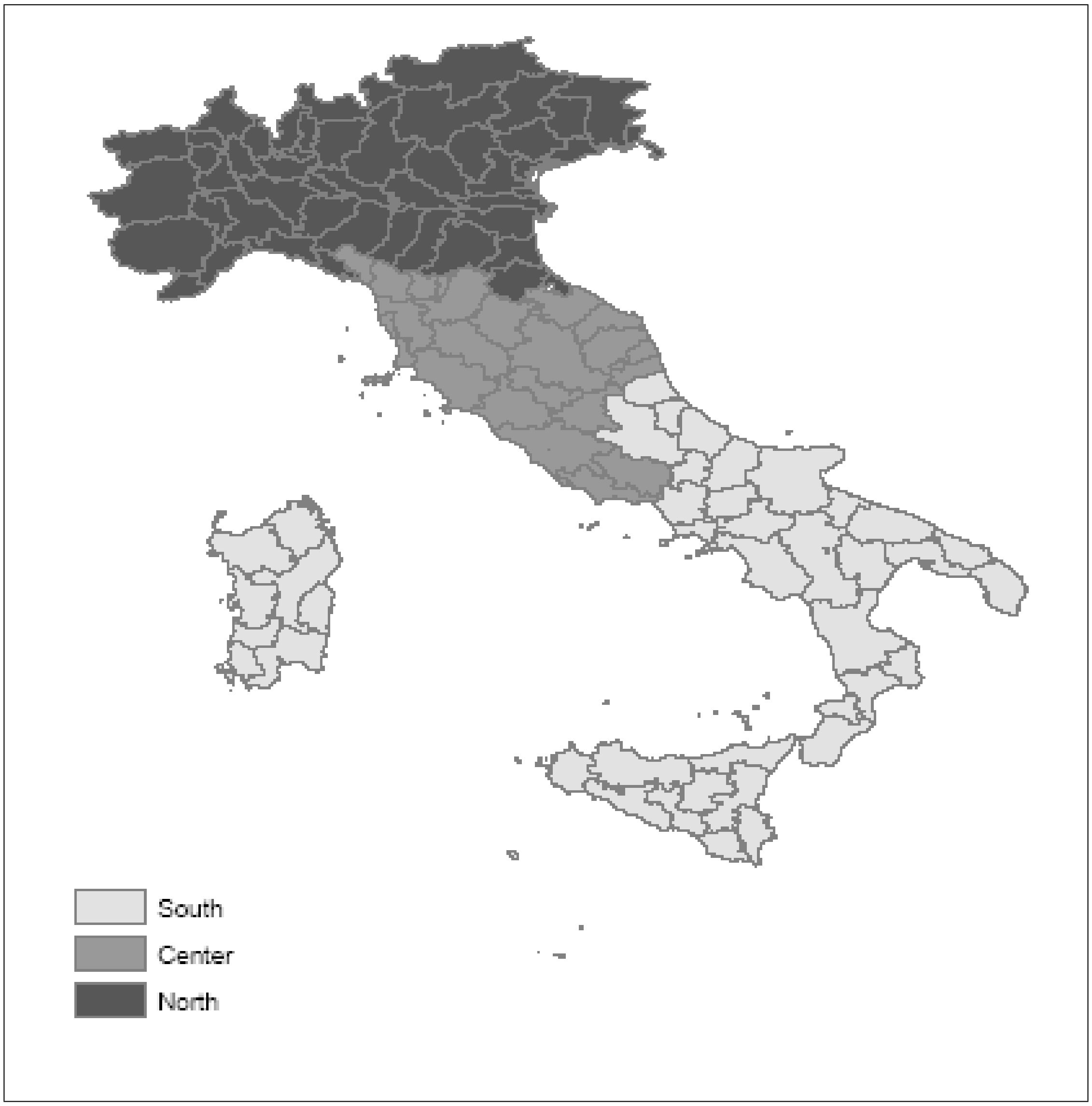

Figure 1-1: Italian Provinces and Macro-Regions

Table 1.2: Summary Statistics for the Samples Used in Estimation

PANEL A	City level data (historical)					
	Mean	*Median*	Standard *Deviation*	*Min*	*Max*	*N*
Small Jewish Community	0.21	0.00	0.41	0.00	1.00	544
Medium Jewish Community	0.10	0.00	0.31	0.00	1.00	544
Large Jewish Community	0.09	0.00	0.29	0.00	1.00	544
Population 1300	4.68	0.00	11.99	0	150	543
Population 1400	2.03	0.00	7.53	0	100	543
Population 1500	3.50	0.00	11.42	0	150	543
Population 1600	5.49	0.00	17.51	0	280	543
Population 1700	5.26	0.00	15.92	0	220	543
Population 1800	8.30	6.00	19.81	0	320	543

PANEL B	City level data					
	Mean	*Median*	Standard *Deviation*	*Min*	*Max*	*N*
GDP per Capita	15.96	14.92	6.38	4.59	37.54	542
Credit /GDP	0.69	0.53	0.60	0.03	4.71	448
Branches /Population	0.0005	0.0004	0.0002	0.0001	0.0012	540
Altimetry Min	0.11	0.06	0.14	-0.00	0.80	542
Altimetry Max	0.76	0.64	0.61	0.00	3.32	542
Altimetry Average	0.29	0.23	0.26	0.00	1.12	542
Seismicity	0.00	0.00	0.00	0.00	0.00	542
Sea	0.13	0.00	0.34	0.00	1.00	541
Close to Sea	0.12	0.00	0.33	0.00	1.00	541
Province Capital	0.17	0.00	0.38	0.00	1.00	542
City Area	0.12	0.09	0.12	0.00	1.50	542

PANEL C	Firm level data: Amadeus database					
	Mean	*Median*	Standard *Deviation*	*Min*	*Max*	*N*
Sales	15900	3530	99700	13	6400000	92316
Net Value of Capital	3120	495	22500	1	2010000	92316
Wages	2230	585	11400	1	619000	92316
Cost of Intermediates	12200	2570	83000	11	6000000	92316
Employees	71.03	24.00	310.94	1	18100	76301

Note:Panel A reports statistics on historical data. The sample is limited to the towns in Italy that had a population of at least 5000 people in 1861. Historical data on urban populations are in thousands (source: Malanima (1998)). Data on the size of Jewish demography refer to the end of the fifteenth century (source: Milano (1963)). In Panel B, "GDP per Capita" is the per capita value added in 2002 in the municipality, expressed in thousands of euros (source: INSTAT). "credit/GDP" is the ratio of claims on nonfinancial private sector to GDP in the municipality (source: Bank of Italy). "Brances/Population" is the ratio of the number of bank branches to residents (source: Bank of Italy). Altimetry is expressed in thousands of meters (source: ISTAT). "Sea" is a dummy that identifies cities on the sea; "Close to Sea" is a dummy that identifies cities that are less than 5 miles from the sea. "Province Capital" is a dummy variable equal to one if the city is the capital of its province (year 2002). "City Area" is the extension of the municipality in square meters (year 1991). Panel C reports statistics for the Amadeus firm-level data in the year 2005. "Sales", "Net value of capital", "Wages" are expressed in thousands of euros.

Table 1.3: Jewish communities in Renaissance and current credit availability

	(1) CREDIT	(2) CREDIT	(3) CREDIT	(4) CREDIT	(5) CREDIT
JEW	40.58***	31.97***	32.01***	15.48***	15.68***
	(3.721)	(3.195)	(3.259)	(3.100)	(3.126)
PROVINCE CAPITAL				53.76***	54.37***
				(3.488)	(3.707)
AREA					-0.00628
					(0.0128)
_cons	53.08***	33.65	33.74	1.999	2.645
	(3.727)	(24.17)	(24.54)	(22.00)	(22.05)
GEO. CHARACTERISTICS	NO	NO	YES	YES	YES
PROVINCE DUMMIES	NO	YES	YES	YES	YES
YEAR DUMMIES	YES	YES	YES	YES	YES
r2_a	0.104	0.566	0.572	0.659	0.659
N	1020	1020	1020	1020	1020
SAMPLE	All	All	All	All	All

The table reports OLS estimates for the years 2002-2004. The unit of observation is the municipality. The left hand side variable, CREDIT, is the ratio of claims on nonfinancial private sector to GDP in the municipality; GDP is GDP per capita and is imputed by looking at the per capita GDP in the "local labor system" to which the municipality belongs to. JEW is a dummy variable equal to one if the municipality hosted a Jewish community in 1450 A.D. PROVINCE CAPITAL is a dummy variable equal to one if the city is the capital of its province. AREA is the extension of the municipality in square meters (data refers to 1991). The set of PROVINCE DUMMIES refers to the Italian provinces in 1992. Geographical variables are elevation of the municipality (minimum, maximum), seismicity (as reported by the Italian national statistical institute) and two dummies for whether the city is located on the coast or close to the cost (less than 5 miles distant). Standard errors are reported in parentheses. *** significant at less than 1 percent; ** significant at 5 percent; * significant at 10 percent.

Table 1.4: Jewish communities in Renaissance and current credit availability

	(1) CREDIT	(2) CREDIT	(3) CREDIT	(4) CREDIT	(5) CREDIT	(6) CREDIT	(7) CREDIT
JEW	5.667	6.246	6.197	-20.73	5.177	2.111	2.473
	(3.701)	(3.801)	(3.866)	(16.53)	(3.704)	(3.600)	(3.750)
JEW*STAY	23.63***	24.57***	24.62***	37.50**	11.45*	11.20*	24.36***
	(5.792)	(5.910)	(5.965)	(18.51)	(5.916)	(5.718)	(5.747)
STAY	-28.43**	-27.10**	-27.14**	-26.83	-17.23	0.147	-8.485
	(12.55)	(12.80)	(12.83)	(17.40)	(12.22)	(12.00)	(12.55)
PROVINCE CAPITAL	52.72***	51.77***	51.67***	49.63***	38.56***	15.79***	25.68***
	(3.436)	(3.511)	(3.753)	(6.885)	(3.771)	(4.599)	(4.738)
AREA			0.000903	0.0342**	-0.0176	-0.0105	0.00952
			(0.0128)	(0.0171)	(0.0124)	(0.0121)	(0.0124)
POP1300					1.289***	1.085***	
					(0.271)	(0.263)	
POP1400					0.390	0.558	
					(0.493)	(0.477)	
POP1500					-0.329*	-0.275	
					(0.173)	(0.168)	
EDUC						23.33***	25.81***
						(2.875)	(3.041)
GEO. CHARACTERISTICS	NO	YES	YES	YES	YES	YES	YES
PROVINCE DUMMIES	YES	YES	YES	YES	YES	YES	YES
YEAR DUMMIES	YES	YES	YES	YES	YES	YES	YES
r2_a	0.664	0.665	0.665	0.549	0.706	0.725	0.689
N	1020	1020	1020	234	1020	1020	1020
SAMPLE	All	All	All	Central Italy	All	All	All

The table reports OLS estimates for the years 2002-2004. The unit of observation is the municipality. The left hand side variable, CREDIT, is the ratio of claims on nonfinancial private sector to GDP in the municipality; GDP is GDP per capita and is imputed by looking at per capita GDP in the "local labor system" to which the municipality belongs to. JEW is a dummy variable equal to one if the municipality hosted a Jewish community in 1450 A.D. STAY is a dummy equal to one if the municipality was under the Aragon crown in 1500 A.D. PROVINCE CAPITAL is a dummy variable equal to one if the city is the capital of its province. AREA is the extension of the municipality in square meters (data refers to 1991). POP1300, POP1400 and POP1500 are respectively the estimated urban population in the municipality in 1300, 1400 and 1500. EDUC is the average number of years of education of the local residents. The set of PROVINCE DUMMIES refers to the Italian provinces in 1992. Geographical variables are elevation of the municipality (minimum, maximum), seismicity (as reported by the Italian national statistical institute) and two dummies for whether the city is located on the coast or close to the cost (less than 5 miles distant). Standard errors are reported in parentheses. *** significant at less than 1 percent; ** significant at 5 percent; * significant at 10 percent.

Table 1.5: Jewish communities in the Renaissance and current economic development (validating the instrument)

	(1) LGDP	(2) LGDP	(3) LGDP	(4) LGDP	(5) LGDP	(6) LGDP
LCREDIT	0.101***	0.0149	0.0135	-0.0187	-0.0459***	-0.102***
	(0.0138)	(0.0169)	(0.0165)	(0.0167)	(0.0163)	(0.0171)
AREA				0.369***	0.198***	0.188***
				(0.0610)	(0.0626)	(0.0577)
REGION CAPITAL					0.259***	
					(0.0363)	
PROVINCE CAPITAL						0.191***
						(0.0187)
GEO. CHARACTERISTICS	NO	NO	YES	YES	YES	YES
PROVINCE DUMMIES	NO	YES	YES	YES	YES	YES
YEAR DUMMIES	YES	YES	YES	YES	YES	YES
r2_a	0.0983	0.380	0.445	0.487	0.539	0.585
N	510	510	510	510	510	510
SAMPLE	Cities not under Aragon crown in 1500 A.D.					

The table reports OLS estimates for the years 2002-2004. The unit of observation is the municipality. The left hand side variable, LGDP is the log of GDP per capita and is imputed by looking at the per capita GDP in the "local labor system" to which the municipality belongs to. LCREDIT, is the log ratio of claims on nonfinancial private sector to GDP in the municipality. REGION CAPITAL is a dummy variable equal to one if the city is the capital of the region. PROVINCE CAPITAL is a dummy variable equal to one if the city is the capital of its province. AREA is the extension of the municipality in square meters (data refers to 1991). The set of PROVINCE DUMMIES refers to the Italian provinces in 1992. Geographical variables are elevation of the municipality (minimum, maximum), seismicity (as reported by the Italian national statistical institute) and two dummies for whether the city is located on the coast or close to the cost (less than 5 miles distant). Standard errors are reported in parentheses. *** significant at less than 1 percent; ** significant at 5 percent; * significant at 10 percent.

Table 1.6: Jewish communities in the Renaissance and current economic development (validating the instrument)

	(1) GDP	(2) GDP	(3) GDP	(4) GDP	(5) GDP	(6) GDP
JEW1130	0.0607	0.0314	-0.0414	-0.0503	0.0264	0.0200
	(0.0491)	(0.0464)	(0.0400)	(0.0409)	(0.0476)	(0.0531)
JEWROMAN						0.0204
						(0.0739)
REGION CAPITAL		0.483***				
		(0.0886)				
PROVINCE CAPITAL			0.493***	0.639***	0.677***	0.670***
			(0.0441)	(0.0466)	(0.0505)	(0.0561)
AREA					0.0126	0.0006
					(0.210)	(0.210)
GEO. CHARACTERISTICS	NO	NO	NO	NO	YES	YES
PROVINCE DUMMIES	NO	NO	NO	YES	YES	YES
YEAR DUMMIES	YES	YES	YES	YES	YES	YES
r2_a	-0.00325	0.118	0.371	0.567	0.619	0.617
N	213	213	213	213	213	213
SAMPLE	Cities that had a Jewish communities in 1450 and were under the Aragons in 1500					

The table reports OLS estimates for the years 2002-2004. The unit of observation is the municipality. The left hand side variable, GDP, is imputed by looking at the per capita GDP in the "local labor system" to which the municipality belongs to. JEW1130 is a dummy variable equal to one if the municipality hosted a Jewish community in 1130 A.D.; JEWROMAN is a dummy variable equal to one if the municipality hosted a Jewish community during the Roman Empire. PROVINCE CAPITAL is a dummy variable equal to one if the city is the capital of its province. REGION CAPITAL is a dummy variable equal to one if the city is the capital of the region. AREA is the extension of the municipality in square meters (data refers to 1991). The set of PROVINCE DUMMIES refers to the Italian provinces in 1992. Geographical variables are elevation of the municipality (minimum, maximum), seismicity (as reported by the Italian national statistical institute) and two dummies for whether the city is located on the coast or close to the cost (less than 5 miles distant). Standard errors are reported in parentheses. *** significant at less than 1 percent; ** significant at 5 percent; * significant at 10 percent.

Table 1.7: Jewish communities in the Renaissance and current economic development (validating the instrument)

	(1) GDP	(2) GDP	(3) GDP	(4) GDP	(5) GDP	(6) GDP	(7) GDP	(8) GDP
JEW	0.0751**	0.0304	-0.0107	-0.0223	-0.0573*	-0.0975	-0.0322	-0.0695**
	(0.0320)	(0.0325)	(0.0328)	(0.0317)	(0.0308)	(0.104)	(0.0328)	(0.0318)
JEW*STAY	0.151***	0.218***	0.224***	0.183***	0.117**	0.421***	0.132**	0.182***
	(0.0526)	(0.0519)	(0.0512)	(0.0496)	(0.0486)	(0.116)	(0.0530)	(0.0490)
STAY	0.169	0.197*	0.170	0.208*	0.158	0.0704	0.215*	0.277***
	(0.115)	(0.112)	(0.111)	(0.107)	(0.104)	(0.108)	(0.110)	(0.106)
AREA			0.604***	0.308***	0.160	0.539***	0.398***	0.423***
			(0.104)	(0.106)	(0.104)	(0.118)	(0.107)	(0.101)
REGION CAP.				0.532***				
				(0.0592)				
PROVINCE CAP.					0.395***			
					(0.0312)			
POP1300							0.00822***	
							(0.00250)	
POP1400							-0.00130	
							(0.00461)	
POP1500							0.000619	
							(0.00163)	
EDUC								0.180***
								(0.0172)
GEO. CHARACT.	NO	YES	YES	YES	YES	YES	YES	YES
PROVINCE DUM.	YES	YES	YES	YES	YES	YES	YES	YES
YEAR DUM.	YES	YES	YES	YES	YES	YES	YES	YES
r2_a	0.678	0.700	0.708	0.728	0.745	0.576	0.723	0.735
N	1185	1185	1185	1185	1185	249	1185	1185
SAMPLE	All	All	All	All	All	Central Italy	All	All

The table reports OLS estimates for the years 2002-2004. The unit of observation is the municipality. The left hand side variable, GDP, is GDP per capita and is imputed by looking at the per capita GDP in the "local labor system" to which the municipality belongs to. LCREDIT, is the log ratio of claims on nonfinancial private sector to GDP in the municipality. JEW is a dummy variable equal to one if the municipality hosted a Jewish community in 1450 A.D. STAY is a dummy equal to one if the municipality was under the Aragon crown in 1500 A.D. PROVINCE CAPITAL is a dummy variable equal to one if the city is the capital of its province. REGION CAPITAL is a dummy variable equal to one if the city is the capital of the region. AREA is the extension of the municipality in square meters (data refers to 1991). POP1300, POP1400 and POP1500 are respectively the estimated urban population in the municipality in 1300, 1400 and 1500. EDUC is the average number of years of education of the local residents. The set of PROVINCE DUMMIES refers to the Italian provinces in 1992. Geographical variables are elevation of the municipality (minimum, maximum), seismicity (as reported by the Italian national statistical institute) and two dummies for whether the city is located on the coast or close to the cost (less than 5 miles distant). Column 6 refers only to municipalities in Central Italy (e.g. in the following regions: Lazio, Toscana, Umbria, Marche). Standard errors are reported in parentheses. *** significant at less than 1 percent; ** significant at 5 percent; * significant at 10 percent.

Table 1.8: credit availability and economic development (IV estimates)

PANEL A	(1) LGDP	(2) LGDP	(3) LGDP	(4) LGDP	(5) LGDP	(6) LGDP
LCREDIT	0.253***	0.259***	0.268***	0.256***	0.212***	0.209*
	(5.71)	(5.92)	(6.35)	(4.93)	(3.80)	(1.86)
AREA				0.0494	0.00261	0.0618
				(0.54)	(0.03)	(0.74)
REGION CAPITAL					0.124**	
					(2.48)	
PROVINCE CAPITAL						0.0291
						(0.47)
GEO. CHARACTERISTICS	NO	NO	YES	YES	YES	YES
PROVINCE DUMMIES	NO	YES	YES	YES	YES	YES
YEAR DUMMIES	YES	YES	YES	YES	YES	YES
sarganp	0.0807	0.250	0.478	0.493	0.780	0.497
N	510	510	510	510	510	510
SAMPLE	Cities not under Aragon crown in 1500 A.D.					

PANEL B	(1) LCREDIT	(2) LCREDIT	(3) LCREDIT	(4) LCREDIT	(5) LCREDIT	(6) LCREDIT
JEWsmall	0.144**	0.245***	0.267***	0.279***	0.261***	0.193***
	(2.41)	(4.47)	(4.70)	(5.00)	(4.70)	(3.62)
JEWmedium	0.441***	0.554***	0.575***	0.503***	0.473***	0.205***
	(5.69)	(7.99)	(8.42)	(7.25)	(6.81)	(2.74)
JEWlarge	0.704***	0.748***	0.728***	0.596***	0.482***	0.106
	(7.05)	(8.21)	(8.20)	(6.39)	(4.85)	(0.99)
R_2	0.0812	0.5523	0.5967	0.6101	0.6192	0.6200

The table reports 2SLS estimates for the years 2002-2004. The unit of observation is the municipality. The sample is limited to municipalities that were not under the Aragon crown in 1500. Panel A reports the second stage estimates. The left hand side variable, LGDP is the log of GDP per capita and is imputed by looking at the per capita GDP in the "local labor system" to which the municipality belongs to. LCREDIT, is the log ratio of claims on nonfinancial private sector to GDP in the municipality. REGION CAPITAL is a dummy variable equal to one if the city is the capital of the region. PROVINCE CAPITAL is a dummy variable equal to one if the city is the capital of its province. AREA is the extension of the municipality in square meters (data refers to 1991). The set of PROVINCE DUMMIES refers to the Italian provinces in 1992. Geographical variables are elevation of the municipality (minimum, maximum), seismicity (as reported by the Italian national statistical institute) and two dummies for whether the city is located on the coast or close to the cost (less than 5 miles distant). Panel B reports the first stage estimates. To save space only the coefficients on instruments are reported. JEWsmall, JEWmedium and JEWlarge are dummy variables equal to one if the municipality hosted respectively a small, a medium or a large Jewish community in 1450. Standard errors are reported in parentheses. *** significant at less than 1 percent; ** significant at 5 percent; * significant at 10 percent.

Table 1.9: Credit availability and economic development (IV estimates - robustness checks)

PANEL A	(1) LGDP	(2) LGDP	(3) LGDP	(4) LGDP	(5) LGDP	(6) LGDP
LCREDIT	0.203***	0.244***	0.260***	0.254***	0.223***	0.241**
	(0.0629)	(0.0562)	(0.0521)	(0.0588)	(0.0634)	(0.122)
AREA				0.0511	-0.00588	0.0487
				(0.0964)	(0.0869)	(0.0884)
REGION CAPITAL					0.119**	
					(0.0531)	
PROVINCE CAPITAL						0.0123
						(0.0667)
GEO. CHARACTERISTICS	NO	NO	YES	YES	YES	YES
PROVINCE DUMMIES	NO	YES	YES	YES	YES	YES
YEAR DUMMIES	YES	YES	YES	YES	YES	YES
N	510	510	510	510	510	510
SAMPLE	Cities not under Aragon crown in 1500 A.D.					

PANEL B	(1) LCREDIT	(2) LCREDIT	(3) LCREDIT	(4) LCREDIT	(5) LCREDIT	(6) LCREDIT
JEW	0.295***	0.389***	0.424***	0.377***	0.339***	0.195***
	(0.0553)	(0.0522)	(0.0524)	(0.0508)	(0.0509)	(0.0508)
r2_a	0.0507	0.531	0.564	0.599	0.612	0.662

The table reports 2SLS estimates for the years 2002-2004. The unit of observation is the municipality. The sample is limited to municipalities that were not under the Aragon crown in 1500. Panel A reports the second stage estimates. The left hand side variable, LGDP is the log of GDP per capita and is imputed by looking at the per capita GDP in the "local labor system" to which the municipality belongs to. LCREDIT, is the log ratio of claims on nonfinancial private sector to GDP in the municipality. REGION CAPITAL is a dummy variable equal to one if the city is the capital of the region. PROVINCE CAPITAL is a dummy variable equal to one if the city is the capital of its province. AREA is the extension of the municipality in square meters (data refers to 1991). The set of PROVINCE DUMMIES refers to the Italian provinces in 1992. Geographical variables are elevation of the municipality (minimum, maximum), seismicity (as reported by the Italian national statistical institute) and two dummies for whether the city is located on the coast or close to the cost (less than 5 miles distant). Panel B reports the first stage estimates. To save space only the coefficients on instruments are reported. JEW is a dummy variable equal to one if the municipality hosted a Jewish community in 1450 A.D. Standard errors are reported in parentheses. *** significant at less than 1 percent; ** significant at 5 percent; * significant at 10 percent.

Table 1.10: Branch density and economic development (IV estimates)

PANEL A	(1) LGDP	(2) LGDP	(3) LGDP	(4) LGDP	(5) LGDP	(6) LGDP
LCREDIT	0.713***	1.336***	1.341***	1.360***	1.166***	1.523
	(0.143)	(0.284)	(0.293)	(0.403)	(0.405)	(1.021)
AREA				-0.0137	-0.0462	-0.0199
				(0.159)	(0.140)	(0.190)
REGION CAPITAL					0.119*	
					(0.0715)	
PROVINCE CAPITAL						-0.0292
						(0.117)
GEO. CHARACTERISTICS	NO	NO	YES	YES	YES	YES
PROVINCE DUMMIES	NO	YES	YES	YES	YES	YES
YEAR DUMMIES	YES	YES	YES	YES	YES	YES
sarganp	0.0354	0.764	0.829	0.835	0.825	0.857
N	519	519	519	519	519	519
SAMPLE	Cities not under Aragon crown in 1500 A.D.					

PANEL B	(1) LCREDIT	(2) LCREDIT	(3) LCREDIT	(4) LCREDIT	(5) LCREDIT	(6) LCREDIT
JEWsmall	0.0525	0.0631**	0.0595**	0.0584**	0.0565**	0.0396
	(0.0321)	(0.0246)	(0.0252)	(0.0250)	(0.0251)	(0.0253)
JEWmedium	0.190***	0.117***	0.116***	0.0941***	0.0897***	0.0294
	(0.0420)	(0.0316)	(0.0319)	(0.0328)	(0.0332)	(0.0374)
JEWlarge	0.228***	0.180***	0.169***	0.127***	0.110**	0.0205
	(0.0541)	(0.0415)	(0.0411)	(0.0442)	(0.0474)	(0.0534)
r2_a	0.0506	0.670	0.679	0.683	0.683	0.690

The table reports 2SLS estimates for the years 2002-2004. The unit of observation is the municipality. The sample is limited to municipalities that were not under the Aragon crown in 1500. Panel A reports the second stage estimates. The left hand side variable, LGDP is the log of GDP per capita and is imputed by looking at the per capita GDP in the "local labor system" to which the municipality belongs to. LCREDIT, is the log ratio of bank branches to residents. REGION CAPITAL is a dummy variable equal to one if the city is the capital of the region. PROVINCE CAPITAL is a dummy variable equal to one if the city is the capital of its province. AREA is the extension of the municipality in square meters (data refers to 1991). The set of PROVINCE DUMMIES refers to the Italian provinces in 1992. Geographical variables are elevation of the municipality (minimum, maximum), seismicity (as reported by the Italian national statistical institute) and two dummies for whether the city is located on the coast or close to the cost (less than 5 miles distant). Panel B reports the first stage estimates. To save space only the coefficients on instruments are reported. JEWsmall, JEWmedium and JEWlarge are dummy variables equal to one if the municipality hosted respectively a small, a medium or a large Jewish community in 1450 Standard errors are reported in parentheses. *** significant at less than 1 percent; ** significant at 5 percent; * significant at 10 percent.

Table 1.11: Credit availability and economic development (IV estimates - robustness checks)

PANEL A	(1) LGDP	(2) LGDP	(3) LGDP	(4) LGDP	(5) LGDP	(6) LGDP	(7) LGDP	(8) LGDP	(9) LGDP
LCREDIT	0.485***	0.460***	0.369*	0.174***	0.177***	0.179*	0.310**	0.309*	0.273*
	(0.148)	(0.163)	(0.192)	(0.0651)	(0.0650)	(0.0997)	(0.157)	(0.160)	(0.150)
AREA	-0.270	-0.287	-0.119	0.0216	-0.00869	0.00569	-0.0606	-0.108	-0.0907
	(0.192)	(0.185)	(0.53)	(0.0881)	(0.0890)	(0.0858)	(0.138)	(0.135)	(0.107)
REGION CAP.		0.0632			0.0894*			0.205***	
		(0.107)			(0.0496)			(0.0706)	
PROVINCE CAP.			-0.0235			0.0152			0.0920
			(0.0999)			(0.0467)			(0.0841)
POP1300				0.000709	0.000502	0.000441	-0.00241	-0.00280	-0.00253
				(0.00178)	(0.00178)	(0.00180)	(0.00352)	(0.00349)	(0.00289)
POP1400				-0.00343	-0.00322	-0.00305	0.00716	0.00639	0.00642
				(0.00362)	(0.00362)	(0.00359)	(0.00569)	(0.00580)	(0.00528)
POP1500				0.00434	0.00349	0.00408	-0.00196	-0.00292	-0.00174
				(0.00272)	(0.00276)	(0.00273)	(0.00184)	(0.00178)	(0.00173)
JEWsmall							-0.0468	-0.0402	-0.0423
							(0.0409)	(0.0419)	(0.0372)
JEWmedium							-0.0678	-0.0669	-0.0779
							(0.0702)	(0.0710)	(0.0506)
JEWlarge							0.00121	-0.00590	-0.0443
							(0.0952)	(0.0951)	(0.0586)
GEO. CHARACT.	YES	YES	YES	YES	YES	YES	YES	YES	YES
PROVINCE DUM.	YES	YES	YES	YES	YES	YES	YES	YES	YES
YEAR DUM.	YES	YES	YES	YES	YES	YES	YES	YES	YES
sarganp	0.154	0.0850	0.0255	0.925	0.892	0.756	0.0605	0.0191	0.0561
N	210	210	210	510	510	510	1020	1020	1020
SAMPLE	Cities not under Aragon in Central Italy			Cities not under Aragon crown in 1500 A.D.				All cities	

PANEL B	(1) LCREDIT	(2) LCREDIT	(3) LCREDIT	(4) LCREDIT	(5) LCREDIT	(6) LCREDIT	(7) LCREDIT	(8) LCREDIT	(9) LCREDIT
JEWsmall	0.246***	0.243***	0.212***	0.205***	0.209***	0.160***			
	(0.0863)	(0.0867)	(0.0785)	(0.0592)	(0.0599)	(0.0570)			
JEWmedium	0.223*	0.222*	0.00244	0.429***	0.429***	0.206***			
	(0.114)	(0.114)	(0.109)	(0.0686)	(0.0687)	(0.0742)			
JEW_large	0.546***	0.507***	0.00263	0.249**	0.249**	-0.0680			
	(0.140)	(0.156)	(0.153)	(0.119)	(0.120)	(0.124)			
JEWsmall*STAY							-0.0615	-0.0439	-0.142
							(0.102)	(0.102)	(0.0987)
JEWmedium*STAY							0.103	0.0905	-0.204
							(0.127)	(0.127)	(0.128)
JEWlarge*STAY							0.529***	0.524***	0.570***
							(0.178)	(0.178)	(0.172)
r2_a	0.655	0.654	0.716	0.646	0.645	0.676	0.627	0.628	0.653

The table reports 2SLS estimates for the years 2002-2004. The unit of observation is the municipality. Panel A reports the second stage estimates. The left hand side variable, LGDP is the log of GDP per capita and is imputed by looking at the per capita GDP in the "local labor system" to which the municipality belongs to. LCREDIT, is the log ratio of claims on nonfinancial private sector to GDP in the municipality. REGION CAPITAL is a dummy variable equal to one if the city is the capital of the region. PROVINCE CAPITAL is a dummy variable equal to one if the city is the capital of its province. AREA is the extension of the municipality in square meters (data refers to 1991). The set of PROVINCE DUMMIES refers to the Italian provinces in 1992. Geographical variables are elevation of the municipality (minimum, maximum), seismicity (as reported by the Italian national statistical institute) and two dummies for whether the city is located on the coast or close to the cost (less than 5 miles distant). JEWsmall, JEWmedium and JEWlarge are dummy variables equal to one if the municipality hosted respectively a small, a medium or a large Jewish community in 1450. Panel B reports the first stage estimates. To save space only the coefficients on instruments are reported. STAY is a dummy equal to one if the municipality was under the Aragon crown in 1500 A.D. Standard errors are reported in parentheses. *** significant at less than 1 percent; ** significant at 5 percent; * significant at 10 percent.

Table 1.12: Credit availability and aggregate productivity (OLS estimates).

	(1) LTFP	(2) LTFP	(3) LTFP	(4) LTFP	(5) LTFP
LCREDIT	0.00725	0.00726	0.00192	-0.0133	-0.000934
	(0.0132)	(0.0137)	(0.0141)	(0.0149)	(0.0144)
REGION CAPITAL			0.0709		
			(0.0443)		
PROVINCE CAPITAL				0.0745***	
				(0.0230)	
POP1300					0.00187
					(0.00184)
POP1400					-0.000702
					(0.00330)
POP1500					-0.0000355
					(0.00121)
GEO. CHARACTERISTICS	NO	YES	YES	YES	YES
PROVINCE DUMMIES	YES	YES	YES	YES	YES
N	381	381	381	381	381
r2_a	0.977	0.977	0.977	0.978	0.977
SAMPLE	Cities not under Aragon crown in 1500 A.D.				

The table reports OLS estimates for the years 2002-2004. The unit of observation is the municipality. The sample is limited to municipalities that were not under the Aragon crown in 1500. The left hand side variable, LTFP is the log of the aggregate productivity of the firms operating in the municipality. It is computed by aggregating the TFP estimated for the firms that operate in the municipality using their share out of total value added as weights. LCREDIT, is the log ratio of claims on nonfinancial private sector to GDP in the municipality. REGION CAPITAL is a dummy variable equal to one if the city is the capital of the region. PROVINCE CAPITAL is a dummy variable equal to one if the city is the capital of its province. POP1300, POP1400 and POP1500 are respectively the estimated urban population in the municipality in 1300, 1400 and 1500. The set of PROVINCE DUMMIES refers to the Italian provinces in 1992. Geographical variables are elevation of the municipality (minimum, maximum), seismicity (as reported by the Italian national statistical institute) and two dummies for whether the city is located on the coast or close to the cost (less than 5 miles distant). Standard errors are reported in parentheses. *** significant at less than 1 percent; ** significant at 5 percent; * significant at 10 percent.

Table 1.13: Credit availability and aggregate productivity (IV estimates)

	(1) LTFP	(2) LTFP	(3) LTFP	(4) LTFP	(5) LTFP
LCREDIT	0.114***	0.112***	0.112***	0.108	0.135***
	(0.0328)	(0.0336)	(0.0392)	(0.0679)	(0.0513)
REGION CAPITAL			-0.0115		
			(0.0499)		
PROVINCE CAPITAL				-0.00506	
				(0.0488)	
POP1300					-0.000519
					(0.00199)
POP1400					0.00238
					(0.00342)
POP1500					-0.00177
					(0.00137)
GEO. CHARACTERISTICS	NO	YES	YES	YES	YES
PROVINCE DUMMIES	YES	YES	YES	YES	YES
N	381	381	381	381	381
sarganp	0.574	0.373	0.343	0.317	0.444
SAMPLE	Cities not under Aragon crown in 1500 A.D.				

The table reports 2SLS estimates for the years 2002-2004. The unit of observation is the municipality. The sample is limited to municipalities that were not under the Aragon crown in 1500. The left hand side variable, LTFP is the log of the aggregate productivity of the firms operating in the municipality. It is computed by aggregating the TFP estimated for the firms that operate in the municipality using their share out of total value added as weights. LCREDIT, is the log ratio of claims on nonfinancial private sector to GDP in the municipality. REGION CAPITAL is a dummy variable equal to one if the city is the capital of the region. PROVINCE CAPITAL is a dummy variable equal to one if the city is the capital of its province. POP1300, POP1400 and POP1500 are respectively the estimated urban population in the municipality in 1300, 1400 and 1500. The set of PROVINCE DUMMIES refers to the Italian provinces in 1992. Geographical variables are elevation of the municipality (minimum, maximum), seismicity (as reported by the Italian national statistical institute) and two dummies for whether the city is located on the coast or close to the cost (less than 5 miles distant). The instruments are three dummies equal to one if the municipality hosted respectively a small, a medium or a large Jewish community in 1450. Standard errors are reported in parentheses. *** significant at less than 1 percent; ** significant at 5 percent; * significant at 10 percent.

Table 1.14: Branch density and aggregate productivity (IV estimates)

	(1) LTFP	(2) LTFP	(3) LTFP	(4) LTFP	(5) LTFP
LCREDIT2	0.523***	0.524***	0.511**	0.614	0.556**
	(0.190)	(0.191)	(0.208)	(0.415)	(0.250)
REGION CAPITAL			0.0109		
			(0.0574)		
PROVINCE CAPITAL				-0.0206	
				(0.0647)	
POP1300					0.00266
					(0.00228)
POP1400					-0.000991
					(0.00408)
POP1500					-0.00228
					(0.00186)
GEO. CHARACTERISTICS	NO	YES	YES	YES	YES
PROVINCE DUMMIES	YES	YES	YES	YES	YES
N	381	381	381	381	381
sarganp	0.922	0.880	0.880	0.911	0.925
SAMPLE	Cities not under Aragon crown in 1500 A.D.				

The table reports 2SLS estimates for the years 2002-2004. The unit of observation is the municipality. The sample is limited to municipalities that were not under the Aragon crown in 1500. The left hand side variable, LTFP is the log of the aggregate productivity of the firms operating in the municipality. It is computed by aggregating the TFP estimated for the firms that operate in the municipality using their share out of total value added as weights. LCREDIT, is the log ratio of bank branches to residents. REGION CAPITAL is a dummy variable equal to one if the city is the capital of the region. PROVINCE CAPITAL is a dummy variable equal to one if the city is the capital of its province. POP1300, POP1400 and POP1500 are respectively the estimated urban population in the municipality in 1300, 1400 and 1500. The set of PROVINCE DUMMIES refers to the Italian provinces in 1992. Geographical variables are elevation of the municipality (minimum, maximum), seismicity (as reported by the Italian national statistical institute) and two dummies for whether the city is located on the coast or close to the cost (less than 5 miles distant). The instruments are three dummies equal to one if the municipality hosted respectively a small, a medium or a large Jewish community in 1450. Standard errors are reported in parentheses. *** significant at less than 1 percent; ** significant at 5 percent; * significant at 10 percent.

Table 1.15: availability and average productivity (IV estimates)

	(1) average LTFP	(2) average LTFP	(3) average LTFP	(4) average LTFP	(5) average LTFP	(6) average LTFP
LCREDIT (credit/GDP)	0.0397 (0.0283)	0.0434 (0.0293)	0.0468 (0.0421)			
LCREDIT2 (branch density)				0.191 (0.133)	0.212 (0.137)	0.231 (0.177)
POP1300			0.000853 (0.00165)			0.00195 (0.00162)
POP1400			-0.000524 (0.00285)			-0.00168 (0.00289)
POP1500			-0.000786 (0.00112)			-0.00107 (0.00126)
GEO. CHARACTERISTICS	NO	YES	YES	NO	YES	YES
PROVINCE DUMMIES	YES	YES	YES	YES	YES	YES
N	381	381	381	381	381	381
sarganp	0.806	0.667	0.580	0.956	0.885	0.868
SAMPLE	Cities not under Aragon crown in 1500 A.D.					

The table reports 2SLS estimates for the years 2002-2004. The unit of observation is the municipality. The sample is limited to municipalities that were not under the Aragon crown in 1500. The left hand side variable, averageLTFP, is the log of the unweighted average of the TFP of the firms that operate in the municipality. LCREDIT is the ratio of claims to non-financial private sector to GDP in the munipality. LCREDIT2, is the log ratio of bank branches to residents. REGION CAPITAL is a dummy variable equal to one if the city is the capital of the region. PROVINCE CAPITAL is a dummy variable equal to one if the city is the capital of its province. POP1300, POP1400 and POP1500 are respectively the estimated urban population in the municipality in 1300, 1400 and 1500. The set of PROVINCE DUMMIES refers to the Italian provinces in 1992. Geographical variables are elevation of the municipality (minimum, maximum), seismicity (as reported by the Italian national statistical institute) and two dummies for whether the city is located on the coast or close to the cost (less than 5 miles distant). The instruments are three dummies equal to one if the municipality hosted respectively a small, a medium or a large Jewish community in 1450. Standard errors are reported in parentheses. *** significant at less than 1 percent; ** significant at 5 percent; * significant at 10 percent.

Table 1.16: Credit availability and reallocation (IV estimates)

	(1) REAL LTFP	(2) REAL TFP	(3) REAL LTFP	(4) REAL LTFP	(5) REAL LTFP	(6) REAL LTFP
LCREDIT (credit/GDP)	0.651**	0.548**	0.540			
	(0.279)	(0.266)	(0.378)			
LCREDIT2 (branch density)				1.893*	1.750*	1.284
				(1.079)	(1.019)	(1.244)
POP1300			0.00711			0.0138
			(0.0136)			(0.0141)
POP1400			-0.0159			-0.0218
			(0.0248)			(0.0254)
POP1500			0.00211			0.00279
			(0.00874)			(0.00942)
GEO. CHARACTERISTICS	NO	YES	YES	NO	YES	YES
PROVINCE DUMMIES	YES	YES	YES	YES	YES	YES
N	242	242	242	242	242	242
sarganp	0.242	0.558	0.472	0.126	0.396	0.318
SAMPLE	Cities not under Aragon crown in 1500 A.D.					

The table reports 2SLS estimates for the years 2002-2004. The unit of observation is the municipality. The sample is limited to municipalities that were not under the Aragon crown in 1500. The left hand side variable, REALL TFP, equals the difference between LTFP and averageLFTP (LTFP is the weighted average of the TFP of those firms that operate in the municipality using their share out of total value added as weights; averageLFTP is the unweighted average). LCREDIT is the log ratio of claims on non-financial private sector to GDP in the munipality. LCREDIT2, is the log ratio of bank branches to residents. REGION CAPITAL is a dummy variable equal to one if the city is the capital of the region. PROVINCE CAPITAL is a dummy variable equal to one if the city is the capital of its province. POP1300, POP1400 and POP1500 are respectively the estimated urban population in the municipality in 1300, 1400 and 1500. The set of PROVINCE DUMMIES refers to the Italian provinces in 1992. Geographical variables are elevation of the municipality (minimum, maximum), seismicity (as reported by the Italian national statistical institute) and two dummies for whether the city is located on the coast or close to the cost (less than 5 miles distant). The instruments are three dummies equal to one if the municipality hosted respectively a small, a medium or a large Jewish communityin 1450. Standard errors are reported in parentheses. *** significant at less than 1 percent; ** significant at 5 percent; * significant at 10 percent.

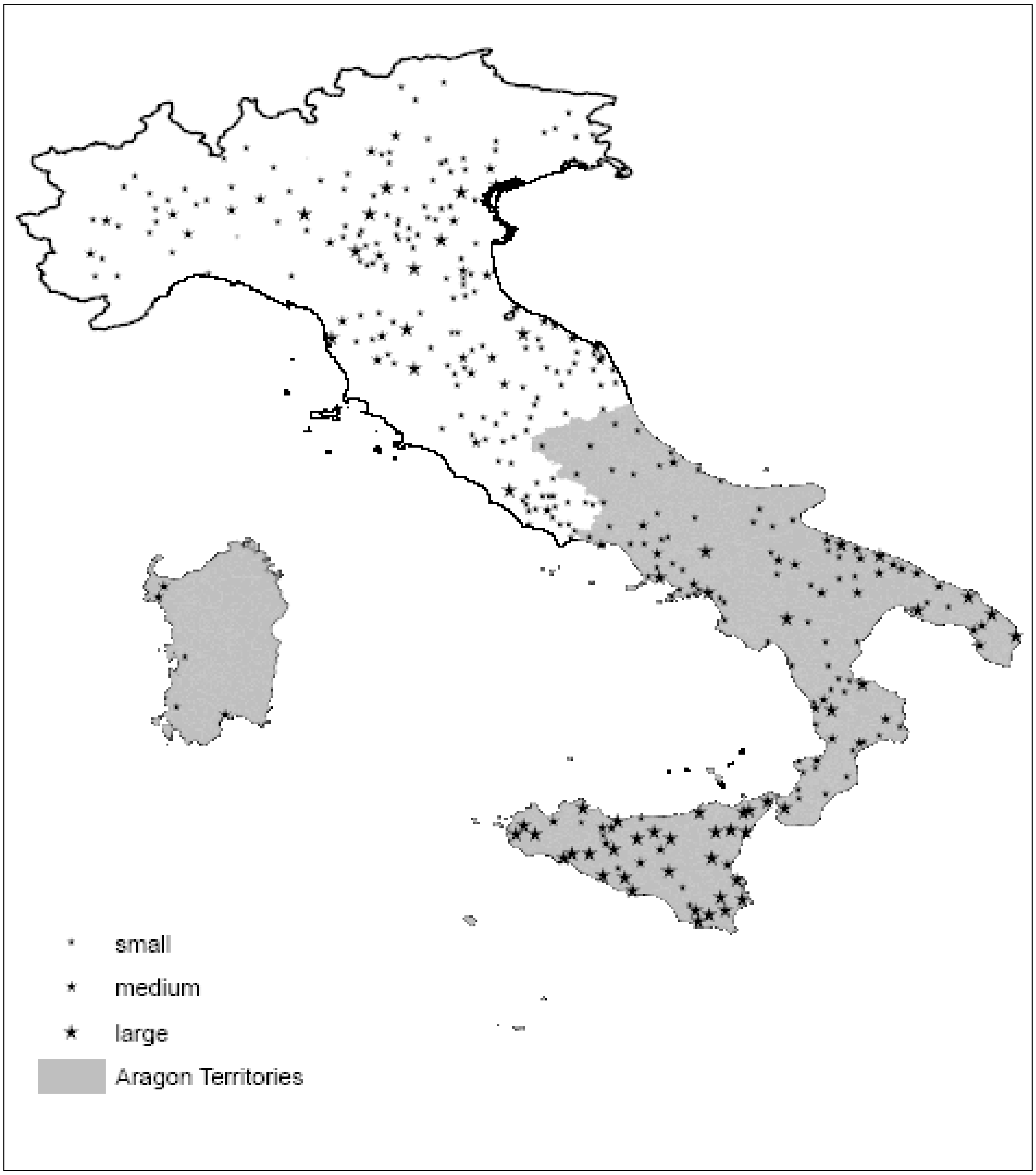

Figure 1-2: Jewish communities in 1450 A.D.

Chapter 2

Contract Incompleteness, Globalization and Vertical Structure: an Empirical Analysis

2.1 Introduction

Anecdotal evidence suggests the presence of substantial heterogeneity in the vertical structure of production across countries. Some observers relate this phenomenon to differences in institutional environments and trade openness. Khanna and Palepu (2000), for example, provide evidence that companies in India are larger and more vertically integrated than in the US and suggest that this happens because trading at arm's length is more costly in developing countries where contract enforcement is weaker. The Economist (1991) notes that the Japanese companies are more vertically integrated than the Western ones, although an increase in foreign competition is leading to a "Japanization" of the Western companies as well[1].

Nevertheless, there has never been a systematic analysis of cross-country differences in vertical integration and their causes. Peter Klein (2005) concludes his overview over the empirical studies on vertical integration in the Handbook of New Institutional Economics by saying:

> "While we know much about transaction cost determinants of vertical integration, we know relatively little about the relation of contracting and organization and the wider legal, political and social environments. The progression from single-industry case studies to cross-industry, within-country analyses, to cross-country investigations is a natural one".

The primary aim of this paper is to make a first attempt at such a cross-country analysis and to investigate the relationship between vertical integration and two important institutional characteristics: contract enforcement and trade openness.

Two well-established theories offer predictions on how differences in contracting institutions among countries could affect the vertical organization of firms. They both relate the vertical structure of firms to the "hold up" problem of underinvestment. Consider the common case where aggregate profits depend on each parties' investment and that these investments are relationship-specific, in the sense that they are sunk outside of the business relationship. If these investments are not contractible, once they have been made,

[1] Other studies have emphasized the differences between Emilia Romagna, an Italian region, and the rest of Europe (Johnston and Lawrence, 1998) and between South Korea and Taiwan (Levy,1991). Fan e al. (2007) documented differences across Chinese regions.

a potential opportunism situation arises. This can lead ex-ante to under-investment and ex-post to inefficient economic performances. Transaction costs economics (TCE) theories, pioneered by the Nobel laureate Oliver Williamson (1975, 1985), assume that vertical integration solves the hold-up problem at a fixed cost and therefore should be prevalent when contracts are harder to write. Nevertheless, this prediction is not entirely ambiguous. Property Rights Theories (PRT), developed by Grossman and Hart (1986) and Hart and Moore (1990), emphasize that vertical integration does not solve the under-investment problem since employees need to be given incentives to invest as well, and the fact that they don't own the tangible asset may weaken their incentives. In the PRT, the effect of better contracting institutions on vertical integration is not entirely clear.

Some recent contributions offer predictions on how differences in international openness affect vertical integration. Part of this literature argues that trade liberalization is a force toward vertical industrial disintegration. For example, McLaren (2000) and Grossman and Helpman (2002) model the effects of trade openness in a TCE structure. In McLaren (2000) buyer-supplier pairs are located in the same country and simultaneously choose whether to vertically integrate or outsource. The integration of a pair produces a negative externality since it thins the secondary market and reduces the outside options for non-integrated firms. In this world, trade openness partially increases the incentives to outsourcing by thickening the secondary market. Market thickening is also a reason for which trade openness increases the advantages of outsourcing in Grossman and Helpman (2002). In this model, thickened secondary markets imply lower matching costs between producer and supplier[2]. However,

[2] Antras (2003) and Antras and Helpman (2004) embeds a property right approach in a general equilibrium, factor proportion model of international trade with imperfect competition and product differentiation. The model pins down the boundaries of multinational firms as well as the international location of produc-

other contributions show that by increasing the gains from becoming a multinational corporation with respect to outsourcing domestically, trade openness may actually increase the vertical integration of domestic firms. (Ornelas and Turner, 2008; Fan et al. 2008).

Therefore, the effects of both contracting institutions and trade openness on vertical integration are potentially ambiguous and a better understanding of these relationships requires an empirical investigation.

In the first part of the paper, I present a simple theoretical model that examines how institutional features of the country contribute to shaping the governance structure of the firm. The model uses the canonical TCE "hold-up" model with some adjustments to adapt it to an international context. A final good producer makes some investments that become fully productive depending on whether the domestic supplier decides to collaborate or not; in case it does not, the domestic producer can turn to a foreign supplier.

The purpose of this model is not to provide a comprehensive theory of vertical integration but to derive a number of simple predictions to confront with the data. Different from previous literature, in an attempt to mimic the real world, the specificity of investments, the quality of contracting institutions and the openness to international trade are classified according to continuous measures. In particular, the level of specificity of the investment is modelled as the part of the investment that is unproductive without the collaboration of the supplier; contract enforcement is modelled as the probability that an ex-ante contract between supplier and producer cannot be enforced; trade barriers are modelled as the fixed cost of turning to a foreign supplier.

---

tion. A reduction of tariffs increases the propensity to international outsourcing relative to multinational vertical integration.

Comparative statics on these three variables produce a set of predictions that can be tested in the data.

First, contracting institutions per se do not affect the vertical structure of the firm. The intuition of this result can be found in the classical "hold up" theory. The fact that ex-ante contracts are rarely enforced does not distort investment decisions under outsourcing as long as the investor has sufficient ex-post bargaining power. However, when investments become very specific, the investor's outside option deteriorates and his bargaining power as well. In conclusion, the combination of greater asset specificity together with lower contract enforcement implies underinvestment under outsourcing and therefore increases the incentive to vertically integrate. In addition, the model predicts another interaction effect of asset specificity and trade barriers. The fact that the investor can find other partners in other countries limits the ability of his domestic partner to hold him up. Therefore lower trade barriers discipline the investor's partner and attenuate the distortions generated by the low quality of domestic institutions.

I test the predictions of the model using detailed data on 13,992 manufacturing firms operating in 45 countries. This dataset comes by aggregating the ICA World Bank Surveys and provides information on several characteristics of the firms in the sample allowing me to associate a measure of vertical integration and asset specificity to each of them. The main dataset is then merged with the Doing Business Database, which provides country data on institutional features, and the TRAIN Database, which provides data on tariffs.

Using cross-country data is particularly appealing for examining the effects of the openness to international trade and the quality of local contracting institutions on vertical integration. In theory, the effects of trade barriers may be studied in a cross-industry analysis since, for example, different industries may face different tariffs on intermediates. However, trade barriers are much more difficult to measure at the industry level (since they are the result of the interaction of trade costs, tariffs and other hidden barriers) than at the country level (where an average can be considered). On the other hand, contracting institutions vary essentially at country level.

The regressions show that vertical integration is less likely when asset specificity is associated with trade openness and high quality contracting institutions, thus confirming the predictions of the theoretical model. In addition, I conduct a number of robustness checks and find that the results are robust to a wide variety of specifications. Finally, to address the potential endogeneity of asset specificity, I run a battery of IV regressions; results are unchanged.

To the best of my knowledge, this is the first paper that tests the effect of trade openness on the vertical integration on a considerable number of countries. On the other side, there are two very recent contributions that use cross country data to estimate the effects of contract institutions on vertical integration. Macchiavello (2009) uses the UNIDO industry-level database to study the effects of contractual institutions and financial development on the vertical integration of firms. Industry data, however, cannot capture the intra-industry trade within vertically disintegrated firms. Acemoglu, Johnson, Mitton (2009) use instead firm level data coming from WorldBase, a database compiled for the primary purpose of providing business contacts that contains information on millions of firms around the world. The problem with this database is that the only information it provides are the firm name, the number of employees, the country where it operates, and the 4-digit SIC code of the primary industries in which the firm operates. No other information is provided. The authors have to impute the level of vertical integration of each firm, using the information coming from the US input-output accounts; asset specificity is imputed as well using US data. Thus, this empirical analysis is based on the strong assumption that technology is common across countries. Moreover, since vertical integration is imputed looking at US data, the variability in the level of vertical integration across countries will depend only on the variability in countries' industry composition. As in Macchiavello (2008), their empirical analysis can help us in understanding why sectors that have higher propensity to vertically integrate are more prominent in certain countries. However, it cannot be used to study why, within the same sector, the propensity to vertically integrate differs across countries. This is unfortunate since we would expect that the "hold up" problem would have greater effects on the level of vertical integration in a country by influencing the level of vertical integration within each industry than by influencing the industrial composition (the latter being more the result of determinants like the country's history, its natural resources and its stage of economic development).

To conclude, my work provides an empirical analysis of the institutional determinants of vertical integration using a cross-country database. It is the first one that uses a cross-country database to evaluate the role of lower tariffs and it adds further evidence on the role of better contract institutions. From both the theoretical and the empirical analysis of the paper, a policy advice emerges. Poor contract enforcement, when associated with specific assets, can distort firms' vertical structure. This can have significant welfare costs. If improving home institutions is not feasible, an equivalent solution is to reduce the trade barriers to the import of intermediates. This would discipline domestic suppliers and increase producers' incentives to invest in specific assets. In other words, reducing trade barriers is a way of "importing" foreign institutions since domestic firms will relate with each other as if the relevant contracting institutions were those of the countries where alternative suppliers are operating.

This paper is organized as follows. Section 1 details the theoretical framework and derives some testable implications. Section 2 presents the main empirical results and several robustness checks. Some concluding remarks close the paper.

2.2 The model

2.2.1 Basic structure

In this section, I present a simple model that examines how contract enforcement institutions, trade barriers and asset specificity interact to define the governance structure of a firm. The purpose of this model is not to provide a comprehensive theory of vertical integration but to derive a number of simple predictions to take to the data.

A final good producer (P) in Home (H) wants to buy an input which enhances the productivity of his investments. There is a specific supplier (HS) whose characteristics are most suitable to provide the input to firm P and is located in H as well. P could either outsource to HS or vertically integrate with her.

Under outsourcing, the two parties write a contract on the price of the intermediate good before the specific investment is realized. However, due to contract incompleteness, there is some probability that this agreement is broken after the specific investment has been realized. At this point, a new agreement has to be reached. However, the bargaining power of HS is much higher than before, because the producer specific investment is sunk without the intermediate good. The amount HS can hold-up P depends on the possible alternatives that the latter has to buy an analogous intermediate. I assume that P can purchase the same intermediate from a foreign supplier FS located in a competitive market. Ex-ante (e.g. before the specific investment has been taken), the price of the foreign intermediate, p_F, is a random variable with randomness reflecting both shocks to the productivity of FS and shocks to the exchange rates. Moreover, when importing an intermediate, the producer has to pay a trade cost t. Ex-ante, the probability of finding an alternative intermediate depends on the trade costs. If this probability is very low, for example because trade barriers are too high and buying in another country is not feasible, then the producer knows that most of the revenues coming from his investment can be expropriated by HS. Then the producer would have lower incentives to invest and this would imply a suboptimal level of investments and ex-post inefficient economic performances.

Under vertical integration, the two parties merge into a single firm. As in Hart and Tirole (1990), this "permits profit-sharing between upstream and downstream units so all conflicts of interest about prices and trading policies are removed". The advantage of this option is that an efficient level of specific investments is realized; the disadvantage is that it requires a fixed cost. The presence of a fixed cost related to the vertical integration choice is a common feature in this literature (see Hart and Tirole, 1990; McLaren, 2000; Ornelas and Turner, 2008) and it can be interpreted as a way to capture all the legal, financial and organizational costs involved when merging two firms.

In sum, the choice between vertical integration and outsourcing solves the trade off between the fixed cost which arises under vertical integration and an inefficient level of specific investments that arises under outsourcing. Better contract enforcement and lower trade barriers attenuate the relevance of the hold up problem and the related investment distortions and therefore increase the incentives to outsource.

The timing of the events in the model is summarized in Figure 2-1. At period 1, the supplier and the domestic producer decide whether to integrate; if they do not they sign an ex ante contract defining the price of the intermediate input[3]. At period 2, the producer makes its relationship-specific investments I. At the beginning of period 3, the state of nature is revealed: the price of the foreign intermediate (p_F) becomes public and parties are informed on whether the initial contract will be enforced or not. If outsourcing was chosen, with probability $(1-\gamma)$ the initial contract is not enforced and the two parties have to bargain again over a new price[4]. At period 4, the producer can decide whether to buy the intermediate input from the foreign supplier.

The producer's production technology has the following form:

$$f(\phi, I, x) = (1-\phi)g(I) + \phi g(I)x \tag{HP1}$$

where I is the producer's investment and x is an indicator variable that is equal to one if the home supplier provides the intermediate good that increases productivity and zero otherwise. $\phi \epsilon (0,1)$ corresponds to the share of the investment which is unproductive without the intermediate good and it captures the specificity of P's investment. The first term of the production function is the output that the producer can eventually generate without any intermediate good. The second term is the additional output generated by the

[3]Notice tha I have ruled out the possibility that the producer could outsource to FS in the first stage of the game. There is no loss of generality in doing this since this alternative would be strictly dominated by the alternative to outsource to DS (because outsourcing to FS has a fixed trade cost).

[4]A similar way of modelling contract incompleteness as a determinant of vertical integration has been used by Acemoglu, Johnson and Mitton (2007).

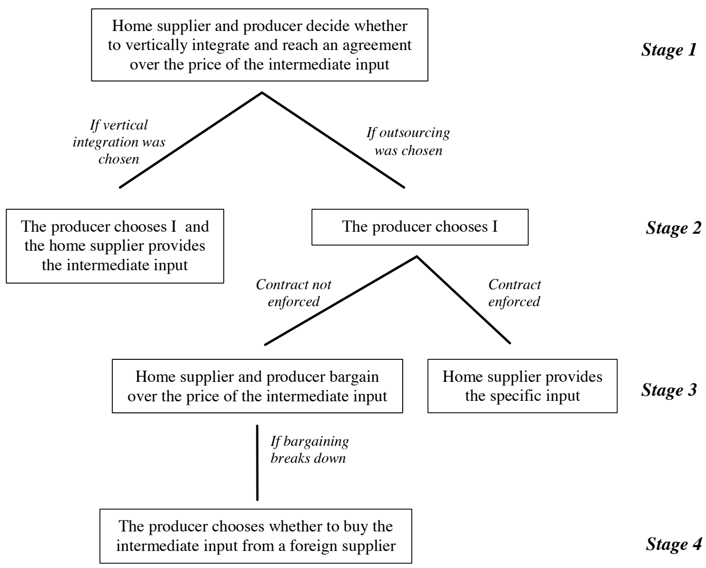

Figure 2-1: The sequence of events

producer conditional on the supplier providing the intermediate good. Assume that:

$$g''(I) < 0 \tag{HP2}$$

Normalize the cost of one unit of specific investment to 1 and assume that the supplier can provide the intermediate at no cost.[5]

The game is solved by backward induction. In stage 4, if the producer still doesn't have the intermediate input he will buy it from the foreign supplier if:

$$p_F < \psi g(I) - t$$

where ψ is a proxy for how appropriate the foreign intermediate input is to the specific

[5] A similar production structure (while somehow simplified) can be find in Acemoglu, Aghion, Griffith, Zilibotti (2005).

investment made by the producer. Let's consider the case where $\psi \leq \phi$ (e.g. the home intermediate is at least as effective as the foreign one).

2.2.2 Stage 3: Expected profits under outsourcing

In the third stage of the game, the producer has already made the investments and is outsourcing the production of the intermediate good to the home supplier. Suppose that the initial contract cannot be enforced and the two parties need to bargain over the price of the input. In the event of disagreement, the two parties receive their outside option. The home supplier would make zero profits while the producer could still find it profitable to use the intermediate input produced by the foreign supplier. Denote the outside option of party i under outsourcing by O_i^O. Then:

$$O_p^O(I_p) = (1-\phi)\, g(I) + [Max\{0; \psi g(I) - t - p_F\}] - I \tag{2.1}$$

$$O_s^O = 0 \tag{2.2}$$

Call p the new price of the intermediate when bargaining is successful and u_i^0 the ex post payoffs of party i:

$$\{ \begin{array}{l} u_p^0(I) = g(I) - I - p \\ \qquad u_s^0 = p \end{array}$$

According to the Nash bargaining solution, the price p satisfies:

$$p = Arg \max_p p^{0.5} \left[\phi g(I) - Max\{0; \psi g(I) - t - p_F\} - p\right]^{0.5} \tag{2.3}$$

which implies:

$$p = \frac{1}{2}\phi g(I) - \frac{1}{2}Max\{0; \psi g(I) - t - p_F\} \tag{2.4}$$

Hence, under the symmetric Nash equilibrium, the surplus accruing to the producer under outsourcing conditional on the fact that the initial contract has not been enforced is:

$$u_p^0(I) = \left(1 - \frac{\phi}{2}\right) g(I) + Max\{0; \psi g(I) - t - p_F\} - I \tag{2.5}$$

Stage 2: Choosing the optimal investments

In the second stage of the game the producer chooses the optimal investments. If the producer and the home supplier are vertically integrated, then the producer will decide the level of investments I in order to maximize the joint variable profits π^{VI} :

$$\pi^{VI} = g(I) - I \tag{2.6}$$

The optimal level of investments under vertical integration I^{VI} is defined by the first order condition $g'(I^{VI}) = 1$. If the producer is outsourcing to the home supplier, then the producer will decide the level of investments I^O in order to maximize the expected profits $E\pi_p^O$:

$$E\pi_p^O = \gamma[g(I) - P - 1] + (1-\gamma)\left(1 - \frac{\phi}{2}\right) g(I) + (1-\gamma)Max\{0; \frac{1}{2}\int_0^{\psi g(I)-t} (\psi g(I) - t - p_F)dF(p_F)\} - I \tag{2.7}$$

where P is the price of the intermediate good as in the initial contract. Intuitively, the first term represents the revenues if the initial contract is enforced; the second term represents the revenues that the producer would have under autarky if the initial contract is not enforced; the third term represents the additional revenues that the producer would have in the presence of international trade if the initial contract is not enforced (due to the improvement in his outside option in the bargaining game with the domestic supplier); the last term represents the investment costs. The optimal level of investments under outsourcing, $I^o(\phi, \gamma, t)$, is the level of investment that maximizes $E\pi_p^O$. In general this function has more than one local maxima. The following hypothesis limits the number of local maxima to two.

$$\frac{g''(I)}{g'(I)^2} < -\frac{(1-\gamma)\frac{1}{2}\psi^2 f(\phi g(I) - t)}{1-(1-\gamma)\frac{1}{2}(\phi - \psi F(\phi g(I) - t))} \tag{HP3}$$

Intuitively $g(I)$ needs to be enough convex (e.g. the marginal productivity of investments needs to fall quickly compared to the hazard rate of the price of foreign supplier).[6] To discuss the local maxima of the function $E\pi_P^O$ is convenient to rewrite it as: $E\pi_P^O = Q(\phi, \gamma, I) + (1-\gamma)\zeta(I, T)$ where $Q(\phi, \gamma, I) \equiv \gamma[g(I) - P - C] + (1-\gamma)\left[\left(1 - \frac{\phi}{2}\right)g(I)\right] - I$ and $\zeta(I, t) \equiv Max\{0; \frac{1}{2}\int_0^{\psi g(I) - t}(\psi g(I) - t - p_F)\, dFp_F\}$. Define $\overline{I}(t)$ the maximum investment for which: $\zeta(\overline{I}(t), t) \equiv 0$ and $I^*(\phi, \gamma)$ the investment that maximizes $Q(\phi, \gamma, I)$. In other words, $\overline{I}(t)$ is the minimum investment necessary to make credible the threat of buying the intermediate from a foreign supplier and $I^*(\phi, \gamma)$ the optimal investment under autarky.

[6] Note that a similar hypothesis is used by Ornelas and Turner (2008)

Lemma 1 *If $I^*(\phi,\gamma) > \overline{I}(t)$, the profit function has a single local maximum in $I^{**}(\phi,\gamma,t)$ identified by the following equation:*

$$\left[1-(1-\gamma)\frac{1}{2}(\phi-\psi F(\psi g(I)-t))\right]g'(I^{**}(\phi,\gamma,t))-1\equiv 0 \tag{2.8}$$

If $I^*(\phi,\gamma) \leq \overline{I}(t)$, the profit function can have an additional single local maximum in $I^*(\phi,\gamma)$ identified by the following equation:

$$\left[1-(1-\gamma)\frac{\phi}{2}\right]g'(I^*(\phi,\gamma))-1\equiv 0 \tag{2.9}$$

and such that: $I^*(\phi,\gamma) < I^{**}(\phi,\gamma,t)$

Thus, the profit function has at most two local maxima $I^*(\phi,\gamma)$ and $I^{**}(\phi,\gamma,t)$. Given the convexity of $g(.)$ it is easy to verify that both $I^*(\phi,\gamma)$ and $I^{**}(\phi,\gamma,t)$ are lower than I^{VI} (e.g. investments are always lower under outsourcing rather than under vertical integration). The entity of underinvestment under outsourcing is proportional to $\left[1-(1-\gamma)\frac{\phi}{2}\right]$ if trade barriers are prohibitive (e.g. if $I^o(\phi,\gamma,t) = I^*(\phi,\gamma)$) and to $\left[(1-\gamma)\frac{1}{2}\phi-(1-\gamma)\frac{1}{2}\psi F(\psi g(I)-t)\right]$ if they are not (e.g. $I^o(\phi,\gamma,t) = I^{**}(\phi,\gamma,t)$). The last expression is very intuitive. The first term represents the classical "hold up" distortion that we find in the transaction cost literature. The interaction between contract incompleteness and asset specificity distorts the incentives to invest of the producer since a part of the surplus generated by the investments can be appropriated by the supplier and this produces a suboptimal investments. The second term represents the effect of opening up the intermediate market and it's the main novelty of the model. The fact that the producer can buy the same intermediate input, with some probability, from a foreign supplier limits the possibility of holding him up and de facto attenuates the distortions created by low quality home institutions. In the limit, if trade barriers and foreign prices are sufficiently low (e.g. $\frac{\psi}{2}F(\psi g(I^O) - t) = 1$), the hold up problem disappears. In this sense, opening up a country with bad contracting institutions to trade is a way of "importing" good institutions. This leads to our first proposition (see the Appendix for the complete proof).

Proposition 1 *Under outsourcing, the producer's optimal investment is non increasing in t.*

Notice that, by applying the implicit function theorem on equations 2.8 and 2.9, it is possible to prove that both $I^* (\phi, \gamma)$ and $I^{**} (\phi, \gamma, t)$ are increasing in contract enforcement, γ, and decreasing in the specificity of the asset, ϕ. This gives an intuition for the following propositions (see Appendix for complete proof):

Proposition 2 *Under outsourcing, the producer's optimal investment is increasing in γ.*

Proposition 3 *Under outsourcing the producer"s optimal investment is decreasing in ϕ.*

Stage 1: Choosing the governance system

Since both parties have access to ex ante transfers, the subgame perfect equilibrium will always pick the organizational form that maximizes their joint surplus. In line with the transaction cost approach make the hypothesis that vertical integration has a fixed cost η. Call $S^{VI}(\eta) \equiv g(I^{VI}) - I^{VI} - \eta$ the joint surplus under vertical integration and $S^O(\gamma, \phi, t) \equiv g(I^O(\gamma, \phi, t)) - I^O(\gamma, \phi, t)$ the joint surplus under outsourcing. The comparison of these values gives the following proposition.

Proposition 4 *Vertical integration is more likely when assets are specific (ϕ high), contracts are incomplete (γ low) and trade barriers are high (t high).*

Proof. *The two parties will vertically integrate as long as $\Delta = S^{VI}(\eta) - S^{O}(\gamma, \phi, t)$ is positive. To obtain an expression for the impact of higher asset specificity on the governance of the firms consider the derivative of the latter with respect to ϕ.*

$$\frac{d\Delta}{d\phi} = \frac{dI^{O}(\gamma, \phi, t)}{d\phi}[1 - g'(I^{O}(\gamma, \phi, t))] \tag{2.10}$$

Notice that $\frac{dI^{O}(\gamma,\phi,t)}{d\phi}$ is not positive by proposition 4 and $[1 - g'(I^{O}(\gamma, \phi, t))]$ is also not positive since $1 - g'(I^{VI}) = 0$ and $I^{O}(\gamma, \phi, t)) < I^{VI}$ (together with the convexity of g). Thus $\frac{d\Delta}{d\phi} \geq 0$. Analogously it can be proven that $\frac{d\Delta}{dt} \geq 0$ and $\frac{d\Delta}{d\gamma} \leq 0$. ■

The intuition behind the last proposition is very straightforward. Higher level of asset specificity, contract incompleteness and trade barriers tend to distort investments under outsourcing making vertical integration more efficient.

The next proposition examines in detail the effect of asset specificity on the governance of the two parties.

Proposition 5 *If the following hypothesis[7] is true:*

$$\left[g''(I)\right]^2 - g'g''' \geq 0 \tag{HP4}$$

the effects of asset specificity on the vertical structure of a firm are magnified by domestic incomplete contracts and dampen by low trade barriers.

[7] *All the most used production functions (Kobb Douglas, CES, Quadratic) do not violate this assumption.*

To obtain the first result, notice that:

$$\frac{d^2\Delta}{d\phi d\gamma} = \frac{dI^O(\gamma,\phi,t)}{d\phi}[-g''(I^O(\gamma,\phi,t))]\frac{dI^O(\gamma,\phi,t)}{d\gamma} + \frac{d^2I^O(\gamma,\phi,t)}{d\phi d\gamma}[1-g'(I^O(\gamma,\phi,t))] \quad (2.11)$$

The first term is negative from propositions 3 and 4 (together with the convexity of g). Appendix A reports the proof that HP4 is sufficient for the cross derivative $\frac{d^2I^O(\gamma,\phi,t)}{d\phi d\gamma}$ being positive which is the final step to show that $\frac{d^2\Delta}{d\phi d\gamma} \leq 0$. Analogously, it can be proved that HP4 is a sufficient condition so that $\frac{d^2\Delta}{d\phi dt} \geq 0$.

The traditional IO literature has emphasized the fact that asset specificity has distorting effects on the governance system of a firm only when it interacts with an institutional environment characterized by incomplete contracts. Thus, it should not come as a surprise that a difficult contract enforcement amplifies the distortive effects of asset specificity.

The contribution that the last proposition offers to this literature is that it proposes an escape clause. In fact, the distortive effects that asset specificity has on the vertical structure of a firm given bad domestic contracting institutions can be dampened if the producer has access to foreign markets for intermediates. Notice that in this model there is no international trade: in equilibrium the producer will always buy the widget from the domestic supplier. However, the threat of being replaced by a foreign supplier helps to discipline the domestic supplier. As the latter cannot hold up the producer anymore, outsourcing does not produce distorted investments and the two parties are less likely to vertically integrate.

In the next sections, the last proposition will be tested empirically.

2.3 Empirical analysis

2.3.1 Data and measurement

My firm level data come from The Investment Climate Assessments Survey (ICA). This is an unbalanced firm level panel of annual data covering 95,320 firms in 105 countries. New additional country surveys are implemented each year so that the data cover different periods for different countries starting from 1999 until 2006. Each survey contains questions on the characteristics of the firm (e.g. four-digits SIC, organizational type, business age), measures of economic performances (e.g. sales, capital, labor, payroll, intermediates, inventories) and measures of the business climate (e.g. questions about trade costs, bribery, corruption, lobbying activity, bureaucratic delays, infrastructure, product and labor market regulations). The sample size varies considerably across countries so that the observations in Cape Verde(2006) are 47 while the observations in China (2004) are 2500. Limited information is provided about how the sample is selected in each country. The survey is often contracted out to a survey firm that has access to some business list. Both by design and given the limitations of maintaining a business list that is fully representative, the typical ICA respondent is a large, mature business relative to the country representative (Haltiwanger and Schweiger, 2004).

I have used a limited subset of the information provided in these large surveys. In particular, I have used data on output value and intermediate costs to measure the degree of vertical integration of the firms in the sample, information on net book value of machinery and equipment, land, buildings and leasehold improvements to measure their asset specificity, information on the number of workers to measure their size.

The Doing Business database (World Bank) provides objective measures of business regulations and their enforcement in 175 countries. It was originally developed to study the regulatory costs of business and to analyze specific regulations that enhance or constrain investments, productivity, and growth; it covers four years (2003-2006). In my empirical analysis, I have used the available information on the trade barriers to imports, the justice system's ability to enforce contracts and the quality of financial institutions.

Regarding the trade barriers to imports, three variables are recorded: the number of documents necessary to import a good, the time necessary to end the import procedures and the cost of importing a 20 foot container. I have used the last measure which includes costs for documents, administrative fees for customs clearance and inland transport but it does not include tariffs or trade taxes.

Regarding the efficiency of the judicial system in resolving a commercial dispute, the database reports the time necessary to enforce a contract when disputing in courts, the cost to do it and the average number of documents needed. In particular, the cost is reported as a percentage of a claim assumed to be equivalent to 200 per cent of the average income per capita. The data are collected through studies of the codes of civil procedure and other court regulations as well as surveys completed by local litigation lawyers (and, in a quarter of the countries, by judges as well). I have used the last two measures to proxy the quality of national contracting institutions.

Finally the Doing Business database reports a legal right index and a credit information index that I have used to infer the quality of financial institutions.

Accurate data on the tariffs are taken from the UNCTAD TRAINS database. As a proxy of trade barriers, I have considered the average of the tariffs on imports of machinery and transport equipment both unweighted and weighted by their corresponding trade value.

Table 2.1 sums up the determinants of the firm vertical structure according to the theoretical model and the variables used to proxy them.

Table 2.1: Measures used in the empirical analysis

Variable	Proxy
Vertical integration	Value added / Total sales (ICA)
Asset specificity	Machinery,equipment/Machinery,equipment,land,buildings (ICA)
Contract incompleteness	Contract enforcement costs (DB)
	Contract enforcement procedures (DB)
Financial development	Legal right index (DB)
	Credit information index (DB)
Trade barriers	Average tariffs on machinery imports (TRAINS)
	Weighted average tariff on machinery imports (TRAINS)
	Average costs to import (DB)
	Longest time to clear customs (ICA)

Data sources in parentheses

Table 2.1 summarizes the distribution of observations across countries. As you can see all the countries considered have low and medium-low per capita income (lower than 9100$). Most of observations are concentrated in Bangladesh, China, Egypt, India, Morocco, Pakistan and Vietnam.

Table 2.1 provides some descriptive statistics for the variables used in the empirical analysis.

The first four rows consider firm level variables coming from the ICA database. The firms considered in the regressions are 13926 distributed in 45 countries and 16 manufacturing industries. Row 1 reports descriptive statistics for the vertical integration index at the firm level. Observations are fairly distributed around 0.5 (mean and median have the same value) starting from values very close to zero to values very close to one. Row 2 reports the average number of permanent workers at the firm level. Both the mean and the median are very high (respectively: 216 and 44). This confirms the fact that the typical ICA respondent is a large, mature business relative to a representative business for a country. Row 3 reports the proxy for asset specificity which is computed by taking the ratio of net book value of machinery and equipment over the same value plus the net book value of lands, buildings and leasehold improvements.

Row 4 reports firm's share of national market of the main product line. Again both mean and median look pretty high (respectively 19% and 5%) confirming the fact that the database concentrates on large mature businesses. Moreover, notice that this information is available for around half of the firms analyzed.

Rows 5 and 6 consider two different measures of contract incompleteness reported in the Doing Business Database: number of procedures and cost for enforcing a commercial contract. In both cases the mean (respectively 32 and 22) is very high compared to US (17 and 7.7). As expected, low income countries are associated with worse contracting institutions.

Row 7 reports a measure of the quality of the financial system which is obtained by running a principal component on a legal right index and a credit information index reported in the Doing Business Database.

The next four rows report the different measures of trade barriers used in my empirical analysis: row 8 and 9 report the average (simple and weighted) tariff to machinery and transport equipment as in the UNCTAD TRAINS database; row 10 reports the cost to import a good as in the Doing Business Database; rows 11 reports the longest time to clear customs as in the ICA surveys averaged at the country level.

Finally, rows 12 and 13 are the countries' 2006 per capita GNI and total population as reported by the Doing Business Database. As can be seen most of the observations are concentrated in countries with low income per capita (the 75th percentile is 2630$).

2.3.2 Main effects

In this section, I study the main effects of asset specificity, contract enforcement costs and trade barriers to intermediates on the vertical structure of firms. Propositions 2, 3 and 4 predict a positive association between asset specificity, contract enforcement costs and trade barriers with the level of firms' vertical integration. In order to test these predictions, I have estimated the following equation:

$$VI_{fsc} = \beta_0 + \beta_1 CI_c + \beta_2 TB_c + \beta_3 AS_f + \varepsilon_{fsc} \tag{2.12}$$

where VI_{fsc} is vertical integration of firm f in country c in sector s, CI_c is the cost to enforce a contract in country c (contract incompleteness), TB_c is the cost to import intermediates in country c (trade barriers), AS_f is the proxy for the specificity of assets of firm f.

Column 1 of Table 4 reports a non significant positive correlation of trade barriers with vertical integration and a non significant negative correlation of contract incompleteness and asset specificity with vertical integration. The data do not seem to support our initial claim. In the second column, I have included a full set of industry dummies. This implies that all cross-country comparisons are relative to the "mean propensity to aggregate" in a particular industry. In other words, this regression looks at, for example, whether firms in a country with high trade barrier are more vertically integrated relative to firms in a country with low trade barrier in the same industry. The results do not change significantly.

In column 3, I include financial development as additional explanatory variable. Macchiavello (2008) and Acemoglu et al (2008) show that financial development tends to be associated with more vertically integrated firms. Since, a large literature documents that good financial institutions are strongly correlated with good contracting institutions and lower trade barriers, I don't want that omitting this variable could lead to spurious correlations in our analysis. The regression confirms the presence of a positive correlation between financial development and vertical integration. Moreover, the negative coefficient on contract incompleteness becomes five times smaller while the positive coefficient on trade barriers doubles.

In column 4, I add country's population and GNP per capita to the regressors. The coefficient on contract incompleteness turns to positive (though still not significant). The negative signs on the coefficients of population and GNP per capita can be explained by the fact that bigger economies have thicker intermediate goods markets and therefore smaller hold-up distortions (and less scope for vertical integration).

In sum, when controlling for financial development and market size, there seem to be some evidence of a positive correlation of vertical integration with both contract incompleteness and trade barriers. However, this evidence is not strong (coefficients are not significant). Moreover, there seem to be a negative correlation between asset specificity and vertical integration, which is at odd with the theory.

There are several reasons that can explain these results. First, the lack of significance is probably due to the fact that contract incompleteness and trade barriers are capturing the effects of many other country level variables that can affect vertical integration and that are missing in my specification. This is a common problem in cross-country analysis where main effects can be very hard to test. The negative sign of asset specificity is difficult to justify. Notice however that the theory predicts asset specificity to have a positive effect on vertical integration only when associated with incomplete contracts.

2.3.3 Interaction effects

The results in the previous section may suggest that there are no robust regularities in cross-country vertical integration patterns. In this section, I turn to interaction effect and show that this is not true.

The problem of unobservables is attenuated when examining interaction effects since eventual omitted variables at country level can be captured by country fixed effects. In this section, I study how the effect of asset specificity on firms' vertical structure varies as contractual institutions and trade barriers vary.

Since I don't want to impose a particular function that define the vertical integration of firms, I take a second order approximation of a general function: $VI = f(AS, CI, TB, FD)$ and add country and industry fixed effects. The following equation is estimated by ols:

$$VI_{fsc} = \beta_1 AS_f + \beta_2 AS_f^2 + \beta_3 CI_c \cdot AS_f + \beta_4 TB_c \cdot AS_f + \beta_5 FD_c \cdot AS_f + S_c + S_s + \varepsilon_{fsc} \quad (2.13)$$

where S_c and S_s are respectively the country and the industry fixed effects. The results are reported in column 1 of Table 5. Notice that the signs are all in line with the theoretical predictions. As predicted by Proposition 6, the effects of asset specificity on the vertical structure of the firm are magnified by domestic incomplete contracts and dampen by low trade barriers. Figure 2.7.4 depicts $\frac{dVI_{fsc}}{dAS_f}$ as a function of trade barriers, contract enforcement costs and financial development. The first graph shows how the marginal effect of asset specificity on vertical integration changes as importing costs change (with contract enforcement costs and financial development fixed at their mean level). Notice that, for low trade barriers (importing costs below 1000 $ for a 20-foot container, 35th percentile) I cannot reject the hypothesis that asset specificity has no significant effect on vertical integration. However, as trade barriers become higher, the effect of asset specificity becomes positive and significant. The second graph depicts the marginal effect of asset specificity as the cost of enforcing a contract vary. When contract enforcing costs are low (court and attorney fees below 35% of the value of the claim, 49th percentile) asset specificity has no significant effect on vertical integration. However, the effect becomes positive and substantial as contract enforcement costs raise. Finally, the third graph depicts the marginal effect of asset specificity on vertical integration as the level of financial development of the country raises. The effect becomes positive and significant, when financial development is above the 82th percentile of the distribution.

In column 2, I add the variable "number of workers" in the regression. The inclusion of firm size as a control variable is due to the fact that a potential concern with the result of this paper is sample selection. As we have seen, the typical ICA respondent is larger than the representative business of the country. It could be that relatively larger companies are more vertically integrated and from country with worst institutions we could only observe larger companies. Controlling for firms size could partially alleviates this sample selection concern.[8] The results are unchanged.

In column 3, I add the interaction between GNI per capita and asset specificity among the regressors. Existing works demonstrate that contract enforcement is correlated with the stage of economic development. I would like to be sure that the interactions $CI_c \cdot AS_f$ and $TB_c \cdot AS_f$ are not just proxying for other factors associated with the stage of development. This new regressors has a very low significance and do not alter substantially our previous results.

Another possible concern is that the ratio between value added and shipments is sensible not only to the firm's vertical integration but also to the firm's market power. This would bias the results if for example, trade barriers provide protection for monopolists or weak contract enforcement is likely associated with weak antitrust policies. In column 6, I have added a variable that could eventually capture firm's market power (market share for main product line). The main results are not affected. Some coefficients change slightly but this is probably due to the fact that I have to limit the analysis for the firms for which I have information on the market power which are less than half of the original sample.

As further robustness checks, I have replicated the analysis above using a different proxy for contract incompleteness: the number of procedures to enforce a commercial contract (instead of the cost). The results are shown in Table 6 and are not qualitatively too different from the previous ones.

---

[8] Following Acemoglu, Johnson and Mitton (2005), I have also experimented with regressions controlling for second, third and fourth order polinomial in firm size and found very similar results.

In Table 7, I have repeated the analysis using three different proxies for trade barriers (trade costs to import, average tariffs and longest time to clear customs) and the two different proxies for contract incompleteness. Again qualitatively, the results are similar.

Finally in Table 8, I have replicated the analysis using an alternative measure of financial development. Instead of looking only at the credit information index provided by the Doing Business Database, I use a principal component of this variable together with other two variables provided in the same database: the legal right index and the private bureau coverage of credit information. The signs are unchanged. However, both the magnitude and the significance of the coefficient on the interaction between asset specificity and contract incompleteness is lower. This is probably related with the high correlation between this new measure of financial development and the usual measure of contract incompleteness that makes it difficult to disentangle the effects of these two variables (interacted with asset specificity) on the vertical integration of firms.

In order to take into account the multilevel dimension of the data, I have redone the analysis using a hierarchical linear model. I estimated the following system of equations by mle:

$$VI_{fsc} = \alpha_0 + \alpha_{4c} AS_f + S_c + S_s + \varepsilon_{fsc} \tag{2.14}$$

$$\beta_{4c} = \delta_0 + \delta_1 CI_c + \delta_2 TB_c + \delta_3 FD_c + \upsilon_{0c} \tag{2.15}$$

The results are reported in Table 9. Both the coefficients and their standard errors do not vary significantly compared to the ols case.

2.3.4 IV regression

Two potential concerns apply to the OLS and HLM estimates that we have seen so far. First, it may be that some omitted firm-level variables are driving both the asset specificity and the vertical integration of the firms in my sample. Second, the estimates may suffer of a potential reverse causality problem. For example, it may be that firms that are more vertically integrate are more likely to perform primary activities that are less technologically intensive and require less specific assets. In both cases, the error term is going to be correlated with the regressors, biasing the estimates.

A more satisfactory approach would be to use an instrumental variable strategy, with instruments that affect asset specificity, without influencing vertical integration through other channels (i.e., they should be orthogonal to the error term, ε_{fsc}, in equation 2.13). In this paragraph, I instrument the firm asset specificity using a measure of intensity in physical investments in the same industry in US[9].

The first-stage equations for the model in equation 2.13 are:

$$AS_f = X'\pi_{11} + Z'\pi_{12} + Z'CI_c\pi_{13} + Z'TB_c\pi_{14} + Z'FD_c\pi_{15} + u_{1f} \tag{2.16}$$

$$AS_f^2 = X'\pi_{21} + Z'\pi_{22} + Z'CI_c\pi_{23} + Z'TB_c\pi_{24} + Z'FD_c\pi_{25} + u_{2f} \tag{2.17}$$

[9] Acemoglu, Aghion, Griffith and Zilibotti (2005) were the first to propose this instrument for asset specificity in a firm-level analysis limited to UK manufacturing plants.

$$AS_f * CI_c = X'\pi_{31} + Z'\pi_{32} + Z'CI_c\pi_{33} + Z'TB_c\pi_{34} + Z'FD_c\pi_{35} + u_{3f} \tag{2.18}$$

$$AS_f * TB_c = X'\pi_{41} + Z'\pi_{42} + Z'CI_c\pi_{43} + Z'TB_c\pi_{44} + Z'FD_c\pi_{45} + u_{4f} \tag{2.19}$$

where Z is the vector of instruments for asset specificity (in other words, investment intensity in the U.S.), and X' is a vector of all the covariates that appear in the second stage as well. In table 10, I report the IV estimates of equation 2.13. The instrumental variable strategy confirms the validity of the previous results. Most of the previous estimates are unchanged; only the effect of financial development combined with asset specificity on vertical integration decreases slightly. The last table reports the first-stage coefficients. The first-stage relationship are highly significant and show a very appealing pattern: physical investment intensity in a particular industry in the US is highly correlated with the asset specificity of the firms in the sample in the same industry. The F-test of the exclusion restriction is 0.66, so my analysis do not suffer from a weak instrument problem.

In conclusion, IV regressions confirms the pattern of previous results. The effects of asset specificity on vertical integration increases in the presence of low quality contracting institutions and high trade barriers.

2.4 Conclusion

This paper investigates the cross-country determinants of vertical integration using a new dataset of over 14,000 manufacturing firms operating in 45 developing countries. In particular, it revisits the effects of the interaction between technology intensity and some specific institutional features on the vertical integration decisions of firms. This focus is motivated by both theory and anecdotal evidence.

A large body of theoretical contributions has highlighted the effects of both international openness and contractual institutions on the vertical structure of firms. Moreover, some empirical works have documented the presence of significant heterogeneity in the propensity of firms to vertically integrate across countries. Nevertheless, there has never been a systematic empirical analysis of the cross-country differences in vertical integration and their causes. First, I develop a simple model that sums up previous theories and allows for some intuitive comparative statics exercise. In particular, it suggests that technology intensity should have greater effects on the vertical structure of firms when combined with low quality contracting institutions and high trade barriers. The empirical results are consistent with these predictions and are robust to different econometric specifications and techniques.

I conclude that poor contract enforcement can distort firms' vertical structure in the presence of specific assets. This can have significant welfare costs especially in developing economies. If improving home institutions is not feasible an equivalent solution is to open to international trade. This would reduce the hold-up problem by disciplining domestic suppliers, thus reducing the need for vertically integrated organizations.

2.5 Appendix A: Proofs and derivations

Proof. [Proof of Lemma 1] We can rewrite the profit function as $E\pi_1^O = Q(\phi, \gamma, I)$ for $I < \overline{I}(t)$ and $E\pi_2^O = Q(\phi, \gamma, I) + (1-\gamma)\zeta(I, t)$ for $I \geq \overline{I}(t)$.

The producer chooses investment in order to maximize this profit function. For $I < \overline{I}(t)$, the profit function is $E\pi_1^O = Q(\phi, \gamma, I)$ and, by HP2, it is locally convex. Thus, if a local maximum exists in this range, then it is defined by the FOC $\frac{dQ(\phi,\gamma,I)}{dI} = 0$, which can be rewritten as:

$$\left[1 - (1-\gamma)\frac{\phi}{2}\right] g'(I^*(\phi, \gamma)) - 1 \equiv 0$$

.

For $I \geq \overline{I}(t)$, the profit function becomes $E\pi_2^O = Q(\phi, \gamma, I) + (1-\gamma)\zeta(I, t)$ and, by the HP2 and HP3 is locally convex. Thus, if a local maximum exists in this range, then it is defined by the FOC $\frac{dQ(\phi,\gamma,I)}{dI} + (1-\gamma)\frac{d\zeta(I,t)}{dI} = 0$, which can be rewritten as:

$$\left[1 - (1-\gamma)\frac{1}{2}(\phi - \psi F(\psi g(I) - t))\right] g'(I^{**}(\phi, \gamma, t)) - 1 \equiv 0$$

Notice that if $I^*(\phi, \gamma) > \overline{I}(t)$, there cannot be a local maximum for $I < \overline{I}(t)$ .In fact, the function $Q(\phi, \gamma, I)$ is strictly increasing in I for $I < \overline{I}(t)$. Therefore, the profit function will have a single local maximum in $I = I^{**}(\phi, \gamma, t)$.

In sum: If $I^*(\phi, \gamma) > \overline{I}(t)$, the profit function has a unique local maximum in $I = I^{**}(\phi, \gamma, t)$. If $I^*(\phi, \gamma) < \overline{I}(t)$, the profit function can have at most two local maxima respectively in $I^*(\phi, \gamma)$ and $I^{**}(\phi, \gamma, t)$. ■

Proof. [Proof of Proposition 2] The producer will choose the level of investment that maximizes the profit function defined by equation 2.7. Given Lemma 1, the profit function has at most two local maxima: $E\pi_1^O(\phi,\gamma) = Q(\phi,\gamma,I^*(\phi,\gamma))$ and $E\pi_2^O(\phi,\gamma,t) = Q(\phi,\gamma,I^{**}(\phi,\gamma,t)) + (1-\gamma)\,\zeta(I^{**}(\phi,\gamma,t),t)$. In order to prove that optimal investment under outsourcing is decreasing in trade barriers t, we will consider three different cases:

1. $E\pi_1^O(\phi,\gamma)$ is the global maximum for every t. In this case the optimal investment is $I^*(\phi,\gamma)$ and does not depend on trade barrriers.

2. $E\pi_2^O(\phi,\gamma,t)$ is the global maximum for every t. In this case the optimal investment is $I^{**}(\phi,\gamma,t)$. Applying the implicit theorem function to equation 2.8, is possible to verify that $I^{**}(\phi,\gamma,t)$ is a strictly decreasing function of t.

3. $E\pi_1^O(\phi,\gamma)$ is the global maximum for some values of t while $E\pi_2^O(\phi,\gamma,t)$ is the global maximum for some others. Notice while the first one is not affected by t, the second one is decreasing and continuos in t. Therefore, in this case it exists a unique $\widehat{t}$ at which the producer is indifferent between $I^*(\phi,\gamma)$ and $I^{**}(\phi,\gamma,t)$. This tariff is implicitly defined by: $E\pi_1^O(\phi,\gamma) \equiv E\pi_2^O\left(\phi,\gamma,\widehat{t}\right)$. Consider an increase in trade costs dt. For $t < \widehat{t}$, the optimal investment is $I^{**}(\phi,\gamma,t)$, which is decreasing in t. For $t > \widehat{t}$, the optimal investment is $I^*(\phi,\gamma)$, which is not affected by t. Finally, when t increases from $\widehat{t}-dt$ to $\widehat{t}+dt$, the optimal investment drops from $I^{**}(\phi,\gamma,t)$ to $I^*(\phi,\gamma)$. Thus, also in this case investment is not decreasing in t.

■

If the profit function has two local maxima $E\pi_1^O(\phi,\gamma)$ and $E\pi_2^O(\phi,\gamma,t)$, then it should be that:

$$\frac{dE\pi_1^O(\phi,\gamma)}{d\gamma} < \frac{dE\pi_2^O(\phi,\gamma,t)}{d\gamma}$$

Proof. Given Lemma 1, the profit function can have two local maxima as long as $I^*(\phi,\gamma) < \overline{I}(t)$.Using the envelope theorem:

$$\frac{dE\pi_1^O(\phi,\gamma)}{d\gamma} = \frac{\partial E\pi_1^O(\phi,\gamma)}{\partial\gamma} = -P + \frac{\phi}{2}g(I^*(\phi,\gamma))$$

$$\frac{dE\pi_2^O(\phi,\gamma,t)}{d\gamma} = \frac{\partial E\pi_2^O(\phi,\gamma,t)}{\partial\gamma} = -P+\frac{\phi}{2}g(I^{**}(\phi,\gamma,t))-\frac{1}{2}\int\limits_0^{\psi g(I^*(t,\gamma))-t}(\psi g(I^{**}(\phi,\gamma,t))-t-p_F)\,dFp_F$$

Define:

$$\begin{aligned}\Sigma &\equiv \frac{dE\pi_2^O(\phi,\gamma,t)}{d\gamma} - \frac{dE\pi_1^O(\phi,\gamma)}{d\gamma} = \\ &= \frac{\phi}{2}(g(I^{**}(\phi,\gamma,t)) - g(I^*(\phi,\gamma)) - \frac{1}{2}\int\limits_0^{\psi g(I^*(t,\gamma))-t}(\psi g(I^{**}(\phi,\gamma,t))-t-p_F)\,dFp_F\end{aligned}$$

Given Lemma 1, if the profit function has two local maxima then $I^*(\phi,\gamma) < \overline{I}(t)$ and $I^{**}(\phi,\gamma,t) > \overline{I}(t)$. Thus:

$$\Sigma > \frac{\phi}{2}(g(I^{**}(\phi,\gamma,t)) - g(\overline{I}(t)) - \frac{1}{2}\int\limits_0^{\psi g(\overline{I}(t))-t}\left(\psi g\left(\overline{I}(t)\right)-t-p_F\right)dFp_F$$

On RHS of the inequality, the first term is positive while the second one is zero by definition of $\overline{I}(t)$. Hence $\Sigma > 0$ (e.g. $\frac{dE\pi_1^O(\phi,\gamma)}{d\gamma} < \frac{dE\pi_2^O(\phi,\gamma,t)}{d\gamma}$). ■

Proof. [Proof of Proposition 3] Given Lemma 1, the profit function has at most two local maxima: $E\pi_1^O(\phi,\gamma) = Q(\phi,\gamma,I^*(\phi,\gamma))$ and $E\pi_2^O(\phi,\gamma,t) = Q(\phi,\gamma,I^{**}(\phi,\gamma,t)) + (1-\gamma)\zeta(I^{**}(\phi,\gamma,t),t)$. In order to prove that optimal investment is increasing in contract enforcement, we can divide our analysis in three cases:

1. $E\pi_1^O(\phi,\gamma)$ is the global maximum for every γ. In this case the optimal investment is $I^*(\phi,\gamma)$ for every γ. Applying the implicit theorem function to equation 2.9, is possible to verify that $I^*(\phi,\gamma)$ is a strictly increasing function of γ.

2. $E\pi_2^O(\phi,\gamma,t)$ is the global maximum for every γ. In this case the optimal investment is $I^{**}(\phi,\gamma,t)$ for every γ. Applying the implicit theorem function to equation 2.8, is possible to verify that $I^{**}(\phi,\gamma,t)$ is a strictly increasing function of γ.

3. $E\pi_1^O(\phi,\gamma)$ is the global maximum for some values of γ while $E\pi_2^O(\phi,\gamma,t)$ is the global maximum for some others.

   Define $\Omega(\phi,\gamma,t) \equiv E\pi_1^O(\phi,\gamma) - E\pi_2^O(\phi,\gamma,t)$. In this case $\Omega(\phi,\gamma,t)$ takes positive values for some γ and negative values for others. Together with the fact that $\Omega(\phi,\gamma,t)$ is continuous in γ (because sum of continuous function) and strictly decreasing in γ (by claim 7), this observation implies that there exists a unique $\widehat{\gamma}$ such that if $\gamma < \widehat{\gamma}$ then $\Omega(\phi,\gamma,t) > 0$, if $\gamma > \widehat{\gamma}$ then $\Omega(\phi,\gamma,t) < 0$ and for $\gamma = \widehat{\gamma}$ we have $\Omega(\phi,\widehat{\gamma},t) = 0$. In other words, if $\gamma < \widehat{\gamma}$ the global maximum is $E\pi_1^O(\phi,\gamma)$, if $\gamma > \widehat{\gamma}$ global maximum is $E\pi_2^O(\phi,\gamma,t)$ and if $\gamma = \widehat{\gamma}$, the profit function has two global maxima $E\pi_1^O(\phi,\widehat{\gamma}) = E\pi_2^O(\phi,\widehat{\gamma},t)$. Consider an increase in γ. For $\gamma < \widehat{\gamma}$, the optimal investment is $I^*(\phi,\gamma)$ and hence is increasing in γ; for $\gamma > \widehat{\gamma}$, the optimal investment is $I^{**}(\phi,\gamma,t)$ and hence is increasing in γ. Finally when γ increases from $\widehat{\gamma} - d\gamma$ to

$\widehat{\gamma}+d\gamma$, the optimal investment jumps up from $I^{*}(\phi,\gamma)$ to $I^{**}(\phi,\gamma,t)$: thus also in this case the optimal investment in increasing in γ.

■

Proof. [Proof of Proposition 4] Given Lemma 1 the profit function has at most two local maxima: $E\pi_1^O(\phi,\gamma)$ and $E\pi_2^O(\phi,\gamma,t)$ for every ϕ. Three subcases are possible:

1. $E\pi_1^O(\phi,\gamma)$ is the global maximum for every ϕ. In this case the optimal investment is $I^{*}(\phi,\gamma)$ for every ϕ. Applying the implicit theorem function to equation 2.9, is possible to verify that $I^{*}(\phi,\gamma)$ is a continuous and strictly decreasing function of ϕ.

2. $E\pi_2^O(\phi,\gamma,t)$ is the global maximum for every ϕ. In this case the optimal investment is $I^{*}(\phi,\gamma,t)$ for every ϕ. Applying the implicit theorem function to equation 2.8, is possible to verify that $I^{**}(\phi,\gamma,t)$ is a continuous and strictly decreasing function of ϕ.

3. $E\pi_1^O(\phi,\gamma)$ is the global maximum for some values of γ while $E\pi_2^O(\phi,\gamma,t)$ is the global maximum for some others.

   Consider $\Omega(\phi,\gamma,t)\equiv E\pi_1^O(\phi,\gamma)-E\pi_2^O(\phi,\gamma,t)$ and notice that, in this case, it takes positive values for some ϕ and negative for others. Together with the fact that $\Omega(\phi,\gamma,t)$ is continuous in ϕ (because sum of continuous function) and strictly increasing in ϕ ($\frac{d\Omega(\phi,\gamma,t)}{d\phi}=-(1-\gamma)\frac{1}{2}\left[g(I^{*}(\phi,\gamma))-g(I^{**}(\phi,\gamma,t)\right])$, this observation implies that there exists a unique $\widehat{\phi}$ such that if $\phi<\widehat{\phi}$ then $\Omega(\phi,\gamma,t)<0$, if $\phi>\widehat{\phi}$ then $\Omega(\phi,\gamma,t)>0$ and for $\phi=\widehat{\phi}$, $\Omega\left(\widehat{\phi},\gamma,t\right)=0$. In other words, if $\phi<\widehat{\phi}$ the global maximum is $E\pi_2^O(\phi,\gamma,t)$, if $\phi>\widehat{\phi}$ global maximum is $E\pi_1^O(\phi,\gamma)$ and if $\phi=\widehat{\phi}$

, the profit function has two global maxima $E\pi_1^O\left(\widehat{\phi},\gamma\right) = E\pi_2^O\left(\widehat{\phi},\gamma,t\right)$. Consider an increase in ϕ. For $\phi < \widehat{\phi}$, the optimal investment is $I^{**}(\phi,\gamma,t)$ and hence is decreasing in ϕ; for $\phi > \widehat{\phi}$, the optimal investment is $I^{*}(\phi,\gamma)$ and hence is decreasing in ϕ. Finally when ϕ increases from $\widehat{\phi} - d\phi$ to $\widehat{\phi} + d\phi$, the optimal investment jumps down from $I^{**}(\phi,\gamma,t)$ to $I^{*}(\phi,\gamma)$: thus also in this case the optimal investment in decreasing in ϕ.

■

Proof. [Proof of Proposition 6] I will prove that $\frac{d^2 I^O(\phi,\gamma,t)}{d\phi d\gamma}$ is positive. Consider two cases:

1. Suppose that: $I^O(\phi,\gamma,t) = I^*(\phi,\gamma)$. Call $SOC1 \equiv \left[1-(1-\gamma)\frac{\phi}{2}\right]g''(I)$ and (notice that this quantity is negative by HP2). Applying the implicit function theorem to equation 2.9, we have:

   $\frac{dI^*(\phi,\gamma)}{d\phi} = \frac{(1-\gamma)\frac{1}{2}g'(I)}{SOC1}$

   $\frac{dI^*(\phi,\gamma)}{d\gamma} = \frac{-\frac{\phi}{2}g'(I)}{SOC1}$

   Notice that $\frac{dI^*(\phi,\gamma)}{d\phi} \leq 0$ while $\frac{dI^*(\phi,\gamma)}{d\gamma} \geq 0$. The cross derivative is then:

   $\frac{d^2 I^*(\phi,\gamma)}{d\phi d\gamma} = \frac{\left[-\frac{1}{2}g' + (1-\gamma)\frac{1}{2}g''\frac{dI}{d\gamma}\right]SOC1 - \left[\frac{\phi}{2}g'' + \left[1-(1-\gamma)\frac{\phi}{2}\right]g'''\frac{dI}{d\gamma}\right](1-\gamma)\frac{1}{2}g'}{SOC1^2}$

   which after some algebra becomes:

   $\frac{d^2 I^*(\phi,\gamma)}{d\phi d\gamma} = \frac{-\frac{1}{2}g'g'' - \frac{\left[(g'')^2 - g'g'''\right]}{g''}\frac{\phi}{4}g'}{SOC1^2}$

   which is not negative as long as $(g'')^2 - g'g''' \geq 0$.

2. Suppose that: $I^O(\phi,\gamma,t) = I^{**}(\phi,\gamma,t)$. Define $SOC2 \equiv \left[1-(1-\gamma)\frac{\phi}{2}\right]g''(I) + (1-\gamma)\frac{\psi}{2}F(\psi g(I) - t)g''(I) + (1-\gamma)\frac{1}{2}\psi^2 g'(I)^2 f(\psi g(I) - t)$ and notice that this

quantity is negative by HP2 and HP3. Applying the implicit function theorem to equation 2.8, we have:

$$\frac{dI^{**}(\phi,\gamma,t)}{d\phi} = \frac{(1-\gamma)\frac{1}{2}g'(I)}{SOC2}$$

$$\frac{dI^{**}(\phi,\gamma,t)}{d\gamma} = \frac{\left[\frac{\phi}{2}-\frac{\psi}{2}F(\psi g(I)-t)\right]g'(I)}{SOC2}$$

$$\frac{dI^{**}(\phi,\gamma,t)}{dt} = \frac{(1-\gamma)\frac{1}{2}\psi g'(I)f(\psi g(I)-t)}{SOC2}$$

Notice that $\frac{dI^{*}(\phi,\gamma)}{d\phi} \leq 0$ while $\frac{dI^{*}(\phi,\gamma)}{d\gamma} \geq 0$. The cross derivative is then:

$$\frac{d^2 I^{**}(\phi,\gamma,t)}{d\phi d\gamma} = \frac{\left[-\frac{1}{2}g'+(1-\gamma)\frac{1}{2}g''\frac{dI}{d\gamma}\right]}{SOC2} + \\ -\frac{\left[(\frac{\phi}{2}-\frac{\psi}{2}F)\left[1-(1-\gamma)(\frac{\phi}{2}-\frac{\psi}{2}F)\right]\left((g'')^2-g'g'''\right)-\frac{\psi^2}{2}(g')^2 f+(1-\gamma)\psi^2 g'g''f\frac{dI}{d\gamma}+(1-\gamma)\frac{\psi^2}{2}(g')^2 f'\psi g'\frac{dI}{d\gamma}\right](1-\gamma)\frac{1}{2}g'}{SOC2^2}$$

which is not positive as long as $(g'')^2 - g'g''' \geq 0$.

■

2.6 Data sources and construction

2.6.1 Value of Output

The ICA survey collects information on "Total market value of the production" (c274c1y) and "Total sales" (c274a1y) but there is a high number of missing values. Therefore I have used the following strategy:

Compute the number of observations about the value of the production and the total sales in each country.

Generate the variable output which, in each country, equals the value of the production when the number of observations about the value of production is at least 110 percent the number of observations about total sales. Otherwise it equals the total sales.

Notice that I could have also adjusted total sales by subtracting the variation in the variable "Inventories and stock" (281k1y). The problem is that inventories data are questionable: too many firms report zero while for Brazil and Ethiopia, when output is computed in this way, it is mostly negative.

2.6.2 Cost of Intermediate Goods

I consider two different measures of raw material costs: "Raw material costs (excluding fuel)" (c274b1y) and "Total purchase of raw material (excluding fuel)" (c274d1y). I use the former variable when available and the latter otherwise.

The cost of energy is computed by summing up the variables "Consumption of electricity" (c274f1y) and "Consumption of fuel" (c274g1y) when both are available and using the variable "Consumption of energy" (c274e1y) otherwise. For the remaining missing values, I impute the share of energy in each sector over the raw material cost.

Finally I compute the cost of intermediate production goods by summing up the cost of material and the cost of energy.

2.6.3 Vertical integration

Vertical integration is measured by the ratio of value added to sales (e.g.: (Total Output-Cost of intermediate)/Total output). This measure has been used in many previous studies but, as already discussed above, is susceptible to bias. This bias increases with the amount of value added by downstream firms. For this reason my analysis is limited to firms producing primarily in manufacturing industries. The observations in the first and the last percentile have been dropped in order to correct for outliers.

2.6.4 Number of workers

The ICA survey collects information on "Average number of permanent workers" (c262a1y) and "Average number of temporary workers" (c263a1y). It is not clear whether missing values for temporary workers indicate that there are no temporary workers in that firm or that the respondent simply gives the total number of workers under the voice permanent workers. I choose to totally disregard data about temporary workers and consider permanent workers as the only measure of the labor used in the production process. No information on hours per worker are collected. The observations in the first and the last percentile have been dropped.

2.7 Appendix B: Data sources and construction

2.7.1 Value of Output

The ICA survey collects information on "Total market value of the production" (c274c1y) and "Total sales" (c274a1y) but there is a high number of missing values. Therefore I have used the following strategy:

Compute the number of observations about the value of the production and the total sales in each country.

Generate the variable output which, in each country, equals the value of the production when the number of observations about the value of production is at least 110 percent the number of observations about total sales. Otherwise it equals the total sales.

Notice that I could have also adjusted total sales by subtracting the variation in the variable "Inventories and stock" (281k1y). The problem is that inventories data are questionable: too many firms report zero while for Brazil and Ethiopia, when output is computed in this way, it is mostly negative.

2.7.2 Cost of Intermediate Goods

I consider two different measures of raw material costs: "Raw material costs (excluding fuel)" (c274b1y) and "Total purchase of raw material (excluding fuel)" (c274d1y). I use the former variable when available and the latter otherwise.

The cost of energy is computed by summing up the variables "Consumption of electricity" (c274f1y) and "Consumption of fuel" (c274g1y) when both are available and using the variable "Consumption of energy" (c274e1y) otherwise. For the remaining missing values, I impute the share of energy in each sector over the raw material cost.

Finally I compute the cost of intermediate production goods by summing up the cost of material and the cost of energy.

2.7.3 Vertical integration

Vertical integration is measured by the ratio of value added to sales (e.g.: (Total Output-Cost of intermediate)/Total output). This measure has been used in many previous studies but, as already discussed above, is susceptible to bias. This bias increases with the amount of value added by downstream firms. For this reason my analysis is limited to firms producing primarily in manufacturing industries. The observations in the first and the last percentile have been dropped in order to correct for outliers.

2.7.4 Number of workers

The ICA survey collects information on "Average number of permanent workers" (c262a1y) and "Average number of temporary workers" (c263a1y). It is not clear whether missing values for temporary workers indicate that there are no temporary workers in that firm or that the respondent simply gives the total number of workers under the voice permanent workers. I choose to totally disregard data about temporary workers and consider permanent workers as the only measure of the labor used in the production process. No information on hours per worker are collected. The observations in the first and the last percentile have been dropped.

Bibliography

[1] Acemoglu, D. , Aghion, P., Griffith, R. and Zibilotti F. (2004), "Vertical integration and technology: theory and evidence", NBER Working Paper 10997.

[2] Acemoglu, D. , Johnson S. and Mitton T. (2005), "Determinants of vertical integration: finance contracts and regulation", NBER Working Paper 11424.

[3] Adelman, M. A. (1995), "Concept and statistical measurement of vertical integration", Business Concentration and Price Policy, Princeton University Press, Princeton.

[4] Aghion, P., Griffith, R. and Howitt, P. (2006), "The U-shaped relationship between vertical integration and competition: theory and evidence", mimeo.

[5] Anderson E. and Schmittlein, D. (1984), "Integration of the sales forces: an empirical examination", Rand Journal of Economics 15, pp. 385-395.

[6] Antras, P. (2003), "Firms, contracts and trade structure", The Quarterly Journal of Economics 118, pp. 1375-1418.

[7] Bain, J. (1956), "Barriers to new competition", Harvard University Press, Cambridge MA.

[8] Bain, J. (1959), "Industrial organization", John Wiley & Sons, New York.

[9] Bamford, J. (1994), "Driving America to tiers", Financial World 8, 163(23).

[10] Barney, J. B., Edwards, F. L. and Ringleb, A. H. (1992), "Organizational responses to legal liability", Academy of Management Journal 35, pp. 328-349.

[11] Bender, C. (2002), "The theory of the firm revisited: changing firm boundaries in a new information and communication environment", Working Paper, Department of International Business, University of Muenster.

[12] Bigelow, L. (2001), "Efficient alignment and survival in the US automobile industry", Working Paper, Olin School of Business, Washington University St Louis.

[13] Buxbaum, P. "Focusing on third parties", Distribution, 1994, 93(12).

[14] Chen Y., Ishikawa, J. and Yu, Z. (2002), "Trade liberalization and strategic outsourcing", Working Paper, Carleton University.

[15] Coase, R. (1937), "The nature of the firm", Economica 4, pp. 386-405.

[16] Desai, M., Gompers, P. and Lerner J. (2003), "Institutions, capital constraints and entrepreneurial firm dynamics: evidence from Europe", NBER Working Paper 10165.

[17] Eckel, C. (2005), "International trade, flexible manufacturing and outsourcing", CEGE Working Paper 45.

[18] The Economist. "Manufacturing, the ins and outs of outing", 1991, 320(7722).

[19] Fan, J., Huang, J., Morck, R., Yeung, B. (2007), "Institutional determinants of Vertical Integration: Evidence form China", Working Paper

[20] Feenstra, R. C. (1998), "Integration of trade and disintegration of production in the global economy", Journal of Economic Perspectives 12, pp. 31-50.

[21] Gonzalez-Diaz, M., Arrunada, B. and Fernandez, A. (2000), "Causes of subcontracting: evidence from panel data on construction firms", Journal of Economic Behavior and Organization 42, pp.167-187.

[22] Gort, M. (1962), "Diversification and integration in American industry", Princeton University Press, Princeton.

[23] Grossman, S. and Hart, O. (1986), "The costs and benefits of ownership: a theory of vertical and lateral integration", Journal of Political Economy 94, pp. 691-719.

[24] Grossman, G. M. and Helpman, E. (2002), "Integration versus outsourcing in industry equilibrium", Quarterly Journal of Economics 117, pp. 85-120.

[25] Hart, O. and Moore, J. (1990), " Property rights and the nature of the firm", Journal of Political Economy 86, pp. 1119-1158.

[26] Joskow, P. (2005), "Vertical integration", in Handbook of Institutional Economics, Menard, C. and Shirley M. editors. Springer.

[27] Klein, P. (2005), The make or buy decision: lessons from empirical studies", in Handbook of New Institutional Economics, Menard C. and Shirley M., Springer.

[28] Kumar, K.B., Rajan, R.G. and Zingales, L. (2001), "What determines firm size?", Working Paper 7208, Graduate School of Business, University of Chicago.

[29] Laffer, A. B. (1969), "Vertical integration by corporations, 1929-1965", The Review of Economics and Statistics 51, pp. 91-93.

[30] Lafontaine, F. and Shaw, K. (1999), "The dynamics of franchise contracting: evidence from panel data", Journal of Political Economy 107, pp. 1041-80.

[31] Lajili, K., Madunic, M. and Mahoney, J. T. (2007), "Testing organizational economics theories on vertical integration", Working Paper, School of Management , University of Ottawa.

[32] Levin, R. C. (1981), "Vertical integration and profitability in oil industry", Journal of Economic Behavior and Organization 2, pp. 215-235.

[33] Levy, D. T. (1985), "The transaction cost approach to vertical integration: an empirical examination", The Review of Economics and Statistics 67, pp438-445.

[34] Lieberman, M. (1991), "Determinants of vertical integration: an empirical test", Journal of Industrial Economics 39, pp. 451-66.

[35] Macchiavello (2008), "Contractual institutions, Financial development and Vertical integration: Theory and Evidence", Working Paper.

[36] Maddigan, R. J. (1981), "The measurement of vertical integration", The Review of Economics and Statistics 63, pp. 328-335.

[37] Masten, S. E. (1984), "The organization of production: evidence from aerospace industry", Journal of Law and Economics 27, pp. 403-17.

[38] Masters, J. K. and Miles, G. (2002), "Predicting the use of external labor arrangements: a test of the transaction cost perspective", Academy of Management Journal 45, pp.431-42.

[39] McLaren, J. (1999), "Supplier relations and the market context: a theory of handshakes", Journal of International Economics 48, pp. 121-138.

[40] McLaren, J. (2000), "Globalization and vertical structure", American Economic Review 90, pp.1239-1254.

[41] Nelson, R. (1963), "Concentration in the manufacturing industries of the United States", New Haven, Yale University Press.

[42] Nugent, J. and Nabli, M. (1992), "Development of financial markets and the size distribution of manufacturing establishments: international comparisons", World Development 20, pp. 1489-1500.

[43] Ohanian, N. K. (1994), "Vertical integration in the US pulp and paper industry, 1900-1940", Review of Economics and Statistics 74, pp. 202-207.

[44] Ornelas, E. and Turner, J. (2008), "Trade liberalization, outsourcing and the hold-up problem", Journal of International Economics 74, pp. 225-241

[45] Penrose, E. (1959), "The theory of the growth of the firm", Basil Blackwell, Oxford.

[46] Pryor, F. (1972), "An international comparison of concentration ratio", The review of Economics and Statistics 54, pp. 130-140.

[47] Rajan, R. and Zingales, L. (1998), "Financial dependence and growth", American Economic Review 88, pp. 559-586.

[48] Schmalensee, R. (1973), "A note on the theory of vertical integration", Journal of Political Economy 81, pp. 442-449.

[49] Scherer, F. (1973), "The determinants of industrial plant sizes in six nations", The Review of Economics and Statistics 55, pp. 135-145.

[50] Spengler, J. (1950), "Vertical integration and anti-trust policy", Journal of Political Economics 58, pp. 347-52.

[51] Tucker, I. B. and Wilder, R. P. (1977), "Trends in vertical integration in the US manufacturing sector", Journal of Industrial Economics 26, pp. 81-94.

[52] Tomiura, E. (2006), "Foreign versus domestic outsourcing: firm level evidence on the role of technology", mimeo.

[53] Tirole, J. (1988), "The theory of industrial organization", MIT press, Cambridge MA.

[54] Walker, G. and Weber, D. (1984), "A transaction cost approach to make or buy decision", Administrative Science Quarterly 29, pp.373-391.

[55] Weiss, A. M. (1994), "Vertical Mergers and firm-specific physical capital: three case studies and some evidence on timing", Journal of Industrial Economics 42, pp. 395-417.

[56] Williamson, O. E. (1970), "Corporate control and business behavior", Prentice-Hall, Englewood Cliffs N.J.

[57] Williamson, O. E. (1991), "Comparative economic organization: the analysis of discrete structural alternatives", Administrative Science Quarterly 36, pp. 269-296.

Table 2.2: Descriptive statistics per country

Country	**GNI per capita**	**Population**	**Observations**
Bangladesh	470	1.42e+08	670
Bolivia	1010	9182015	40
Cambodia	380	1.41e+07	4
Chile	5870	1.63e+07	643
China	1740	1.30e+09	931
Costa Rica	4590	4327228	251
Dominican Republ	2370	8894907	98
Ecuador	2630	1.32e+07	198
Egypt	1250	7.40e+07	1391
El Salvador	2450	6880951	338
Ethiopia	160	7.13e+07	312
Guatemala	2400	1.26e+07	314
Guyana	1010	751218	146
Honduras	1190	7204723	360
India	720	1.09e+09	2088
Kyrgyz Republic	440	5156000	44
Lebanon	6180	3576818	58
Lesotho	960	1794769	14
Madagascar	290	1.86e+07	110
Malawi	160	1.29e+07	127
Mauritius	5260	1248000	81
Moldova	880	4205747	22
Mongolia	690	2554000	135
Morocco	1730	3.02e+07	1090
Nicaragua	910	5486685	354
Oman	9070	2566981	25
Pakistan	690	1.56e+08	850
Philippines	1300	8.31e+07	535
Poland	7110	3.82e+07	41
South Africa	4960	4.52e+07	437
Sri Lanka	1160	1.96e+07	276
Syria	1380	1.90e+07	46
Tajikistan	330	6506980	46
Tanzania	340	3.83e+07	58
Thailand	2750	6.42e+07	609
Uzbekistan	510	2.66e+07	48
Vietnam	620	8.30e+07	1032
Zambia	490	1.17e+07	91
kosovo	3280	9993904	13
Total	1659.367	2.96e+08	912.186

Notes: Data on per capita GNI and Population refer to 2006 (Source: Doing Business Database)

Table 2.3: Descriptive statistics

	Mean	**St.Dev**	**Min**	**25th Pctile**	**50th Pctile**	**75th Pctile**	**Max**	**N**
Firm level data								
Value added to sales	.49	.23	.00067	.32	.49	.65	1	13926
Number of workers	216	646	0	15	44	166	19047	13926
Asset specificity	.6	.29	0	.38	.6	.85	1	13926
Mkt share	19	28	0	1	5	25	100	6999
Country level data								
Enforcement contract costs	37	10	20	29	36	42	58	39
Enforcement contract procedure	27	26	10	15	18	28	137	39
Financial development	-.22	1.7	-2.3	-1.6	-.67	.87	4.9	39
Average tariffs	9.5	6	1.4	4.5	9	13	28	39
Weighted average tariffs	10	6.9	.72	5	9	15	28	39
Trade costs to import	1477	928	375	850	1230	1962	3970	39
Max time to clear customs	17	9.6	4.3	10	15	23	40	39
GNI per capita	2044	2178	160	510	1160	2630	9070	39
Population	8.9e+07	2.7e+08	751218	5486685	1.3e+07	4.5e+07	1.3e+09	39

Table 2.4: Main Effects

	(1)	(2)	(3)	(4
	VI	VI	VI	VI
AS	-0.0135	-0.0233*	-0.0191	-0.0116
	(-1.16)	(-2.03)	(-1.47)	(-0.90)
CI	-0.000587	-0.000682	-0.000191	0.00121
	(-0.58)	(-0.69)	(-0.19)	(1.04)
TB	0.0000307	0.0000270	0.0000466	0.0000255
	(1.13)	(1.02)	(1.36)	(0.78)
FD			0.0134	0.0112
			(1.42)	(1.56)
lpop				-0.0246**
				(-3.13)
lgni				-0.00615
				(-0.37)
INDUSTRY DUMMIES	NO	YES	YES	YES
r2_a	0.00673	0.0282	0.0332	0.0516
N	13926	13926	13926	13926

Notes. The dependent variable, VI, is the ratio of value added to total sales. AS is the ratio of the value of machinery and equipment to the value of machinery and equipment, lands and buildings. CI is the cost to enforce a contract in terms of court and attorney fees as a percentage of a claim assumed to be 2 times the average income per capita (source: Doing Business Database). TB is the cost of importing a 20 foot container (source: Doing Business Database). FD is a credit information index that measures rules affecting the scope, access and the quality of credit information (source: Doing Business Database). t-statistics in parentheses. * $p<0.05$; ** $p<0.01$, *** $p<0.001$.

Table 2.5: Interaction Effects: Main Results

	(1) VI	(2) VI	(3) VI	(4) VI
AS	-0.239***	-0.239***	-0.258*	-0.231
	(-4.39)	(-4.41)	(-2.59)	(-1.46)
AS·CI	0.00134*	0.00134*	0.00135*	0.00203*
	(2.07)	(2.04)	(2.13)	(2.30)
AS·FD	0.0181***	0.0181***	0.0175**	0.0120
	(3.93)	(3.92)	(3.53)	(1.69)
AS·TB	0.0000327**	0.0000327**	0.0000337**	0.0000285
	(3.29)	(3.29)	(3.51)	(1.71)
AS^2	0.0547	0.0546	0.0542	0.0443
	(1.64)	(1.61)	(1.62)	(0.87)
AS·GNI			0.00284	-0.000774
			(0.25)	(-0.04)
MkT Share				0.0000177
				(0.12)
Workers		-0.000000120	-0.000000115	-0.000000919
		(-0.02)	(-0.02)	(-0.11)
INDUSTRY DUMMIES	NO	YES	YES	YES
COUNTRY DUMMIES	YEs	YES	YES	YES
r2_a	0.126	0.125	0.125	0.123
N	13926	13926	13926	6999

Notes. The dependent variable, VI, is the ratio of value added to total sales. AS is the ratio of the value of machinery and equipment to the value of machinery and equipment, lands and buildings. CI is the cost to enforce a contract in terms of court and attorney fees as a percentage of a claim assumed to be 2 times the average income per capita (source: Doing Business Database). TB is the cost of importing a 20 foot container (source: Doing Business Database). FD is a credit information index that measures rules affecting the scope, access and the quality of credit information (source: Doing Business Database). t-statistics in parentheses. * $p<0.05$; ** $p<0.01$, *** $p<0.001$.

Table 2.6: Interaction Effects: Alternative Measure of Contract Enforcement

	(1) VI	(2) VI	(3) VI	(4) VI
AS	-0.184***	-0.184***	-0.240	-0.291
	(-3.73)	(-3.62)	(-2.02)	(-1.54)
AS·CI2	0.000478	0.000478	0.000637*	0.00122*
	(1.53)	(1.53)	(2.06)	(2.21)
AS·FD	0.0153***	0.0153***	0.0139**	0.0113
	(3.72)	(3.72)	(3.40)	(1.80)
AS·TB	0.0000277**	0.0000277**	0.0000309**	0.0000360*
	(2.74)	(2.74)	(3.07)	(2.44)
AS^2	0.0556	0.0557	0.0546	0.0517
	(1.60)	(1.57)	(1.55)	(0.99)
AS·GNI			0.00764	0.0114
			(0.61)	(0.53)
Mkt Share				0.0000130
				(0.09)
Workers		8.70e-08	0.000000119	-0.000000363
		(0.01)	(0.02)	(-0.04)
INDUSTRY DUMMIES	YES	YES	YES	YES
COUNTRY DUMMIES	YES	YES	YES	YES
r2_a	0.839	0.839	0.839	0.850
N	13926	13926	13926	6999

Notes: The dependent variable, VI, is the ratio of value added to total sales. AS is the ratio of the value of machinery and equipment to the value of machinery and equipment, lands and buildings. CI2 is the average number of procedures for enforcing a contract from the moment the plaintiff files a lawsuit in court until the moment of payment (source: Doing Business Database). TB is the cost of importing a 20 foot container (source: Doing Business Database). FD is a credit information index that measures rules affecting the scope, access and the quality of credit information (source: Doing Business Database). t-statistics in parentheses. * p<0.05; ** p<0.01, *** p<0.001.

Table 2.7: Interaction Effect: Alternative Measure of Trade Barriers

	(1)	(2)	(3)	(4)	(5)	(6)
	VI	VI	VI	VI	VI	VI
AS	-0.164***	-0.164***	-0.177***	-0.177***	-0.188***	-0.188***
	(-3.72)	(-3.73)	(-3.85)	(-3.84)	(-3.96)	(-3.94)
AS·CI	0.000300	0.000300	0.000861	0.000861	0.000744	0.000744
	(0.38)	(0.38)	(1.45)	(1.44)	(1.32)	(1.31)
AS·FD	0.0129**	0.0129**	0.0109**	0.0109**	0.00987*	0.00987*
	(3.48)	(3.47)	(2.88)	(2.87)	(2.59)	(2.59)
AS· Tariffs	0.00192	0.00192				
	(1.44)	(1.43)				
AS· Ave cust			0.00228	0.00228		
			(1.98)	(1.97)		
AS· Max cust					0.00228*	0.00228*
					(2.39)	(2.39)
AS[1]	0.0520	0.0520	0.0548	0.0548	0.0573	0.0572
	(1.51)	(1.49)	(1.61)	(1.58)	(1.71)	(1.67)
Workers		-0.000000106		-6.43e-08		-3.58e-08
		(-0.02)		(-0.01)		(-0.01)
INDUSTRY DUM	YES	YES	YES	YES	YES	YES
COUNTRY DUM	YES	YES	YES	YES	YES	NO
r2_a	0.839	0.839	0.839	0.839	0.839	0.839
N	13926	13926	13922	13922	13926	13926

Notes: The dependent variable, VI, is the ratio of value added to total sales. AS is the ratio of the value of machinery and equipment to the value of machinery and equipment, lands and buildings. CI is the cost to enforce a contract in terms of court and attorney fees as a percentage of a claim assumed to be 2 times the average income per capita (source: Doing Business Database). Tariffs in the average tariff on machineries and equipments (source: TRAINS). Ave Cust (Max Cust) is the average (maximum) time to clear customs (source: ICA). FD is a credit information index that measures rules affecting the scope, access and the quality of credit information (source: Doing Business Database). t-statistics in parentheses. * p<0.05; ** p<0.01, *** p<0.001.

Table 2.8: Interaction Effect: Alternative Measure of Financial Development

	(1)	(2)	(3)	(4)	(5)	(6)
	VI	VI	VI	VI	VI	VI
AS	-0.157**	-0.157**	-0.173	-0.127**	-0.127**	-0.183
	(-3.50)	(-3.46)	(-1.60)	(-2.90)	(-2.83)	(-1.42)
AS·CI	0.00104	0.00104	0.00105			
	(1.53)	(1.53)	(1.59)			
AS·CI2				0.000503	0.000504	0.000633
				(1.74)	(1.74)	(2.00)
AS·FD2	0.0179*	0.0179*	0.0171	0.0167*	0.0167*	0.0147*
	(2.44)	(2.44)	(1.89)	(2.65)	(2.64)	(2.18)
AS·TB	0.0000252*	0.0000252*	0.0000259**	0.0000232*	0.0000232*	0.0000259*
	(2.59)	(2.59)	(2.76)	(2.33)	(2.33)	(2.50)
AS^2	0.0515	0.0515	0.0513	0.0526	0.0527	0.0519
	(1.49)	(1.46)	(1.47)	(1.49)	(1.46)	(1.45)
AS·GNI			0.00216			0.00705
			(0.16)			(0.51)
Workers		2.55e-08	2.54e-08		0.000000189	0.000000207
		(0.00)	(0.00)		(0.03)	(0.04)
INDUSTRY DUM	YES	YES	YES	YES	YES	YES
COUNTRY DUM	YES	YES	YES	YES	YES	NO
r2_a	0.125	0.125	0.125	0.125	0.125	0.125
N	13926	13926	13926	13926	13926	13926

Notes: The dependent variable, VI, is the ratio of value added to total sales. AS is the ratio of the value of machinery and equipment to the value of machinery and equipment, lands and buildings. CI is the cost to enforce a contract in terms of court and attorney fees as a percentage of a claim assumed to be 2 times the average income per capita (source: Doing Business Database). CI2 is the average number of procedures for enforcing a contract from the moment the plaintiff files a lawsuit in court until the moment of payment (source: Doing Business Database). TB is the cost of importing a 20 foot container (source: Doing Business Database). FD2 is the principal component of a legal right index (which measures the degree to which collateral and bankruptcy laws facilitate lending) and a credit information index (source: Doing Business Database). t-statistics in parentheses. * p<0.05; ** p<0.01, *** p<0.001.

Table 2.9: Interaction Effect: HLM Regressions

	(1) VI	(2) VI	(3) VI
AS	-0.245***	-0.245***	-0.264**
	(-4.47)	(-4.47)	(-2.75)
AS·CI	0.00146*	0.00146*	0.00154*
	(2.11)	(2.11)	(2.09)
AS·FD	0.0181***	0.0181***	0.0175**
	(3.39)	(3.39)	(2.82)
AS·TB	0.0000333*	0.0000333*	0.0000345*
	(2.29)	(2.29)	(2.23)
AS^2	0.0558*	0.0558*	0.0562*
	(2.38)	(2.37)	(2.37)
AS·GNI			0.00237
			(0.21)
Workers		-1.00e-07	0
		(-0.03)	(0.00)
INDUSTRY DUMMIES	YES	YES	YES
COUNTRY DUMMIES	YES	YES	YES
r2_a	0.112	0.113	0.115
N	13926	13926	13926

Notes: The dependent variable, VI, is the ratio of value added to total sales. AS is the ratio of the value of machinery and equipment to the value of machinery and equipment, lands and buildings. CI is the cost to enforce a contract in terms of court and attorney fees as a percentage of a claim assumed to be 2 times the average income per capita (source: Doing Business Database). TB is the cost of importing a 20 foot container (source: Doing Business Database). FD is a credit information index that measures rules affecting the scope, access and the quality of credit information (source: Doing Business Database). t-statistics in parentheses. * p<0.05; ** p<0.01, *** p<0.001.

Table 2.10: Interaction Effects: IV regressions

	(1) VI	(2) VI	(3) VI	(4) VI
AS	-0.211***	-0.212***	-0.202*	-0.205
	(-5.67)	(-5.61)	(-2.39)	(-1.64)
AS·CI	0.00141*	0.00141*	0.00143*	0.00212*
	(2.04)	(2.03)	(2.15)	(2.24)
AS·FD	0.0111***	0.0111***	0.0115**	0.0092
	(3.86)	(3.85)	(3.43)	(1.62)
AS·TB	0.0000366**	0.0000366**	0.0000321**	0.0000296
	(3.39)	(3.39)	(3.59)	(1.72)
AS^2	0.0599	0.0599	0.0597	0.0493
	(1.74)	(1.69)	(1.669)	(0.65)
AS·GNI			0.00184	-0.000674
			(0.12)	(-0.04)
MkT Share				0.0000217
				(0.15)
Workers		-0.000000220	-0.000000215	-0.000000119
		(-0.99)	(-0.87)	(-0.85)
INDUSTRY DUMMIES	NO	YES	YES	YES
COUNTRY DUMMIES	YES	YES	YES	YES
N	13926	13926	13926	6999

Notes: The dependent variable, VI, is the ratio of value added to total sales. AS is the ratio of the value of machinery and equipment to the value of machinery and equipment, lands and buildings. CI is the cost to enforce a contract in terms of court and attorney fees as a percentage of a claim assumed to be 2 times the average income per capita (source: Doing Business Database). TB is the cost of importing a 20 foot container (source: Doing Business Database). FD is a credit information index that measures rules affecting the scope, access and the quality of credit information (source: Doing Business Database). t-statistics in parentheses. * $p<0.05$; ** $p<0.01$, *** $p<0.001$.

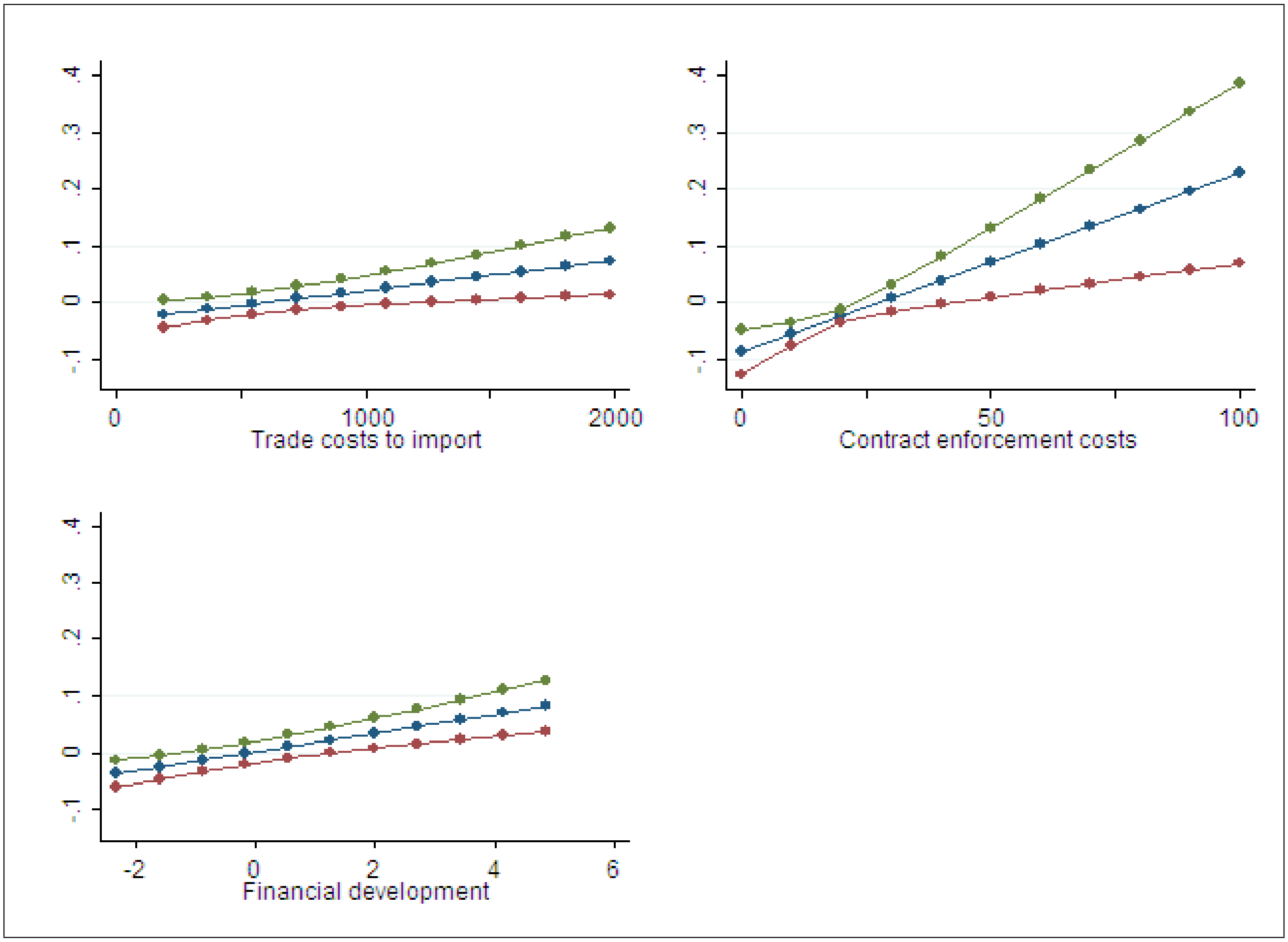

Figure 2: Marginal Effect of Asset Specificity on Vertical Integration

Notes: The central line depicts the estimated marginal effect of vertical integration (measured as the ratio of value added to total sales) on asset specificity (measured as the ratio of the value of machinery and equipment to the value of machinery and equipment, lands and buildings). The other two lines define the 5 percent confidence boundaries. " Trade costs to import" is the cost of importing a 20 foot container; "Financial development" is a credit information index that measures rules affecting the scope, access and the quality of credit information. "Contract enforcement costs" are the cost to enforce a contract in terms of court and attorney fees as a percentage of a claim assumed to be 2 times the average income per capita.

Chapter 3

Productivity, Welfare and Reallocation: Theory and Firm-Level Evidence

3.1 Introduction

How much of growth comes from innovation and technical advances, and how much from changes in allocative efficiency? This question arises in a variety of contexts, in fields as diverse as growth and development, trade, and industrial organization. Yet, despite the importance of the question, there is no consensus regarding the answer. A large number of papers have proposed a bewildering variety of methods to measure the importance of allocative efficiency, leading to a wide range of numerical estimates. Much of the confusion is due to the lack of an organizing conceptual framework for studying this issue. We propose such a framework, and then provide a quantitative answer using one particular set of data.

In starting such a project, one immediately faces the question: What do we mean by allocative efficiency? Indeed, what do we mean by growth? We take the view that growth is an improvement in social well-being. While growth is commonly described in terms of GDP per worker or consumption per capita, these statistics are usually viewed as indicators of some deeper target. Their virtue, a considerable one, is that they can be generated from aggregate data, which are usually readily available. We ask if we can produce a more complete description of economic welfare and its change, while also restricting ourselves to aggregate data. Given an empirical method for characterizing aggregate welfare, allocative efficiency is naturally defined as the increment to welfare achieved by reallocating productive resources to more efficient uses, holding constant the aggregate quantities of resources used in production.

We undertake three tasks. First, we begin from a utility-maximization problem that is standard in the economics of growth and business cycles: We assume that a representative household with an infinite horizon values both consumption and leisure, and maximizes utility subject to a standard intertemporal budget constraint.[1] We show that, in a closed economy without distortionary taxes, this standard specification of the objective function implies, to a first order, that welfare depends on the present discounted value of total factor productivity (TFP) for the aggregate economy and on the initial level of the capital stock.

[1]While the valuation of leisure is not common in a growth context, it is quantitatively very important. Reviewing a large number of social goods that are valued by consumers but not counted in GDP, Nordhaus and Tobin (1973) found the omission of leisure the most significant (with another imputation for the use of non-market time in home production the second most important). Our household maximization framework also corrects automatically for two other gaps that Nordhaus and Tobin find are significant: The need to subtract depreciation (moving to a NDP rather than a GDP framework), and the need to adjust for a growing population.

This result is "TFP without firms"—it is derived purely from the standard model of a price-taking, competitive household. Thus, our result holds for all specifications of technology and market structure, including ones where TFP does not measure technology, as long as consumers are free to choose the quantities of goods they purchase at prices they take as being outside their control. Here we follow the intuition of Basu and Fernald (2002), and supply a general proof of their basic proposition that TFP is relevant for welfare. We then extend this result in several directions by allowing for distortionary taxes, public spending, and an open economy. We discuss how these extensions modify our fundamental result by introducing additional terms in the first order representation of welfare.

Second, we use this result to show that we can calculate the welfare contributions of particular sectors of the economy—which can be as large as industries and as small as individual firms. We present industry and firm contributions to welfare for a set of European OECD countries (Belgium, France, Great Britain, Italy, Spain), using industry data from EU-KLEMS and the Amadeus firm-level data set. Among other things, we use these data to compare the distributions of firm-level productivities relative to the country means across the countries in our sample, and ask how much welfare would increase if, for example, Italian firms had the same relative productivity distribution as those in Great Britain. This analysis is akin to that of Hsieh and Klenow (2009), but it has a direct welfare interpretation and is more general because it does not require assumptions about the production technology.

Third, we show how to decompose welfare—aggregate TFP—into components due to technology, aggregate distortions, and allocative inefficiency. Any such decomposition *does* depend on assumptions about production technology, adjustment costs, and industrial organization. Different assumptions will lead to different decompositions, but within the same overarching social-welfare framework. Finally, we implement one specific decomposition, again based on Basu and Fernald (2002), using firm-level data from a number of European countries represented in the Amadeus data set. We find that welfare grows significantly faster than technology changes, but improvements in allocative efficiency usually account for a modest fraction of the gap between the two.

Our first result clarifies the nature of the important link between welfare on the one hand and aggregate productivity and national income measurement on the other.[2] Our main goal in this section is to provide a clear objective for any decomposition of productivity. To have an economic interpretation, any such decomposition should indicate how productivity contributes to the ultimate target, which is social welfare. Under the usual assumptions and to a first-order approximation, that target *is* a measure of productivity, aggregate TFP. But the method is more important than the specific result. A different specification of the consumer's problem may deliver a different result about the relationship between welfare and productivity. (In fact, we derive results in the paper showing that under certain conditions—for example, if there are distortionary taxes—the correct welfare measure may differ substantially from the usual Solow residual.) But it is still important for researchers interested in decomposing productivity or studying allocative efficiency to relate their empirical method to the solution to some well-specified maximization problem so that the implications of their decompositions for some ultimate welfare objective, which are

[2] Earlier works also make a connection between the two. Some of the most important are Nordhaus and Tobin (1973), Weitzman (1976, 2003) and Hulten (1978). Our approach closely follows that of Basu and Fernald (2002). In an independent and simultaneous work, Hulten and Schreyer (2009) obtain several similar results in a continuous-time setting under the assumptions of perfect foresight and perfect competition.

usually left implicit in any such study, can be made explicit, and the necessary assumptions can be examined closely.

One benefit of starting from a well-defined objective function is that it enables the researcher to take consistent, model-based positions on a variety of issues that bedevil the measurement of productivity and allocative efficiency. For example, Baker and Rosnick (2007), reasoning that the ultimate object of growth is consumption, make the reasonable conjecture that one should deflate nominal productivity gains by a consumption price index to create a measure they call "usable productivity." We begin from the assumption that consumption (and leisure) at different dates are the only inputs to economic wellbeing, but nevertheless show that output should be calculated in the conventional way, rather than being deflated by consumer prices.[3] To take another example, there is no consensus in the literature about the proper treatment of scale economies. Most researchers examine allocative efficiency by asking whether firms with higher levels of Hicks-neutral technology produce more output. Others pose the same question in terms of labor productivity, which includes scale economies but does not subtract capital's contribution to output. Using our framework, it is easy to show that only the Solow TFP index gives the correct welfare accounting. Unlike a pure technical change measure, the Solow residual includes scale effects, which do contribute to welfare by producing more output for given inputs. Unlike labor productivity, the TFP residual subtracts the change in capital input valued at its opportunity cost to the consumer.

Our analytical results create several links between productivity and welfare. One im-

[3] The other main adjustment by Baker and Rosnick, moving to a net measure of output as a starting point for productivity measurement, follows a long tradition of research on this topic, and is fully supported by our derivation.

portant message is that welfare depends on the entire expected future path of TFP. Not surprisingly, the same size change in current TFP has very different effects on welfare if it is expected to be persistent than if it is expected to be transitory. This result suggests new ways of assessing the importance of reallocation. To our knowledge, the literature does not examine the time-series properties, especially the persistence, of measures of allocative efficiency. But our derivation shows that to understand the contribution that reallocation makes to growth, it is important to know the persistence as well as the mean. In principle, the allocative efficiency component of TFP might be either more or less persistent than total TFP, making reallocation either more or less important than its average share would suggest.

So far we have been vague about whether our results relate to TFP in levels or in growth rates. In fact, our results apply to both. We show that the *level* of welfare for a representative consumer is, up to a first-order approximation, proportional to the present discounted value of expected log levels of TFP. Welfare *change* for the consumer, on the other hand, is proportional to the change in log levels, i.e., to the present discounted value of TFP growth as we define it (equal to the standard Solow productivity residual if there are zero economic profits), plus an "expectation revision" term that depends on the difference in expectations of future log levels of TFP between time t-1 and time t. Under perfect foresight, the expectation revision term is identically zero, and the change in welfare is proportional to the present discounted value of current and future Solow residuals alone.

Starting from a well-posed optimization problem also forces us to confront two issues in national income and welfare measurement. First, our derivation shows that "consumption" should be defined as any good or service that consumers value, whether or not it is included in GDP. Similarly, "capital" should include all consumption that is foregone now in order to raise consumption possibilities for the future. These items include, for example, environmental quality and intangible capital. Of course, both are hard to measure and even harder to value, since there is usually no explicit market price for either good. But our derivation is quite clear on the principle that the environment, intangibles and other non-market goods should be included in our measure of "welfare TFP." We follow conventional practice in restricting the output measure for our TFP variable to market output (and the inputs to measured physical capital and labor), but in so doing we, and almost everyone else, are mismeasuring real GDP and TFP. Second, our starting point of a representative-consumer framework implies that we automatically ignore issues of distribution that intuition says should matter for social welfare. We believe that distributional issues are very important. However, our objective of constructing a better welfare measure from aggregate data alone implies that we cannot incorporate measures of distribution into our framework. Thus, we maintain the representative-consumer framework, but without in any way minimizing the importance of issues that cannot be handled within that framework.

Having established that aggregate TFP is the natural measuring stick for aggregate welfare, we then ask the next natural question: Can one show what contribution a subset of the economy (which may be as small as a single firm) makes to the aggregate welfare index? The answer is yes, as shown by Domar (1961). Domar established that a correctly-weighted average of sectoral TFP residuals sums to Solow's familiar aggregate index. We use a variant of his result to present the welfare contributions made by large sectors of the economy using the EU-KLEMS dataset. We compare the sources of welfare differences across countries, asking what fraction of cross-country differences are due to differences in industrial structure as opposed to differences in the welfare contributions of the same sector across countries. We then do a similar exercise using our firm-level data over the period 1998-2004, and investigate the extent to which differences in the relative productivity distribution of firms across countries contributes to differences in welfare.

Finally, we decompose aggregate TFP into components. As we noted, while TFP is itself meaningful in welfare terms without any additional assumptions, we need to make assumptions about firm technology and behavior in order to decompose TFP in a meaningful way. We use a variant of the decomposition of Basu and Fernald (2002), which is derived by assuming that firms minimize costs and are price-takers in factor markets, but may have market power for the goods they sell and might produce with increasing returns to scale. As we also noted, different assumptions about technology would give different decompositions, without changing the essential features of the results. For example, Basu and Fernald (2002) assume that factors are freely mobile across firms, without adjustment costs, while Basu, Fernald and Shapiro (2001) extend the framework to include costly factor adjustment. Abel (2003) and Basu et al. (2001) show that adjustment costs are a special type of intangible capital, of a sort that needs to be accumulated in fixed proportions with physical capital. Thus, accounting for adjustment costs in the empirical results would require us to impute an addition to measured output, which is conceptually the same issue as accounting for non-market consumption goods or for more general forms of intangible capital accumulation.

Some of the components in the decomposition we use can be clearly identified as being due to reallocation, since they depend on marginal products of identical inputs not being equalized across firms. Other components depend on aggregate distortions, such as the average degree of market power and various tax rates. In order to estimate the reallocation terms, we need to estimate firm-level marginal products. We do so using firm-level data for a number of manufacturing industries across six European countries, as represented in the Amadeus data base.[4] We extend the existing decomposition to study reallocation both within and between industries, since the two kinds of reallocation may have different policy implications.

We use the Amadeus data to estimate production functions for firms within a number of manufacturing industries across six countries. We experimented with a variety of estimation methods to ensure that our main results were robust. We found that there is usually a substantial gap between our estimates of technical change for each manufacturing industry and that industry's contribution to aggregate TFP growth (and hence welfare). However, for most countries, the majority of this gap is due to the aggregate distortions (especially when taxes are included in the decomposition). Reallocation strictly defined usually accounts for a small fraction of the gap.

The paper is organized as follows. We present the key equations linking productivity and welfare in Section 2, with the full derivation presented in an appendix. While our derivations link welfare to both TFP levels and growth rates, we choose to work mostly in growth rates, since there are well-known difficulties in comparing TFP levels across industries and countries. In Section 3, we show how to identify firm and sector level contributions to the productivity residual. In Section 4 we present our data and discuss measurement issues. Section 5 assesses the contribution of different sectors and groups of firm to the productivity residual in five European countries. Section 6 shows how to decompose the productivity

[4] Petrin, White and Reiter (2009) also use firm-level data to implement a variant of the Basu-Fernald (2002) decomposition. They use U.S. Census data for manufacturing industries. We compare our results to theirs in Section 6.

residual into components that reflect reallocation, technology, and aggregate distortions. We than discuss in Section 7 the econometric framework we use to estimate these sources of welfare change, and present the results in Section 8. We conclude in Section 9 with some reflections and suggestions for future research.

3.2 The Productivity Residual and Welfare

Intuition suggests that technological progress is responsible for the secular increase in the standard of living. The usual justification for studying the Solow residual is that, under perfect competition and constant returns to scale, it measures technological change. However, should we care about the Solow residual in an economy with non-competitive output markets, non-constant returns to scale, and possibly other distortions where the Solow residual is no longer a good measure of technological progress? Here we build on the intuition of Basu and Fernald (2002) that a slightly modified form of the Solow residual is welfare relevant even in those circumstances and derive rigorously the relationship between a modified version of the productivity residual (in growth rates or log levels) and the intertemporal utility of the representative household. The fundamental result we obtain is that, to a first-order approximation, utility reflects the present discounted values of productivity residuals.

Our results are complementary to those in Solow's classic (1957) paper. Solow established that if there was an aggregate production function then his index measured its technical change. We now show that under a very different set of assumptions, which are disjoint from Solow's, the familiar TFP index is also the correct welfare measure. The results are parallel to one another. Solow did not need to assume anything about the consumer side of the economy to give a technical interpretation to his index, but he had to make assumptions about technology and firm behavior. We do not need to assume anything about the firm side (which includes technology, but also firm behavior and industrial organization) in order to give a welfare interpretation, but we do need to assume the existence of a representative consumer. Both results assume the existence of a potential function (Hulten, 1973), and show that TFP is the rate of change of that function. Which result is more useful depends on the application, and the trade-off that one is willing to make between having a result that is very general on the consumer side but requires very precise assumptions on technology and firm behavior, and a result that is just the opposite.

3.2.1 Approximating around the steady state

More precisely, assume that the representative household maximizes intertemporal utility:

$$V_t = E_t \sum_{s=0}^{\infty} \frac{1}{(1+\rho)^s} \frac{N_{t+s}}{H} U(C_{1,t+s}, .., C_{Z,t+s}; \overline{L} - L_{t+s}) \tag{3.1}$$

where $C_{i,t}$ is the capita consumption of good i at time t, L_t are hours of work per capita, $\overline{L}$ is the time endowment, and N_t population. H is the number of households, assumed to be fixed and normalized to one from now on. X_t denotes Harrod neutral technological progress, assumed to be common across all sectors. Population grows at constant rate n and X_t at rate g. For a well defined steady state in which hours of work are constant while consumption and real wage share a common trend, we assume that the utility function has the King, Plosser and Rebelo (1988) form:

$$U(C_{1,t+s},..,C_{Z,t+s};L-\overline{L}_s)=\frac{1}{1-\sigma}C(C_{1,t+s},..,C_{Z,t+s})^{1-\sigma}\nu(\overline{L}-L_{t+s})$$

with $0<\sigma<1$ or $\sigma>1$.[5] We assume that $C()$ has constant returns to scale. Define $c_{i,t+s}=\frac{C_{i,t+s}}{X_{t+s}}$. We can rewrite the utility function in a normalized form as follow:

$$v_t=\frac{V_t}{N_tX_t^{(1-\sigma)}}=E_t\sum_{s=0}^{\infty}\beta^sU(c_{1,t+s},..,c_{Z,t+s};\overline{L}-L_{t+s}) \tag{3.2}$$

where $\beta=\frac{(1+n)(1+g)^{1-\sigma}}{(1+\rho)}$ is assumed to be less than one. The budget constraint (with variables scaled by N_tX_t) is:

$$k_t+b_t=\frac{(1-\delta)+p_t^K}{(1+g)\,(1+n)}k_{t-1}+\frac{(1+r_t)}{(1+g)\,(1+n)}b_{t-1}+p_t^LL_t+\pi_t-\sum_{i=1}^{Z}p_{i,t}^Cc_{i,t} \tag{3.3}$$

New capital goods are the numeraire, $k_t=\frac{K_t}{X_tN_t}$ denotes capital per effective worker, $b_t=\frac{B_t}{P_t^IX_tN_t}$ are real bonds. $p_t^K=\frac{P_t^K}{P_t^I}$, $p_t^L=\frac{P_t^L}{P_t^IX_t}$, $p_{i,t}^C=\frac{P_{i,t}^C}{P_t^I}$ denote, respectively, the user cost of capital, the wage per hour of effective worker, and the price of consumption goods. $(1+r_t)$ is the real interest rate (again in terms of new capital goods) and $\pi_t=\frac{\Pi_t}{P_t^IX_tN_t}$ denotes profits.

Log linearizing around the non stochastic steady state, intertemporal household utility

[5] If $\sigma=1$, then the utility function must be $U(C_1,..,C_G;\overline{L}-L)=\log(C)-\nu(\overline{L}-L)$. See King, Plosser and Rebelo (1988).

can be written (to a first order approximation) as:

$$v_t = v + E_t \sum_{s=0}^{\infty} \beta^s \lambda \left[\sum_{i=1}^{Z} p_i^C c_i \widehat{c}_{i,t+s} + i\widehat{i}_{t+s} - p^L L \ \widehat{L}_{t+s} - \frac{p^K k}{(1+g)(1+n)} \ \widehat{k}_{t+s-1} \right] + \lambda \frac{(1-\delta)+p^K}{(1+g)(1+n)} k \widehat{k}_{t-1} \tag{3.4}$$

where v is the steady state value of utility, $\widehat{x} = \log x_t - \log x$ denote log deviation from the steady state. In obtaining this result we have used the FOC of the household maximization problem:

$$U_{c_{i,t}} - \lambda_t p_{i,t}^C = 0 \tag{3.5}$$

$$U_{L_t} + \lambda_t p_t^L = 0 \tag{3.6}$$

$$-\lambda_t + \beta E_t \frac{(1-\delta)+p_{t+1}^K}{(1+g)(1+n)} \lambda_{t+1} = 0 \tag{3.7}$$

$$-\lambda_t + \beta_t \frac{1}{(1+g)(1+n)} E_t (1+r_{t+1}) \lambda_{t+1} = 0 \tag{3.8}$$

and the log linear approximation of the budget constraint around the steady state:

$$k \ \widehat{k}_t + b\widehat{b}_t - \frac{(1-\delta)+p^K}{(1+g)(1+n)} k\widehat{k}_{t-1} - \frac{(1+r)}{(1+g)(1+n)} b\widehat{b}_{t-1} - p^L L \widehat{L}_t - p^L L \widehat{p}_t^L - \frac{p^K k}{(1+g)(1+n)} \widehat{p}_t^K$$
$$-\pi\widehat{\pi}_t + \sum_{i=1}^{Z} p_i^C c_i \widehat{c}_{i,t} + \sum_{i=1}^{Z} p_i^C c_i \widehat{p}_{i,t} = 0$$

Equation (3.4) says that intertemporal utility (in log deviation from the steady state) equals the expected present discounted value of terms that represent the sum of the components of final demand (in log deviation from the steady state), weighted by their steady state contribution to demand, minus primary inputs (in log deviation from the steady state) times their respective steady state factor prices.

3.2.2 Connecting with the productivity residual

We are now close to relating utility to a modified version of the Solow residual. There are two options here. The first one is to obtain a first order approximation for the log level of utility in terms of the log level productivity residual. Simple manipulations allow us to rewrite log level utility as a function of expected future Solow residuals plus an initial (log) level productivity residual. The second one focuses instead on approximating the log change in utility over time.

To connect utility with the Solow residual, we will rely on the following (Divisia) definition of growth in normalized value added:

$$\Delta \log y_t = \sum_{i=1}^{Z} \frac{p_i^C c_i}{p^Y y} \Delta \log c_{i,t+s} + \frac{i}{p^Y y} \Delta \log i_t \tag{3.9}$$

Using the fact that nominal value added $P_t^Y Y_t = \sum_{i=1}^{Z} P_{i,t}^C C_{i,t} N_t + P_t^I I_t$, it is also true that non-normalized value added growth, $\Delta \log Y_t$, equals:

$$\Delta \log Y_t = \sum_{i=1}^{Z} \frac{P_i^C C_i N}{P^Y Y} \Delta \log(C_{i,t} N_t) + \frac{P^I I}{P^Y Y} \Delta \log I_t \tag{3.10}$$

where the growth rate of each demand component is aggregated using constant steady state shares.[6]

[6] Here we are departing slightly from convention, as value added is usually calculated with time varying

To establish a relationship with the (log) level of productivity, we will, instead, use the fact that, to a first order approximation, the level of value added (in terms of normalized variables) is given by:

$$\widehat{y}_t = \log y_t - \log y = \sum_{i=1}^{Z} \frac{P_i^C C_i N}{P^Y Y} \widehat{c}_{it} + \frac{P^I I}{P^Y Y} \widehat{i}_t = \sum_{i=1}^{Z} s_{c_i} \widehat{c}_{it} + s_i \widehat{i}_t \tag{3.11}$$

Starting from this latter case, using (3.11), intertemporal utility in (3.4) can be written as:

$$v_t = v + E_t \sum_{s=0}^{\infty} \beta^s \lambda p^Y y \left[\widehat{y}_t - \frac{p^L L}{p^Y y} \widehat{L}_{t+s} - \frac{p^K k}{p^Y y (1+g)(1+n)} \widehat{k}_{t+s-1} \right] + \lambda \frac{(1-\delta) + p^K}{(1+g)(1+n)} k \widehat{k}_{t-1} \tag{3.12}$$

which, after some manipulations detailed in the appendix can be rewritten as:

$$v_t - v = \left(\lambda p^Y y\right) E_t \sum_{s=0}^{\infty} \beta^s \log pr_{t+s} + \lambda \frac{(1-\delta) + p^K}{(1+g)(1+n)} k \widehat{k}_{t-1} + f(t) \tag{3.13}$$

where:

$$\log pr_t = \log Y_t - s_L \log N_t L_t - s_K \log K_{t-1} \tag{3.14}$$

is the log level of aggregate value added, $\log Y_t$, minus aggregate factor inputs, $\log N_t L_t$ and $\log K_t$ multiplied by their respective distributional shares, s_L and s_K. $s_C = \frac{\sum_{i=1}^{Z} p_i^C c_i}{p^y y}$ is the

shares.

share of consumption goods in value added and $f(t)$ is a deterministic function of time:

$$\begin{aligned} f(t) &= -\frac{\lambda p^y y}{1-\beta}\left[\log y - s_L \log L - s_K \log k + \frac{\beta}{(1-\beta)}\left[g(1-s_K)+n(1-s_L-s_K)\right]\right] \\ &\quad -\frac{\lambda p^y y}{1-\beta}\left[(1-s_K)\log X_t + (1-s_L-s_K)\log N_t\right] - \frac{\lambda p^y y s_K}{1-\beta}(n+g) \end{aligned} \tag{3.15}$$

Utility, therefore, is an increasing function of the sequence of (log) level aggregate productivity residuals, appropriately discounted.[7]It also depends upon the log deviation of the initial level of the capital stock, $\widehat{k}_{t-1}$, since for any sequence of productivity, welfare is higher if the consumer starts with an higher initial endowment of capital.

To establish the relationship with the Solow residual (a growth rate concept) there are two options. One option is to use the fact that, for any variable x :

$$E_t \widehat{x}_{t+s} = E_t(\log x_{t+s} - \log x) = E_t \sum_{i=1}^{s} (\log x_{t+i} - \log x_{t+i-1}) + \log x_t - \log x$$

In the appendix we show that log level utility (3.4), implies that per capita (log) intertemporal utility can also be written as:

$$v_t - v = \frac{\lambda p^Y y}{(1-\beta)} E_t \sum_{s=1}^{\infty} \beta^s \Delta \log pr_{t+s} + \frac{\lambda p^Y y}{(1-\beta)} \log pr_t + \lambda \frac{(1-\delta)+p^K}{(1+g)(1+n)} k \widehat{k}_{t-1} + f(t) \tag{3.16}$$

where $\Delta \log pr_t$ denotes the "modified" Solow productivity residual:

[7] Note that the utility index v is positive for $0 < \sigma < 1$ and negative for $\sigma > 1$.

$$\Delta \log pr_{t+s} = \Delta \log Y_{t+s} - s_L \Delta \log N_{t+s} L_{t+s} - s_K \Delta \log \ K_{t+s-1} \tag{3.17}$$

We use the word "modified," for two reasons. First, we do not assume that the distributional shares of capital and labor add to one, as they would if there were zero economic profits. Zero profits are guaranteed in the benchmark case with perfect competition and constant returns to scale, but can also arise with imperfect competition and increasing returns to scale, as long as there is free entry, as in the standard Chamberlinian model of imperfect competition. Second, the distributional shares are calculated at their steady state values and, hence, are not time varying. Rotemberg and Woodford (1991) argue that in a consistent first-order log-linearization of the production function the shares of capital and labor should be taken to be constant, and Solow's (1957) use of time-varying shares amounts to keeping some second-order terms while ignoring others. Now log level productivity has been written as a combination of expected future Solow residuals and one initial productivity term in levels. Assume one is willing to make the assumption that an economy at time $t-1$ was at the steady state, so that $\log y_{t-1} - s_L \log L_{t-1} - s_K \log k_{t-1} = \log y - s_L \log L - s_K \log k$ In this special case simple algebra shows that v_t depends upon the expected present discounted value of Solow residuals (from the present to infinity).

$$v_t - v = \frac{\lambda p^Y y}{(1-\beta)} E_t \sum_{s=0}^{\infty} \beta^s \Delta \log pr_{t+s} + \lambda \frac{(1-\delta) + p^K}{(1+g)(1+n)} k \widehat{k}_{t-1} + f_0 \tag{3.18}$$

where:

$$f_0 = \frac{\lambda p^Y y}{(1-\beta)^2}\left[g(1-s_K)+n(1-s_K-s_L)\right] \tag{3.19}$$

An alternative and more satisfactory way to illustrate the relationship between welfare and the Solow residual (with no level term) is to return to (3.4) and take its difference through time ($\Delta v_t = v_t - v_{t-1}$). Using only the definition of value added in growth terms, equation (3.9), the growth rate of per capita utility can be written as follows:

$$\begin{aligned}\Delta v_t &= \lambda p^y y E_t \sum_{s=0}^{\infty} \beta^s \Delta \log pr_{t+s} \\ &\quad + \lambda p^y y \sum_{s=0}^{\infty} \beta^s \left[E_t \log pr_{t+s} - E_{t-1} \log pr_{t+s}\right] \\ &\quad + \lambda \frac{(1-\delta)+p^K}{(1+g)(1+n)} k \Delta k_{t-1} + f_1\end{aligned} \tag{3.20}$$

where $E_t \Delta \log pr_{t+s}$ represents the expected Solow residual while $E_t \log pr_{t+s} - E_{t-1} \log pr_{t+s}$ represents the revision in expectations of the log level of the productivity residual, based on the new information received between t-1 and t. In addition, $\Delta \log K_{t-1}$ captures the change in the initial endowment of capital. Finally, the constant f_1 is:

$$f_1 = \frac{\lambda p^Y y}{(1-\beta)}\left[g(1-s_K)+n(1-s_K-s_L)\right] \tag{3.21}$$

Note that the revision term in the second summation will reduce to a linear combination of the innovations in the stochastic shocks affecting the economy at time t. Moreover, if we assume that the modified Solow residual follows a simple stable first order autoregressive process, then the current Solow residual, $\Delta \log pr_t$, is a sufficient statistic for all the terms in the first summation. In this case, the growth in expected per capita utility is a linear function of today's actual Solow residual, of innovations at time t in the stochastic processes driving the economy and of the change in initial endowment of capital.

3.2.3 Extensions

We now show that our method of using TFP to measure welfare can be extended to cover multiple types of capital and labor, taxes, and government expenditure. The first extension modifies our baseline results in only a trivial way, but the others all require more substantial changes to the formulas above. These results show that the basic idea of using TFP to measure welfare holds in a variety of economic environments, but also demonstrate the advantage of deriving the welfare measure from an explicit dynamic model of the household. The model shows exactly what modifications to the basic framework are required in each case, and demonstrates that some of these modifications are quantitatively significant.

Multiple Types of Capital and Labor

The extension to the case of multiple types of labor and capital is immediate. For simplicity, we could assume that each individual is endowed with the ability to provide different types of labor services, $L_{h,t}$ and that the utility function can be written as:

$$U(C_{1,t+s},..,C_{Z,t+s},\overline{L},L_{1,t+s},...,L_{H_L,t+s}) = \frac{1}{1-\sigma}C(C_{1,t+s},..,C_{Z,t+s})^{1-\sigma}\nu\left[\overline{L} - L(L_{1,t+s},...,L_{H_L,t+s})\right] \tag{3.22}$$

where $L(.)$ is an homogenous function of degree 1, H_L is the number of types of labor and $P_t^{L_h}$denotes the payment to a unit of $L_{h,t}$.[8] Similarly consumers can accumulate different types of capitals $K_{h,t}$ and rent them out at $P_t^{K_h}$. Take capital good 1 as the numeraire. Equation (3.4) now becomes:

$$\begin{aligned} v_t - v &= E_t \sum_{s=0}^{\infty} \beta^s \lambda \left[\sum_{i=1}^{Z} p_i^C c_i \widehat{c}_{i,t+s} + \sum_{h=1}^{H_K} p_h^I i_h \widehat{i}_{h,t+s} - \sum_{h=1}^{H_K} p_h^L L_{h,t+s} \; \widehat{L}_{h,t+s} - \sum_{h=1}^{H_K} \frac{p_h^K k_h}{(1+g)(1+n)} \; \widehat{k}_{h,t+s-1} \right] \\ &+ \sum_{h=1}^{H_K} \lambda \frac{(1-\delta_h) + p_h^K}{(1+g)(1+n)} k_h \widehat{k}_{h,t-1} \qquad (3.23) \end{aligned}$$

Redefine the normalized real GDP in deviation from SS as:

$$\widehat{y}_t = \sum_{i=1}^{Z} s_{c_i} \widehat{c}_{i,t+s} + \sum_{h=1}^{H_K} s_{i_h} \widehat{i}_{h,t+s} \qquad (3.24)$$

Using the two equations above, we get:

$$\begin{aligned} v_t &= v + E_t \sum_{s=0}^{\infty} \beta^s \lambda p^Y y \left[\widehat{y}_t - \sum_{h=1}^{H_K} \frac{p_h^L L_{h,t+s}}{p^Y y} \; \widehat{L}_{h,t+s} - \sum_{h=1}^{H_K} \frac{p_h^K k_h}{(1+g)(1+n) p^Y y} \; \widehat{k}_{h,t+s-1} \right] \\ &+ \sum_{h=1}^{H_K} \lambda \frac{(1-\delta_h) + p_h^K}{(1+g)(1+n)} k_h \widehat{k}_{h,t-1} \qquad (3.25) \end{aligned}$$

Proceeding exactly as in the previous section , the same equations will characterize the relationship between utility and the productivity residual, with the only difference that the

[8] We assume that the nature of the utility function is such that positive quantities of all types of labors are supplied.

latter is defined now as:

$$\log pr_t = \log Y_t - \sum_{h=1}^{H_K} s_{L_h} \log N_t L_{h,t} - \sum_{h}^{H_L} s_{K_h} \log K_{h,t-1} \tag{3.26}$$

Taxes

Our derivation of section 2.2 requires only reinterpretation to apply exactly to an environment with either distortionary and/or lump-sum taxes. The reason is that all prices in the budget constraint, equation (3.4), are from the point of view of the consumer. Thus, if there are taxes, the prices should all be interpreted as after-tax prices. Therefore our derivation implicitly allows for proportional taxes on capital and labor income as well as sales or value-added taxes levied on consumption and/or investment goods. The variable that we have been calling "profits," π, is really any transfer of income that the consumer takes as exogenous. Thus, it can be interpreted to include lump-sum taxes or rebates.

However, for the sake of exposition, we shall interpret all prices in equation (3.4) as being from the point of view of a firm, and thus before all taxes. To modify (3.4) to allow for taxes, we define some notation. Let τ^K be the tax rate on capital income, τ^L be the tax rate on labor income, τ_i^C be the *ad valorem* tax on consumption goods of type i, and τ^I be the corresponding tax on investment goods[9]. We assume that the revenue so raised is distributed back to individuals using lump-sum transfers. (We consider government expenditures in the next sub-section.) Then it is apparent that we arrive at the following

[9] For simplicity, we are assuming no capital gains taxes and no expensing for depreciation. These could obviously be added at the cost of extra notation.

modified version of equation (3.4):

$$v_t - v = \lambda E_t \sum_{s=0}^{\infty} \beta^s [\sum_{i=1}^{Z} \left(1+\tau_i^C\right) p_i^C c_i \widehat{c}_{i,t+s} + \left(1+\tau^I\right) i\widehat{i}_{t+s} - \left(1-\tau^L\right) p^L L \widehat{L}_{t+s}$$
$$- \left(1-\tau^K\right) \frac{p^K k}{(1+g)(1+n)} \widehat{k}_{t+s-1}] + \lambda \frac{(1-\delta) + p^K \left(1-\tau^K\right)}{(1+g)(1+n)} k\widehat{k}_{t-1} \quad (3.27)$$

To make contact with the data, note that the national accounts define nominal expenditure using prices as perceived from the demand side. Thus, equation (3.11) can be written exactly as before and still be consistent with standard national accounts data:

$$\widehat{y}_{t+s} = \sum_{i=1}^{Z} s_{c_i} \widehat{c}_{it} + s_i \widehat{i}_t \quad (3.28)$$

where s_{c_i} and s_i are inclusive of indirect taxes (subsidies) on consumption and investment. On the other hand, the national accounts define factor prices as perceived by firms, before income taxes. Thus, the data-consistent definition of the welfare residual with taxes needs to be based on a new definition of $\log pr_t$. Rewrite equation (3.14) as:

$$\log pr_{t+s} = \log Y_{t+s} - \frac{\left(1-\tau^L\right) p^L L N}{p^Y y} \log N_{t+s} L_{t+s} - \frac{\left(1-\tau^K\right) p^K k}{p^Y y (1+g)(1+n)} \log K_{t+s-1} \quad (3.29)$$
$$= \log Y_{t+s} - \left(1-\tau^L\right) s_L \log N_{t+s} L_{t+s} - \left(1-\tau^K\right) s_K \log K_{t+s-1}$$

This new definition of $\log pr_t$ then needs to be applied to equations such as (12) and (13) in section 2.2.

While it is easy to incorporate taxes into the analysis—as noted above, they are present implicitly in the basic expressions derived in section 2.2—the quantitative impact of modeling taxes explicitly can be large. Suppose that output is produced using an aggregate, constant-returns-to-scale production function of capital, labor and technology, as in Solow's classic (1957) paper. Then, without distortionary taxes, only changes in technology change welfare.

Now suppose the average marginal tax rate on both capital and labor income is 30 percent, and the share of consumption in output is 0.60. Suppose the government manages to raise aggregate capital and labor inputs by 1 percent permanently without a change in technology (perhaps via a small cut in tax rates). Then the flow increase in utility is equivalent to an increase in steady-state consumption of 0.5 percent. If the discount factor is 0.95 on an annual basis, the present value of this policy change is equivalent to a one-year increase in consumption of 10 percent of the steady-state level!

Government expenditure

With some minor modification, our framework can be extended to allow for the provision of public goods and services. We illustrate this under the assumption that government activity is financed with lump-sum taxes. Using the results from the previous subsection, it is straightforward to extend the argument to the case of distortionary taxes.

Assume that government spending takes the form of public consumption valued by consumers. We rewrite the instantaneous utility function as

$$U(C_{1,t+s},..,C_{Z,t+s},\overline{L},L_{1,t+s},...,L_{H_L,t+s}) = \frac{1}{1-\sigma}C(C_{1,t+s},..,C_{Z,t+s};C_{G,t+s})^{1-\sigma}\nu(\overline{L}-L_{t+s}) \tag{3.30}$$

where C_G denotes per-capita public consumption, and we continue to assume that $C(.)$ is homogenous of degree one in its arguments. Equation (3.4) now becomes:

$$v_t - v = \lambda E_t \sum_{s=0}^{\infty} \beta^s \left[\frac{U_{c_G} c_G \widehat{c}_{G,t+s}}{\lambda} + \sum_{i=1}^{Z} p_i^C c_i \widehat{c}_{i,t+s} + i\widehat{i}_{t+s} - p^L L \widehat{L}_{t+s} - \frac{p^K k}{(1+g)(1+n)} \widehat{k}_{t+s-1} \right] + \lambda \frac{(1-\delta) + p^K}{(1+g)(1+n)} k \widehat{k}_{t-1} \quad (3.31)$$

where $c_{G,t} = \frac{C_{G,t}}{X_t}$. The definition of GDP in deviation from steady state is now:

$$\widehat{y}_t = \sum_{i=1}^{Z} s_{c_i} \widehat{c}_{it} + s_i \widehat{i}_t + s_{c_G} \widehat{c}_{G,t}$$

where $s_{c_G} = \frac{P^G C_G}{P^Y Y}$ and P^G is the public consumption deflator. Let $s^*_{c_G} = \frac{U_{c_G} c_G \widehat{c}_{G,t+s}}{\lambda}$. Then we can write:

$$v_t - v = \lambda p^Y y E_t \sum_{s=0}^{\infty} \beta^s \left[\widehat{y}_{t+s} - s_L \widehat{L}_{t+s} - s_K \widehat{k}_{t+s-1} + \left(s^*_{c_G} - s_{c_G} \right) \widehat{c}_{G,t+s} \right] + \lambda \frac{(1-\delta) + p^K}{(1+g)(1+n)} k \widehat{k}_{t-1} \quad (3.32)$$

Hence in the presence of public consumption the Solow residual needs to be adjusted up or down depending on whether public consumption is under- or over-provided (i.e., $s^*_{c_G} > s_{c_G}$ or $s^*_{c_G} < s_{c_G}$ respectively). If the government sets public consumption exactly at the utility-maximizing level, $s^*_{c_G} = s_{c_G}$ and no correction is necessary. In turn, in the standard neoclassical case in which public consumption is pure waste $s^*_{c_G} = 0$, the welfare residual is computed on the basis of private final demand – i.e., GDP minus government purchases.

What if government purchases also yield productive services to private agents? This could be the case if, for example, the government provides education or health services, or public infrastructure, which may be directly valued by consumers and may also raise private-sector productivity. In such case, the above expression remains valid, but it is important to note that the net contribution of public expenditure to welfare would not be fully by captured by $\left(s_{c_G}^* - s_{c_G}\right)\widehat{c}_{G,t+s}$. To this term we would need to add a measure of the productivity of public services, which in the expression is implicitly included in the productivity residual $\widehat{y}_{t+s} - s_L\widehat{L}_{t+s} - s_K\widehat{k}_{t+s}$.

3.2.4 Open Economy

We now discuss how our fundamental result changes when we allow the economy to be open. Start from the definition of a country's current account:

$$CA_t = B_t - B_{t-1} = i_t B_{t-1} + P_t^{EXP} EXP_t - P_t^{IMP} IMP_t \tag{3.33}$$

where B_t is now the value of the net foreign assets, EXP_t and IMP_t are total exports and total imports and P_t^X and P_t^M are their respective prices. In a normalized form (3.33) becomes:

$$b_t = \frac{(1+r_t)}{(1+g)(1+n)} b_{t-1} + p_t^{EXP} \exp_t - p_t^{IMP} imp_t \tag{3.34}$$

Loglinearizing we obtain:

$$b\widehat{b}_t = \frac{(1+r)\,b}{(1+g)\,(1+n)}\widehat{b}_{t-1}+\frac{br}{(1+g)\,(1+n)}\widehat{r}_t+p^{EXP}\exp\widehat{p}_t^{EXP}-p^{IMP}imp\widehat{p}_t^{IMP}+p^{EXP}\exp\widehat{\exp}_t-p^{IMP}imp\widehat{im} \tag{3.35}$$

where $\exp_t = \frac{EXP_t}{P_t^I X_t N_t}$; $imp_t = \frac{IMP_t}{P_t^I X_t N_t}$; $p_t^{EXP} = \frac{P_t^{EXP}}{P_t^I}$ and $p_t^{IMP} = \frac{P_t^{IMP}}{P_t^I}$. Since $B_t \neq 0$, equation (3.4) can now be rewritten as:

$$\begin{aligned} v_t &= v + E_t \sum_{s=0}^{\infty} \beta^s \lambda \left[\sum_{i=1}^{Z} p_i^C c_i \widehat{c}_{i,t+s} + i\widehat{i}_{t+s} + b\widehat{b}_{t+s} - \frac{(1+r)\,b\widehat{b}_{t+s-1}}{(1+g)\,(1+n)} - p^L L\ \widehat{L}_{t+s} - \frac{p^K k\widehat{k}_{t+s-1}}{(1+g)\,(1+n)} \right] \\ &\quad +\lambda \frac{(1-\delta)+p^K}{(1+g)\,(1+n)} k\widehat{k}_{t-1} + \lambda \frac{(1+r)}{(1+g)\,(1+n)} b\widehat{b}_{t-1} \end{aligned} \tag{3.36}$$

Redefine the normalized real GDP in deviation from the SS as:

$$\widehat{y}_t = \sum_{i=1}^{Z} s_{c_i}\widehat{c}_{it} + s_i\widehat{i}_t + s_{\exp}\widehat{\exp}_t - s_{imp}\widehat{imp}_t \tag{3.37}$$

where s_x and s_m are respectively the share of exports and imports out of total value added.

Using the equations (3.35) and (3.37) into (3.36), we get:

$$\begin{aligned} v_t &= v + E_t \sum_{s=0}^{\infty} \beta^s \lambda p^y y \left[\widehat{y}_{t+s} - s_L \widehat{L}_{t+s} - s_K\ \widehat{k}_{t+s-1} + \left(\frac{br/p^Y y}{(1+g)\,(1+n)}\widehat{r}_t + s_x\widehat{p}_t^X - s_m\widehat{p}_t^M\right) \right] \\ &\quad +\lambda \frac{(1-\delta)+p^K}{(1+g)\,(1+n)} k\widehat{k}_{t-1} + \lambda \frac{(1+r)}{(1+g)\,(1+n)} b\widehat{b}_{t-1} \end{aligned} \tag{3.38}$$

Hence, in an open economy the standard Solow residual needs to be adjusted for the returns on net foreign assets, $\frac{br/p^Y y}{(1+g)(1+n)}\widehat{r}_t$, and for terms capturing the terms of trade, $s_x\widehat{p}_t^X - s_m\widehat{p}_t^M$. An improvement in the terms of trade has effects analogous to an increase in TFP - both give the consumer higher consumption for the same input of capital and labor (and therefore higher welfare). See Kohli (2004) for a static version of this result.

The terms in $\widehat{r}_{t+s}$ also capture present and expected future capital gains and losses on net foreign assets due either to exchange rate movements or to changes in the foreign currency prices of the assets. Finally, the initial conditions include not only the (domestic) capital stock, but also the net stock of foreign assets. An higher initial value of either asset expands the consumer's budget set and allows him to attain higher welfare.

3.3 Decomposing the Productivity Residual: Firm and Sector Level Contributions

The fundamental result from the previous section is that the growth in welfare is related to the expected present discounted value of the aggregate (modified) Solow productivity residual. In this section we will argue that this aggregate effect can be decomposed into the contribution of individual firms (or subset of firms). In order to do this we will look at aggregate value added, not from the expenditure side as we have done so far, but from the product side. More specifically, define aggregate value added as the following Tornqvist/Divisia index of firm-level value added:

$$\Delta \log Y_t = \sum_i w_i \Delta \log Y_{i,t} \tag{3.39}$$

The corresponding index for producer prices is:

$$\Delta \log {P^Y}_t = \sum_i w_i \Delta \log P^Y_{i,t} \tag{3.40}$$

Moreover, one can easily show that the following is true as an approximation:

$$s_K \Delta \log K_{t-1} = \sum_i w_i s_{K,i} \Delta \log K_{i,t-1}$$

and

$$s_L \Delta \log N_t L_t = \sum_i w_i s_{L,i} \Delta \log N_t L_{i,t}$$

As a result the aggregate Solow residual can be written as the weighted sum (with value-added weights) of the firm-level Solow residuals. More specifically:

$$\Delta \log pr_t = \sum_i w_i \Delta \log pr_{it}$$

where $\Delta \log pr_{it}$ is defined as:

$$\Delta \log pr_{it} = \Delta \log Y_{i,t} - s_{K,i} \Delta \log K_{i,t-1} - s_{L,i} \Delta \log N_t L_{i,t} \tag{3.41}$$

We can use this result to examine the sectoral sources of productivity growth, which is key to welfare change, within a country. We can ask a variant of the same question for firms, as we explain in the results sub-section. Finally, we can compare cross-sectional summary statistics. For example, we can ask whether small or large firms contribute more to national welfare improvement.

3.4 Data and Measurement

Our main source of information is Amadeus, a comprehensive firm-level pan-European database developed by Bureau Van Dijk. For every firm it provides data on the industry where it operates (at the 4-digit NACE level), its location, the year of incorporation, the ownership structure and the number of employees, in addition to the complete balance sheets and the profit and loss accounts. The data set includes both publicly traded and non traded companies. We limit our analysis to a subset of countries: Belgium, France, Great Britain, Italy, and Spain. We focus on manufacturing companies with operating revenues greater than or equal to 2 million Euros and continuous observations within the period of analysis. (We restrict ourselves to the balanced panel because Amadeus does not supply census data; there is no way to distinguish between entry into the sample and actual entry into the economy.)

We also use industry-level yearly data from the EU-KLEMS project, which provides output, input and price data for industries at roughly the 2-digit level of aggregation across a large number of countries up to 2005. These countries are mostly, but not exclusively, European; the project also gives data for non-EU countries like Australia, Japan, Korea and the United States. The EU-KLEMS data are extensively documented by O'Mahony and Timmer (2009).

In addition to the non-parametric welfare-relevant index numbers presented in the next section, we will also estimate production functions using firm level data, allowing the coefficients to vary across 2-digit industries for the period 1998-2005.[10] Before 1998 the number

[10] The use of a finer sectoral disaggregation is questionable if one wants to have enough firms in each sector for estimation purposes.

of firms in the survey is significantly smaller in most countries. Between 1998 and 2000 many firms enter in the data set. The coverage provided by the dataset varies across these countries. In 2005 the aggregated sales of the firms represented in Amadeus represent between 20 percent and 45 percent of the manufacturing sector's total production value, as documented in EU-KLEMS.

Our gross output proxy is (firm level) revenues deflated by the sectoral value added deflator obtained from the EU-KLEMS data set, at the 2 digit level. All deflators used here will be at the 2 digit level and are obtained from EU-KLEMS. We are aware that using industry deflators in place of firm-level prices can cause problems (Klette and Grilliches (1996)), but firm-level price data for output are not available in Amadeus. Our proxy for labor input is manpower costs deflated by the labor services deflator. (For some countries, such as Italy, the number of employees figure is not reliable, since there is not a reporting requirement for the number of employees in the main section of the balance sheet.) Capital is the historical value of tangible fixed assets divided by the price index for investment. We have also experimented with the perpetual inventory method, obtaining similar results. A measure for materials, intermediates and other services used in production has been computed using the following formula: materials = Operating Revenues - (Operating Profits+Manpower costs+Depreciation). The figure obtained in this way is then deflated by the materials and services deflator. Given gross output and materials input, value added is constructed as a Tornqvist/Divisia index.

3.5 Sources of Welfare Differences

Welfare change depends on the expected present discounted value of TFP growth as shown by equations (3.18) and (3.20). It is therefore important to investigate the time-series property of TFP growth. We do so in Table 3.1, using annual data from EU-KLEMS up to 2005 for the entire private economy for Belgium, France, Great Britain, Italy, and Spain. We use the measure of TFP developed in EU-KLEMS, based on the assumption of zero profits and time varying distributional shares and present both country by country and pooled results. The persistence of TFP growth is a key statistic, since it shows how the entire summation of expected productivity residuals changes as a function of the new information about the TFP growth rate. For most countries the log level of the TFP index is well described by an AR(1) stationary process around a country-specific linear trend. Additional lags of log TFP are not significant and the residual is white noise, as suggested by the Lagrange Multiplier test for residual serial correlation. The only exception is Spain, where the coefficient of log TFP (t-3) is significantly different from zero at the 5% level and the LM test rejects the hypothesis of no serial correlation (up to the third order). Thus, for most countries the growth rate of TFP is well described by an ARMA(1,1) model. We henceforth focus on the current TFP growth rate, since for most countries the data do not reject the proposition that the current growth rate (or its innovation) gives all necessary information about the entire future path of TFP, and hence welfare.

We first ask which sectors contributed the most to welfare change in these countries over the period of our study through their contribution to aggregate TFP growth. The results are in Table 3.2. We look at the contributions of five major industry groups:

Manufacturing, Utilities, Construction, Wholesale and Retail Trade and FIRE. For each country, we present in line 1 the mean of the Tornqvist index of TFP growth for these industries, which represent the overwhelming majority of private output. Interestingly, average TFP growth over this period is less than 1 percent per annum, even for the leading economies, France and Britain. The sectoral decompositions are also interesting. The next five lines give average sectoral TFP growth rates (not growth contributions, which would multiply the growth rates by the respective sectoral weights, and give a mechanical advantage to large sectors). Manufacturing makes a positive contribution for each country. The contribution of FIRE (Finance, Insurance, Real Estate), on the other hand, is often negative, especially in Britain, which has become a financial hub for the world.[11] But the humble utility sectors are the largest source of productivity growth on average (in every country other than Italy). Alesina et al (2005) suggest an explanation for this pattern based on deregulation of the utilities sectors in many European countries (with Italy a laggard in terms of the timing and pace of deregulation).

In Table 3.3, we look at the contributions of different groups of firms to welfare growth, now using our firm-level data from Amadeus for these countries. We now look at the average TFP growth rates of small and large firms, from 1998-2004. No very clear pattern emerges. Large firms have higher TFP growth rates in two countries (Belgium and Spain); small firms have higher growth rates in two others (Italy and Great Britain), and the contributions are basically identical in the remaining country, France.

[11] However, measures of both nominal and real financial sector output are often unreliable. See Wang, Basu and Fernald (forthcoming) for a model-based method for constructing financial sector output. Basu, Inklaar and Wang (forthcoming) apply this theory to construct nominal bank output measures, and Inklaar and Wang (2007) provide a theory-consistent measure of real bank output in the United States.

We can further decompose productivity differences across countries by applying the following decomposition, based on Griliches and Regev (1995). We wish to ask whether the difference in the productivity growth rate of any pair of countries is due to differences in their sectoral compositions or to differences in the growth rates for each sector. Let i now index sectors (not firms) and C be one of the countries in our sample other than the UK.

$$\begin{aligned}\sum_i w_i^C \Delta \log pr_{it}^C - \sum_i w_i^{UK} \Delta \log pr_{it}^{UK} &= \sum_i \frac{(w_i^C + w_i^{UK})}{2} (\Delta \log pr_{it}^C - \Delta \log pr_{it}^{UK}) \\ &+ \sum_i \frac{(\Delta \log pr_{it}^C + \Delta \log pr_{it}^{UK})}{2} (w_i^C - w_i^{UK})\end{aligned}$$

In Table 3.4, we examine the results of the Griliches-Regev (1995) decomposition, investigating the sources of growth of each country's TFP relative to that of Britain, which is the TFP growth leader over our period. The first column describes the difference in productivity change between Great Britain and the other economies in our sample (and is of course negative in all cases). The second column gives the amount of the difference accounted for by cross country differences in TFP growth for each sector, while the third column gives the amount of the difference due to differences in industrial structure (the share of each industry in the aggregate for that country). In most cases, cross country differences in the growth rate of the same sector account for the great majority of the gap with the UK. The exception is France, which actually grows faster than Britain comparing the same sector in the two countries, but loses nearly two-tenths of a percentage point of TFP growth per year due to differences in industrial structure.

In Table 3.5, we do an exercise designed to show whether the productivity patterns in each country are related to cross-country differences in the shape of the distribution of productivity growth rates across firms. This is an exercise in the spirit of Hsieh and Klenow (2009). However, Hsieh and Klenow expended considerable effort (and had to make a number of strong assumptions) in order to isolate firm-level technology within each country-sector. Our results show that if the object is to investigate the reasons for differences in welfare change across countries, it is not necessary (and indeed not sufficient) to understand how technology differs across firms; we should concentrate on differences in the Solow residual instead. We do the following exercise. For our full sample of firms within each country, we calculate TFP, and then subtract from firm-level TFP growth the TFP growth for the aggregate of the firms in that country. We then divide the range of productivity growth rates into 10 bins, and ask what percentage of firm value-added is produced by firms in each standardized productivity decile. (We experimented with dividing the range of growth rates more finely, into 20 bins, with qualitatively similar results.) Finally, we ask how much faster or slower aggregate TFP would have grown if the standardized distribution for the country had been replaced by the standardized distribution for Great Britain.

The results are in Table 3.5. For ease of viewing the results, we also plot the distributions for each country and the distribution for Britain in Figure 1. We find that replacing the distributions in Belgium and Spain with the British distribution would actually have caused those two countries to grow slightly more slowly. However, the same exercise for France and Italy shows that those two countries would each have had half a percentage point higher TFP growth per year over the full six years. This is a significant difference, especially for

Italy where it approximately doubles the annual TFP growth for our aggregate of firms. Thus, there is some evidence that a portion of the TFP growth differences relative to Britain, which is the probably the least regulated and most "US-like" of the countries in our sample, is driven by differences in institutions that allow weak firms to linger or prevent strong firms from expanding. The evidence is particularly strong in the case of Italy, which has been a conspicuous laggard in its rate of productivity growth over the last decade.

3.6 Decomposing the Productivity Residual: The Role of Reallocation and Technology

The great benefit of an index-number approach, such as the one we take in the previous section, is that it provides interesting results without requiring formal econometrics. The cost is that we cannot then identify the components of productivity growth, such as technical change or scale economies. Having established that aggregate TFP is, under some assumptions, the natural measuring stick for aggregate welfare, we now proceed to decompose aggregate TFP into components. We choose to work in growth rates, since there are well-known difficulties in comparing TFP levels across industries and countries. As we noted, while TFP growth is itself meaningful in welfare terms without any additional assumptions, we need to make assumptions about firm technology and behavior in order to decompose it in a meaningful way. We use the decomposition of Basu and Fernald (2002), which is derived by assuming that firms minimize costs and are price-takers in factor markets, but may have market power for the goods they sell and might produce with increasing returns to scale. Some of the components in the decomposition we use can be clearly identified as being due to reallocation, since they depend on marginal products of identical inputs not being equalized across firms. Other components depend on aggregate distortions, such as the average degree of market power and various tax rates.

3.6.1 Summary of the Basu and Fernald decomposition

Following Basu and Fernald (2002), in this paragraph we decompose changes in aggregate productivity into changes in aggregate technologies and changes in three non-technological components reflecting imperfections and frictions in output and factor markets. Suppose that each firm i has the following production function:

$$Q_i = F^i(K_i, L_i, M_i, T_i^Q) \tag{3.42}$$

where Q_i is the gross output, K_i, L_i and M_i are inputs of capital, labor and materials, T_i^Q is a technology index and F^i is an homogenous function. Assume that firms are price takers in factor markets but have market power in the output markets. Call P_{Ji} the price for factor J faced by firm i and μ_i^Q the mark up that firm i imposed over marginal costs. For any input J, let F_J^i be the marginal product. Firm i's first order condition implies:

$$P_i F_J^i = \mu_i^Q P_{Ji} \tag{3.43}$$

Output growth, $d\log Q_i$, can be written as:

$$d\log Q_i = \mu_i^Q \left[s_{L,i}^Q d\log L_i + s_{K,i}^Q d\log K_i + s_{M,i}^Q d\log M_i\right] + \frac{F_{T^Q}^i T_i^Q}{F^i} d\log T_i^Q = \mu_i^Q d\log X_i^Q + \frac{F_{T^Q}^i T_i^Q}{F^i} d\log T_i^Q$$

where $s_{J,i}^Q$ is the revenue share of input J out of gross output, $d\log T_i^Q$ denotes technology growth and $d\log X_i^Q$ is revenue share weighted total input growth. Remember that our ultimate goal is decomposing the aggregate Solow residual. In the national account identity in closed economy, total expenditure equals the sum of firms' value added. Consider the standard Divisia index of firm level value added:

$$d\log Y_i = \frac{d\log Q_i - s_{M,i}^Q d\log M_i}{1-s_{M,i}^Q} = d\log Q_i - \frac{s_{M,i}^Q}{1-s_{M,i}^Q}\left(d\log M_i - d\log Q_i\right)$$

and define the change in aggregate primary inputs, $d\log X_i$, as the share-weighted sum of the growth rates of capital and labor:

$$d\log X_i = \frac{s_{K,i}^Q}{1-s_{M,i}^Q} d\log K_i + \frac{s_{L,i}^Q}{1-s_{M,i}^Q} d\log L_i = s_{K,i} d\log K_i + s_{L,i} d\log L_i$$

After some algebra, taking into account that the firms' value added productivity residual $d\log pr_i$ equals $d\log Y_i - d\log X_i$, we obtain:

$$d\log pr_i = (\mu_i - 1) d\log X_i + (\mu_i - 1)\frac{s_{M,i}^Q}{1-s_{M,i}^Q}\left(d\log M_i - d\log Q_i\right) + d\log T_i$$

where:

$\mu_i = \mu_i^Q \frac{1-s_{M,i}^Q}{1-\mu_i^Q s_{M,i}^Q}$

$d\log T_i = \frac{F_{T^Q}^i T_i^Q}{F^i} \frac{d\log T_i^Q}{1-\mu_i^Q s_{M,i}^Q}$

Let us move now to aggregate quantities. Define aggregate inputs as the simple sums of firm-level quantities: $K = \sum_{i=1}^{I} K_i$ and $L = \sum_{i=1}^{I} L_i$.

Now define aggregate output growth as a Divisa index of firm level value added:

$$d \log Y = \sum_{i=1}^{I} w_i d \log Y_i$$

where w_i is firm i's share of nominal value added: $w_i = P_i^Y Y_i / P^Y Y$ and define aggregate primary input growth as:

$$d \log X = \frac{s_K^Q}{1 - s_M^Q} d \log K + \frac{s_L^Q}{1 - s_M^Q} d \log L = s_K d \log K + s_L d \log L$$

where s_J is the share of input J out of total value added. After some algebraic manipulation, $d \log X$ can be written in terms of the weighted average of firm level primary input growth: $d \log X = \sum_{i=1}^{I} w_i d \log X_i$. Aggregate productivity growth, $d \log pr$, is the difference between aggregate output growth $d \log Y$ and aggregate inputs growth $d \log X$. Basu and Fernald shows that after some manipulations, $d \log pr$ can be decomposed in the following way:[12]

$$d \log pr = (\overline{\mu} - 1) d \log X + (\overline{\mu} - 1) d \log M/Q + R_\mu + R_M + d \log T \qquad (3.44)$$

where:

$\overline{\mu} = \sum_{i=1}^{I} w_i \mu_i$

$d \log M/Q = \sum_{i=1}^{I} w_i \frac{s_{M,i}^Q}{1 - s_{M,i}^Q} \left(d \log M_i - d \log Q_i \right)$

[12] We are assuming here that the price paid by each firm for capital and labor is the same. If it is allowed to differ, Basu and Fernald (2002) show that two additional terms should be added to the right hand side of (3.44): $R_K \equiv \overline{\mu} \sum_{i=1}^{I} w_i s_{K,i} \left[\frac{P_{Ki} - P_K}{P_{Ki}} \right] d \log K_i$ and $R_L \equiv \overline{\mu} \sum_{i=1}^{I} w_i s_{L,i} \left[\frac{P_{Li} - P_L}{P_{Li}} \right] d \log L_i$. These input reallocation terms represent gains from directing primary inputs towards firms where they have higher social valuation.

$R_\mu = \sum_{i=1}^{I} w_i \left(\mu_i - \overline{\mu}\right) d\log X_i$

$R_M = \sum_{i=1}^{I} w_i \left(\mu_i - \overline{\mu}\right) \frac{s_{M,i}^{Q}}{1-s_{M,i}^{Q}} \left(d\log M_i - d\log Q_i\right)$

$d\log T = \sum_{i=1}^{I} w_i d\log T_i$

It is easy to provide an intuition for the welfare relevance of each term in which we have decomposed aggregate productivity. The first term, $(\overline{\mu} - 1)d\log X$, is a direct consequence of imperfect competition. Consumers would prefer to provide more labor and capital and consume the extra goods produced, since their utility value exceeds the disutility of producing them. Hence aggregate productivity and welfare increases with aggregate primary input growth, and this is true even if firms have the same markup. In this sense, $(\overline{\mu} - 1)d\log X$ reflects an aggregate distortion and should not be counted as part of "reallocation," which we use as shorthand for allocative efficiency.

The third term, R_μ, represents the increase in productivity and welfare coming from the fact that primary inputs are directed towards firms with higher-than-average markups, since higher prices and markups express higher social valuation.

The terms $(\overline{\mu} - 1)d\log M/Q$ and R_M reflect the fact that a markup greater than one reduces the use of materials as well as primary inputs below the socially optimal level. This distortion is greater the greater is the markup. Note that if materials had to be used in fixed proportion to output, $d\log M_i - d\log Q_i$ would equal zero and so would both $(\overline{\mu}-1)d\log M/Q$ and R_M. (In other words, the distortions regarding primary inputs would summarize fully the distortions in input use due to markups that exceed one.) More specifically, $(\overline{\mu}-1)d\log M/Q$ reflects the distortion generated by an average markup above unity and R_M reflects reallocation across firms with different markups (relative to $\overline{\mu}$). Only the latter should be counted as part of reallocation. Finally the term $d\log T$ represents the contribution to productivity and welfare of changes in aggregate technology.

The Basu and Fernald decomposition can be extended by disaggregating R_μ into a within sectors and between sectors component. This is useful in assessing whether the gain from reallocation (if any) occur because resources are reallocated across industries or within industries across firms. Basu and Fernald used industry level data in their empirical exercise so they could at best evaluate the between component. (We say "at best" because if there are within-industry reallocation terms, then Basu and Fernald's estimation using industry-level data would not give a consistent estimate of even the average industry markup, $\overline{\mu}$. In general, one can estimate $\overline{\mu}$ correctly only by taking the average of firm-level markups, estimated using firm-level data.) If one uses firm-level data, one can discuss the relative importance of the within and between components. R_M can also be decomposed into a within and between component, but there is a residual term.

Let $P^{YJ}Y^J = \sum_{i\epsilon J} P_i^Y Y_i^J$ be the total value added produced in industry J, $w^J = P^{YJ}Y^J/P^Y Y$ the share of industry J out of aggregate output and $w_i^J = P_i^Y Y_i^J/P^{YJ}Y^J$ the share of value added of firm i in industry J. Denote with Q_i a firm gross output and with P_i^Q its price. Then $w_i^{QJ} = P_i^Q Q_i^J/P^{QJ}Q^J$, where $P^{QJ}Q^J = \sum_{i\epsilon J} P_{ii}^Q Q^J$, represents the firm share of industry gross output. Finally, the primary inputs growth in industry J is $d\log X^J = s_K^J d\log K^J + s_L^J d\log L^J$, where $s_K^J = \frac{\sum_{i\epsilon J} P_{K,i}K_i}{P^{YJ}Y^J}$ and $s_L^J = \frac{\sum_{i\epsilon J} P_{L,i}L_i}{P^{YJ}Y^J}$. Define R_μ^J and R_μ^Jas the industry equivalent of the reallocation terms R_μ and R_M when aggregating over industry J rather than the entire economy, i.e. $R_\mu^J = \sum_{i\epsilon J} w_i^J \left(\mu_i - \overline{\mu}^J\right) d\log X_i$ and $R_M^J = \sum_{i\epsilon J} w_i^J \left(\mu_i - \overline{\mu}^J\right) \frac{s_{M,i}^Q}{1-s_{M,i}^Q} \left(d\log M_i - d\log Q_i\right)$, where $\overline{\mu}^J = \sum_{i\epsilon J} w_i^J \mu_i$. We can decompose the reallocation term for primary inputs, R_μ, into a within and a between component (denoted by superscripts W and B, respectively) as follows:

$$R_\mu = R_\mu^W + R_\mu^B$$

where $R_\mu^W = \sum_{J=1}^{K} w^J R_\mu^J$ and $R_\mu^B = \sum_{J=1}^{K} w^J \left(\overline{\mu}^J - \overline{\mu}\right) d\log X^J$.Note that the between component can be calculated on the basis of industry data only.

The decomposition for the reallocation term for materials, R_M, is instead:

$$R_M = R_M^W + R_M^B + R_{w-w^Q}$$

where $R_M^W = \sum_{J=1}^{K} w^J R_M^J$, $R_M^B = \sum_{J=1}^{K} w^J \left(\overline{\mu}^J - \overline{\mu}\right)\left(d\log Q^J - d\log Y^J\right)$ and $R_{w-w^Q} = \sum_{J=1}^{K} w^J \left(\overline{\mu}^J - \overline{\mu}\right)\left(w_i^J - w_i^{QJ}\right) d\log Q_i$. In the between component, $d\log Y^J = \sum_{i\epsilon J} w_i^J d\log Y_i$ is the Divisia index of industry value added. $d\log Q^J = \sum_{i\epsilon J} w_i^{QJ} d\log Q_i$ is the Divisia index of gross output (using w_i^{QJ} as weights). The residual term, R_{w-w^Q}, reflects the difference between value added weights and gross output weights in aggregating firm level gross output within an industry.

3.7 Econometric Framework

The modified Solow productivity residual can be essentially calculated from the data and requires no estimation if the distributional shares are observable (or if we observe the labor share and assume approximately zero profits). However, in order to break down the productivity residual into components that reflect aggregate distortions, reallocation and technology growth we must obtain estimates of the markups and of technology growth. We will do that by assuming that the (gross) production function in sector j is Cobb-Douglas:

$$\log Q_{it} = \varepsilon_L^j \log L_{it} + \varepsilon_K^j \log K_{it} + \varepsilon_M^j \log M_{it} + \eta_{jt} + \alpha_i + \omega_{it} \tag{3.45}$$

where i denotes firms ($i = 1, .., I_j$), t time ($t = 1, .., T_j$), and small case variables logs. η_{jt} is an industry specific common component of productivity, α_i a time-invariant firm-level component and ω_{it} an idiosyncratic component. In our application using the Amadeus data set, T_j is small and N_j large.

We will experiment with different estimation methods: OLS, LSDV, Olley and Pakes, Difference and System GMM (assuming that ω_{it} is either serially uncorrelated, or that it follows an AR(1) process). The advantages and disadvantages of each choice are well known, although there is no agreement on which estimator one should ultimately choose. One fundamental estimation problem is the endogeneity of the input variables, which are likely to be correlated both with α_i and ω_{it}. Correlation with ω_{it} may reflect both simultaneity of input choices or measurement errors. Given the shortness of the panel, elimination of α_i through a within transformation is not the appropriate strategy. Differencing of (3.45) and application of the difference GMM estimator (Arellano and Bond (1991)) is a possibility, but appropriately lagged values of the regressors may be poor instruments if inputs are very persistent. Application of the GMM System estimator (Blundell and Bond (1998) and Blundell and Bond *(*2000*)* is probably a better option. An alternative approach is the one proposed by Olley and Pakes (1996). This estimator addresses the simultaneity (and selection) problem by using firm investment as a proxy for unobserved productivity and requires the presence of only one unobserved state variable at the firm level and monotonicity of the investment function. We are not interested to take a stand in this paper on which one is the preferable estimation strategy. Fortunately for us, the results of the decomposition are insensitive to the choice of a particular estimator.

Having obtained estimates of the output elasticity for each factor we will recover the firm specific markup from the first order conditions for materials, equation (3.43). In the Cobb Douglas case, this can be expressed as:

$$\widehat{\mu}_i^Q = \frac{\widehat{\varepsilon}_M^j}{s_{M,i}^Q} \tag{3.46}$$

where $s_{M,i}^Q$ is the time average of the firm specific revenue share of materials for firm i. A hat denotes estimated values. We have chosen to focus on the FOC for materials because they are likely to be the most a flexible input. Whereas the labor share, $s_{L,i}^Q$, can be easily recovered from the data, the same is not true for the capital share, $s_{K,i}^Q$, unless one is willing to make assumptions about the user cost of capital, which is problematic in the presence of firm heterogeneity in the cost of finance. We have recovered the capital share from estimates of the markup described above and of the elasticity of output with respect to capital, using:

$$s_{K,i}^Q = \frac{\widehat{\varepsilon}_K^j}{\widehat{\mu}_i^Q} \tag{3.47}$$

Alternatively we have obtained $s_{K,i}^Q$ from:

$$s_{ki} = 1 - s_{L,i}^Q - s_{M,i}^Q - \frac{\Pi_i}{Y_i} = 1 - s_{L,i}^Q - s_{M,i}^Q - (1 - \frac{\widehat{\theta}^j}{\widehat{\mu}_i^Q}) \tag{3.48}$$

where $\widehat{\theta}^j = \widehat{\varepsilon}_K^j + \widehat{\varepsilon}_L^j + \widehat{\varepsilon}_M^j$ is the degree of returns to scale in sector j. The result are robust to this choice.

3.8 Results

We will discuss now the empirical results obtained when the production function is estimated on the firm level data contained in Amadeus for Belgium, France, Great Britain, Italy, Spain over the period 1998-2005. To avoid overburdening the reader, we report results for selected estimators (OLS, System GMM, and Olley and Pakes) for only one of our countries, Belgium.

The estimation results for the elasticity of output with respect to each factor, for constant returns to scale and for average markups are reported in the tables 3.6, 3.7, and 3.9. Estimates are pretty standard and vary somewhat across estimators. Recall that materials include services together with materials and intermediates. The degree of returns to scale is very close to one in most sectors using OLS and System GMM, while it is slightly smaller, but still close to one, with the Olley and Pakes estimator. The estimate of ε_K^j is greater for the OLS estimator and the smallest for the Olley and Pakes estimator. For five sectors it is negative using the GMM System estimator with serially uncorrelated errors, although not significantly so. The test of overidentifying restrictions and the test of second order serial correlation for the GMM System do not suggest major misspecification issues for most sectors, which leads us to focus on this version of the GMM estimator, instead of the one allowing for an AR(1) error component in the level equation. The average estimated markup, obtained using (3.46), exceeds one in all sectors, whatever the estimator used. Moreover it is strictly greater than one for 64% of firms, using OLS, 70% using System GMM, and 63%

using Olley and Pakes.

We find markup estimates that are quite reasonable compared to existing estimates in the micro-econometric literature [13], albeit somewhat high relative to the macro literature. The numerical estimates in Tables 3.6 through 3.9, usually in the range of 1.10 to 1.25, seem quite small, but one needs to remember that these are markups on gross output. Converting to markups on value added using a representative materials share of 0.7, the markups are in the range of 1.43 to 3. Similarly, the implied profit rates are a bit on the high side. Using equation 3.48, the profit rate can be calculated as $(1 - \frac{\widehat{\theta}^j}{\widehat{\mu}_i^Q})$. Taking constant returns as our modal estimate, the markup range just discussed corresponds to profit rates in the range of 9 to 20 percent, expressed as a percentage of gross output.

Our estimates of the markup and thus of the profit rate are probably upper bounds. We do not control for variations in firm-level input utilization (changes in the number of shifts or variations in labor effort), except through our use of time fixed effects. Thus, we remove variations in utilization due to common industry effects but not due to firm-specific demand variation over time. Basu (1996) suggests that variable utilization is likely to bias upward the output elasticity of materials in particular, which is the parameter that has the largest impact on our estimates of markups and profit rates. Unfortunately, we do not have the firm-level data on hours worked per employee that would be necessary to implement the utilization control derived from the optimizing model of Basu, Fernald and Kimball (2006). Thus, our estimates of the average distortions coming from markup pricing, as summarized by the first two terms in equation (3.44) are likely to be on the high side. But the fact

[13] For example, Dobbelaure and Mairesse (2008) find very similar markups using panel data for French firms.

that our estimated average markups are large does not create any particular bias in our estimates of the reallocation terms, which are our particular focus, since the reallocation terms involving markups depend on the gaps between firm-level and average markups.

In light of this discussion, it is interesting to look at the estimates of the various reallocation terms for our sample of six countries, which are presented in Table 3.10 and in 3.11. In Table 3.10 we report for each country in our sample, average productivity growth, $d\log pr$, the sum of aggregate distortions, $(\overline{\mu}-1)d\log X + (\overline{\mu}-1)d\log M/Q$, the sum of the reallocation terms for primary factors and materials, $R_\mu + R_M$, and technology growth, $d\log T$. The last column reports as residual the difference between productivity, on the one hand, and the sum of aggregate distortions, of the reallocation terms, and of technological progress, on the other, i.e. the difference between the left hand side and the right hand side of (3.44). This equation may not hold as an equality for three reasons: first we do not observe the true value of the markup, but only its estimated value; (ii) whereas the labor share is observed in the data, calculations of the capital share depends upon a zero profit assumptions or an estimate of the markup and of the degree of returns to scale; (iii) as Basu and Fernald (2002) show, if the price paid for capital and labor differs across firms, additional terms involving the difference of factor prices for each firm from the average, multiplied by each factor growth rate will appear on the right hand side of (3.44).[14]

First of all, we see from Table 3.10 robust average annual productivity growth for all countries in our sample of large firms. The case of Italy is particularly striking, since our

[14] See footnote 11. Petrin, Reiter and White (2009) argue that changes in fixed costs create yet another gap between the two sides of equation *(3.44)*. However, changes in fixed costs are equivalent to an additive technology shock, and to a first-order approximation both additive and multiplicative technology shocks are already incorporated into the estimate of $dlogT$. Thus, changes in fixed costs are not an additional gap between productivity growth and technological change.

sample of firms has an average productivity growth rate, $d \log pr$, of 2.8 percent, while the EU-KLEMS database shows that for all of Italian manufacturing average TFP declined at a rate of 1.2 percent per year over our sample period. Second, we see that technical change was also positive for all countries, and over 1 percent per year in all countries except Spain, where it averaged 0.5 percent. The strongest rates of technical change, in excess of 4 percent per year, were registered in France, which is usually found to be a high-productivity country in most cross-country studies, and in the United Kingdom, which had 2.2 percent average TFP growth in manufacturing over this time period.

Before discussing the results on reallocation, note that the residual is sizeable and we decide to allocate it to the aggregate distortion, reallocation, and technology growth component in proportion to their relative size. In Table 3.11 we report the proportion of aggregate productivity accounted for by each component, after this adjustment. The results suggest, first, that in most countries most of productivity growth is accounted by technology growth. More specifically, technological progress accounts for the totality of productivity growth in Great Britain and in France, for a large fraction in Italy (.66%) and for a sizeable, but smaller fraction in Belgium and Spain (43% and 21% respectively). Second, aggregate distortions are quite important in Spain Belgium, and Italy, where they account for 85%, 55%, and 33% of productivity growth respectively. They are, instead rather small in Great Britain and in France. The reallocation terms for primary factors or materials accounts for a small proportion of productivity growth in all countries.[15] It follows that, unless one is willing to treat the entire residual as part of reallocation term, factor reallocation does not

[15] Because the reallocation term is so small, not much is learned from presenting its the decomposition in a within and between component.

appear to be an important component of productivity growth.[16] Here the nature of the sample may work against finding strong results, since most of the firms are quite large in all the years they are observed. Reallocation effects are most clearly apparent when firms that are small initially grow to a large size due to their superior productivity. There are probably fewer such firms in our sample than in the population, thus reducing the quantitative impact of reallocation. Petrin, White and Reiter (2009) come to the different conclusion that reallocation represents a large fraction of productivity growth, using manufacturing plant level data for the US. They calculate their reallocation term as the difference between a Divisia index of firm level productivity growth and a Divisia index of technology growth. Thus, they include aggregate distortions as part of reallocation, which should not be the case if one wants to estimate and index of allocative efficiency strictly defined. We also find that aggregate distortions can be substantial for some countries.[17]

Finally, although reallocation of factors towards uses where they have a higher social valuation has not been a large part of the improvement in productivity and welfare for the sample period we have analyzed, does not mean that a benevolent central planner could not achieve large welfare improvement from factor reallocation. This distinction between the historical decomposition we have presented and what could be potentially obtained should be kept in mind when drawing inferences from these results.

[16] If we treat the residual as reflecting the difference in primary factor prices faced by firms and treat it entirely as part of the reallocation term, as in Basu and Fernald (2002), reallocation would account for approximately a third of productivity growth in Great Britain.

[17] Petrin, White and Reiter (2009) implement the decomposition proposed by Petrin and Levinsohn (2008) which is a variant of the Basu and Fernald (2002) decomposition.

3.9 Conclusions

We show that the present value of aggregate TFP growth, for a given initial endowment of capital, is a complete welfare measure for a representative consumer, up to a first-order approximation. This result rigorously justifies TFP, rather than technical change or labor productivity, as the central statistic of interest in any exploration of productivity, at all levels of aggregation. Importantly, the result holds even when TFP is not a correct measure of technical change, for example due to increasing returns, externalities, or imperfect competition. It also suggests that productivity decompositions should be oriented towards showing how particular features or frictions in an economy either promote or hinder aggregate TFP growth, which is the key to economic welfare. Our theoretical results point to a key role for the persistence of aggregate TFP growth, since welfare change is related to the entire expected time path of productivity growth in addition to the current growth rate. Moreover, our derivation shows that in order to create a proper welfare measure, TFP has to be calculated using prices faced by households rather than prices facing firms. In modern, developed economies with high rates of income and indirect taxation, the gap between household and firm TFP can be considerable. Finally, in an open economy, the change in welfare will also reflect present and future changes in the returns on net foreign assets and in the terms of trade.

We use these central results to show that one can explore the sources of welfare change using both non-parametric index numbers and formal econometrics. The non-parametric approach has the great advantage of simplicity, and avoids the need to address issues of econometric identification. Many interesting cross-country comparisons can be performed using the index-number approach, including calculating summary statistics of allocative efficiency for each country based on firm-level data. However, if one wants to ask how much of aggregate TFP growth is due to technical change, as opposed to scale economies or allocative inefficiency, one does need to make additional assumptions and estimate production functions at the firm level. We show how one can decompose aggregate TFP growth in such a manner using firm-level data.

The results suggest that in the majority of OECD countries we have analyzed (Belgium, France, Great Britain, Italy, and Spain) most of productivity growth in manufacturing is accounted for by technology growth. This is particularly true for Great Britain and France. Moreover, aggregate distortions are quite important in many countries, such as Spain Belgium, and Italy. Finally, the reallocation terms for primary factors or materials account for a small proportion of productivity growth in all countries over the period 1995-2005. We will explore in future research whether this results extends to other countries or time periods, or to other data sets less biased towards larger firms.

3.10 Appendix A: Derivations

3.10.1 Making the problem stationary

The representative household maximizes intertemporal utility:

$$V_t = \sum_{s=0}^{\infty} \frac{1}{(1+\rho)^s} \frac{N_{t+s}}{H} U(C_{1,t+s}, .., C_{Z,t+s}; \overline{L} - L_{t+s}) \tag{3.49}$$

where $C_{i,t+s}$ is the capita consumption of good i at time t+s, L_{t+s} are hours of work per capita, $\overline{L}$ is the time endowment, and N_{t+s} population. H is the number of households, assumed to be fixed and normalized to one from now on. Consider the laws of motion for N_t and for X_t, where the latter denotes Harrod neutral technological progress (so that total labor input in efficiency units is $(N_t X_t L_t)$:

$$N_t = N_0(1+n)^t \tag{3.50}$$

$$X_t = X_0(1+g)^t \tag{3.51}$$

and normalize $H = 1$.

We can rewrite the utility function as:

$$V_t = N_t \sum_{s=0}^{\infty} \frac{(1+n)^s}{(1+\rho)^s} U(C_{1,t+s}, .., C_{Z,t+s}; \overline{L} - L_{t+s}) \tag{3.52}$$

For a well defined steady state in which hours of work are constant we assume that the utility function has the King Plosser and Rebelo form(1988):

$$U(C_{1,t+s},..,C_{Z,t+s};L-\overline{L}_s)=\frac{1}{1-\sigma}C(C_{1,t+s},..,C_{Z,t+s})^{1-\sigma}\nu(\overline{L}-L_{t+s})$$

We assume that $C()$ is homogenous of degree 1. Define $c_{i,t+s}=\frac{C_{i,t+s}}{X_{t+s}}$. We can rewrite the utility function in the following form:

$$U(C_{1,t+s},..,C_{Z,t+s};L-\overline{L}_{t+s})=\frac{1}{1-\sigma}X_{t+s}^{(1-\sigma)}C(c_{1,t+s},..,c_{Z,t+s})^{1-\sigma}\nu(\overline{L}-L_{t+s})$$

or

$$U(C_{1,t+s},..,C_{Z,t+s};L-\overline{L}_s)=(1+g)^{s(1-\sigma)}X_t^{(1-\sigma)}\frac{1}{1-\sigma}C(c_{1,t+s},..,c_{Z,t+s})^{1-\sigma}\nu(\overline{L}-L_{t+s})$$

Inserting this into V_t, we get:

$$V_t=N_tX_t^{(1-\sigma)}\sum_{s=0}^{\infty}\beta^sU(c_{1,t+s},..,c_{Z,t+s};\overline{L}-L_{t+s}) \tag{3.53}$$

where: $\beta=\frac{(1+n)(1+g)^{1-\sigma}}{(1+\rho)}$.

3.10.2 Budget constraint

Start from the usual budget constraint:

$$P_t^IK_t+B_t=(1-\delta)\,P_t^IK_{t-1}+(1+i_t)\,B_{t-1}+P_t^LL_tN_t+P_t^KK_{t-1}+\Pi_t-\sum_{i=1}^{Z}P_{i,t}^CC_{i,t}N_t \tag{3.54}$$

Divide both sides by $P_t^I X_t N_t$ to get:

$$\frac{K_t}{X_t N_t} + \frac{B_t}{P_t^I X_t N_t} = (1-\delta)\frac{K_{t-1}}{X_{t-1}N_{t-1}}\frac{X_{t-1}N_{t-1}}{X_t N_t} + (1+i_t)\frac{B_{t-1}}{P_{t-1}^I X_{t-1}N_{t-1}}\frac{P_{t-1}^I}{P_t^I}\frac{X_{t-1}N_{t-1}}{X_t N_t}$$
$$+\frac{P_t^L}{P_t^I}\frac{L_t N_t}{X_t N_t} + \frac{P_t^K}{P_t^I}\frac{K_{t-1}}{X_{t-1}N_{t-1}}\frac{X_{t-1}N_{t-1}}{X_t N_t} + \frac{\Pi_t}{P_t^I X_t N_t} - \sum_{i=1}^{Z}\frac{P_{i,t}^C}{P_t^I}\frac{C_{i,t}N_t}{X_t N_t}$$

Define: $k_t = \frac{K_t}{X_t N_t}$, $b_t = \frac{B_t}{P_t^I X_t N_t}$, $p_t^K = \frac{P_t^K}{P_t^I}$, $p_t^L = \frac{P_t^L}{P_t^I X_t}$, $p_{i,t}^C = \frac{P_{i,t}^C}{P_t^I}$, $(1+r_t) = \frac{(1+i_t)}{(1+\pi_t)}$, $\pi_t = \frac{\Pi_t}{P_t^I X_t N_t}$. The budget constraint can be rewritten as:

$$k_t + b_t = \frac{(1-\delta)+p_t^K}{(1+g)(1+n)}k_{t-1} + \frac{(1+r_t)}{(1+g)(1+n)}b_{t-1} + p_t^L L_t + \pi_t - \sum_{i=1}^{Z} p_{i,t}^C c_{i,t} \qquad (3.55)$$

3.10.3 Optimality conditions

The representative household solves the following maximization:

$$Max\ v_t = Max\frac{V_t}{N_t X_t^{(1-\sigma)}} = E_t\sum_{s=0}^{\infty}\beta^s\{U(c_{1,t+s},..,c_{Z,t+s};\overline{L}-L_{t+s})$$

$$+\lambda_{t+s}\left(-k_{t+s}-b_{t+s}+\frac{(1-\delta)+p_{t+s}^K}{(1+g)(1+n)}k_{t+s-1}+\frac{(1+r_{t+s})}{(1+g)(1+n)}b_{t+s-1}+p_{t+s}^L L_{t+s}+\pi_{t+s}-\sum_{i=1}^{Z}p_{i,t+s}^C c_{i,t+s}\right)\}$$

where $v_t = \frac{V_t}{N_t X_t^{(1-\sigma)}}$ is normalized intertemporal utility. The FOCs are:

$$U_{c_{i,t}} - \lambda_t p_{i,t}^C = 0 \qquad (3.56)$$

$$U_{L_t} + \lambda_t p_t^L = 0 \tag{3.57}$$

$$-\lambda_t + \beta E_t \frac{(1-\delta) + p_{t+1}^K}{(1+g)(1+n)} \lambda_{t+1} = 0 \tag{3.58}$$

$$-\lambda_t + \beta_t \frac{1}{(1+g)(1+n)} E_t (1 + r_{t+1}) \lambda_{t+1} = 0 \tag{3.59}$$

3.10.4 Approximation around SS

Define with $\widehat{x} = \log x_t - \log x$ the log deviation from the steady state of a variable (x is the steady state value of x_t). Loglinearizing the normalized value function around the steady state, one obtains (to a first order approximation):

$$\begin{aligned}
v_t - v \quad = \quad & E_t[\sum_{s=0}^{\infty} \beta^s (\sum_{i=1}^{Z} U_{c_i} c_i \widehat{c}_{i,t+s} + U_{L_t} L \widehat{L}_{i,t+s} \\
& + \lambda p^L L \widehat{L}_{i,t+s} - \lambda \sum_{i=1}^{Z} p_i^C c_i \widehat{c}_{i,t+s} - \lambda k \widehat{k}_{t+s} - \lambda b \widehat{b}_{t+s}) \\
& + \sum_{s=0}^{\infty} \beta^{s+1} (\lambda \frac{(1-\delta) + p^K}{(1+g)(1+n)} k \widehat{k}_{t+s} + \lambda \frac{(1+r)}{(1+g)(1+n)} b \widehat{b}_{t+s}) \\
& + \sum_{s=0}^{\infty} \beta^s \widehat{\lambda}_{t+s} \, (-k_{t+s} - b_{t+s} + \frac{(1-\delta) + p_{t+s}^K}{(1+g)(1+n)} k_{t+s-1} + \frac{(1+r_{t+s})}{(1+g)(1+n)} b_{t+s-1} \\
& + p_{t+s}^L L_{t+s} + \pi_{t+s} - \sum_{i=1}^{Z} p_{i,t+s}^C c_{i,t+s})] \\
& + \sum_{s=0}^{\infty} \beta^s (\lambda p^L L \widehat{p}_{t+s}^L + \frac{p^K k}{(1+g)(1+n)} \, \widehat{p}_{t+s}^K - \sum_{i:1}^{Z} p_{i,t}^C c_i \widehat{p}_{i,t+s} + \pi \widehat{\pi}_{t+s} + \frac{rb}{(1+g)(1+n)} r_{t+s}) \\
& + \lambda \frac{(1-\delta) + p^K}{(1+g)(1+n)} k \widehat{k}_{t-1} + \lambda \frac{(1+r)}{(1+g)(1+n)} b \widehat{b}_{t-1}
\end{aligned}$$

Using the first order conditions, the first four lines equal zero. Moreover since $b_t = 0$, we get:

$$v_t = v + E_t \sum_{s=0}^{\infty} \beta^s \lambda \left[p^L L \widehat{p}^L_{t+s} + \frac{p^K k}{(1+g)(1+n)} \widehat{p}^K_{t+s} + \pi \widehat{\pi}_{t+s} - \sum_{i=1}^{Z} p^C_{i,t} c_i \widehat{p}_{i,t+s} \right] + \lambda \frac{(1-\delta)+p^K}{(1+g)(1+n)} k \widehat{k}_{t-1} \tag{3.60}$$

Now log linearize the budget constraint:

$$k\,\widehat{k}_t + b\widehat{b}_t - \frac{(1-\delta)+p^K}{(1+g)(1+n)} k\widehat{k}_{t-1} - \frac{(1+r)}{(1+g)(1+n)} b\widehat{b}_{t-1} - p^L L \widehat{L}_t - p^L L \widehat{p}^L_t - \frac{p^K k}{(1+g)(1+n)} \widehat{p}^K_t$$
$$-\pi\widehat{\pi}_t + \sum_{i=1}^{Z} p^C_i c_i \widehat{c}_{i,t} + \sum_{i=1}^{Z} p^C_i c_i \widehat{p}_{i,t} = 0$$

Using this result and the fact that $b = 0$ in (3.60) gives us:

$$v_t = v + E_t \sum_{s=0}^{\infty} \beta^s \lambda \left[\sum_{i=1}^{Z} p^C_i c_i \widehat{c}_{i,t+s} + k\widehat{k}_{t+s} - \frac{(1-\delta)+p^K}{(1+g)(1+n)} k\widehat{k}_{t+s-1} - p^L L \widehat{L}_{t+s} \right] + \lambda \frac{(1-\delta)+p^K}{(1+g)(1+n)} k\widehat{k}_{t-1} \tag{3.61}$$

Notice that the law of motion of capital: $K_t = (1-\delta)K_{t-1} + I_t$, can be rewritten as: $\frac{K_t}{X_t N_t} = (1-\delta)\frac{K_{t-1}}{X_{t-1}N_{t-1}}\frac{X_{t-1}N_{t-1}}{X_t N_t} + \frac{I_t}{X_t N_t}$ which after some algebra becomes:

$$k_t = \frac{(1-\delta)}{(1+g)(1+n)} k_{t-1} + i_t$$

Differentiating it around the steady state, we get:

$$k\widehat{k}_t = \frac{(1-\delta)}{(1+g)(1+n)} k\widehat{k}_{t-1} + i\widehat{i}_t$$

Inserting this equation into equation 3.61 we get:

$$v_t = v + E_t \sum_{s=0}^{\infty} \beta^s \lambda \left[\sum_{i=1}^{Z} p_i^C c_i \widehat{c}_{i,t+s} + i\widehat{i}_{t+s} - p^L L \ \widehat{L}_{t+s} - \frac{p^K k}{(1+g)(1+n)} \ \widehat{k}_{t+s-1} \right] + \lambda \frac{(1-\delta)+p^K}{(1+g)(1+n)} k\widehat{k}_{t-1} \tag{3.62}$$

3.10.5 Connecting the level of productivity to the level of welfare

Define value added (for normalized variables in deviation from steady state) as:

$$\widehat{y}_t = \log y_t - \log y = \sum_{i=1}^{Z} \frac{P_i^C C_i N}{P^Y Y} \widehat{c}_{it} + \frac{P^I I}{P^Y Y} \widehat{i}_t = \sum_{i=1}^{Z} s_{c_i} \widehat{c}_{it} + s_i \widehat{i}_t \tag{3.63}$$

Inserting this equation into 3.62, and noticing that $\frac{p^K k}{p^Y y(1+g)(1+n)}$is the SS value of $s_{K,t} \equiv \frac{P_t^K K_{t-1}}{P_t^Y Y_t}$ we get:

$$v_t = v + E_t \sum_{s=0}^{\infty} \beta^s \lambda p^Y y \left[\widehat{y}_t - s_L \ \widehat{L}_{t+s} - s_K \ \widehat{k}_{t+s-1} \right] + \lambda \frac{(1-\delta)+p^K}{(1+g)(1+n)} k\widehat{k}_{t-1} \tag{3.64}$$

Using the definition of the normalized variable, this can be rewritten as:

$$v_t = v + \left(\lambda p^Y y\right) E_t \sum_{s=0}^{\infty} \beta^s \left[(\log \frac{Y_{t+s}}{N_{t+s}X_{t+s}} - \log y) - s_L(\log L_{t+s} - \log L) - s_K(\log \frac{K_{t+s-1}}{N_{t+s-1}X_{t+s-1}} - \log\right.$$

$$+\lambda \frac{(1-\delta)+p^K}{(1+g)(1+n)} k\widehat{k}_{t-1} \qquad (3$$

or:

$$v_t = \left(\lambda p^Y y\right) E_t \sum_{s=0}^{\infty} \beta^s \left[\log Y_{t+s} - s_L \log\ N_{t+s}L_{t+s} - s_K \log K_{t+s-1}\right] + \lambda \frac{(1-\delta)+p^K}{(1+g)(1+n)} k\widehat{k}_{t-1} + f(t) \tag{3.66}$$

where:

$$f(t) = -\frac{\lambda p^y y}{1-\beta}\left[\log y - s_L \log L - s_K \log k + \frac{\beta}{(1-\beta)}\left[g(1-s_K) + n(1-s_L-s_K)\right]\right]$$
$$-\frac{\lambda p^y y}{1-\beta}\left[(1-s_K)\log X_t + (1-s_L-s_K)\log N_t\right] - \frac{\lambda p^y y s_K}{1-\beta}(n+g)$$

Define aggregate productivity (in log level) as: $\log pr_t = \log Y_t - s_L \log N_t L_t - s_K \log K_{t-1}$. Notice that we are taking a definition with constant shares. Using this definition, the equation above can be rewritten as:

$$v_t - v = \left(\lambda p^Y y\right) E_t \sum_{s=0}^{\infty} \beta^s \log pr_{t+s} + \lambda \frac{(1-\delta)+p^K}{(1+g)(1+n)} k\widehat{k}_{t-1} + f(t) \tag{3.67}$$

3.10.6 Connecting the aggregate Solow residual with the level of welfare

Now define: $\Phi_{t+s} = \sum_{i:1}^{Z} p_i^C c_i \log c_{i,t+s} + i \log i_{t+s} - p^L L \log L_{t+s} - \frac{p^K k}{(1+g)(1+n)} \log k_{t+s-1}$ and note that for a variable x:

$$x_{t+s} - x = \sum_{i=1}^{s} (x_{t+i} - x_{t+i-1}) + x_t - x$$

Using this property, equation 3.62 can be rewritten as:

$$v_t = v + E_t \sum_{s=0}^{\infty} \beta^s \lambda \left[E_t \sum_{i=1}^{s} (\Phi_{t+i} - \Phi_{t+i-1}) + \Phi_t - \Phi \right] + \lambda \frac{(1-\delta) + p^K}{(1+g)(1+n)} k \widehat{k}_{t-1}$$

If we are willing to make the hypothesis that the period before the shock the system was in its steady state, so that: $\Phi_{t-1} = \Phi$ and $\widehat{k}_{t-1} = 0$, then the equation above can be rewritten as: $v_t = v + E_t \sum_{s=0}^{\infty} \beta^s \lambda \left[\sum_{i:0}^{s} \Delta \Phi_{t+i} \right]$ or alternatively:

$$v_t = v + \frac{\lambda}{(1-\beta)} E_t \sum_{s=0}^{\infty} \beta^s \Delta \Phi_{t+s} + \lambda \frac{(1-\delta) + p^K}{(1+g)(1+n)} k \widehat{k}_{t-1}$$

Substituting back the definition of Φ_{t+s} , we get:

$$\begin{aligned} v_t \quad = \quad & v + \frac{\lambda}{(1-\beta)} E_t \sum_{s=0}^{\infty} \beta^s (\sum_{i=1}^{Z} p_i^C c_i \Delta \log c_{i,t+s} + i \Delta \log i_{t+s} - p^L L \Delta \log L_{t+s} \\ & - \frac{p^K k}{(1+g)(1+n)} \Delta \log k_{t+s-1}) + \lambda \frac{(1-\delta) + p^K}{(1+g)(1+n)} k \widehat{k}_{t-1} \end{aligned} \tag{3.68}$$

where Δ denotes difference over time. Define value added growth (at constant shares) as:[18]

$$\Delta \log y_t = \sum_{i=1}^{Z} \frac{p_i^C c_i}{p^Y y} \Delta \log c_{i,t+s} + \frac{i}{p^Y y} \Delta \log i_t \tag{3.69}$$

Using the fact that nominal value added $P_t Y_t = \sum_{i:0}^{Z} P_{i,t}^C C_{i,t} N_t + P_t^I I_t$, it is also true that:

$$\Delta \log Y_t = \sum_{i=1}^{Z} \frac{P_i^C C_i N}{P^Y Y} \Delta \log(C_{i,t} N_t) + \frac{P^I I}{P^Y Y} \Delta \log I_t \tag{3.70}$$

Now, insert this into equation 3.68 and factor out $p^Y y$ to obtain:

$$\begin{aligned} v_t \quad = \quad & v + \frac{\lambda p^Y y}{(1-\beta)} E_t \sum_{s=0}^{\infty} \beta^s \left(\Delta \log y_{t+s} - \frac{p^L L}{p^Y y} \Delta \log L_{t+s} - \frac{p^K k}{(1+g)(1+n) p^Y y} \Delta \log k_{t+s-1} \right) \\ & + \lambda \frac{(1-\delta) + p^K}{(1+g)(1+n)} k \widehat{k}_{t-1} \end{aligned} \tag{3.71}$$

Using the fact that:

$$\Delta \log y_t = \Delta \log(\frac{Y_t}{X_t N_t}) = \Delta \log Y_t - g - n$$

$$\Delta \log L_t = \Delta \log(N_t L_t \frac{1}{N_t}) = \Delta \log N_t L_t - n$$

$$\Delta \log k_t = \Delta \log(\frac{K_t}{X_t N_t}) = \Delta \log K_t - g - n$$

[18] Here we are departing slightly from convention, as value added is usually calculated with time varying shares.

we can rewrite equation 3.71 as:

$$\begin{aligned} v_t &= v + \frac{\lambda p^Y y}{(1-\beta)} E_t \sum_{s=0}^{\infty} \beta^s \left[\Delta \log Y_{t+s} - s_L \Delta \log N_{t+s} L_{t+s} - s_K \Delta \log \ K_{t+s-1}\right] \\ &+ \frac{\lambda p^Y y}{(1-\beta)} \sum_{s=0}^{\infty} \beta^s \left[g(1-s_K) + n(1-s_K-s_L)\right] + \lambda \frac{(1-\delta)+p^K}{(1+g)(1+n)} k\widehat{k}_{t-1} \end{aligned}$$

Denote $\Delta \log pr_{t+s}$ the (modified) Solow productivity residual:

$$\Delta \log pr_{t+s} = \Delta \log Y_{t+s} - s_L \Delta \log N_{t+s} L_{t+s} - s_K \Delta \log \ K_{t+s-1}$$

Using this definition, we get:

$$v_t - v = \frac{\lambda p^Y y}{(1-\beta)} E_t \sum_{s=0}^{\infty} \beta^s \Delta \log pr_{t+s} + \lambda \frac{(1-\delta)+p^K}{(1+g)(1+n)} k\widehat{k}_{t-1} + f_0 \tag{3.72}$$

where:

$$f_0 = \frac{\lambda p^Y y}{(1-\beta)^2} \left[g(1-s_K) + n(1-s_K-s_L)\right]$$

Now suppose we are not willing to assume that: $\Phi_{t-1} = \Phi$ and $\widehat{k}_{t-1} = 0$. We are back to the case:

$$v_t = v + E_t \sum_{s=0}^{\infty} \beta^s \lambda \left[\sum_{i=1}^{s} (\Phi_{t+i} - \Phi_{t+i-1}) + \Phi_t - \Phi\right] + \lambda \frac{(1-\delta)+p^K}{(1+g)(1+n)} k\widehat{k}_{t-1}$$

which can be rewritten as:

$$
\begin{aligned}
v_t &= v + \frac{\lambda}{(1-\beta)} E_t \sum_{s=1}^{\infty} \beta^s \Delta \Phi_{t+s} + \frac{\lambda}{(1-\beta)} (\Phi_t - \Phi) \\
&\quad + \lambda \frac{(1-\delta) + p^K}{(1+g)(1+n)} k \widehat{k}_{t-1}
\end{aligned} \tag{3.73}
$$

If we assume that:

$$
\sum_{i=1}^{G} p_i^C c_i \widehat{c}_{i,t+s} + i \widehat{i}_{t+s} = p^y y \widehat{y}
$$

then $\Phi_t - \Phi = p^y y \widehat{y}_t - p^L L \widehat{L}_t - \frac{p^K k}{(1+g)(1+n)} \widehat{k}_{t-1}$ and after some algebra:

$$
\Phi_t - \Phi = p^y y \left[\log pr_t - (\log y - s_L \log L - s_K \log k) - (1 - s_K) \log X_t - (1 - s_L - s_K) \log N_t + s_K g + s_K n \right]
$$

Substituting this result into equation (3.73) and rearranging some terms, we get:

$$
v_t - v = \frac{\lambda p^Y y}{(1-\beta)} E_t \sum_{s=1}^{\infty} \beta^s \Delta \log pr_{t+s} + \frac{\lambda p^Y y}{(1-\beta)} \log pr_t + \lambda \frac{(1-\delta) + p^K}{(1+g)(1+n)} k \widehat{k}_{t-1} + f(t) \tag{3.74}
$$

3.10.7 Connecting the aggregate Solow residual with the change in welfare

Take the difference between the expected level of intertemporal utility v_t defined in (3.62) and v_{t-1}.

$$\begin{aligned}
\Delta v_t &= E_t \sum_{s=0}^{\infty} \beta^s \lambda \left[\sum_{i=1}^{Z} p_i^C c_i \log c_{i,t+s} + i \log i_{t+s} - p^L L \log L_{t+s} - \frac{p^K k}{(1+g)(1+n)} \log \; k_{t+s-1} \right] \\
&\quad - E_{t-1} \sum_{s=0}^{\infty} \beta^s \lambda \left[\sum_{i=1}^{Z} p_i^C c_i \log c_{i,t+s-1} + i \log i_{t+s-1} - p^L L \log L_{t+s-1} - \frac{p^K k}{(1+g)(1+n)} \log \; k_{t+s-2} \right] \\
&\quad + \lambda \frac{(1-\delta) + p^K}{(1+g)(1+n)} k \Delta \widehat{k}_{t-1}
\end{aligned}$$

The right hand side, after adding and subtracting, for each variable x_{t+s}, $E_t x_{t+s}$, can be written as:

$$\begin{aligned}
\Delta v_t &= E_t \sum_{s=0}^{\infty} \beta^s \lambda [\sum_{i=1}^{Z} p_i^C c_i \Delta \log c_{i,t+s} + i \Delta \log i_t - p^L L \Delta \log L_{t+s} - \frac{p^K k}{(1+g)(1+n)} \Delta \log \; k_{t+s-1}] \\
&\quad + \sum_{s=0}^{\infty} \beta^s \lambda [\sum_{i=1}^{Z} p_i^C c_i \left(E_t \log c_{i,t+s} - E_{t-1} \log c_{i,t+s} \right) + i \left(E_t \log i_{t+s} - E_{t-1} \log i_{t+s} \right) \\
&\quad - p^L L E_t (\log L_{t+s} - E_{t-1} \log L_{t+s}) - \frac{p^K k}{(1+g)(1+n)} \left(E_t \log k_{t+s-1} - E_{t-1} \log k_{t+s-1} \right)] \\
&\quad + \lambda \frac{(1-\delta) + p^K}{(1+g)(1+n)} k \Delta \widehat{k}_{t-1}
\end{aligned}$$

Rearranging the RHS, we obtain:

$$\begin{aligned}\Delta v_t &= \lambda p^y y E_t \sum_{s=0}^{\infty} \beta^s \Delta \log pr_{t+s} \\ &+ \lambda p^y y \sum_{s=0}^{\infty} \beta^s \left[E_t \log pr_{t+s} - E_{t-1} \log pr_{t+s} \right] \\ &+ \lambda \frac{(1-\delta) + p^K}{(1+g)(1+n)} k \Delta k_{t-1} + f_1 \end{aligned} \tag{3.75}$$

where $E_t \log pr_{t+s} - E_{t-1} \log pr_{t+s}$ represents the revision in expectations of the level of the productivity residual (normalized by population and Harrod neutral technological progress) based on the new information received between t-1 and t and:

$$f_1 = \frac{\lambda p^Y y}{(1-\beta)} \left[g(1-s_K) + n(1-s_K-s_L) \right]$$

Bibliography

[1] Abel, A.B., 2003. The Effects of a Baby Boom on Stock Prices and Capital Accumulation in the Presence of Social Security. Econometrica 71(2), 551-578

[2] Alesina, A., Ardagna, S., Nicoletti, Schiantarelli, F., 2005. Regulation and investment. Journal of European Economic Association 3(4), 1-35

[3] Arellano, M., Bond, S., 1991. Some tests of specification for panel data: Monte Carlo evidence and an application to employment equations. Review of Economic Studies 14, 328-352

[4] Baker, D., Rosnick, D., 2007. "Usable productivity" growth in the US: an international comparison, 1980-2005. International Productivity Monitor

[5] Blundell, R., Bond, S., 1998. Initial conditions and moment restrictions in dynamic panel models. Journal of Econometrics 87, 115-143

[6] Blundell, R., Bond, S., 2000. GMM estimation with persistent panel data: an applications to production functions. Econometric Review 87, 115-143

[7] Basu, S., Fernald, J.G., 2002. Aggregate productivity and aggregate technology. European Economic Review 46, 963-991

[8] Basu, S., Fernald J., Kimball M., 2006. Are Technology Improvements Contractionary?. American Economic Review 96 (5), 1418-1448

[9] Basu, S., Fernald, J.G., Shapiro, M.D., 2003. Productivity Growth in the 1990s: Technology, Utilization, or Adjustment? NBER Working Paper 8359

[10] Basu, S., Inklaar, R., Wang C.J, forthcoming. The Value of Risk: Measuring the Service Output of U.S. Commercial Banks. Economic Inquiry.

[11] Dobbelaere S., Mairesse J., 2008. Panel data estimates of the production function and product and labor market imperfections. NBER Working Paper 13975

[12] Domar, E.D., 1961. On the measurement of technical change. Economic Journal 71, 710-729

[13] Grilliches, Z., Regev, H., 1995. Productivity and firm turnover in Israeli industry. Journal of Econometrics 65, 175-203

[14] Hellerstein, J.K., Neumark, D., 2008. Workplace Segregation in the United States: Race, Ethnicity, and Skill. The Review of Economics and Statistics 90(3), 459-477

[15] Hsieh, C.T., Klenow, P.J., 2007. Misallocation and Manufacturing TFP in China and India. NBER Working Papers 13290

[16] Hulten, C.R, 1978. Growth accounting with intermediate inputs, The Review of Economic Studies 45, 511-518

[17] Hulten, C.R., Schreyer, P., 2009. GDP, Technical Change and the Measurement of Net Income: the Weitzman Model Revisited. NBER Working Paper 16010

[18] Inklaar, R., Wang, J.C., 2007. Not Your Grandfather's Bank Any More? Consistent Measurement of Non-Traditional Bank Output. mimeo, University of Groningen and Federal Reserve Bank of Boston

[19] King, R.G., Plosser, C.I., Rebelo, S.T., 1988. Production, growth and business cycles. Journal of Monetary Economics 21, 195-232

[20] Klette, T.J., Griliches, Z., 1996. The Inconsistency of Common Scale Estimators When Output Prices Are Unobserved and Endogenous. Journal of Applied Econometrics 11(4), 343-61

[21] Kohli, U., 2004. Real GDP, Real Domestic Income, and Terms-of-Trade Changes. Journal of International Economics 62, 83-106

[22] Nordhaus, W., Tobin, J., 1972. Is growth obsolete? NBER General Series 96

[23] O'Mahony, M., Timmer, M.P., 2009. Output, Input and productivity measures at the industry level: the EU KLEMS database. Economic Journal 119, pp.F374-F403

[24] Olley, S., Pakes, A., 1996, The dynamics of productivity in the telecommunications equipment industry. Econometrica 64, 1263-1297

[25] Petrin, A., Levinsohn, J., 2009. Measuring Aggregate Productivity Growth Using Plant-Level Data. Unpublished.

[26] Petrin, A., White, K., Reiter, J.P., 2009. The impact of plant-level resource reallocations and technical progress on U.S. macroeconomic growth. Unpublished.

[27] Wang, C.J., Basu, S., Fernald, J.G., forthcoming In W. Erwin Diewert, John Greenlees and Charles Hulten, eds., Price Index Concepts and Measurement

[28] Weitzman, M., 1976. On the welfare significance of national product in a dynamic economy. Quarterly Journal of Economics 90, 156-162.

[29] Weitzman, M., 2003. Income capital and the maximum principle. Harvard University Press, Cambridge.

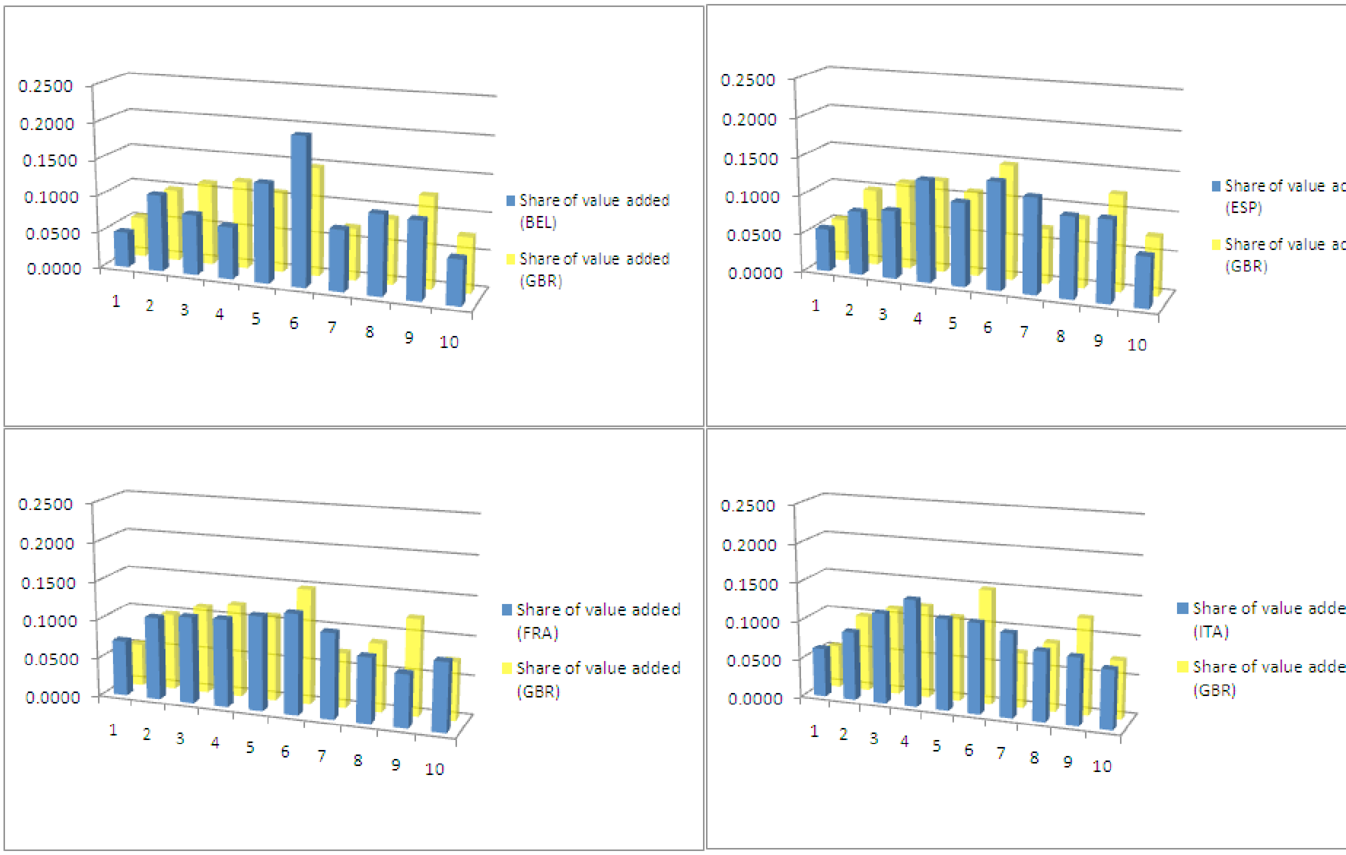

Figure 3-1: Accounting for differences with respect to Great Britain in aggregate productivity change: firm-level data

Table 3.1: Time series of the Solow residual

Dep. Variable = $\log pr_t^J$	(1) BEL	(2) ESP	(3) FRA	(4) ITA	(5) GBR
$\log pr_{t-1}^J$	0.790***	0.875***	0.694***	0.847***	0.739***
	(0.0957)	(0.0973)	(0.100)	(O.133)	(0.0984)
N	25	25	25	35	35
LM1 (p-value)	0.491	0.926	0.215	0.927	0.396
LM3 (p-value)	0.290	0.0166	0.4740	0.992	0.118

The table reports OLS estimates for the years up to 2005. A time trend is included in the regression. LM1 (LM3) is the Lagrange Multiplier tests for residual serial correlation up to the first (third) order. Standard errors are reported in parentheses. *** significant at less than 0.1 percent; ** significant at 1 percent; * significant at 5 percent.

Table 3.2: Accounting for aggregate productivity change between 1984 and 2004: aggregate and industry productivity change (annual rates)

	BEL	**ESP**	**FRA**	**ITA**	**GBR**
	$d\log pr$	$d\log pr$	$d\log pr$	$d\log pr$	$d\log pr$
	-0.0014	*-0.0024*	*0.0076*	*0.0027*	*0.0088*
Industry:					
Manufacturing	0.0068	0.0048	0.0112	0.0041	0.0103
Electricity, Gas, Water supply	0.0159	0.0327	0.0287	-0.0071	0.0149
Construction	0.0100	-0.0091	-0.0034	-0.0120	0.0000
Wholesale and Retail	-0.0203	-0.0069	0.0065	0.0051	0.0058
Finance, Insurance, Real estate	0.0007	-0.0009	-0.0103	-0.0030	-0.0132

Table 3.3: Accounting for aggregate productivity change between 1998 and 2005: small and large firms (annual rates)

	$d\log pr$ **(aggregate)**	**group**	**Share(group)**	$d\log pr$ **(group)**
BEL	0.0251	*LARGE FIRMS*	0.9241	0.0256
		SMALL FIRMS	0.0759	0.0182
ESP	0.0006	*LARGE FIRMS*	0.7044	0.0031
		SMALL FIRMS	0.2956	-0.0053
FRA	0.0293	*LARGE FIRMS*	0.7964	0.0294
		SMALL FIRMS	0.2036	0.0288
ITA	0.0057	*LARGE FIRMS*	0.7019	0.0046
		SMALL FIRMS	0.2981	0.0083
GBR	0.0523	*LARGE FIRMS*	0.9417	0.0519
		SMALL FIRMS	0.0583	0.0587

Table 3.4: Accounting for differences with respect to Great Britain in aggregate productivity change between 1984 and 2003: industry-level data (annual rates)

	Difference wrt GBR $d\log pr$	**Difference accounted by:**	
		differ. in average $d\log pr$	*differ. in industrial comp*
BEL	-0.0101	-0.0099	-0.0002
ESP	-0.0112	-0.0117	0.0005
FRA	-0.0011	0.0005	-0.0017
ITA	-0.0061	-0.0066	0.0005

Table 3.5: Accounting for differences with respect to Great Britain in aggregate productivity change between 1998 and 2005: firm-level data (annual rates)

BEL

Decile	$d \log pr$ (average)	Share value added (BEL)	Share (BEL)*$d \log pr$		Share(GBR)*$d \log pr$
1	-0.1593	0.0480	-0.0076		-0.0087
2	-0.0665	0.1045	-0.0069		-0.0066
3	-0.0321	0.0823	-0.0026		-0.0036
4	-0.0083	0.0713	-0.0006		-0.0010
5	0.0106	0.1343	0.0014		0.0012
6	0.0323	0.2006	0.0065		0.0048
7	0.0536	0.0834	0.0045		0.0038
8	0.0726	0.1089	0.0079		0.0064
9	0.1065	0.1056	0.0112		0.0131
10	0.1859	0.0611	0.0114		0.0139
		aggregate $d \log pr$=	0.0251	aggregate $d \log pr$ using GBR shares =	0.0232

ESP

Decile	$d \log pr$ (average)	Share value added (ESP)	Share (ESP)*$d \log pr$		Share(GBR)*$d \log pr$
1	-0.2116	0.0555	-0.0117		-0.0116
2	-0.1101	0.0826	-0.0091		-0.0109
3	-0.0671	0.0883	-0.0059		-0.0076
4	-0.0354	0.1316	-0.0047		-0.0042
5	-0.0121	0.1076	-0.0013		-0.0013
6	0.0112	0.1384	0.0015		0.0016
7	0.0319	0.1229	0.0039		0.0023
8	0.0562	0.1040	0.0058		0.0050
9	0.0928	0.1050	0.0097		0.0114
10	0.1907	0.0643	0.0123		0.0143
		aggregate $d \log pr$=	0.0006	aggregate $d \log pr$ using GBR shares =	-0.0011

FRA

Decile	$d \log pr$ (average)	Share value added (FRA)	Share (FRA)*$d \log pr$		Share(GBR)*$d \log pr$
1	-0.1552	0.0716	-0.0111		-0.0085
2	-0.0697	0.1063	-0.0074		-0.0069
3	-0.0334	0.1113	-0.0037		-0.0038
4	-0.0048	0.1129	-0.0005		-0.0006
5	0.0186	0.1212	0.0023		0.0020
6	0.0413	0.1285	0.0053		0.0061
7	0.0632	0.1091	0.0069		0.0045
8	0.0896	0.0837	0.0075		0.0079
9	0.1284	0.0675	0.0087		0.0159
10	0.2440	0.0878	0.0214		0.0182
		aggregate $d \log pr$=	0.0293	aggregate $d \log pr$ using GBR shares =	0.0348

ITA

Decile	$d \log pr$ (average)	Share value added (ITA)	Share (ITA)*$d \log pr$		Share(GBR)*$d \log pr$
1	-0.2138	0.0628	-0.0134		-0.0117
2	-0.0979	0.0886	-0.0087		-0.0097
3	-0.0565	0.1170	-0.0066		-0.0064
4	-0.0262	0.1383	-0.0036		-0.0031
5	-0.0031	0.1179	-0.0004		-0.0003
6	0.0189	0.1171	0.0022		0.0028
7	0.0418	0.1077	0.0045		0.0030
8	0.0682	0.0892	0.0061		0.0060
9	0.1094	0.0863	0.0094		0.0135
10	0.2148	0.0751	0.0161		0.0160
		aggregate $d \log pr$=	0.0057	aggregate $d \log pr$ using GBR shares =	0.0101

GBR

Decile	$d \log pr$ (average)	Share value added (GBR)	Share (GBR)*$d \log pr$
1	-0.1477	0.0549	-0.0081
2	-0.0418	0.0990	-0.0041
3	-0.0076	0.1126	-0.0009
4	0.0170	0.1195	0.0020
5	0.0354	0.1091	0.0039
6	0.0555	0.1477	0.0082
7	0.0766	0.0710	0.0054
8	0.1051	0.0881	0.0093
9	0.1430	0.1234	0.0176
10	0.2536	0.0747	0.0189
		aggregate $d \log pr$=	0.0523

Table 3.6: Estimate of the production function for Belgium using OLS.

industry	ϵ_k	se(ϵ_k)	ϵ_l	se(ϵ_l)	ϵ_m	se(ϵ_m)	θ	μ
15	.024254	.0021049	.1715899	.002782	.7892767	.0021862	.9851206	1.095934
16	.0950597	.0137134	.0665313	.0187802	.7669978	.026175	.9285887	1.110088
17	.0345663	.0026726	.2034413	.0039416	.7408124	.0036503	.97882	1.11444
18	.0098506	.0050477	.1718834	.0065111	.8142364	.0058267	.9959704	1.119567
19	.0366267	.021146	.2478886	.0275733	.7290825	.0223094	1.013598	1.037782
20	.0331624	.0032934	.1650923	.0050643	.7872586	.0044691	.9855132	1.12492
21	.018816	.0042258	.2189154	.0084079	.7431836	.0068755	.980915	1.116072
22	.0242819	.0024171	.2473774	.0046304	.7000399	.0044372	.9716992	1.185321
23	.014024	.0175733	.2790716	.0273318	.7549599	.0124866	1.048056	1.172826
24	.0350945	.0028137	.2011176	.0048504	.7674007	.0043999	1.003613	1.126789
25	.0327325	.0026066	.2045444	.0043409	.7536795	.0040567	.9909564	1.132583
26	.0297956	.0025126	.1893844	.0036685	.7797855	.0039222	.9989654	1.170156
27	.0424432	.003352	.1620249	.0052216	.7756518	.0044041	.9801198	1.179669
28	.0307327	.0018343	.248825	.0031259	.701847	.0029655	.9814047	1.169207
29	.0417338	.0026518	.2236885	.004274	.725044	.0044404	.9904663	1.147926
30	.0438384	.0097131	.2133444	.0152757	.7542191	.0179286	1.011402	1.1183
31	.0185072	.0038281	.2576855	.007376	.7055325	.0069817	.9817252	1.167024
32	.0293824	.0113148	.1988501	.0195012	.7742139	.0164682	1.002446	1.262404
33	.0285132	.0050092	.187353	.009275	.7525092	.0091761	.9683754	1.257869
34	.0152044	.0044613	.1854333	.0063027	.7913007	.0053243	.9919384	1.153288
35	.0283478	.0063816	.224206	.0111818	.7547176	.0114414	1.007271	1.216691
36	.0168709	.0027038	.1537802	.003637	.8301196	.0034038	1.000771	1.235662
37	.0496096	.0052469	.142829	.009983	.7814211	.0076006	.9738598	1.133132

Table 3.7: Estimate of the production function for Belgium using system GMM (without AR(1) errors)

industry	ϵ_k	se(ϵ_k)	ϵ_l	se(ϵ_l)	ϵ_m	se(ϵ_m)	θ	μ
15	.0215719	.0143043	.116369	.0533856	.845057	.0234978	.9829979	1.173386
16	.1015663	.0310604	.0631688	.0307818	.7815633	.0627052	.9462985	1.131169
17	.0197039	.0122925	.2157265	.0192355	.7701956	.0248191	1.005626	1.158642
18	-.0056251	.0206674	.1726151	.035114	.8318266	.0329617	.9988165	1.143753
19	.0311772	.0317691	.2345751	.0333582	.7474033	.0138473	1.013156	1.06386
20	.0339828	.0176306	.1430129	.0319975	.8317935	.0344319	1.008789	1.188557
21	.0109747	.0141275	.2621731	.0495077	.7361645	.0470625	1.009312	1.105531
22	-.0020648	.01703	.2060769	.0415456	.7255507	.0572205	.9295627	1.228516
23	.0090164	.0240922	.275057	.0588486	.7610017	.0387986	1.045075	1.182212
24	-.0025681	.0212298	.2419967	.0459446	.7398928	.0336436	.9793214	1.086399
25	.0253915	.0109819	.2186054	.034797	.7245669	.0391779	.9685638	1.088834
26	.0415794	.0154214	.1784004	.0352567	.7706105	.0488047	.9905902	1.156388
27	.0359403	.0272913	.1275111	.0291271	.7962665	.0325451	.9597178	1.211021
28	.0082466	.0110756	.221101	.0223465	.7515167	.0249234	.9808643	1.251952
29	.003992	.0129839	.1750448	.0241039	.8069009	.027443	.9859377	1.277526
30	.0428727	.0109362	.2259799	.0234276	.7459196	.0248053	1.014772	1.105995
31	.0098459	.0318262	.2204551	.0485979	.7511831	.0353373	.9814841	1.242535
32	-.0129264	.0155392	.2198884	.0259918	.7900414	.0319096	.9970034	1.288212
33	.0143141	.0157171	.2116466	.0439033	.7581939	.0447299	.9841546	1.267371
34	.0058168	.0141078	.1667487	.0274858	.8067272	.0201717	.9792928	1.175772
35	.0199072	.0128357	.1912204	.0391206	.7724271	.0389743	.9835548	1.24524
36	-.0145004	.0134578	.1721131	.0259554	.8601587	.0200723	1.017771	1.280377
37	.0538229	.0186533	.1383617	.0440445	.7744551	.0277837	.9666397	1.123031

Table 3.8: Estimate of the production function for Belgium using system GMM (without AR(1) errors): validating tests

industry	hansen (p-value)	ar1 (p-value)	ar2p (p-value)
15	.1526701	2.48e-08	.2259859
16	1	.1564406	.7954623
17	.0493479	.0102783	.080242
18	.7306365	.0457279	.6914725
19	1	.9864044	.0796718
20	.6683506	.001477	.089917
21	.2565103	.0130565	.6296991
22	.1097888	.0124187	.9321187
23	1	.8665326	.1526774
24	.6437734	.0004697	.1085116
25	.5183361	.0621041	.2388145
26	.1234163	.0059662	.437268
27	.605441	.0003418	.3046295
28	.0032045	.0080472	.4424166
29	.0916351	.0000682	.5258806
30	1	.3186199	.2074303
31	.1405106	.0128848	.8066118
32	.9991825	.4230046	.0968406
33	.853533	.0147613	.6647028
34	.2278766	.0112883	.4655574
35	.9994887	.0619346	.1509752
36	.956588	.0000944	.2486518
37	.6836056	.054695	.4062182

Table 3.9: Estimate of the production function for Belgium using Olley Pakes.

industry	ϵ_k	ϵ_l	ϵ_m	θ	μ
15	.0053129	.1607034	.7857859	.9518021	1.091087
16	.006006	.0524176	.7730815	.8315051	1.118893
17	.0042158	.195162	.737206	.9365838	1.109014
18	.0054147	.1729344	.8007517	.9791009	1.101026
19	.0053613	.2503178	.7430528	.9987319	1.057668
20	.0111285	.1595692	.7815555	.9522532	1.116771
21	.0111529	.2168774	.7270212	.9550515	1.0918
22	.0097333	.2480117	.6909198	.9486648	1.169878
23	.0099718	.281923	.7470524	1.038947	1.160542
24	.0110562	.1813903	.766848	.9592946	1.125978
25	.0110247	.1982173	.7443261	.9535682	1.118527
26	.0143439	.189329	.757315	.9609879	1.136437
27	.0138856	.1633305	.7672731	.9444892	1.166926
28	.0129493	.2405701	.6961676	.949687	1.159746
29	.0136856	.2278383	.7157144	.9572383	1.133155
30	.0138737	.2379677	.7353308	.9871722	1.090294
31	.0145976	.2384386	.6918178	.944854	1.144339
32	.0144183	.2252213	.7305138	.9701534	1.191149
33	.0147442	.1848012	.7544926	.9540381	1.261184
34	.0146547	.1883912	.7944484	.9974943	1.157876
35	.0152902	.2039668	.7447227	.9639798	1.200578
36	.0152605	.1524164	.8277795	.9954565	1.232179
37	.0156083	.1322395	.7942417	.9420895	1.151723

Table 3.10: Decomposition of the change in aggregate productivity (estimates using System GMM without AR(1) errors)

country	$\mathbf{d} \log \mathbf{pr}$	$(\overline{\mu}-1)(d\log X + d\log M/Q)$	$R_\mu + R_M$	$d \log T$	Residual
BEL	.0352278	0.01446	0.00048	.0114122	0.00888
ESP	.0311934	0.02011	-0.00132	.0048491	0.00755
FRA	.0478567	-0.00055	0.00169	.0405101	0.00621
GBR	.0601621	-0.00084	0.00083	.0490316	0.01114
ITA	.0280874	0.00695	0.00025	.0141505	0.00674

Notes: The estimates of firm productivity are obtained by estimating a production function with year fixed effects using system GMM.

Table 3.11: Decomposition of the change in aggregate productivity. Additional results.

country	**Productivity growth**	Aggregate distortions	Reallocation	Technological change
BEL	1	0.5488	0.0182	0.4331
ESP	1	0.8505	-0.0558	0.2051
FRA	1	-0.0132	0.0406	0.9727
GBR	1	-0.0171	0.0169	1.0002
ITA	1	0.3256	0.0117	0.6629

Notes: The entries in the table represent the percentage of productivity growth accounted by aggregate distortions, reallocation and technical change after reallocating the residual in proportion to the size of each of these components.